



Assessing the diversity of agrobacterial populations

Malek Shams

► To cite this version:

Malek Shams. Assessing the diversity of agrobacterial populations. Agricultural sciences. Université Claude Bernard - Lyon I, 2012. English. NNT : 2012LYO10293 . tel-00984505

HAL Id: tel-00984505

<https://theses.hal.science/tel-00984505>

Submitted on 28 Apr 2014

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.



N° d'ordre : 293-2012

Année 2012

THESE DE L'UNIVERSITE DE LYON

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE
E2M2

DIPLOME DE DOCTORAT

(arrêté du 7 août 2006)

Assessing the diversity of agrobacterial populations

par

M. Malek SHAMS

soutenue publiquement le 19 décembre 2012

Directeur de thèse : Xavier NESME
Co-Encadrante de thèse : Céline LAVIRE

Jury :

Pascal Simonet
Charles Manceau
Yves Dessaux
Yvan Moenne-Loccoz
Céline Lavire
Xavier Nesme

Directeur de recherche CNRS
Ingénieur de recherche INRA
Directeur de recherche CNRS
Professeur université de Lyon
Maître de conférence université de Lyon
Ingénieur de recherche INRA

Rapporteur
Rapporteur
Examinateur
Examinateur
Examinateur
Examinateur

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Vice-président du Conseil d'Administration
Vice-président du Conseil des Etudes et de la Vie
Universitaire
Vice-président du Conseil Scientifique
Secrétaire Général

M. François-Noël GILLY

M. le Professeur Hamda BEN HADID
M. le Professeur Philippe LALLE

M. le Professeur Germain GILLET
M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard
Faculté de Médecine et de Maïeutique Lyon Sud –
Charles Mérieux
UFR d'Odontologie
Institut des Sciences Pharmaceutiques et Biologiques

Institut des Sciences et Techniques de la Réadaptation
Département de formation et Centre de Recherche en
Biologie Humaine

Directeur : M. le Professeur J. ETIENNE
Administrateur provisoire : M. le Professeur G.
KIRKORIAN
Directeur : M. le Professeur D. BOURGEOIS
Directeur : Mme la Professeure C.
VINCIGUERRA.
Directeur : M. le Professeur Y. MATILLON
Directeur : M. le Professeur P. FARGE

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies
Département Biologie
Département Chimie Biochimie
Département GEP
Département Informatique
Département Mathématiques
Département Mécanique
Département Physique
Département Sciences de la Terre
UFR Sciences et Techniques des Activités Physiques
et Sportives
Observatoire de Lyon
Polytech Lyon
Ecole Supérieure de Chimie Physique Electronique
Institut Universitaire de Technologie de Lyon 1
Institut Universitaire de Formation des Maîtres
Institut de Science Financière et d'Assurances

Directeur : M. le Professeur F. De MARCHI
Directeur : M. le Professeur F. FLEURY
Directeur : Mme le Professeur H. PARROT
Directeur : M. N. SIAUVE
Directeur : M. le Professeur S. AKKOUCHE
Directeur : M. le Professeur A. GOLDMAN
Directeur : M. le Professeur H. BEN HADID
Directeur : Mme S. FLECK
Directeur : Mme la Professeure I. DANIEL
Directeur : M. C. COLLIGNON

Directeur : M. B. GUIDERDONI
Directeur : M. P. FOURNIER
Directeur : M. G. PIGNAULT
Directeur : M. C. VITON
Directeur : M. R. BERNARD
Directeur : Mme la Professeure V. MAUME-
DESCHAMPS

Avant-propos

Je voudrais tout d'abord cordialement remercier les membres du jury, pour avoir accepté d'évaluer mes travaux de thèse, et en particulier P. Simonet et C. Manceau qui ont fait le rapport sur ce mémoire, Y. Dessaux qui a apporté son expertise sur la biologie d'*Agrobacterium*, C. Lavire et X. Nesme qui m'ont encadré durant ce travail et Y. Moenne-Loccoz directeur de l'UMR 5557 qui m'a accepté dans l'unité. Je remercie aussi les membres de mon comité de pilotage de thèse, F. Hommais, T. Pommier et D. Muller. Mes plus sincères remerciements vont à mes collègues qui m'ont activement aidé durant ce travail. J. Reynaud qui a co-encadré ce travail avant son départ en retraite. J. Pothier pour les études Maldi-TOF et B. Duffy qui m'a accueilli pour cela à Agroscope à Wädenswil pour un stage de recherche dans le cadre de l'action COST 873, ainsi que P. Portier qui nous a gentiment fourni les souches d'*A. vitis* incluses dans cette étude. De même, je remercie A. Rhouma, M. Bouri (Institut de l'Olivier, Tunis) et F. Bertolla pour le travail d'échantillonnage des sols de Tunisie; C. Mougel et A. Zancarini (INRA, Dijon) pour m'avoir aimablement fourni les ADN métagénomiques de différents génotypes de *M. truncatula*; et A. Chosson, L. Vial et la plate-forme Ibio en général pour les premières analyses de données de pyroséquençage. Je remercie aussi bien sûr D. Chapulliot, R. Briquaire et I. Kerzaon pour leur assistance efficaces dans les "grosses manip". Je remercie aussi J. Pulawska de m'avoir convié à co-écrire un article de revue sur les méthodes de diagnostic des agrobactéries à l'issue de l'atelier international sur le sujet que C. Lavire et l'équipe avaient organisé dans le cadre du COST 873. Mes pensées amicales vont aussi à mes super co-doctorants T. Campillo et F. Lassalle de l'équipe, avec lesquels j'ai participé à un article important, et tous les autres étudiants et permanents de l'unité ou des environs avec qui j'ai beaucoup échangé des idées (et bien d'autres choses bien sûr!) lors de soirées mémorables.

Cette thèse a été possible grâce à une bourse du Ministère des Affaires Etrangères français complétée par un contrat avec le CNRS dans le cadre du projet EcoGenome financé par l'ANR, mais aussi grâce à l'aide et aux conseils efficaces du personnel administratif de l'UMR et en particulier de C. Delorme que je remercie ici chaleureusement.

Je pense aussi à mes frères et sœurs, Maisam, Somayeh et Saeedeh, et à mes amis d'Iran et d'ailleurs qui m'ont soutenu durant cette période. Enfin, bien sûr, je remercie mes parents adorés, Nosrat et Ayoub, qui m'ont encouragé sans réserve à poursuivre de longues études et à faire une thèse à l'étranger, et à qui je dédie ce travail.

Table of Contents

Summary.....	1
--------------	---

General Introduction.....	3
---------------------------	---

Part 1: Bibliography

1-1: Rapid and efficient methods to isolate, type strains and determine species of <i>Agrobacterium</i> spp. in pure culture and complex environment.....	7
1-2: Detection and identification methods and new tests as developed and used in the framework of COST873 for bacteria pathogenic to stone fruits and nuts: Tumorigenic <i>Agrobacterium</i> spp.	9
1-3: <i>Agrobacterium</i> biodiversity studies.....	11

Part 2: Experimental

2-1: <i>Agrobacterium</i> identification using whole-cell MALDI-TOF mass spectrometry.....	23
2-2: Rapid and accurate species identification and exhaustive population diversity assessment of <i>Agrobacterium</i> spp. using <i>recA</i> -based PCR.....	41
2-3: Assessing the presence of agrobacteria and Ti plasmids in soil at a country scale.....	65
2-4: Assessing the pyrosequencing approach to analyze <i>Agrobacterium</i> microdiversity.....	77
Discussion and perspectives.....	91

Annexes

1. Shams M., Campillo T., Lavire C., Muller D., Nesme X., Vial L. 2012. Rapid and efficient methods to isolate, type strains and determine species of *Agrobacterium* spp. in pure culture and complex environments. In : Biochemical Testing. Book edited by: Dr. José C. Jiménez-López, Purdue University, USA. pp-3-20. ISBN 979-953-307-507-4. doi: 10.5772/2250
2. Campillo T., Lavire C., Shams M., Pothier J.F., Pulawska J. 2012. Detection and identification methods and new tests as developed and used in the framework of COST873 for bacteria pathogenic to stone fruits and nuts: Tumorigenic *Agrobacterium* spp. J. Plant Pathol. 94 :S1.97-S1.104. doi: 10.4454/jpp.v94i1sup.016
3. Lassalle F., Campillo T., Vial L., Baude J., Costechareyre D., Chapulliot D., Shams M., Abrouk D., Lavire C., Oger-Desfeux C., Hommais F., Guéguen L., Daubin V., Muller D., Nesme X. 2011. Genomic species are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. Genome Biol. Evol. 3:762–781. doi:10.1093/gbe/evr070

Résumé

Les bactéries du genre *Agrobacterium* forment un ensemble taxonomiquement diversifié composé de nombreuses espèces, présent dans la plupart des sols et dans la rhizosphère des plantes. Les agrobactéries sont le plus souvent anodines voire stimulatrices de la croissance des plantes mais celles qui hébergent un plasmide Ti induisent la maladie de la galle du collet à de nombreuses plantes d'intérêt agronomique. Les plasmides Ti peuvent se transmettre par conjugaison aux autres agrobactéries qui deviennent à leur tour pathogènes. Ainsi, du point de vue écologique comme du point de vue épidémiologique, il est impératif de connaître la diversité des populations d'agrobactéries du sol, hôtes potentiels des plasmides Ti. Dans ce contexte, nous avons dans notre travail de thèse d'une part donné l'état actuel des connaissances sur la taxonomie du genre *Agrobacterium*, et nous avons fait une revue des méthodes d'isolement et de typage de ces bactéries. D'autre part, nous avons cherché à mettre au point des méthodes d'identification rapides et fiables des différentes espèces. Une première méthode est basée sur la technique MALDI-TOF MS. Elle permet d'identifier facilement les espèces de bactéries cultivées mais la méthode n'est pas assez résolutive pour typer des souches et encore moins la présence de plasmides Ti dans les isolats. Nous avons alors développé des amorces de PCR spécifiques de chaque espèces (17), pour le genre *Agrobacterium* et la famille Rhizobiaceae. Ces amorces se sont révélées efficaces pour identifier les bactéries cultivées mais aussi pour détecter leur présence dans des communautés microbiennes. Nous avons ensuite utilisé ces outils pour étudier la répartition des agrobactéries à l'échelle d'un pays, d'une station et entre sols nus et sols rhizosphériques en utilisant soit des isolats soit des ADN métagénomiques extraits des différents environnements. Enfin, nous avons montré que le clonage-séquençage ou le séquençage à haut débit d'amplicons obtenus à partir d'ADN de communautés microbiennes nous permettaient de connaître facilement et avec une grande profondeur la diversité des populations d'agrobactéries. Les outils que nous avons développés sont fiables et faciles d'accès. Ils permettront d'étudier l'effet des facteurs de l'environnement sur la structure génétique des populations d'*Agrobacterium* pour répondre, par exemple, à la question de l'effet sélectif des plantes sur les espèces d'agrobactéries du sol recrutées dans la rhizosphère.

General Introduction

Agrobacterium are Alphaproteobacteria common in most soils that closely interact with plants in two respects. Firstly, and as a general trait of the whole taxon, they are rhizospheric bacteria saprophytically living in the root environment (i.e. rhizosphere) of numerous plants. Rhizospheric interactions are generally considered to be of commensal type with no detrimental effect to the plant, but in most instances they are likely beneficial to plants (Hao et al., 2011). For evident agronomic purposes it is thus worthwhile to explore the genetic diversity of agrobacteria of potentially important plant growth-promoting rhizobacteria (PGPR). Nevertheless, it is not yet known if the PGPR ability is a general trait among *Agrobacterium* or just a strain- or species-specific trait. Such a study requires an expert knowledge of the *Agrobacterium* taxonomy. Secondly, when they harbor a dispensable Ti plasmid (i.e. tumor inducing plasmid), agrobacteria are plant pathogens able to cause the crown gall disease to most dicots and gymnosperms and some monocots (Pitzscke & Hirt, 2010). Ti plasmids are conjugative and can easily spread in indigenous soil agrobacteria. As a result transconjugant agrobacteria become in turn pathogenic, contributing both to disease spread and perennial soil contaminations. An epidemiological survey of crown gall thus also requires expert determination of the *Agrobacterium* taxonomy. However, the *Agrobacterium* taxonomy was historically based upon pathogenicity, while in modern bacterial taxonomy, the criteria to delineate taxa include genomic and phylogenetic informations. This recently brought considerable modifications to the *Agrobacterium* nomenclature that need to be clearly explicated to base ecological works on most recent advances. In parallel, a point about biochemical and molecular methods set up to establish the taxonomic status of agrobacteria is required.

Agrobacteria are considered as common soil and rhizosphere inhabitants. However, to our knowledge, there is no exhaustive data about biotic or abiotic factors that influence their presence in particular environments. Our hypothesis is that determining factors that are shaping agrobacterial population genetic structures will have important issues for crown gall control itself. Indeed finding factors that limit the survival of agrobacteria and especially of the Ti plasmid harboring ones may provide us new ways to control perennial soil contaminations by pathogenic agrobacteria. As we will show later, indigenous agrobacterial populations are composites, generally with several species and strains per species. It is however difficult to extensively assess the taxonomic status of *Agrobacterium* populations by standard microbial approaches which are expensive and time-consuming. Hence, we were setting up affordable methods that will allow the simultaneous, rapid and accurate identification of numerous

agrobacterial isolates and/or undoubtedly detect *Agrobacterium* species directly in their common biotopes, soil, root and tumors, without bacterial isolations.

In a first part, the present thesis establishes the state of the art about the current agrobacterial taxonomy (Chapter 1-1), and current methods to isolate, type strains and determine species of *Agrobacterium* (Chapter 1-2). In a second part, we present development and evaluation of new tools for large scale *Agrobacterium* biodiversity studies. First, we evaluated the usefulness of MALDI-TOF MS technology as a high throughput tool to characterize and classify agrobacteria (Chapter 2-1). Then we set up a *recA*-based PCR method to accurately and exhaustively assess agrobacterial diversity either of isolated agrobacteria or directly in various biotopes (Chapters 2-2). We applied standard biochemical, and *recA*-based and Ti plasmid-based identification/detection methods to study the prevalence of pathogenic and non-pathogenic (i.e. Ti plasmid-free) agrobacteria at the country and local scales (Chapter 2-3). Finally, we tested whether analyzing the internal composition of *recA* amplicons could be a way to directly assess the micro-diversity of agrobacterial populations using cloning sequencing or pyrosequencing approaches. The later methodology was applied to establish the actual field diversity of *Agrobacterium* and to evaluate whether plant genotypes differentially select agrobacteria in their root systems, providing first data upon biotic factors shaping the population structure of agrobacteria (Chapter 2-4).

Part 1: Bibliography

Chapter 1-1

Rapid and efficient methods to isolate, type strains and determine species of *Agrobacterium* spp. in pure culture and complex environments.

Malek Shams, Tony Campillo, Céline Lavire,
Daniel Muller, Xavier Nesme and Ludovic Vial
*Université de Lyon, Ecologie Microbienne Lyon,
UMR CNRS 5557, USC INRA 1193
France*

Published in:

Biochemical Testing

Edited by Jose C. Jimenez-Lopez

Summary

In this book chapter, we reviewed: i) the current state of the agrobacterial taxonomy, ii) methods presently available for isolation of agrobacteria from complex environments, iii) methods for a rapid presumptive genus determination by minimal chemical tests, iiiv) methods to determine species and type strains with marker genes, at both the chromosome and the Ti plasmid levels with applications to the epidemiological study of a model crown gall outbreak¹; v) methods to detect agrobacteria and Ti plasmids in complex environments.

The full paper is given in annexe 1.

¹ exemplified by works of Ponsonnet (1994) Structure génétique des populations d'*Agrobacterium*. Apport à l'écologie des plasmids Ti. Thèse de doctorat, Université Claude Bernard Lyon 1.

Chapter 1-2

Detection and identification methods and new tests as developed and used in the framework of COST873 for bacteria pathogenic to stone fruits and nuts: Tumorigenic *Agrobacterium* spp.

T. Campillo^{1,2}, C. Lavire¹, M. Shams¹, J.F. Pothier³ and J. Pulawska⁴

¹ Université de Lyon, Université Lyon 1, CNRS, INRA, Laboratoire Ecologie Microbienne Lyon, UMR 5557, USC 1193, F-69622 Villeurbanne, France

² Université de Lyon, Université Lyon 1, CNRS; INSA de Lyon, Bayer Crop Science, UMR 5240, Laboratoire Microbiologie, Adaptation et Pathogénie, F-69622 Villeurbanne, France

³ Agroscope Changins-Wädenswil ACW, Plant Protection Division, Schloss 1, 8820 Wädenswil, Switzerland

⁴ Research Institute of Horticulture, Pomology Division, ul. Pomologiczna 18, 96-100 Skierniewice, Poland

Summary

Crown gall, caused by bacteria belonging to different species of the genus *Agrobacterium*, is one of the most serious diseases affecting nursery production of fruit trees and nuts. From a practical point of view, determination if the soil of fields designated for nursery plantations is free from tumour-inducing agrobacteria is very important. During the infection process, after transfer of bacterial DNA to the plant cell, the presence of bacteria is no longer required for gall development. Therefore, infections caused by *Agrobacterium* remain difficult to detect although many methods are available for diagnosis of crown gall and identification of agrobacteria. In the present minireview, methods for isolation, identification and detection of tumorigenic agrobacteria developed within COST873 are described and ready-to-use protocols based on both classical and molecular methods are provided.

The full paper is given in annexe 2.

Chapter 1-3

Agrobacterium biodiversity studies

First and as mentioned in chapter 1.1, *Agrobacterium* are soil and rhizospheric bacteria. The population size of agrobacteria in soil has been estimated to vary between 10^3 and 10^7 CFU/g (Bouzar and Moore, 1987; Bouzar et al., 1993; Burr et al., 1987; Schroth et al., 1971). Additionally, seasonal fluctuations of the total agrobacterial population were observed, with higher population size during spring and summer than during fall and winter (Krimi et al., 2002). Apart from diseased foci, the great majority of soil agrobacteria are not pathogenic (Bouzar and Moore, 1987; Vogel et al., 2003; Pulawska and Sobiczewski, 2005; Costechareyre et al., 2010; Shams et al., 2012). However some soils were found to be permissive to agrobacteria and to crown gall in turn. These soils allowed the long term persistency of pathogenic agrobacteria for several decades, rendering the soils inappropriate for the culture of susceptible crops (Krimi et al., 2002; Rhouma et al., 2007).

Agrobacterium can also be isolated from areas never exploited agriculturally. Bouzar and Moore (1987) reported biovar 2 strains isolated in natural oak savanna and tallgrass prairie. D'Hondt et al. (2004) found bacteria closely related to *A. tumefaciens* complex (biovar 1) in deeply buried sediments of the eastern Pacific Ocean, while Süss et al. (2006) showed that bacteria of the same taxon —with a *rrs* closely related to those of G2 and G9 of *A. tumefaciens*²— are widespread in Mediterranean sediments, and could be isolated from up to 200,000-year-old marine sapropels. Remarkably, members of *Agrobacterium* were also isolated from human clinical samples. *Agrobacterium* infections in humans are frequently associated with the use of plastic catheters or with immunocompromising diseases like HIV (Edmont et al., 1993; Landron et al., 2002). None of these strains carried a Ti/Ri plasmid. Studies on these strains showed that they belonged to *A. pusense* —genomovar G2 of the *A. tumefaciens* complex³— (Panday et al., 2011).

Diversity of agrobacteria is a general question, which can be analyzed from different point of views, such as genomic or plasmidic, genomic and plasmidic together, or diversity studies regarding host specificity. In chapter 1.1 and 1.2 different techniques to isolate, type and discriminate strains of *Agrobacterium* have been presented. Some techniques are available and are used since many years to assess agrobacterial diversity, others are quite new. Depending on classification method used, studies about *Agrobacterium* can consider strain, species or biovar level and are useful to answer different kind of questioning. Current techniques are divided into

² taxonomic indications determined by ourself using nucleotide accessions provided in the cited paper.

³ id.

those that grossly identify agrobacteria at the biovar level and those that operate very fine discriminations at the strain/isolate level. We asked which technique was the most used and at which taxonomic level *Agrobacterium* diversity was mostly studied. Table 1-3-1 summarized methods used and taxonomic assignation obtained from some of the publications dealing with *Agrobacterium* biodiversity studies. Hereafter some interesting results obtained using those methods are highlighted.

Table 1-3-1. *Agrobacterium* diversity studies and taxonomic level of identification.

Host Plant or Soil	Work	Technique	Taxonomic determination ^o
Grapevines <i>Vitis Vinifera</i>	Otten et al., 1996 Ridé et al., 2000 Szegedi and Bottka, 2002 Kawaguchi et al., 2005 Kawaguchi et al., 2008 Palacio-Bielsa et al., 2009 Kawaguchi, 2011	16S-23S intergenic spacer region Phenotypic tests, Plasmid typing PCR on; Polygalacturonase primers and PTi Multiplex PCR (16S rDNA+VirC1-VirC2) rep-PCR, MLSA, Serological tests 16S rRNA sequencing, PCR on Ti Plasmid MLSA	Strain Biovar Biovar Biovar Strain Strain Strain
Poplar Ppoulus (<i>Leuce</i> section)	Nesme et al., 1987 Michel et al., 1990 Nesme et al., 1992	PCR-RFLP of the 16S+ITS serological, biochemical, DNA-DNA hybridization DNA-DNA hybridization	Strain Strain Strain
Fruit trees	Michel et al., 1990 Albiach and Lopez, 1992 Pulawska et al., 1998 Ridé et al., 2000 Peluso et al., 2003 Pulawska et al., 2006 Pulawska et al., 2012	serological, biochemical, DNA-DNA hybrid plasmid profile analysis Plasmid profile analysis Phenotypic tests, Plasmid typing Physiological, biochemical ,PCR-16S + IGS ribosomal Physiopathological, biochemical & molecular analyses RAPD, gyrB gene sequencing 16S rRNA gene sequence, DNA-DNA hybridization, housekeeping gene sequences	Strain Strain Plasmid Δ Biovar Strain Biovar Strain
Chrysanthemum <i>Chrysanthemum indicum</i>	Miller, 1975 Bush and Pueppke, 1991 Vaudequin-Dransart et al., 1995 Ogawa et al., 2000 Kawaguchi et al., 2005	Morphologically, Physiologically, Pathogenicity Biochemical tests vir and T-DNA probes, biochemical test physiological & biochemical tests Multiplex PCR (16S rDNA+VirC1-VirC2)	Biovar Biovar Strain Biovar Biovar
weeping fig <i>Ficus benjamina</i>	Bouzar et al., 1995 Vaudequin-Dransart et al., 1995 Zoina et al., 2001 Raio et al., 2004	Biochemical tests & 16S rRNA vir and T-DNA probes, biochemical test morphological,physiological phytopathological PCR-RFLP on 16S rRNA, (16S+IGS), Ti plasmid	Species Strain Biovar Strain
Roses	Marti et al., 1999 Pionnat et al., 1999 Kawaguchi et al., 2005	PCR on vir region, Phenotypic tests Physiological, PCR-RFLP on 16S rRNA Multiplex PCR (16S rDNA+VirC1-VirC2)	Biovar Biovar Biovar
Pecan <i>Carya illinoensis</i>	Bouzar et al., 1983	Biochemical tests, Pathogenicity assays	Biovar
Lippia	Unger et al., 1985	Biochemical tests	Biovar
Melon	Kawaguchi et al., 2005	Multiplex PCR (16S rDNA+VirC1-VirC2)	Biovar
Soil	Unger et al., 1985	Biochemical tests	Biovar
Soil	Bouzar and Moore, 1987	Biochemical tests	Biovar
Soil	Nesme et al., 1987	PCR-RFLP of the 16S+ITS	Strain

Soil	Bouzar et al., 1993	Pathogenicity assays, Biochemical tests, RFLP	Biovar
Soil	Pionnat et al., 1999	Physiological, PCR-RFLP on 16S rRNA	Biovar
Soil	Ogawa et al., 2000	physiological & biochemical tests	Biovar
Soil	Raio et al., 2004	PCR–RFLP of 16S rRNA, (16S+IGS), Ti plasmid	Strain
Soil	Pulawska et al., 2005	semi-nested PCR and sunflower biotest	Biovar
Soil	Kawaguchi et al., 2005	Multiplex PCR (16S rDNA+VirC1-VirC2)	Biovar
Soil	Palacio-Bielsa et al., 2009	16S rRNA sequencing, PCR on Ti Plasmid	Strain
Soil	Pulawska et al., 2012	RAPD, gyrB gene sequencing 16S rRNA gene sequence, DNA–DNA hybridization, housekeeping gene sequences	Strain

Biovar 1 / biovar 2 and genomic studies.

Puławska et al. (2012) analyzed the genetic diversity of 47 strains of *Agrobacterium* originating from different host plants and geographical locations in Poland, together with 12 strains from other countries. By using PCR-RFLP of a flanking part of *gyrB* and *parE* genes, *gyrB* sequencing and randomly amplified polymorphic DNA (RAPD) techniques, they showed that the majority of agrobacteria isolated in Poland belong to biovar 2. Bacteria of this taxon dominated also among crown gall causal agents on fruit trees in many other areas, for example, in the USA on peach (Alconero 1980), on fruit trees in Greece (Panagopoulos and Psallidas 1973), Hungary (Süle 1978) and Spain (Lopez et al. 1988). However, in some areas and on some hosts, biovar 1 predominated, as with pecan in the USA (Bouzar et al. 1983) or Australia and New Zealand (Keane et al. 1970). Biovar 2 members were found to be homogenous group (Mougel et al., 2002; Popoff et al., 1984) and conversely, the highest diversity was observed within strains of biovar 1 (Puławska et al. 2012).

Biodiversity and genomic studies.

RAPD analysis of tumorigenic agrobacteria isolated from tumors on fruit trees in six nurseries in Poland showed a high heterogeneity of bacteria. No correlation between genetic diversity and geographical origin, or host plant was found. Although some bacteria were isolated from the same nursery and the same species of host plant at the same time, they showed different RAPD patterns. The RAPD patterns generated with 5 RAPD primers were so diverse that it was

possible to find a unique DNA fingerprint for each strain tested (Pulawska et al., 2006). Similar observations were made by Llop et al. (2003). When looking at smaller environments, a high diversity of agrobacterial population could still be found. Differences in agrobacterial populations have indeed been observed on an area as small as 1cm³ of an agricultural soil (Vogel et al., 2003). Out of 55 isolates, 42 different ribotypes dispatched into 3 genomovars (G1, G4 and G8) of the *A. tumefaciens* complex were found by using the 16S + ITS PCR-RFLP (Vogel et al., 2003).

Epidemiological investigations combining genomic and plasmidic studies.

Isolates from a given host species usually show high biochemical and genetic diversity. For example, rose isolates belonged to biovars 1 or 2 with a nearly equal occurrence of succinamopine and nopaline Ti plasmids. They could be further subclassified into several chromosomal and Ti plasmid groups using PCR-RFLP analysis (Pionnat et al., 1999). Chromosomal and Ti plasmid diversity was also observed among poplar isolates (Nesme et al., 1987; Nesme et al., 1992). PCR-RFLP of chromosomal (16S plus the 16S-232S intergenic regions) and Ti plasmid regions of 41 *Agrobacterium* strains from various origins showed that Ti plasmids belonging to the same PCR-RFLP group were found in strains with different chromosomal backgrounds and conversely different Ti plasmids were harbored by strains of the same chromosomal group (Ponsonnet and Nesme, 1994). *A. vitis*, the cause of crown gall on grapevine is an exception. In a study based on 16S-23S intergenic spacer region, a collection of 76 *A. vitis* isolates was analyzed and direct correspondence between chromosomal information and opine type of Ti plasmid was found (Otten et al., 1996). This correspondence suggests a strong epidemiological link between analyzed strains. The same observation was made by Palacio-Bielsa et al. (2009) on Spanish *A. vitis* isolates. It can be hypothesized that the reason for such diverse combinations of chromosomal background and Ti plasmid type in the case of biovar 1 and 2 and *A. larrymoorei* strains, which is the opposite of the more homogenous structure of *A. vitis* population, could be the result of very different pathosystems. *A. vitis* strains are systemic and they have little chance of contact with a large and diverse pool of *Agrobacterium* strains and in consequence, of accessory plasmids. Such a phenomenon is the opposite for biovar 1 and 2 involved in common crown galls on fruit trees, as the tumor surface is largely in contact with soil where a large pool of *Agrobacterium* and plasmid diversity can be found. Furthermore, tumors

are opine-rich environments that must facilitate conjugation and thus modification, of the progeny plasmid content.

Conclusion

Identification at the strain/isolate level was found to be very relevant for epidemiological investigations because finding the same strains in different crown gall outbreaks is a clue that those outbreaks are epidemiologically related. However identification at the strain/isolate level is often too detailed for agro-ecological studies. Indeed, the environmental diversity of agrobacterial strains is so huge that finding the same strains in different environment is a rare event except of course in the case of epidemics as mentioned above. Although new techniques are available, most studies still focus on classical biochemical method that only give results at the biovar level. However this level is not informative enough, especially for biovar 1 that is more diverse than other taxa within *Agrobacterium*. Hence, except for biovar 3 (i.e. *A. vitis*) that is infeodated to vine and found only in association with this plant, biovars 1 and 2 are so common that they are consistently found in almost all environments. Biovar typing is therefore of limited interest to elucidate factors that shape agrobacterial population genetic structures. The reason why it is still often used may be linked to the lack of affordable techniques that could be used to extensively analyze agrobacterial diversity at wide scale and at different taxonomic levels.

References

- Albiach MR, Lopez MM (1992) Plasmid heterogeneity in Spanish isolates of *Agrobacterium tumefaciens* from thirteen different hosts. *Appl Environm Microbiol* 58: 2683-2687
- Alconero R., 1980. Crown gall of peaches from Maryland, South Carolina, and Tennessee and problems with biological control. *Plant Disease* 64: 835-838.
- Bouzar H, Chilton WS, Nesme X, Dessaux Y, Vaudequin V, Petit A, Jones JB, Hodge NC (1995) A new *Agrobacterium* strain isolated from aerial tumors on *Ficus benjamina* L. *Appl Environ Microbiol* 61: 65-73
- Bouzar H, Moore LW (1987) Isolation of different *Agrobacterium* biovars from a natural oak savanna and tallgrass prairie. *Appl Environm Microbiol* 53: 717-721
- Bouzar H., Moore L.W., Schaad N.W., 1983. Crown gall on pecan: A survey of *Agrobacterium* strains and potential for biological control in Georgia. *Plant Disease* 67: 310-312.
- Bouzar H, Quadah D, Krimi Z, Jones JB, Trovato M, Petit A, Dessaux Y (1993) Correlative association between resident plasmids and the host chromosome in a diverse *Agrobacterium* soil population. *Appl Environm Microbiol* 59:1310-1317
- Burr TJ, Katz BH, Bishop AL (1987) Populations of *Agrobacterium* in vineyard and non vineyard soils and grape roots in vineyards and nurseries. *Plant Disease* 71: 617-620

- Bush AL, Pueppke SG (1991a) Characterization of an unusual new *Agrobacterium tumefaciens* strain from *Chrysanthemum morifolium* Ram. Appl Environ Microbiol 57: 2468-2472
- Vaudequin-Dransart V, Petit A, Poncet C, Ponsonnet C, Nesme X, Jones JB, Bouzar H, Chilton WS, Dessaux Y (1995) Novel Ti plasmids in *Agrobacterium* strains isolated from fig tree and *chrysanthemum* tumors and their opinelike molecules. Mol Plant-Microbe Interact 8: 311-321
- Costechareyre, D., et al., Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis. Microb. Ecol., 2010. 60(4): p. 862-872.
- D'Hondt S., Jorgensen B.B., Miller D.J., Batzke A., Blake R., Cragg B.A., 2004. Distributions of microbial activities in deep subseafloor sediments. Science 306: 2216-2221.
- Edmont MB, Riddler SA, Baxter CM, Wicklund BM, Pasculle AW (1993) *Agrobacterium radiobacter*: a recently recognized opportunistic pathogen. Clin Infect Dis 16: 388-391
- Kawaguchi A. 2011. Genetic diversity of *Rhizobium vitis* strains in Japan based on multilocus sequence analysis using the sequences of *pyrG*, *recA* and *rpoD*. J. Gene. Plant. Pathol. 77:299–303.
- Kawaguchi A, Sawada H, Ichinose Y (2008b) Phylogenetic and serological analyses reveal genetic diversity of *Agrobacterium vitis* strains in Japan. Plant Pathol 57:747–753
- Kawaguchi A, Sawada H, Inoue K, Nasu H, 2005. Multiplex PCR for the identification of *Agrobacterium* biovar 3 strains. Journal of General Plant Pathology 71, 54–9.
- Keane PJ, Kerr A, New PB (1970) Crown gall of stone fruit. II. Identification and nomenclature of *Agrobacterium* isolates. Aus J Biol Sci 23: 585-595 Süle S (1978) Biotypes of *Agrobacterium tumefaciens* in Hungary. J Appl Bacteriol 44: 207-213
- Krimi Z, Petit A, Mougél 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
- Landron C, Le Moal G, Roblot F, Grignon B, Bonnin A, Becq-Giraudon B (2002) Central venous catheter-related infection due to *Agrobacterium radiobacter*: a report of 2 cases. Scand J Infect Dis 34: 693-694
- Llop P, Lastra B, Marsal H, Murillo J, Lopez MM (2003) Tracking *Agrobacterium* strains by a RAPD system to identify single colonies from plant tumors. Eur J Plant Pathol 109: 381-389
- Marti R, Cubero J, Daza A, Piquer J, Salcedo CI, Morente C, Lopez MM (1999) Evidence of migration and endophytic presence of *Agrobacterium tumefaciens* in rose plants. Eur J Plant Pathol 105: 39-50
- Michel MF, Brasileiro ACM, Depierreux C, Otten L, Delmotte F, Jouanin L (1990) Identification of different *Agrobacterium* strains isolated from the same forest nursery. Appl Environ Microbiol 56: 3537-3545
- Miller HN (1975) Leaf, stem, crown, and root galls induced in *Chrysanthemum* by *Agrobacterium tumefaciens*. Phytopathology 65: 805-811
- Mougél C, Thioulouse J, Perrière G, Nesme X. 2002. A mathematical method for determining genome divergence and species delineation using AFLP. Int. J. Syst. Evol. Microbiol. 52:573–586.
- Nesme X, Michel MF, Digat B (1987) Population heterogeneity of *Agrobacterium tumefaciens* in galls of *Populus L.* from a single nursery. Appl Environ Microbiol 53: 655-659
- Nesme X, Ponsonnet C, Picard C, Normand P (1992) Chromosomal and pTi genotypes of *Agrobacterium* strains isolated from *Populus* tumors in two nurseries. FEMS Microbiol Ecol 101: 189-196

- Ogawa Y, Ishikawa K, Mii M (2000) Highly tumorigenic *Agrobacterium tumefaciens* strain from crown gall tumors of *chrysanthemum*. Arch Microbiol 173:311-315
- Otten L, De Ruffray P, Momol EA, Momol MT, Burr TJ (1996) Phylogenetic relationships between *Agrobacterium vitis* isolates and their Ti plasmids. Mol Plant-Microbe Interact 9: 782-786
- Palacio-Bielsa A., González-Abolafio R., Álvarez B., Lastra B., Cambra M.A., Salcedo C.I., López M.M., Penyalver R., 2009. Chromosomal and Ti plasmid characterization of tumorigenic strains of three *Agrobacterium* species isolated from grapevine tumours. Plant Pathology 58: 584-593.
- Panagopoulos CG, Psallidas PG (1973) Characteristics of Greek Isolates of *Agrobacterium tumefaciens* (E. F. Smith & Townsend) Conn. J Appl Bacteriol 36: 233-240
- Panday D, Schumann P, Das SK. 2011. *Rhizobium pusense* sp. nov., isolated from the rhizosphere of chickpea (*Cicer arietinum* L.). Int. J. Syst. Evol. Microbiol. 61:2632–2639.
- Peluso R, Raio A, Morra F, Zoina A (2003) Physiological, biochemical and molecular analyses of an Italian collection of *Agrobacterium tumefaciens* strains. Eur J Plant Pathol 109: 291-300
- Pionnat S., Keller H., Hélicher D., Bettachini A., Dessaux Y., Nesme X., Poncet C. 1999. Ti plasmids from *Agrobacterium* characterize rootstock clones that initiated an outspread of crown gall disease in Mediterranean countries. Appl. Environ. Microbiol. 65:4197-4206.
- Ponsonnet, C. and X. Nesme, Identification of *Agrobacterium* strains by PCR-RFLP analysis of pTi and chromosomal regions. Arch. Microbiol., 1994. 161(4): p. 300-309.
- Popoff MY, Kersters K, Kiredjian M, Miras I, Coynault C (1984) Position taxonomique de souches de *Agrobacterium* d'origine hospitalière. Ann Microbiol 135A: 427-442
- Pulawska J, Sobiczewski P (2005) Development of a semi-nested PCR based method for sensitive detection of tumorigenic *Agrobacterium* in soil. J Appl Microbiol 98: 710-721
- Pulawska J., Sulikowska M., Sobiczewski P., 2006. Genetic diversity of *Agrobacterium* strains determined by RAPD and *gvrB* gene sequence analysis. In: Elphinstone J., Weller
- Puławska J, Willems A, De Meyer SE, Süle S. 2012. *Rhizobium nepotum* sp. nov. isolated from tumors on different plant species. Syst. Appl. Microbiol. 35:215–220.
- Raio A, Peluso R, Nesme X, Zoina A (2004) Chromosomal and plasmid diversity of *Agrobacterium* strains isolated from *Ficus benjamina* tumors. Eur J Plant Pathol 110: 163-174
- Rhouma A., Boubaker A., Nesme X., Dessaux Y. 2007. Plasmid and chromosomal diversity of a Tunisian collection of *Agrobacterium tumefaciens* strains. Tunisian journal of plant protection 1:73-84.
- Ridé M, Ridé S, Petit A, Bollet C, Dessaux Y, Gardan L (2000) Characterization of plasmid-borne and chromosome-encoded traits of *Agrobacterium* biovar 1, 2, and 3 strains from France. Appl Environ Microbiol 66: 1818-1825
- Schroth M.N., Weinhold A.R., McCain A.H., Hildebrand D.C., Ross N., 1971. Biology and control of *Agrobacterium tumefaciens*. Hilgardia 40: 537-552.
- Shams M, Campillo T, Lavire C, Muller D, Nesme X, Vial L. 2012. Rapid and Efficient Methods to Isolate, Type Strains and Determine Species of *Agrobacterium* spp. in Pure Culture and Complex Environments. In Dr. Jose C. Jimenez-Lopez (ed.)-Biochemical Testing. InTech. DOI: 10.5772/2250
- Süss J., Schubert K., Sass H., Cypionka H., Overmann J., Engelen B., 2006. Widespread distribution and high abundance of *Rhizobium radiobacter* within Mediterranean subsurface sediments. Environmental Microbiology 8: 1753-1763.

- Szegedi E, Bottka S (2002) Detection of *Agrobacterium vitis* by polymerase chain reaction in grapevine bleeding sap after isolation on a semiselective medium. *Vitis* 41: 37-42
- Thwaites S, R., Parkinson N., Stead D, Saddler G. (eds). Proceedings of the 11th International Conference on Plant Pathogenic Bacteria, Edinburgh 2006: 31-32
- Unger L, Ziegler SF, Huffman GA, Knauf VC, Peet R, Moore LW, Gordon MP, Nester EW (1985) New class of limited-host-range *Agrobacterium* megatumor-inducing plasmids lacking homology to the transferred DNA of a widehost-range, tumor-inducing plasmid. *J Bacteriol* 164: 723-730
- Vogel J., Normand P., Thioulouse J., Nesme X., Grundmann G.L., 2003. Relationship between spatial and genetic distance in *Agrobacterium* spp. in one cubic centimeter of soil. *Applied and Environmental Microbiology* 69: 1482-1487.
- Zoina A., Raio A., Peluso R., Spasiano A., 2001. Characterization of agrobacteria from weeping fig (*Ficus benjamina*). *Plant Pathology* 50: 620-627.

Part 2: Experimental

Chapter 2-1

***Agrobacterium* identification using whole-cell MALDI-TOF mass spectrometry⁴**

⁴ This study was done in collaboration with J. Pothier and B. Duffy in the frame of the European COST 873 action.

Fine typing of agrobacteria strains is an absolute prerequisite to assess the environmental diversity of agrobacteria for epidemiological and agro-ecological investigations especially to determine the epidemic or endemic status of crown gall outbreaks. As a reminding point, current techniques were divided into those who that grossly identify agrobacteria at the biovar level and those that operate very fine discriminations at the strain/isolate level. Methodologies delineating bacteria at the biovar level are not accurate enough for that purposes. Indeed, except for biovar 3 (i.e. *A. vitis*) that is infeodated to vine and therefore found only in association with this plant, biovars 1 and 2 are so common that they are consistently found in almost all environments (Chapter 1-3). Biovar typing is therefore of limited interest to elucidate factors that shape agrobacterial population genetic structures. At the opposite, the identification at the strain/isolate level was found to be very relevant for epidemiological investigations because finding the same strains in different crown gall outbreaks is a clue that those outbreaks are epidemiologically related (Chapter 1-3; Ponsonnet et al., 1994; Pionnat et al., 1999). However, in contrast, identification at the strain/isolate level is often too detailed for agro-ecological studies. Indeed, the environmental diversity agrobacterial strain is generally so huge that finding the same strains in different environment is a rare event except of course in the case of epidemics as described above. In this respect, the MLSA approach developed by Costechareyre (thesis 2009) for agrobacteria is relevant methodology to define the strain, species and genus status of any agrobacterial isolates. However, sequencing of housekeeping genes is labor-intensive, time-consuming, and impractical for high-throughput investigations. Thus, there is still a valuable need of simplified and rapid approach to perform mass agrobacterium identification.

Whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) is an emerging technology (Fig. 2-3-1) in microbial diagnostics (Krader and Emerson, 2004; Sauer et al., 2008). Identity is based on unique mass/charge ratio (m/z) fingerprints of proteins, which are ionized using short laser pulses directed to bacterial cells obtained from a single colony embedded in a matrix. After desorption, ions are accelerated in vacuum by a high electric potential and separated on the basis of the time taken to reach a detector, which is directly proportional to the mass-to-charge ratio of an ion. This technique has been shown to deliver reproducible protein mass fingerprints starting from an aliquot of a single bacterial colony within minutes and without any prior separation, purification, or concentration of samples. Whole-cell MALDI-TOF MS is a reliable technique across broad conditions (e.g., different growth media, cell growth states), with limited variability in mass-peak signatures

within a selected mass range ($2,000 < m/z < 20,000$) that does not affect reliability of identification (Lay, 2001). MALDI-TOF MS profiles primarily represent ribosomal proteins, which are the most abundant cellular proteins and are synthesized under all growth conditions (Ryzhov et al., 2001). However, this approach is not informative as it relies upon the availability of a reference database. Current databases are weighted to clinical taxa (Bessède et al., 2011; Stephan et al., 2010) and only a few references are available for plant-associated and environmental bacteria (Sauer et al., 2008; Rezzonico et al., 2010; Wensing et al., 2010).

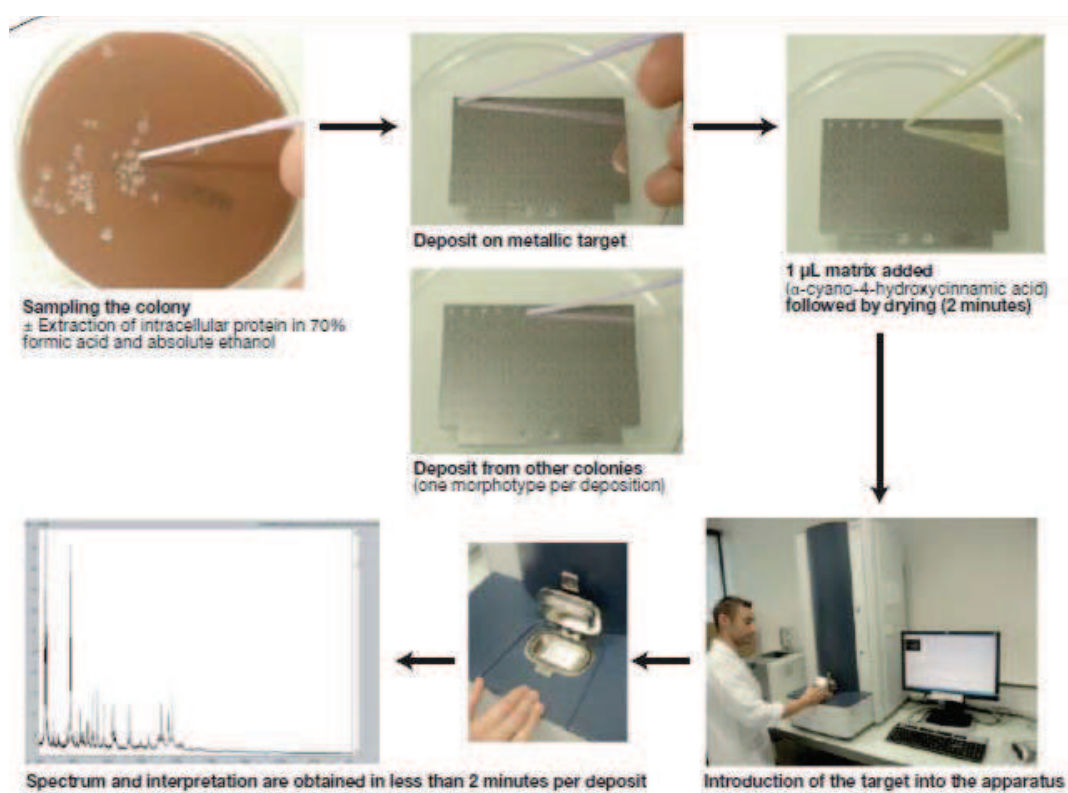


Figure 2-1-1. Sequence of identification by MALDI-TOF mass spectrometry of a bacterial colony in a clinical microbiology laboratory. (Photographs: Olivier Gaillot, Laboratoire de bactériologie-hygiène, CHRU de Lille)

In this chapter, we evaluated the usefulness of MALDI-TOF MS technology as a high throughput tool to characterize and classify agrobacteria. For this aim, strains covering all *Agrobacterium* species known so far as well as outgroup strains belonging to the Rhizobiaceae family were analyzed using the whole cell MALDI-TOF MS approach. To evaluate the quality of the taxonomic assignation done by the method, the MALDI-TOF MS data clustering was compared to the *recA* phylogeny done with the same strains. In addition, we evaluated whether the technique was able to provide information about strain pathogenicity by looking for consistant and remarkable differences between MALDI-TOF MS patterns of a standard pathogenic strain (C58) and its plasmid free derivatives.

Material and Method

Bacterial strains and growth conditions

A total of 114 strains covering all the genomic species richness of the genus *Agrobacterium* known so far as well as 4 outgroup strains belonging to the Rhizobiaceae family were investigated with whole-cell MALDI-TOF MS with 1 to 19 strains per species (Table 2-1-1). A special emphasis was given to *A. vitis*, with a selection of 19 strains done by P. Portier on the basis of their *recA* diversity. Along to the standard pathogenic strain C58, C58 derivative cured of pTiC58 and pAtC58 plasmids or containing other Ti plasmids were included in the analysis (Table 2-1-2). Bacteria were grown on MG agar (Ophel and Kerr, 1990) and LPG agar medium (yeast extract, 5 g/L; Bacto peptone, 5 g/L; glucose, 10 g/L) at 28°C for 42-50 h.

IC MALDI-TOF analysis

Colonies harvested on LPG plates were directly smeared onto target spots of a polished ground steel 48-position MALDI-TOF MS sample target (Industrietechnik mab AG, Basel, Switzerland) with two distinct spot per strain using a blind random spotting design. Each spot was then overlaid with 1 µl of a saturated solution of sinapinic acid (SA – 49508, Sigma-Aldrich, Buchs, Switzerland) in 60% acetonitrile (154601, Sigma-Aldrich, Buchs, Switzerland) – 0.3% trifluoroacetic acid (T6508, Sigma-Aldrich, Buchs, Switzerland), and air-dried for some minutes at room temperature.

Table 2-1-1. Strains and species used in this study.

Strain designation ^a	Other designation(s) ^a	Country ^b	AFLP ^c	<i>recA</i> allele ^d
<i>A. tumefaciens</i> genomovar G1				
ICPB TT111	CFBP 5767/LMG 196	USA	G1	<i>recA</i> -G1-1
S377	CFBP 5768/LMG 326	N/A	G1	<i>recA</i> -G1-1
ATCC 4720	CFBP 5493/LMG 182	USA	G1	<i>recA</i> -G1-1
NCPPB 396	CFBP 5765/LMG 176	Germany	G1	<i>recA</i> -G1-2
CFBP 5771		France	G1	<i>recA</i> -G1-2
S56	CFBP 5491/LMG 321	N/A	G1	<i>recA</i> -G1-3
S4	CFBP 5492/LMG 318	N/A	G1	<i>recA</i> -G1-3
CFBP 5622		France	G1	<i>recA</i> -G1-4
CFBP 2712		France	G1	<i>recA</i> -G1-5
CFBP 2517		France	G1	<i>recA</i> -G1-6
CFBP 7067		France	ND	<i>recA</i> -G1-7
CFBP 7064		France	ND	<i>recA</i> -G1-8
CFBP 7066		France	ND	<i>recA</i> -G1-9
Fb243	CFBP 5479	Italy	G1	<i>recA</i> -G1-10
<i>A. tumefaciens</i> genomovar G2 (<i>A. pusense</i>)				
Bernaerts L2/2/1	CFBP 5875/LMG 146	Belgium	ND	<i>recA</i> -G2-1
Bernaerts M 2/1	CFBP 5876/LMG 147	Belgium	G2	<i>recA</i> -G2-2
CIP 497-74	CFBP 5494/CFBP 2884	France	G2	<i>recA</i> -G2-3
CIP 28-75	CFBP 5495	France	G2	<i>recA</i> -G2-4
CIP 43-76	CFBP 5496	France	G2	<i>recA</i> -G2-5
MKS-03		France	ND	<i>recA</i> -G2-7
<i>A. tumefaciens</i> genomovar G3				
CIP 107443	CFBP 6623/CIP 493-74	France	G3	<i>recA</i> -G3-1
CIP 107442	CFBP 6624/CIP 111-78	France	G3	<i>recA</i> -G3-2
<i>A. tumefaciens</i> genomovar G4 (<i>A. radiobacter</i>)				
B6*	CFBP 2413/LMG 187	USA	G4	<i>recA</i> -G4-1
CIP 67-1	Strain B6 at CIP	USA	G4	<i>recA</i> -G4-1
ATCC 4452	CFBP 5766/LMG 181	USA	G4	<i>recA</i> -G4-1
DC07-012	CFBP 7273	Tunisia	ND	<i>recA</i> -G4-1
ATCC 4718	CFBP 5764/LMG 139	USA	G4	<i>recA</i> -G4-2
ATCC 19358 ^T	CFBP 5522/LMG 140	N/A	G4	<i>recA</i> -G4-2
CFBP 5621		France	G4	<i>recA</i> -G4-2
LMG 340	CFBP 5769/ICPB TT11	USA	G4	<i>recA</i> -G4-2
LMG 62	CFBP 5763/ICPB TR6	N/A	G4	<i>recA</i> -G4-2
CFBP 5627		France	G4	<i>recA</i> -G4-2
CFBP 2514		Spain	G4	<i>recA</i> -G4-2
Kerr 14	CFBP 5761/LMG 15	Australia	G4	<i>recA</i> -G4-3
Hayward 0322	CFBP 5770/LMG 1687	Australia	G4	<i>recA</i> -G4-3
Myr3	CFBP 7128	Tunisia	G4	<i>recA</i> -G4-4
AR9	CFBP 7126	Tunisia	G4	<i>recA</i> -G4-5
CH14	CFBP 7199	Tunisia	ND	<i>recA</i> -G4-6
<i>A. tumefaciens</i> genomovar G5				
CIP 107445	CFBP 6625/CIP 291-77	France	G5	<i>recA</i> -G5-1
CIP 107444	CFBP 6626/CIP 120-78	France	G5	<i>recA</i> -G5-2
<i>A. tumefaciens</i> genomovar G6				
NCPPB 925	CFBP 5499/LMG 225	South Africa	G6	<i>recA</i> -G6-1

Zutra F/1	CFBP 5877/LMG 296	Israel	G6	<i>recA</i> -G6-1
AR125		Tunisia	G6	<i>recA</i> -G6-1
<i>A. tumefaciens</i> genomovar G7				
RV3	CFBP 5500/LMG 317	N/A	G7	<i>recA</i> -G7-1
NCPPB 1641	CFBP 5502/LMG 228	UK	G7	<i>recA</i> -G7-2
Zutra 3/1	CFBP 6999/LMG 198	Israel	G7	<i>recA</i> -G7-3
DC07-042	CFBP 7274	Tunisia	ND	<i>recA</i> -G7-3
AR78 OR GF78	CFBP 7001	Tunisia	G7	<i>recA</i> -G7-4
AR29	CFBP 7125	Tunisia	G7	<i>recA</i> -G7-5
AR94	CFBP 7127	Tunisia	G7	<i>recA</i> -G7-6
AR92	CFBP 7124	Tunisia	G7	<i>recA</i> -G7-7
AR118	CFBP 7129	Tunisia	G7	<i>recA</i> -G7-8
CH6	CFBP 7198	Tunisia	ND	<i>recA</i> -G7-9
DC07-029	CFBP 7200	Tunisia	ND	<i>recA</i> -G7-10
<i>A. tumefaciens</i> genomovar G8 (<i>A. fabrum</i>)				
ICPB TT9	CFBP 5504/LMG 64	USA	G8	<i>recA</i> -G8-1
T37	CFBP 5503/LMG 332	USA	G8	<i>recA</i> -G8-1
J-07	CFBP 5773	France	G8	<i>recA</i> -G8-1
Mushin 6	CFBP 6550/LMG 201	Australia	G8	<i>recA</i> -G8-1
LMG 75	CFBP 6549	USA	G8	<i>recA</i> -G8-2
LMG 46	CFBP 6554	USA	G8	<i>recA</i> -G8-3
C58*	CFBP 1903/LMG 287	USA	G8	<i>recA</i> -G8-4
AW137	LMG R-10181	N/A	G8	<i>recA</i> -G8-4
DC07-004	CFBP 7272	Tunisia	ND	<i>recA</i> -G8-4
MKS-01		France	ND	<i>recA</i> -G8-5
<i>A. tumefaciens</i> genomovar G9				
Hayward 0362	CFBP 5507/LMG 26	Australia	G9	<i>recA</i> -G9-1
Hayward 0363	CFBP 5506/LMG 27	Australia	G9	<i>recA</i> -G9-2
<i>A. tumefaciens</i> genomovar G13				
PGF 01-22	CFBP 6927	France	G13	<i>recA</i> -G13-1
<i>A. tumefaciens</i> genomovar G14 (<i>A. nepotum</i>)				
C3.4.1		Poland	G14	<i>recA</i> -G14-1
<i>A. larrymoorei</i>				
AF 3.10 ^T	CFBP 5473/LMG 21410	USA	<i>A. larrymoorei</i>	<i>recA</i> -Ala-1
Fb33	CFBP 5476	Italy	ND	<i>recA</i> -Ala-2
Fb23	CFBP 5475	Italy	ND	<i>recA</i> -Ala-2
Fb72	CFBP 5477	Italy	ND	<i>recA</i> -Ala-2
LT4	CFBP 5480	Italy	ND	<i>recA</i> -Ala-3
NL7	CFBP 5481	The Netherlands	ND	<i>recA</i> -Ala-4
NL25	CFBP 5482	The Netherlands	ND	<i>recA</i> -Ala-4
NL360D	CFBP 5483	The Netherlands	ND	<i>recA</i> -Ala-4
<i>A. rubi</i>				
LMG 17935 ^T	CFBP 5509/CFBP 6448	USA	<i>A. rubi</i>	<i>recA</i> -Aru-1
TR2	CFBP 999	USA	ND	<i>recA</i> -Aru-1
LMG 294		N/A	ND	<i>recA</i> -Aru-2

<hr/>				
<i>A. vitis</i>				
K309 ^T	CFBP 5523/LMG 8750	Australia	<i>A. vitis</i>	<i>recA-Avi-1</i>
AV08.155	CFBP 7224	France	ND	<i>recA-Avi-1</i>
AV07.004	CFBP 7196	France	ND	<i>recA-Avi-2</i>
AV08.011	CFBP 7228	France	ND	<i>recA-Avi-3</i>
AV07.006	CFBP-LABO12945	France	ND	<i>recA-Avi-4</i>
S4	CFBP2660	Hungary	ND	<i>recA-Avi-5</i>
AV07.175	CFBP 7217	France	ND	<i>recA-Avi-6</i>
AV07.147	CFBP 7216	France	ND	<i>recA-Avi-7</i>
AV08.120	CFBP 7222	France	ND	<i>recA-Avi-8</i>
AV08.017	CFBP 7218	France	ND	<i>recA-Avi-9</i>
AV08.022	CFBP 7219	France	ND	<i>recA-Avi-10</i>
AV08.007	CFBP-LABO13296	France	ND	<i>recA-Avi-11</i>
AV08.013	CFBP-LABO13297	France	ND	<i>recA-Avi-12</i>
AV08.214	CFBP-LABO13298	France	ND	<i>recA-Avi-13</i>
AV09.020	CFBP-LABO13300	France	ND	<i>recA-Avi-15</i>
AV07.018	CFBP-LABO13295	France	ND	<i>recA-Avi-16</i>
AV08.172	CFBP-LABO13074	France	ND	<i>recA-Avi-17</i>
AV08.260	CFBP-LABO13299	France	ND	<i>recA-Avi-20</i>
<i>Agrobacterium</i> sp.				
NCPPB 1650	CFBP 4470	South Africa	ND	<i>recA-1650-1</i>
CH11		Poland	ND	<i>recA-CH11</i>
81K		Poland	ND	ND
<i>Rhizobium rhizogenes</i>				
LMG 150 ^T	CFBP 2408/CFBP 5520	N/A	<i>R. rhizogenes</i>	<i>recA-Rrh-1</i>
K84	CFBP 1937/LMG 138	Australia	<i>R. rhizogenes</i>	<i>recA-Rrh-1</i>
<i>Rhizobium etli</i> CFN42 ^T	LMG 17827	Mexico	ND	<i>recA-Ret-1</i>
<i>Ensifer meliloti</i> 1021		N/A	ND	<i>recA-Eme-1</i>
<i>Rhizobium undicola</i> ORS 991	LMG 11874	Senegal	ND	<i>recA-Aun-1</i>
<hr/>				

^a genomovar G1 to G9 were named according to the nomenclature of Mougél et al.(2002) Culture collections providing strains are abbreviated in the strain names as CFBP (Collection Française de Bactéries associées aux Plantes), CIP (Collection de l'Institut Pasteur), ICPB (International Collection of Phytopathogenic Bacteria), LMG (Collection of the Laboratorium voor Microbiologie en Microbiele Genetica), NCPPB (National Collection of Plant Pathogenic Bacteria), NCPPB (National Collection of Plant Pathogenic Bacteria). Superscript T following strain designation indicate the type strain of a species. ^b N/A, information not available. ^c Genomospecies determined by AFLP as described by Portier et al. (2006). ND, not determined. ^d Allele code as established by Costechareyre et al. (2010)+*A. vitis* + Acc. Numbers

Protein mass fingerprints were obtained using an AXIMA Confidence MALDI-TOF Mass Spectrometry (Shimadzu-Biotech Corp., Kyoto, Japan), with a detection in the linear, positive mode at a laser frequency of 50 Hz and within a mass range from 2,000-20,000 Daltons. The acceleration voltage was 20 kV, and the extraction delay time was 200 ns. MALDI-TOF MS spectra were the average of at least four replicate measurements using slides prepared at least in two different instances. Protein mass fingerprints were determined using 100 different laser

direction and at least 10 laser shots per direction, then averaged and processed to provide a single raw spectrum using the Launchpad v2.8 software (Shimadzu-biotech, Kyoto, Japan). The software was also used for peak processing using the following settings: the Advanced Scenario was chosen from the Parent Peak Cleanup menu, Peak Width was set to 80 channels, Smoothing Filter Width to 50 channels, Baseline Filter Width to 500 channels and Threshold Apex was chosen as the peak detection method. For the Threshold Apex Peak Detection, the Threshold Type was set to Dynamic, the Threshold Offset to 0.020 mV and the Threshold Response Factor

Table 2-1-2. C58 derivatives used to test plasmid content effect upon MALDI-TOF MS spectra.

C58 derivatives	Ti plasmid content	At plasmid content
C58	pTiC58	pAtC58
C58C1	No Ti plasmid	pAtC58
CFBP 1896	pTiH100alpha	pAtC58
CFBP 1897	pTiR10C21	pAtC58
CFBP 1898	pTiA66	pAtC58
GMI 9023	No Ti plasmid	NoAt plasmid
UIA5	No Ti plasmid	No At plasmid
AB 150	pTiC58	pAtC58
AB 152	pTiC58	pAtC58
AB 153	pTiC58	pAtC58
AW137	pTiC58	pAtC58
DC07-004	pTiC58	pAtC58

AW137 and DC07-004 are strains found to be exactly similar to C58 by MLSA using 8 housekeeping genes, received respectively as a laboratory strain from Ghent University or as a recent isolate from a Tunisian soil (Costechareyre et al., 2010).

to 1.2. Each target plate was externally calibrated using spectra of reference strain *Escherichia coli* DH5 alpha. Processed spectra were reduced to the 150 most intensive peaks with an *ad hoc* PERL script written by J.F. Pothier, V. Pflüger, B. Duffy (unpublished), and exported as peak lists with *m/z* values and signal intensity in the ASCII format. A binary matrix was then generated using the SARAMISTTM (Spectral ARchive And Microbial Identification System, AnagnosTec, Potsdam-Golm, Germany) SuperSpectrum tool. Peak lists were trimmed to a mass range of 2-20 kDa. Peak lists were binned and average masses were calculated using the SARAMISTTM SuperSpectrum tool with an error of 800 ppm. Final consensus spectra for each strain were generated with a customized VBA script (J.F. Pothier, V. Pflüger, B. Duffy, unpublished) by eliminating masses present within fewer than half of the replicate

measurements. Multivariate cluster analysis using the UPGMA algorithm with Dice coefficient was performed in PAST v2.14 (Hammer et al., 2001) and the resulting dendrogram was visualized with FigTree v1.3.1.

For clustering, a dendrogram was constructed using strain consensus spectra selecting peaks that were found in at least half of 4 independent measurements, with the weighted pair-group average (UPGMA) clustering algorithm with the Dice coefficient. Similarity distances are expressed as percentage. The tree topology robustness was evaluated using the 1000 bootstrap resembling of peak values.

Results

Relevant growth medium for efficient agrobacteria detection using MALDI-TOF MS

To assess the growth condition influence on spectra, we compared results obtained using two standard growth media. There was a good cell multiplication for all strains on MG agar in both 24 and 48 hours of culture, while *Agrobacterium vitis* as well as *Rhizobium* strains had a very slower growth using LPGA (data not shown). Spectra were however clearer and more readable using LPG than MG medium, likely because polysaccharides intensely produced with the MG medium were blurring the peak spectra (data not shown). The best informative spectra that allowed unambiguous *Agrobacterium* identification used in further studies were obtained using 24-hour old colonies grown in LPG plate.

***Agrobacterium* species identification using MALDI-TOF MS**

With the conditions described above, a typical MALDI-TOF MS spectrum of agrobacteria contained about 150 peaks between 2,000 and 20,000 Da, with the highest intensity peaks found between 4,000 and 20,000 Da. Spectra displayed high overall levels of similarity within species (not shown), but slightly different patterns between different species as exemplified Fig. 2-1-2. MALDI-TOF MS spectrum similarity analysis showed that agrobacteria belonging to the same species gathered in single clusters in all instances (Fig. 2-1-3). Rare mis-classifications occurred however. In all instances, they concerned mis-named strains. A definitive and fully conclusive dendrogram is thus currently in progress and will be delivered soon (draft given in Fig. 2-1-3'). Result overall showed the potential interest of the methodology to rapidly identify agrobacterial isolates. However, within homogenous species clusters, strains with the same *recA* allele were

frequently found to be intermingled with other allele, showing that the method is inappropriate in delineating agrobacteria at the strain/isolate level.

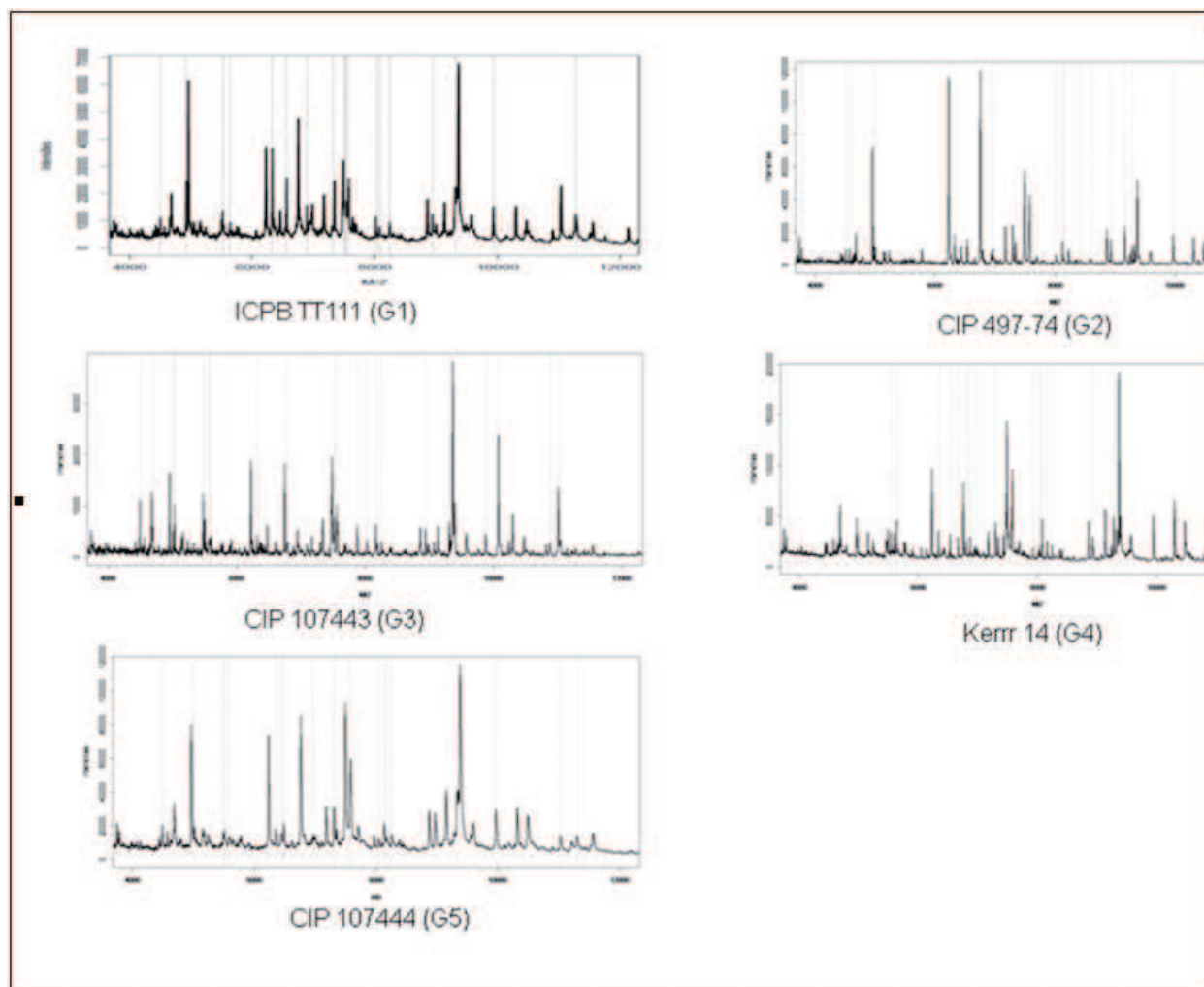


Figure 2-1-2. Typical MALDI-TOF MS spectra of *Agrobacterium* species. *A. tumefaciens* ICPB TT111 (G1), *A. tumefaciens* CIP 497-74 (G2), *A. tumefaciens* CIP 107443 (G3), *A. tumefaciens* Kerr 14 (G4) and *A. tumefaciens* CIP 107444 (G5). Absolute intensity of ions (mV) is shown on the y axis and the masses (in Daltons) of the ions are shown on the x axis. The m/z values represent mass-to-charge ratios. The biomarker masses identified in this work are indicated with vertical dashed lines.

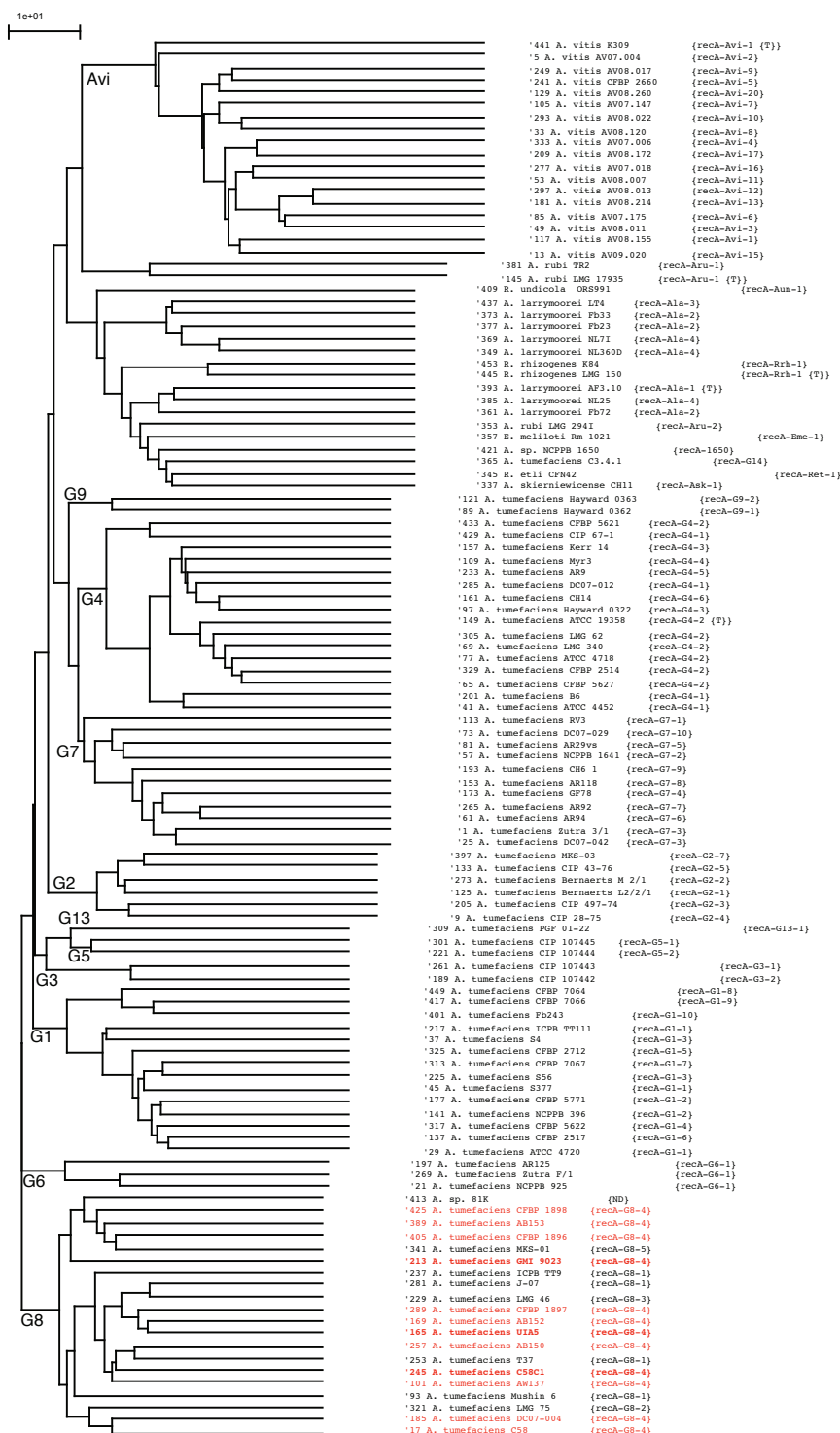


Figure 2-1-3. Similarity analysis of whole cell MALDI-TOF MS protein mass fingerprints for strains of *Agrobacterium* species and related taxa. C58 and C58 derivatives were in red, and Ti plasmid free C58 derivative were in bold. G1-G9, *A. tumefaciens* genomic species.



Figure 2-1-3'. Similarity analysis of whole cell MALDI-TOF MS protein mass fingerprints for strains of *Agrobacterium* species and related taxa. C58 and C58 derivatives were in red, and Ti plasmid free C58 derivative were in bold. G1-G9, *A. tumefaciens* genomic species. Corrected version.

Plasmid content assessment using MALDI-TOF MS

To discriminate pathogenic agrobacteria from non-pathogenic bacteria (i.e. Ti plasmid containing vs. Ti plasmid free), we used different derivatives of the standard pathogenic strain C58. Derivatives have an identical chromosomal background but differ in their plasmid content (Table 2-1-2). MALDI-TOF MS spectra of Ti plasmid free derivatives were however intermingled those of C58 (Fig. 2-1-3). Analyses were also performed with agrobacteria incubated with ferulic acid in order to induce Ti plasmid virulence genes and thus stimulate the production of typical Ti plasmid protein, but no more marked differences were obtained (data not shown). Taken together, these results indicated that with the current experimental conditions, the MALDI-TOF MS technology did not allow the identification of pathogenic strains.

Discussion

Whole cell fingerprinting by MALDI-TOFMS is an emerging and powerful technique for rapid identification of bacteria. In the present study, the IC MALDI-TOF MS analysis was able to separate most *Agrobacterium* species (Fig. 2-1-3). The segregation showed the same level of accuracy as *recA* sequencing, DNA-DNA hybridization or AFLP methodologies and was more sensitive than either morpho-biochemical tests or 16S rRNA gene sequencing approaches. MALDI-TOF MS is set up generally at a range between 2 and 20 kDa of biomolecules in order to obtain the stable and strong signal to noise ratio. This size range is dominated by ribosomal proteins which ionize well and provide accurate spectra (Stevenson et al., 2010; Wieser et al., 2012). Ribosomal proteins are less influenced by culture conditions allowing MALDI-TOF MS to be consistently used to differentiate some bacterial species (Claydon et al. 1996; Demirev et al. 1999; Fenselau and Demirev 2001). Interestingly, *Agrobacterium* species not distinguishable using 16S rRNA sequencing technique (Mougel et al., 2002) were accurately identified and classified in closed but completely separated clades as exemplified with genomovars G8 and G6 member. By the same way, genomovars G4 and G2 that have shown identical 16S rRNA sequences are grouped in close species clusters using the MALDI-TOF MS technique. The methodology still require improvements for some species such as *A. rubi* and *A. larrymoorei*, but it was found to be a promising tool for mass species identification of agrobacteria.

The MALDI-TOF MS technique has been reported to fail in identifying species that have very similar ribosomal proteins such as *Shigella* spp. and *Escherichia coli*, or *Streptococcus pneumoniae* and members of the *S. oralis/mitis* group (Wieser et al., 2012). In our opinion, it is

likely because the later so-called different species are not readily different genomic species, contrary to the bona fide species of *Agrobacterium* analyzed in the present study. The present work offers us a great chance to utilize the MALDI-TOF MS technology for the species identification of single isolates and indeed for extensive agrobacterial population studies.

The protein mass fingerprints can be used to generate a Superspectrum by identifying specific mass peaks for each *Agrobacterium* species. Comparison of these protein mass fingerprints allows the definition of a set of markers present in fingerprints of all strains from a given species. Even though subsets of masses are shared within different *Agrobacterium* species, a combination of discriminatory signals is thought to provide unique *Agrobacterium* species signatures. As ribosomal protein mass can be inferred from the bioinformatic analysis of sequenced genomes, such an analysis is currently attempted by our collaborators (J. Pothier and Valentin Pflüger), using genome sequence of members of all *A. tumefaciens* species done by the Agrobacterscope(http://www.cns.fr/agc/microscope/about/collabprojects.php?P_id=51&wwwpkgdb=8f9fc33c35a024a8f6bf12ea71d7ff4d).

The MALDI-TOF MS technique is however not precise enough to discriminate strains within a species. Furthermore, and unfortunately, we did not find specific spectra for Ti plasmid even in presence of Ti plasmid protein production stimulator (Ferulic acid). We nevertheless wonder if the methodology can still be improved for this purpose.

In conclusion, results presented in this work clearly showed that MALDI-TOF MS is a reliable and rapid method for agrobacteria species identification, comparable to what is obtainable with housekeeping gene sequencing. The technology is moreover more rapid and affordable than MLSA. Indeed, by using this methodology, it is possible to prepare over 100 bacterial cell extracts and to analyze them in an 8 h working day, a number which could significantly be increased when coupled to process automation. Cell extracts themselves can be stored in ethanol for months without significant loss in spectrum quality. This in turn could allow pooling much larger numbers of samples. MALDI-TOF MS thus opens a new and very useful way for diversity and ecological studies applicable to analysis of large populations of isolates allowing the differentiation of strains, species and genera of agrobacteria. Finally, we are hopeful to use MALDI-TOF MS on meta-proteome samples obtained from different environments instead of just pure isolate identification, as it had been used by Ziegler et al. (2012). It would then allow us

to analyze composite communities of bacteria and understand their microbial construction in a near future.

References

- Aujoulat, F., et al., Multi-locus sequence-based analysis delineated a clonal population of *Agrobacterium (Rhizobium) radiobacter / tumefaciens* from human origin. J. Bacteriol., 2011. **In press**: p. DOI:10.1128/JB.00107-11.
- Bessède, E., et al., Identification of *Campylobacter* species and related organisms by matrix associated laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry. Clin. Microbiol. Infect., 2011. **17**: p. DOI:10.1111/j.1469-0691.2011.03468.x.
- Croxatto, A., et al., Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiology Reviews. 2012. **36**:380–407
- Costechareyre, D., et al., Rapid and efficient identification of *Agrobacterium* species by recA allele analysis. Microb. Ecol., 2010. **60**(4): p. 862-872.
- Krader, P. and D. Emerson, Identification of archaea and some extremophilic bacteria using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. Extremophiles, 2004. **8**(4): p. 259-268.
- Krimi, Z., et al., Seasonal fluctuations and long-term persistence of pathogenic populations of *Agrobacterium* spp. in soils. Appl. Environ. Microbiol., 2002. **68**(7): p. 3358-3365.
- Lay, J.O., Jr. , MALDI-TOF mass spectrometry of bacteria. Mass Spectrom. Rev., 2001. **20**(4): p. 172-194.
- Maiden, M.C.J., et al., Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA, 1998. **95**(6): p. 3140-3145.
- Marvina LF., et al., Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry. 2003. Clinica Chimica Acta 337: 11 –21
- Mougel, C., et al., A mathematical method for determining genome divergence and species delineation using AFLP. Int. J. Syst. Evol. Microbiol., 2002. **52**(2): p. 573-586.
- Ophel, K. and A. Kerr, *Agrobacterium vitis* sp. nov. for strains of *Agrobacterium* biovar 3 from grapevines. Int. J. Syst. Bacteriol., 1990. **40**(3): p. 236-241.
- Ponsonnet, C. and X. Nesme, Identification of *Agrobacterium* strains by PCR-RFLP analysis of pTi and chromosomal regions. Arch. Microbiol., 1994. **161**(4): p. 300-309.
- Portier, P., et al., Identification of genomic species in *Agrobacterium* biovar 1 by AFLP genomic markers. Appl. Environ. Microbiol., 2006. **72**(11): p. 7123-7131.
- Rezzonico, F., et al., Whole cell MALDI-TOF mass spectrometry application for rapid identification and clustering analysis of *Pantoea* species. Appl. Environ. Microbiol., 2010. **76**: p. 4497-4509.
- Rhouma, A., et al., Susceptibility of some stone and pome fruit rootstocks to crown gall. Phytopathol. Mediterr., 2005. **44**(3): p. 275-284.
- Ryzhov, V. and C. Fenselau, Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Anal. Chem., 2001. **73**(4): p. 746-750.
- Sauer, S., et al., Classification and identification of bacteria by mass spectrometry and computational analysis. PLoS One, 2008. **3**(7): p. e2843.

- Stephan, R., et al., Rapid genus- and species-specific identification of *Cronobacter* spp. by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J. Clin. Microbiol., 2010. **48**(8): p. 2846-2851.
- Stevenson LG, Drake SK, Shea YR, Zelazny AM, Murray PR. 2010. Evaluation of Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry for Identification of Clinically Important Yeast Species. Journal of Clinical Microbiology. doi:10.1128/JCM.00687-09
- Verroken, A., et al., Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS) for species identification of *Nocardia*. J. Clin. Microbiol., 2010: p. JCM.01234-10.
- Wensing, A., S. Zimmermann, and K. Geider, Identification of the corn pathogen *Pantoea stewartii* by mass spectrometry of whole cell extracts and its detection with novel PCR primers. Appl. Environ. Microbiol., 2010: p. AEM.01032-10.
- Wieser A, Schneider L, Jung J, Schubert S. 2012. MALDI-TOF MS in microbiological diagnostics-identification of microorganisms and beyond (mini review). Appl Microbiol Biotechnol. 93:965-974
- Ziegler et al., In Situ Identification of Plant-Invasive Bacteria with MALDI-TOF MassSpectrometry. 2012. PLoS ONE doi:10.1371/journal.pone.0037189

Chapter 2-2

Rapid and accurate species identification and exhaustive population diversity assessment of *Agrobacterium* spp. using *recA*-based PCR

Shams M.^{1,2,3}, Vial L.^{1,2,3}, Chapulliot D.^{1,2,3,4}, Nesme X.^{1,2,3,4}, Lavire C.^{1,2,3}

¹Université de Lyon, F-69622, Lyon, France

²Université Lyon 1, F-69622 Villeurbanne, France

³CNRS, UMR5557 Ecologie Microbienne, F-69622 Villeurbanne, France

⁴INRA, USC 1364 Ecologie Microbienne, F-69622 Villeurbanne, France

Summary

Agrobacteria are common soil bacteria that interact with plants as commensals, plant growth promoting rhizobacteria or alternatively as pathogens. Understanding factors that shape their populations is a prerequisite for ecological studies. Indigenous agrobacterial populations are composites, generally with several species and strains per species. It would, however, be hard to thoroughly assess such diversity by standard microbial approaches. We therefore developed a *recA*-based PCR approach to accurately identify and specifically detect agrobacteria at various taxonomic levels. Specific primers were designed for the 16 bona fide genomic *Agrobacterium* species presently known, including 11 species of the *Agrobacterium tumefaciens* complex (G1-G9, G13 and G14), *A. larrymoorei*, *A. rubi*, *A. skirniawicse*, *A. sp.* 1650, and *A. vitis*, and for the close relative *Allorhizobium undicola*. Specific primers were also designed for superior taxa, *Agrobacterium* spp. and the Rhizobiaceae family. Primer specificities were assessed with target and non-target pure culture DNAs as well as with DNAs extracted from composite agrobacterial communities. In addition, we showed that the amplicon cloning-sequencing approach used with *Agrobacterium*-specific or Rhizobiaceae-specific primers is a way to assess the agrobacterial diversity of an indigenous agrobacterial population. Hence, the *Agrobacterium*-specific primers designed in the present study enabled the first accurate and rapid identification of all *Agrobacterium* species, as well as their direct detection in environmental samples. Analysis of *recA*-amplicon compositions thus provides a way to accurately and exhaustively assess agrobacterial diversity in various biotopes for ecological, agro-ecological and epidemiological studies.

Agrobacteria are members of the Alphaproteobacteria subclass that is common in almost all soils where they closely interact with plants in two respects. Firstly, and as a primary trait, agrobacteria generally live saprophytically in the vicinity of plant roots. This rhizospheric interaction is commensal and has no detrimental effects on plants. It may even be beneficial to plants by promoting plant growth (Hao et al., 2012). Secondly, it is well known that agrobacteria are plant pathogens that can cause crown gall disease in most dicots, gymnosperms and some monocots. However, the pathogenicity of agrobacteria is only the consequence of the acquisition of a Ti (i.e. tumor-inducing) plasmid. Fortunately for agriculture, indigenous soil agrobacteria populations are generally Ti plasmid free and most soils are uncontaminated and thus suitable for growing crops susceptible to crown gall, such as stone and pip fruit trees, raspberry, rose, poplar, etc. However, pathogenic agrobacteria may be fortuitously introduced with diseased plant material in any soils that may become contaminated. Contamination can be caused by introduced pathogenic strains. Alternatively, as Ti plasmids are conjugative, they can be easily transferred to indigenous soil agrobacteria that in turn become pathogenic, contributing to disease spread and perennial soil contamination (Krimi et al., 2002; Costechareyre et al., 2010). Even for crown gall epidemiology, it is thus necessary to gain further insight into the ecology of actual and potential Ti plasmid reservoirs such as agrobacteria in their primary habitats, soils and rhizospheres.

As explained above, agrobacteria is a term used to describe bacteria based on a functional trait (i.e. their actual or potential pathogenicity). It is not a valid taxonomic term. According to the latest taxonomic proposals, agrobacteria are now divided into two genera: *Agrobacterium* and *Rhizobium* (Costechareyre et al., 2010; Lindström and Young, 2011; Shams et al., 2012). In this revised nomenclature, biovar 2 agrobacteria (as defined by Keane et al., 1970) are now grouped into *Rhizobium rhizogenes*, while other agrobacteria are grouped into the genus *Agrobacterium*, now defined as a monophyletic taxon. The monophyletic *Agrobacterium* genus currently includes: *A. vitis*, *A. rubi*, *A. larrymoorei*, *A. skierniewicense*, *A. sp.* NCPPB 1650, likely *Allorhizobium undicola*, and a cluster of several genomic species corresponding to biovar 1 agrobacteria. Biovar 1 genomic species are collectively called the *Agrobacterium tumefaciens* complex, and individually called genomovar G1, G2, G3, G4, G5, G6, G7, G8, G9, G13 and G14. We added G14 to the collection of *Agrobacterium tumefaciens* genomic species after the recent description of a novel species in biovar 1 by Puławska et al. (2012). Only some of these genomic species have been given a Latin epithet: *puseense* for G2 (Panday et al., 2011), *radiobacter* for G4, *fabrum* for G8 (Lassalle et al., 2011) and *nepotum* for G14 (Puławska et al.,

2012). *R. skierniewicense*, recently described by Puławska et al. (2012), is a close relative of *A. rubi* and *A. larrymoorei* outside the *A. tumefaciens* complex and is called *A. skierniewicense* in this paper according to the present *Agrobacterium* genus definition.

Agrobacterium genomic species were initially determined by DNA–DNA hybridizations (Popoff et al., 1984), but easier methodologies were proposed to achieve the same goal. Alternative methodologies also provide access to the infra-specific diversity required for population genetic studies. Among the methodologies, AFLP (Vos et al., 1995) was found to be useful for delineating *Agrobacterium* genomic species (Mougel et al., 2002), but the method is not dedicated for the rapid identification of a novel isolate. Standard method based upon 16S rRNA gene (*rrs*) analyses is insufficient to distinguish *Agrobacterium* species, since several ones exhibit identical *rrs* (Mougel et al., 2002). On the contrary, we showed that *recA* gene sequencing can discriminate *Agrobacterium* species and also revealed the presence of numerous *recA* alleles within most species (Costechareyre et al., 2010). Sequencing this gene is now therefore routinely done to give a first idea of the species and allele status of novel isolates. The availability of numerous sufficiently long sequences of *recA* alleles offers scope for the development of specific primers to identify each *Agrobacterium* species, the *Agrobacterium* genus and the Rhizobiaceae family. We think that designing primers targeting specific taxa would permit identification of target taxa by a fast and inexpensive approach. To assess the diversity of numerous populations so as to elucidate ecological factors that shape their genetic structures, we proposed to set up a PCR-based approach to evaluate the species and allelic diversity of agrobacterial populations based on individual sequences present in *recA* amplicons obtained from complex communities.

The present study was aimed at developing species-, genus- and family-specific *recA*-based PCR primers suitable: i) for rapid identification of genomic species of *Agrobacterium* spp. and ii) to assess the allelic and species diversity of indigenous communities of *Agrobacterium* and Rhizobiaceae.

Materials and Methods

Bacteria

The *Rhizobium* and *Agrobacterium* strains used in this study listed in Table 2-2-1 are available at CFBP (Collection Française de Bactéries associées aux Plantes, INRA, Angers, France, <http://www-intranet.angers.inra.fr/cfbp/>). *Agrobacteria* isolated from an environmental community were obtained from a Poaceae rhizosphere at La-Côte-Saint-André (LCSA, Isère, France) using selective media amended with potassium tellurite and cycloheximide, as previously described (Mougel et al., 2001; Shams et al., 2012). All strains were routinely grown on mannitol glutamate (MG) medium (Ophel and Kerr, 1990) at 28°C and stored at -80°C in glycerol.

DNA extraction

Pure culture genomic DNAs were extracted from bacteria grown overnight at 28°C in MG broth using the NucleoSpin® Tissue kit (Macherey-Nagel GmbH & Co., UK). Metagenomic soil DNA was extracted according to instructions of the PowerSoil® DNA isolation kit (Mo Bio, Carlsbad, CA, US). DNA sample quantity and quality were determined with NanoDrop 2.5.4. (NanoDrop Technologies, Wilmington, DE) and after electrophoresis in 1% agarose gel (Euromedex, Souffelweyersheim, France).

Primer design and PCR reactions

Degenerated primers F2898 5'-TAA TAC GAC TCA CTA TAG GGT CTT TGC GKC TCG TAG AGG AYA-3' and F2899 5'-ATT AAC CCT CAC TAA AGG GAT GCA GGA AGC GGT CGG CRA TSA G-3' designed by Costechareyre et al. (2010) were used to amplify and sequence a 1028-bp *recA* region spanning from position 13 to 1040 of Atu1874 (*recA*) in C58. PCR were performed at 57°C for *A. tumefaciens* species, but for bacteria outside this complex the annealing temperatures required adjustments

Table 2-2-1. *recA* alleles of *Agrobacterium* spp. and related taxa used to design specific PCR primers.

Allele name	Allele reference strain	CFBP / LMG n°	Accession n°	Length	<i>recA</i> allele description
<i>Agrobacterium tumefaciens</i> genomovar G1					
<i>recA</i> -G1-1	ICPB TT111 *	CFBP 5767 / LMG 196	FM164286	969	Costechareyre et al., 2010
<i>recA</i> -G1-2	CFBP 5771 *		FM164294	969	Costechareyre et al., 2010
<i>recA</i> -G1-3	S56	CFBP 5491 / LMG 321	FM164289	969	Costechareyre et al., 2010
<i>recA</i> -G1-4	CFBP 5622		FM164291	969	Costechareyre et al., 2010
<i>recA</i> -G1-5	CFBP 2712		FM164292	969	Costechareyre et al., 2010
<i>recA</i> -G1-6	CFBP 2517		FM164293	969	Costechareyre et al., 2010
<i>recA</i> -G1-7	CFBP 7067		FM164295	969	Costechareyre et al., 2010
<i>recA</i> -G1-8	CFBP 7064		FM164297	969	Costechareyre et al., 2010
<i>recA</i> -G1-9	CFBP 7066		FM164298	969	Costechareyre et al., 2010
<i>recA</i> -G1-10	Fb243 #	CFBP 5479	deposit in progress	966	This study
<i>recA</i> -G1-11	H13-3	not available	AGROH133_07016	1092	Wibberg et al., 2011
<i>recA</i> -G1-12	Foundi	in progress at CFBP	deposit in progress	1092	This study
<i>recA</i> -G1-13	5A	not available	AGVZ00000000	1092	Hao et al., 2012
<i>recA</i> -G1-14	R-46040	not available	FR772678	550	De Meyer et al. (2011)
<i>A. tumefaciens</i> genomovar G2 (<i>A. pusense</i>)					
<i>recA</i> -G2-1	Bernaerts L2/2/1	CFBP 5875 / LMG 146	FM164299	969	Costechareyre et al., 2010
<i>recA</i> -G2-2	Bernaerts M 2/1	CFBP 5876 / LMG 147	FM164300	969	Costechareyre et al., 2010
<i>recA</i> -G2-3	CIP 497-74 *	CFBP 5494 / CFBP 2884	FM164301	969	Costechareyre et al., 2010
<i>recA</i> -G2-4	CIP 28-75 *	CFBP 5495	FM164302	969	Costechareyre et al., 2010
<i>recA</i> -G2-5	CIP 43-76	CFBP 5496	FM164303	969	Costechareyre et al., 2010
<i>recA</i> -G2-6	MKS-02 #	in progress at CFBP	deposit in progress	936	This study
<i>recA</i> -G2-7	MKS-03 #	in progress at CFBP	deposit in progress	953	This study
<i>recA</i> -G2-8	AGR20 #	in progress at CFBP	deposit in progress	969	This study
<i>A. tumefaciens</i> genomovar G3					
<i>recA</i> -G3-1	CIP 107443 *	CFBP 6623 / CIP 493-74	FM164304	969	Costechareyre et al., 2010
<i>recA</i> -G3-2	CIP 107442 *	CFBP 6624 / CIP 111-78	FM164305	969	Costechareyre et al., 2010
<i>A. tumefaciens</i> genomovar G4 (<i>A. radiobacter</i>)					
<i>recA</i> -G4-1	B6 *	CFBP 2413 / LMG 187	FM164306	969	Costechareyre et al., 2010
<i>recA</i> -G4-2	ATCC 19358 ^T	CFBP 5522 / LMG 140	FM164311	969	Costechareyre et al., 2010
<i>recA</i> -G4-3	Kerr 14 *	CFBP 5761 / LMG 15	FM164309	969	Costechareyre et al., 2010
<i>recA</i> -G4-4	Myr3	CFBP 7128	FM164334	969	Costechareyre et al., 2010
<i>recA</i> -G4-5	AR9	CFBP 7126	FM164335	969	Costechareyre et al., 2010
<i>recA</i> -G4-6	CH14	CFBP 7199	FN432354	969	Costechareyre et al., 2010
<i>A. tumefaciens</i> genomovar G5					
<i>recA</i> -G5-1	CIP 107445 *	CFBP 6625	FM164317	969	Costechareyre et al., 2010
<i>recA</i> -G5-2	CIP 107444 *	CFBP 6626	FM164318	969	Costechareyre et al., 2010
<i>recA</i> -G5-3	F2	not available	AFSD00000000	1092	Li et al., 2011

<i>A. tumefaciens</i> genomovar G6					
<i>recA</i> -G6-1	NCPPB 925 *	CFBP 5499 / LMG 225	FM164319	969	Costechareyre et al., 2010
<i>A. tumefaciens</i> genomovar G7					
<i>recA</i> -G7-1	RV3 *	CFBP 5500 / LMG 317	FM164321	969	Costechareyre et al., 2010
<i>recA</i> -G7-2	NCPPB 1641	CFBP 5502/ LMG 228	FM164322	969	Costechareyre et al., 2010
<i>recA</i> -G7-3	Zutra 3/1 *	CFBP 6999 / LMG 198	FM164323	969	Costechareyre et al., 2010
<i>recA</i> -G7-4	GF78	CFBP 7001	FM164339	969	Costechareyre et al., 2010
<i>recA</i> -G7-5	AR29	CFBP 7125	FM164337	969	Costechareyre et al., 2010
<i>recA</i> -G7-6	AR94	CFBP 7127	FM164338	969	Costechareyre et al., 2010
<i>recA</i> -G7-7	AR92	CFBP 7124	FM164340	969	Costechareyre et al., 2010
<i>recA</i> -G7-8	AR118	CFBP 7129	FM164336	974	Costechareyre et al., 2010
<i>recA</i> -G7-9	CH6	CFBP 7198	FM955141	969	Costechareyre et al., 2010
<i>recA</i> -G7-10	DC07-029	CFBP 7200	FM164341	957	Costechareyre et al., 2010
<i>recA</i> -G7-11	MKS-05	in progress at CFBP	deposit in progress	957	This study
<i>recA</i> -G7-12	MKS-04	in progress at CFBP	deposit in progress	957	This study
<i>A. tumefaciens</i> genomovar G8 (<i>A. fabrum</i>)					
<i>recA</i> -G8-1	T37 *	CFBP 5503 / LMG 332	FM164325	969	Costechareyre et al., 2010
<i>recA</i> -G8-2	LMG 75	CFBP 6549	FM164328	969	Costechareyre et al., 2010
<i>recA</i> -G8-3	LMG 46	CFBP 6554	FM164329	969	Costechareyre et al., 2010
<i>recA</i> -G8-4	C58 *	CFBP 1903 / LMG 287	AE007869	1092	Wood et al., 2001 Goodner et al., 2001
<i>recA</i> -G8-5	MKS-01	CFBP 7336	deposit in progress	965	This study
<i>recA</i> -G8-6	ATCC 31749	not available	AECL00000000	1092	Ruffing et al., 2011
<i>A. tumefaciens</i> genomovar G9					
<i>recA</i> -G9-1	Hayward 0362 *	CFBP 5507 / LMG 26	FM164331	969	Costechareyre et al., 2010
<i>recA</i> -G9-2	Hayward 0363 *	CFBP 5506 / LMG 27	FM164332	969	Costechareyre et al., 2010
<i>recA</i> -G9-3	AGR7 [#]	in progress at CFBP	deposit in progress	969	This study
<i>A. tumefaciens</i> genomovar G13					
<i>recA</i> -G13-1	PGF 01-22 *	CFBP 6927	FM164333	969	Costechareyre et al., 2010
<i>A. tumefaciens</i> genomovar G14 (<i>A. nepotum</i>)					
<i>recA</i> -G14-1	C3.4.1	in progress at CFBP	HE610312	740	This study
<i>recA</i> -G14-2	AGR18 [#]	in progress at CFBP	deposit in progress	969	This study
<i>recA</i> -G14-3	AGR19 [#]	in progress at CFBP	deposit in progress	969	This study
<i>A. larrymoorei</i>					
<i>recA</i> -Ala-1	AF 3.10 ^T *	CFBP 5473	FN432355	966	Costechareyre et al., 2010
<i>recA</i> -Ala-2	Fb33	CFBP 5476	FN432393	966	Costechareyre et al., 2010
<i>recA</i> -Ala-3	LT4	CFBP 5480	FN813464	957	Costechareyre et al., 2010
<i>recA</i> -Ala-4	NL7	CFBP 5481	FN813465	948	Costechareyre et al., 2010
<i>A. rubi</i>					
<i>recA</i> -Aru-1	TR3 ^T *	CBP 5509 / LMG 17935	AM182122	498	Martens et al., 2007
<i>recA</i> -Aru-2	Braun EU6	LMG 294	deposit in progress	852	This study

<i>A. skirniwicense</i>					
<i>recA</i> -Ask-1	Ch11 ^T	CFBP 7420 / LMG 26191	HE610311	869	This study
<i>A. sp</i>					
<i>recA</i> -1650-1	NCPPB 1650	CFBP 4470 / LMG 230	FN813466	896	Costechareyre et al., 2010
<i>A. vitis</i>					
<i>recA</i> -Avi-1	K309 ^T *	CBP 5523 / LMG 8750	FR847962	957	This study
<i>recA</i> -Avi-2	S4	not available	CP000633	1089	Slater et al., 2009
<i>recA</i> -Avi-3	NKZ-2	not available	AB543861	437	Kawaguchi, 2011
<i>recA</i> -Avi-4	IS552-1	not available	AB543847	437	Kawaguchi, 2011
<i>recA</i> -Avi-5	G-Ag-67	not available	AB543860	437	Kawaguchi, 2011
<i>recA</i> -Avi-6	UK-2	not available	AB543845	437	Kawaguchi, 2011
<i>recA</i> -Avi-7	MAFF 211677	not available	AB543837	437	Kawaguchi, 2011
<i>recA</i> -Avi-8	ARK-2	not available	AB608984	437	Kawaguchi, unpublished
<i>recA</i> -Avi-9	VAR06-30	not available	AB285087	437	Kawaguchi, 2011
<i>Allorhizobium undicola</i>					
<i>recA</i> -Aun-1	ORS 992 ^T	CFBP 6719 / LMG 11875	HE610313	760	This study
<i>Rhizobium rhizogenes</i>					
<i>recA</i> -Rrh-1	K84 *	CFBP 1937	CP000628	1086	Slater et al., 2009
<i>R. leguminosarum</i>					
<i>recA</i> -Rleg	3841 ^T	LMG 14904	AM236080	1095	Young et al., 2006
<i>R. etli</i>					
<i>recA</i> -Ret-1	CFN42 ^T	LMG 17827	CP000133	1089	González et al., 2006
<i>Ensifer meliloti</i>					
<i>recA</i> -Sme-1	RM 1021 *	LMG 4251	AL591688	1086	Galibert et al., 2001

As several strains may have the same *recA* allele, only one with a long sequence was arbitrarily selected as *recA* allele reference strain. Reference strains were made available when possible at CFBP (Collection Française de Bactéries associées aux Plantes, INRA, Angers, France) or at LMG (Laboratorium voor Microbiologie, Ghent University, Belgium). Species assignments were done using standard genomic methods (DNA/DNA hybridization, AFLP, complete genome sequence analysis) or phylogenetic analyses performed with at least two different genes, except for strains indicated by # which were assigned using only *recA* sequences. ^T indicates species type-strains. Large *recA* were obtained from complete genome sequences. * indicates strains used to prepare the synthetic community.

Specific primers listed in Table 2-2-2 were designed by visual inspection of *recA* allele sequences aligned with BioEdit software (version 7.0.9.0). The level of specificity of each primer against currently known sequences was checked on line at NCBI (<http://www.ncbi.nlm.nih.gov/>) on the nucleotide collection (nr/nt) database with blastn (somewhat similar sequences) using 7 as word size.

Standard PCRs were carried out in a 30 µl reaction volume containing 2.4 mM MgCl₂, 0.4 mM (each) dNTPs, 1U of Taq DNA polymerase (Invitrogen, France), 1 µM of each primer and 30 ng of DNA template. The optimal annealing temperatures of each primer set given in Table 2-2-2 were determined after running gradient PCRs using the Biometra TGradient Thermocycler. After

an initial 5 min denaturation at 95°C, the PCR conditions consisted of denaturation at 95°C for 45 sec, annealing at optimal primers temperatures for 45 sec and extension at 72°C for 1 min for 35 cycles.

Table 2-2-2. Species-, *Agrobacterium*- and Rhizobiaceae-specific primer sets.

Forward primer	Position	Reverse primer	Position	Taxon	bp	°C
Primer sets experimentally validated with target and non-target pure culture DNA and complex DNA mixtures						
F7386 AGCAAGGCACTGGAAGCGG	49-67	F7387 CCATACATGATGTCGAATTC	808-827	Rhizobiaceae	779	52
F8360 AGCTCGGTCCAATGAAA	113-130	F8361 GCTTGCGCAGCGCCTGGCT	547-565	<i>Agrobacterium</i>	453	52
F6782 CGTGGTTGAAGTAGAGACCA	132-151	F6786 TTTTCACGCCCTGCCCA	926-944	Genomovar G1	813	60
F6790 TGACGCCGAACATGCGCTT	312-330	F6786 TTTTCACGCCCTGCCCA	926-944	Genomovar G2	633	67
F8232 CGTTCGGCAAGGGATCGATC	89-108	F8227 CCGCGGCGTCAGTGCCGCG	471-489	Genomovar G3	401	65
F8533 GCGTTCGGTGCGGTTGATG	432-451	F8827 CACCTGCTTGAAGGGCGGA	789-807	Genomovar G4	511	60
F8232 CGTTCGGCAAGGGATCGATC	89-108	F8228 AGCGTATCGGTGATTTCGAGG	408-428	Genomovar G5	290	60
F8392 TATCGCGCTCGGTATCGGT	177-195	F8229 AGTCGACCACGAGCACGTCC	447-466	Genomovar G6	442	65
F6788 TGGCATTGCAGACCATTGCC	257-276	F8413 CGAACGGACGCATAGAAC	681-698	Genomovar G7	502	60
F6784 CGGTTGACGTTCTTGTTCATCGAT	443-465	F6786 TTTTCACGCCCTGCCCA	926-944	Genomovar G8	502	60
F8359 CTGCGCAAGCTGACGGCT	556-573	F8230 CACGCCGAGATCGACCAAC	852-870	Genomovar G9	315	65
F8232 CGTTCGGCAAGGGATCGATC	231-249	F8231 GGAAGTGGAGATCGAGGCT	570-588	Genomovar G13	358	60
Primer sets experimentally validated with target and non-target pure culture DNA						
F9033 GGTATCGGCGGTTTGCCGCG	187-206	F9034 TCTTGACGACCTTGACACG	760-778	Genomovar G14	592	67
F9186 TTTCCACAGGGTCTCTCAGT	152-171	F9187 TCTCGCCTTCGATTTCCGCA	489-508	<i>A. larrymoorei</i>	357	60
G0009 GTTCCAATGAAAGTATCGTC	119-138	G0001 GAAGACCCGGAAGGCTATCA	513-532	<i>A. rubi</i>	414	57
G0006 CCGTAAGACAACGCTCGCT	243-261	G0007 CGCGTGTCTGTTACCGACG	744-763	<i>A. sp.</i> 1650	530	60
F9190 TGAGGTCGAGACCGTTTCTACA	138-159	G0008 CTTCGGGCGAACCCTAGGAA	636-655	<i>A. skirniawicenses</i>	518	60
G0002 TGAAAAGGTCGTGGAGATCGAT	126-147	G0003 TCAGTGCTGCAACGGAGTCA	462-481	<i>A. undicola</i>	356	60
G0004 GATATCGCGCTCGGCATTGGT	175-195	G0005 CCTTCGATTTACGTTTCG	485-503	<i>A. vitis</i>	329	55

Position indicates the primer position in the complete *recA* sequence of C58 (Atu1874). Taxon indicates the target taxon specifically amplified with primer sets. bp, amplicon length in base pairs. °C, annealing temperature.

Primer specificity and analysis of complex communities

Primer specificity was tested on DNA of strains representing all *recA* alleles of target species indicated in Table 2-2-1 and one representative of other species, and then in conditions mimicking the presence of several species in template DNAs. Firstly, to check the ability of primers to specifically amplify the target species even in the presence of other species, template DNAs were prepared by vigorously mixing 30 ng DNA of one strain per species for all species, including the target. Secondly, and conversely, to check the absence of amplification in the absence of the target species, template DNAs were prepared by mixing 30 ng DNA of one strain per species for all species except the target.

The synthetic community was prepared by mixing 23 strains harboring readily different *recA* alleles (Table 2-2-1). For this purpose, 2 ml of overnight cultures of each strain were adjusted to an OD₆₀₀ of 0.5, mixed by vortexing and maintained on ice to stop growth until DNA extraction as described above. Further PCRs were performed using 30 ng of this composite DNA as template.

PCR product cloning, sequencing and sequence analysis

PCR products were cloned into the pGEM-T easy vector as indicated by the manufacturer (Promega, Madison, USA) and clones were sequenced by Biofidal (Villeurbanne, France) using M13 forward and reverse universal primers. PCR products were sequenced by Genoscreen (Lille, France).

Bellerophon software was used to detect and eliminate chimera as described by Huber et al. (2004), with the Kimura correction and a 200 bp sliding window. Sequences were then compared to a locally curated *recA* sequence database made with the accessions given Table 2-2-1 in order to assign each sequence to the closest *recA* allele.

Results

The present study was aimed at designing primers in the conserved *recA* gene to identify agrobacteria and related bacteria at different taxonomic levels, including all known *Agrobacterium* species (eleven within and five outside of the *A. tumefaciens* complex) plus the close relative species *Allorhizobium undicola*, the *Agrobacterium* genus and the Rhizobiaceae family. We thus first attempted to obtain the most exhaustive view of *recA* diversity in *Agrobacterium* spp. Indeed, since the work of Costechareyre et al. (2010) describing 50 *recA* alleles amongst 14 species in *Agrobacterium* spp., two new species (*A. skierniewicense*, *A. nepotum*) have been described and 27 new strains with novel *recA* alleles have been found (Table 2-2-1). By visual inspection of 77 aligned *Agrobacterium* allele sequences, we then designed sets of specific primers for all species as well as for superior taxa (Table 2-2-2).

Species-specific primers

Species-specific primers similarity analysis performed with blastn in GenBank showed that all primer sets likely amplify only their target taxon. Moreover, all agrobacteria that belonged to a particular species were recognized and amplified using relevant species-specific primers, thus generating amplicons of expected lengths. For instance, G2 delivered a 633 bp amplicon with

G2-specific primers with all G2 agrobacteria irrespective of their *recA* allele, while no amplification was obtained with these primers using non-target DNAs (Fig. 2-1-1). In addition, with all species-specific primer sets, no amplification occurred in the absence of target DNAs, even when template DNAs consisted of mixtures of DNAs of different but closely related species (Fig. 2-1-2). Finally, we verified that no aspecific amplifications (i.e. amplicons of unexpected length) were obtained when using a DNA mixture mimicking a complex community formed by the pooling of 23 different strains belonging to different *Agrobacterium* and *Rhizobium* species (Table 2-2-1; Fig. 2-2-3). This indicated that the primer set specificities are conserved even in the presence of numerous non-target but closely related alleles in template DNAs, as consistently occurs in environmental communities.

***Agrobacterium*- and Rhizobiaceae-specific primers**

Using a moderately stringent temperature (52°C), *Agrobacterium*-specific primers F8360-F8361 amplified a fragment of expected length (453 bp) with all isolates of the *A. tumefaciens* complex, *A. sp.* NCPPB 1650, *A. rubi*, *A. skirniawicence* and *A. larrymoorei* (with the exception of AF3.10, *recA*-Ala-1 allele) as well as with the close relative species *Allorhizobium undicola* (Fig. 2-2-4A). Amplifications were reliably obtained at 52°C in spite of the occurrence of a mismatch between one primer and *recA* alleles such as *recA*-G6-1, *recA*-G7-11, *recA*-G9-2, *recA*-Ala-2, *recA*-Ala-4. *recA*-Ala-1 with an antepenultimate mismatch at 3' required a lower annealing temperature for reliable amplifications (data not shown). Conversely, no amplifications were obtained with *A. vitis* or other Rhizobiaceae isolates that all showed at least four mismatches with primers (Fig. 2-2-4A).

Rhizobiaceae-specific primers F7386-F7387 amplified a fragment of the expected length (779 bp) with all tested Rhizobiaceae species except *A. undicola*, while no amplification was obtained with any Gammaproteobacteria, Betaproteobacteria or other Alphaproteobacteria families (Fig. 2-2-4A and data not shown). As a final confirmation, sequences of PCR products obtained with pure culture DNAs always matched the template strain sequences (data not shown).

Species and infra-specific diversity of complex communities

Determining the internal species and allele diversity of complex amplicons would be a way to characterize the actual diversity of complex communities. To evaluate the usefulness of *Agrobacterium*- or Rhizobiaceae-specific primers to assess the allelic and species diversity of

indigenous communities of *Agrobacterium* and Rhizobiaceae we first made a synthetic community of known allele composition (i.e. 23 different strains with alleles perfectly and non perfectly matching primer sequences). We then used a cloning-sequencing approach to decypher the allele diversity of amplicons obtained using *Agrobacterium*- or Rhizobiaceae-specific primers

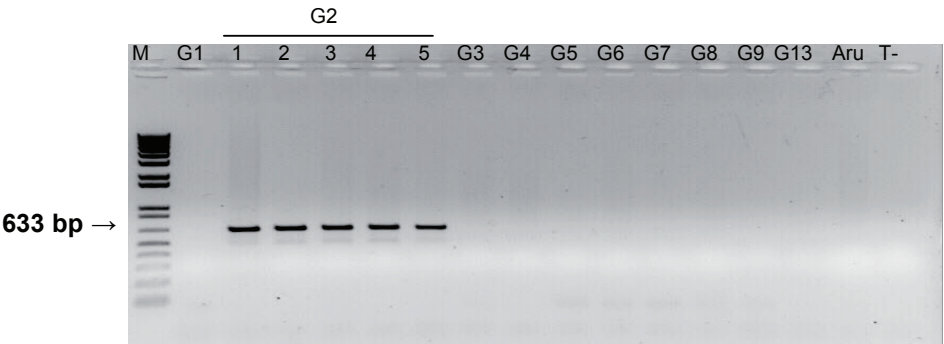


Figure 2-2-1. Specificity of G2-specific primers F6786-F6790.
 PCRs were performed with template DNAs extracted from strains that had different G2 *recA* alleles or from other species. G1, TT111; 1, Bernaerts L2/2/1 (*recA*-G2-1); 2, Bernaerts M 2/1 (*recA*-G2-2); 3, CIP 497-74 (*recA*-G2-3); 4, CIP 28-75 (*recA*-G2-4); 5, CIP 43-76 (*recA*-G2-5); G3, CIP 107443; G4, B6; G5, CIP 107445; G6, NCPPB 925; G7, RV3; G8, C58; G9, Hayward 0362; G13, PGF 01-22; Aru, *A. rubi* LMG 17935^T; T-, negative control; M, 1 kb+ Invitrogen DNA marker.

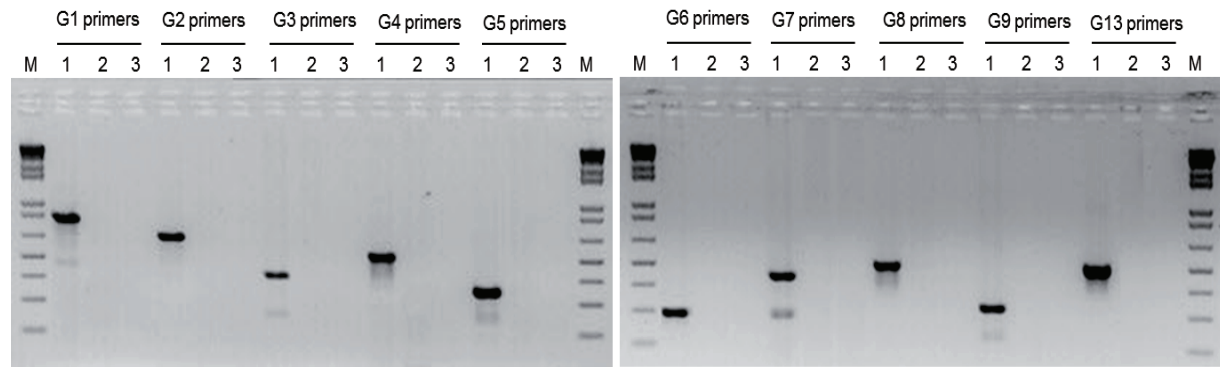


Figure 2-2-2. Specificity of amplification with species-specific primers.
 PCRs were performed with species-specific primers and DNA mixtures containing (1) or not containing (2) target species DNAs or without DNA (3). Species-specific primers used are specified on the figure. DNAs used for DNA mixture preparation were extracted from TT111 (G1 species), Bernaerts L2/2/1 (G2); CIP 107443 (G3); B6 (G4); CIP 107445 (G5); NCPPB 925 (G6); RV3 (G7); C58 (G8); Hayward (G9); PGF 01-22 (G13); M, 1kb+ Invitrogen .

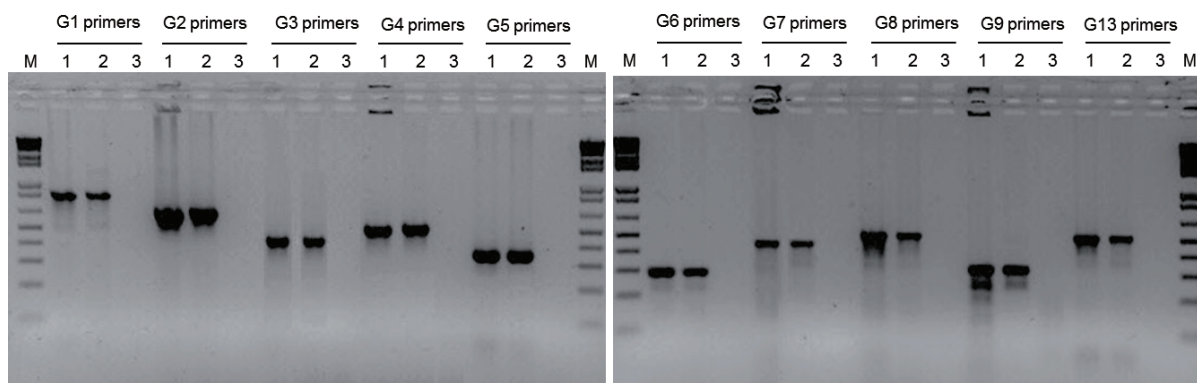


Figure 2-2-3. Absence of aspecific amplification in the presence of a composite DNA using species-specific primers.

PCRs were performed with species-specific primers using a complex synthetic community DNA as template (1) or the DNA of a representative of target species as positive control (2) or without DNA as negative control (3). Species-specific primers used are specified above wells. The list of DNAs used as positive controls are presented in Fig. 2-2-1. The composition of synthetic community is indicated Table 2-2-2. M, 1kb+ Invitrogen DNA marker.

For *Agrobacterium*- and Rhizobiaceae-specific primers, respectively, 47 and 41 colonies were selected and sequenced. Unexpected sequences matching none of the alleles of the initial community were obtained and suspected as chimera sequences. Consequently, out of the 47 sequences obtained with *Agrobacterium*-specific primers, 18 sequences (38%) were identified as chimeras by Bellerophon software and removed from subsequent analyses. The 29 remaining sequences were divided into 14 alleles among the 20 potentially amplifiable by the *Agrobacterium*-specific primer set comprising *recA*-G6-1 and *recA*-G9-2, two alleles containing primer mismatches (Table 2-2-4). At the species level, 9 species out of the 12 present in different amounts in the initial community were detected, indicating the relevance of this primer set for analyzing both allelic and species diversities of *Agrobacterium* communities.

Out of the 41 sequences obtained with Rhizobiaceae-specific primers, 12 sequences (29%) were identified as chimeras. The 29 remaining sequences were divided into 17 alleles among the 23 that were amplifiable by the Rhizobiaceae-specific primer set. Ultimately, eleven species divided into three genera were detected. This indicates that this primer set is relevant for analyzing the allele, species and genus diversity of a Rhizobiaceae community.

Diversity analysis of an environmental community

In order to test whether primers targeting *Agrobacterium* or Rhizobiaceae can be used to assess the agrobacterial diversity of an environmental community, we used the direct approach based on

the cloning and sequencing of amplicons obtained with metagenomic DNA extracted from the whole bacterial community of a wild Poaceae rhizosphere and compared results obtained with the standard microbiological approach.

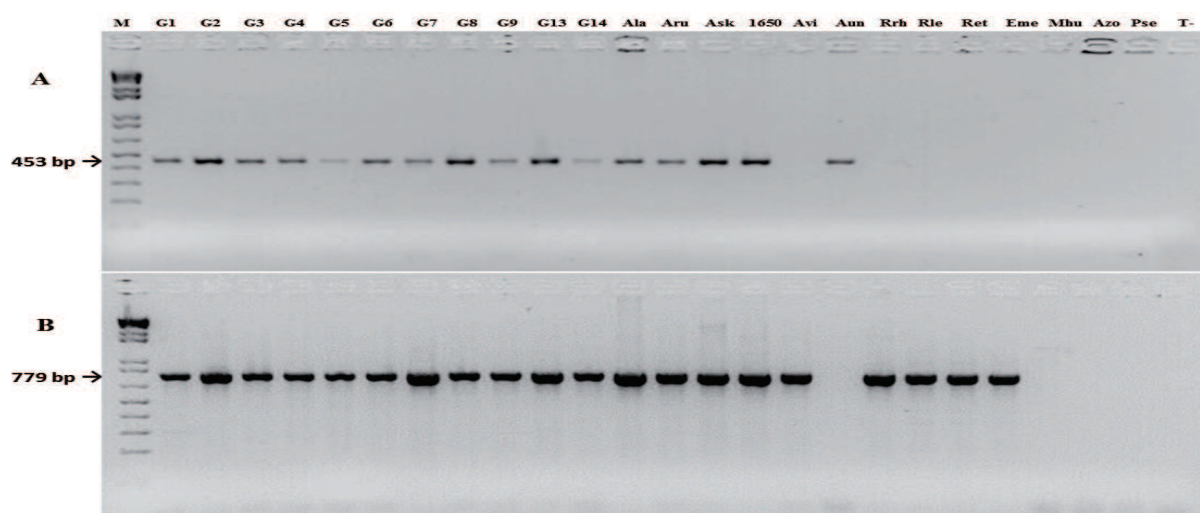


Figure 2-2-4. Range of species amplified using *Agrobacterium*-specific primers F8360-F8361 (A) or Rhizobiaceae-specific primers F7386-F7387 (B). G1, TT111; G2, CIP 497-74; G3, CIP 107443; G4, B6; G5, CIP 107445; G6, NCPPB 925; G7, RV3; G8, C58; G9, Hayward 0362; G13, PGF 01-22; Ala, Fb33 (*A. larrymoorei*); Aru, LMG 17935 (*A. rubi*); Ask, Ch11 (*A. skieniewicenses*); 1650, NCPPB 1650 (*A. sp.*); Avi, K309^T (*A. vitis*); Aun, ORS 992 (*A. undicola*); Rrh, K84 (*R. rhizogenes*); Rle, A34 (*R. leguminosarum*); Ret, CFN 42^T (*R. etli*); Eme, RM1021 (*E. meliloti*); Mhu, *Mesorhizobium huakuii*; Azo, B510 (*Azospirillum sp.*); Pse, A1032 (*Pseudomonas sp.*); T-, negative control; M, 1kb+ Invitrogen DNA marker.

From the 30 sequences of 779 bp obtained with Rhizobiaceae-specific primers, 11 (34%) were identical and showed almost total identity with *recA*-G8-1 (Fig. 2-2-5). No other sequences gathered within the *A. tumefaciens recA* cluster. Seven sequences significantly gathered in the *recA* cluster grouping *A. rubi*, *A. sp.* NCPPB 1650, *A. skieniewicenses* and *A. larrymoorei*. All of these sequences were of bona fide *Agrobacterium* species but they were not from the *A. tumefaciens* species complex. Outside the *Agrobacterium* genus cluster, eight *recA* sequences gathered significantly with the cluster including *Ensifer* species, *R. herbae* and *R. giardini*, and three gathered with other *Rhizobium* species such as *R. tubonenese*, *R. leguminosarum*, or *R. rhizogenes*. As expected, none gathered outside of Rhizobiaceae. This clearly showed that all sequences obtained from an amplicon performed with Rhizobiaceae-specific primers could be readily assigned as bona fide Rhizobiaceae bacteria. In addition, as a remarkable biological result, with 62% of the sequences, G8 *recA* clearly dominated in the Rhizobiaceae community

amplified from this rhizosphere, while no *R. rhizogenes* sequences (e.g. biovar 2 agrobacteria) were found.

Table 2-2-3. Experimental validation of species-, *Agrobacterium*- and Rhizobiaceae-specific primer sets.

Primer set	Template DNA																
	G1	G2	G3	G4	G5	G6	G7	G8	G9	G13	G14	Ala	Aru	Asp	Ask	Aun	Eme
F7386-F7387	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
F8360-F8361	+	+	+	+	+	+	+	+	+	+	+	(+)	+	+	+	+	-
F6786-F6782	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F6786-F6790	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F8227-F8232	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F8533-F8827	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
F8228-F8232	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
F8229-F8392	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
F6788-F8413	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
F6786-F6784	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
F8230-F8359	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-
F8231-F8232	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
F9033-F9034	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
F9186-F9187	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
G0009-G0001	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
G0006-G0007	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
F9190-G0008	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
G0002-G0003	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
G0004-G0005	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

G1 to G9, G13 and G14 indicate species of the *A. tumefaciens* complex. Ala, *A. larrymoorei*; Aru, *A. rubi*; Asp, NCPPB 1650; Aun, *A. undicola*; Avi, *A. vitis*; Rrh, *R. Rhizogenes*; Ret, *R. etli*; Rle, *R. leguminosarum*; Eme, *E. meliloti*. For PCRs performed with annealing temperatures indicated in Table 2: +, yielded a fragment of the expected length with all alleles described; (+), yielded a fragment of the expected length with all alleles described except for AF3.10 (*recA*-Ala-1) that required a lower temperature for reliable amplifications; -, no amplification.

In a second step, screening for particular *recA* alleles in the metagenomic DNA was performed using the cloning-sequencing approach with *Agrobacterium*-specific primers. From the 47 sequences obtained 44 (93.6%) were identical and showed 100% identity with *recA*-G8-1 (or *recA*-G8-3 because these two alleles are identical in the short 453 bp region targeted by F8360 and F8361), one was identical to *recA*-G1-7, and two were identical to *recA*-G1-14. All sequences obtained are in agreement with primers specificities. Notably, Bellerophon detected no chimera among the *recA* sequence sets obtained from this community (probably because the

environmental communities were far less diversified than the synthetic one tested above and thus less prone to chimera formation).

Table 2-2-4. *Agrobacteria recA* alleles detected in a synthetic community of known composition using the amplicon cloning-sequencing approach.

Species	Synthetic community		Cloning-sequencing	
		<i>recA</i> allele	Rhizobiaceae	<i>Agrobacterium</i>
<i>A. tumefaciens</i> G1	ICPB TT111	<i>recA</i> -G1-1	1	0
	CFBP 5771	<i>recA</i> -G1-2	5	0
<i>A. tumefaciens</i> G2	CIP 497-74	<i>recA</i> -G2-3	1	2
	CIP 28-75	<i>recA</i> -G2-4	1	1
<i>A. tumefaciens</i> G3	CIP 107443	<i>recA</i> -G3-1	1	2
	CIP 107442	<i>recA</i> -G3-2	2	3
<i>A. tumefaciens</i> G4	B6	<i>recA</i> -G4-1	0	1
	Kerr 14	<i>recA</i> -G4-3	3	0
<i>A. tumefaciens</i> G5	CIP 107445	<i>recA</i> -G5-1	2	4
	CIP 107444	<i>recA</i> -G5-2	1	2
<i>A. tumefaciens</i> G6	NCPPB 925	<i>recA</i> -G6-1	1	1
<i>A. tumefaciens</i> G7	RV3	<i>recA</i> -G7-1	0	2
	Zutra 3/1	<i>recA</i> -G7-3	1	0
<i>A. tumefaciens</i> G8	T37	<i>recA</i> -G8-1	1	3
	C58	<i>recA</i> -G8-4	3	1
<i>A. tumefaciens</i> G9	Hayward 0362	<i>recA</i> -G9-1	2	3
	Hayward 0363	<i>recA</i> -G9-2	2	3
<i>A. tumefaciens</i> G13	PGF 01-22	<i>recA</i> -G13-1	0	1
<i>A. rubi</i>	TR3 ^T	<i>recA</i> -Aru-1	0	0
<i>A. larrymoorei</i>	AF 3.10 ^T	<i>recA</i> -Ala-1	0	0
<i>A. vitis</i>	K309 ^T	<i>recA</i> -Avi-1	1	0
<i>R. rhizogenes</i>	K84	<i>recA</i> -Rrh	1	0
<i>E. meliloti</i>	RM 1021	<i>recA</i> -Eme	0	0
Total			29	29

Strains with different *recA* alleles were pooled in equal amounts to form a synthetic community prior to DNA extraction. Numbers indicate the number of cloned fragments obtained with F7386-F7387 or F8360-F8361 primer sets that specifically amplify Rhizobiaceae or *Agrobacterium* spp., respectively.

Agrobacteria isolated from the rhizosphere using the *A. tumefaciens*-elective 1A-Te medium of Mougél et al. (2001) were first presumptively assigned as G1 bacteria, G4 and G8 using species-specific primers described in Table 2. Amplicon sequencings then confirmed these previous species assignments, and allowed us to identify the presence of alleles *recA*-G1-6, *recA*-G4-2 and *recA*-G8-5 amongst isolates (data not shown).

Table 5. Agrobacterial species detected in an environmental community according to detection methods.

Detected species	Isolation	Cloning-sequencing		Direct PCR
		Rhizobiaceae	<i>Agrobacterium</i>	
<i>A. tumefaciens</i> G1	3	0	3	4/6
<i>A. tumefaciens</i> G2	0	0	0	3/6
<i>A. tumefaciens</i> G3	0	0	0	0/6
<i>A. tumefaciens</i> G4	1	0	0	3/6
<i>A. tumefaciens</i> G5	0	0	0	0/6
<i>A. tumefaciens</i> G6	0	0	0	5/6
<i>A. tumefaciens</i> G7	0	0	0	4/6
<i>A. tumefaciens</i> G8	1	11	44	3/6
<i>A. tumefaciens</i> G9	0	0	0	0/6
<i>A. tumefaciens</i> G13	0	0	0	0/6
Other <i>Agrobacterium</i> sp.	0	8	0	nt
<i>Rhizobium</i> spp.	2	11	0	nt

Numbers indicate the number of bacterial colonies isolated from the community (Isolation), of cloned fragments (Cloning-sequencing) or of positive amplification out a total of six assays (Direct PCR) done with the community metagenomic DNA. nt, not tested. Cloning-sequencing was done with F7386-F7387 or F8360-F8361 primer sets that specifically amplify Rhizobiaceae or *Agrobacterium* spp., respectively. Direct PCRs were done with the species-specific primers described in Table 2.

Presence of *Agrobacterium* species in the community was then investigated by direct PCR using species-specific primers with metagenomic DNA. As expected from the microbiological and cloning-sequencing results presented above, G1 and G8 were repeatedly detected in this way (Table 5). The microbiological approach has already pointed the presence of a G4 isolate, which is confirmed by visualisation of a positive signal with G4 specific primers. However, the method also revealed the probable presence of G2, G6 and G7 in this community, whereas they were not detected by the standard microbiological or by cloning-sequencing approaches based on primers for superior taxa. No amplification was obtained with other tested species-specific primers (G3, G5, G9, G13).

Discussion

To achieve our dual objective of rapid agrobacterial species identification and agrobacterial diversity assessment in complex environments, we designed *recA* based PCR primers with specificities at different taxonomic levels. The housekeeping *recA* gene has been established as a useful target for the identification of species inside closely related taxa (e.g. Rhizobiaceae, *Burkholderia cepacia* complex, *Brucella* spp; Gaunt et al. 2001, Payne et al. 2006, Scholz et al. 2008). Costechareyre et al. (2010) showed that *recA* gene sequencing can discriminate *Agrobacterium* species, contrary to 16S rDNA sequences (Mougel et al. 2002).

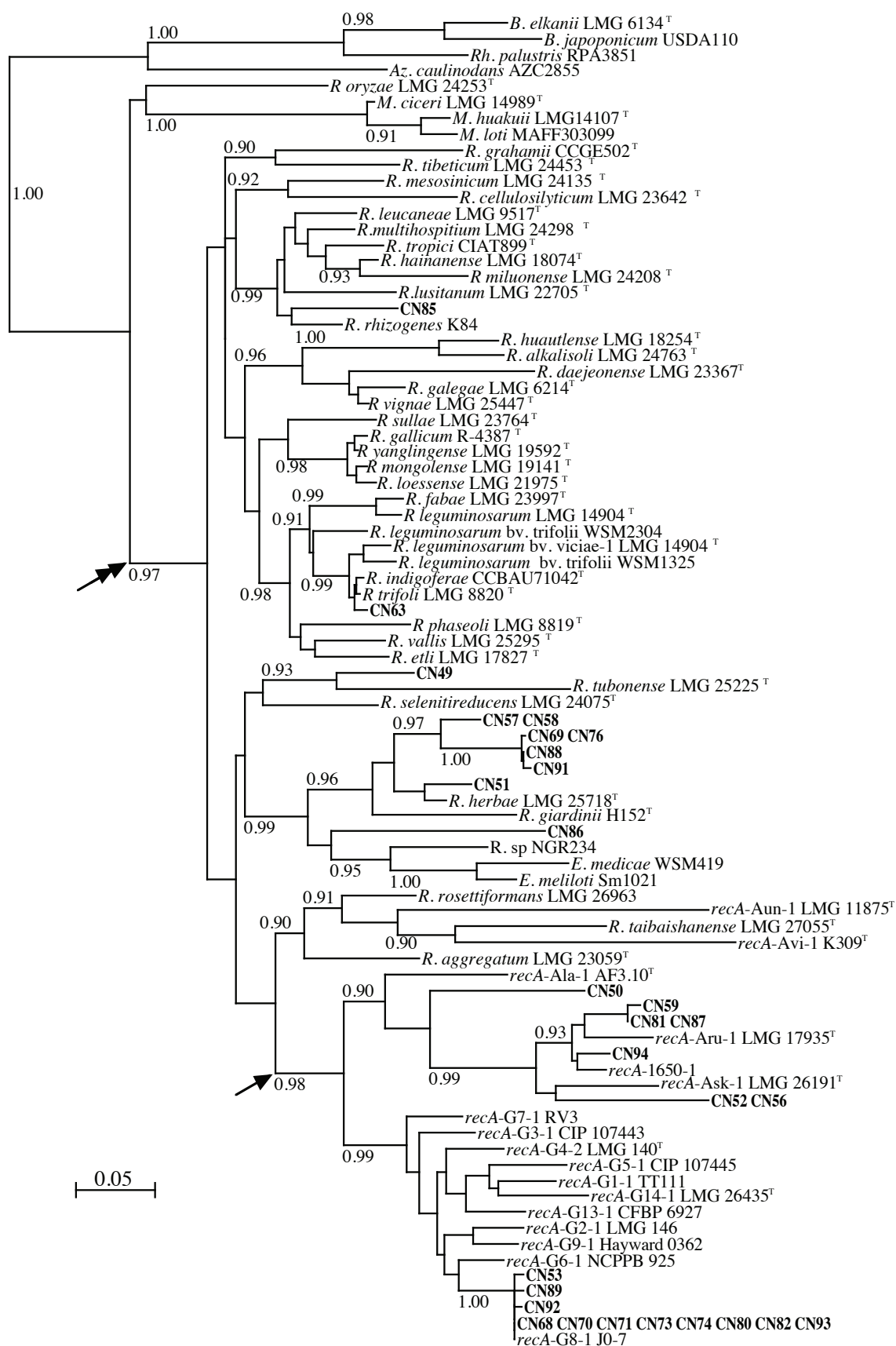


Figure 2-2-5. Phylogenetic position of *recA* sequence features delivered by cloning-sequencing of an amplicon obtained using Rhizobiaceae-specific primers F7386-F7387 with an environmental DNA extracted from a Poaceae rhizosphere community.

The maximum-likelihood method (ln(L)=-15771.8, 1101 sites, GTR 4 rate classes) was used to analyze *recA* sequences obtained by cloning-sequencing (red) or from GenBank selecting one sequence per Rhizobiaceae species (black). *B.*, *Bradyrhizobium*; *Rh.*, *Rhodopseudomonas* *Az.*, *Azorhizobium*; *R.*, *Rhizobium*; *E.*, *Ensifer*; *Avi.*, *Agrobacterium vitis*; *Aun.*, *Allorhizobium undicola*; *Aru.*, *A. rubi*; 1650, *A. sp.*; *Ask.*, *A. skieniewiczense*; G1 to G14, *A. tumefaciens* genomovars; G14, *A. nepotum*; G4, *A. radiobacter*; G2, *A. pusense*; G8, *A. fabrum*. Double arrow indicates Rhizobiaceae potentially amplifiable with Rhizobiaceae-specific primers F7386-F7387. Single arrow indicates the main species subset amplified using *Agrobacterium*-specific primers F8360-F8361. Numbers indicates clade probabilities. Branch length unit: number of substitutions per nucleotide site.

Reliable taxonomic identification would require further investigations such as the sequencing of other housekeeping genes, as proposed with the multi-locus sequencing analysis (MLSA) approach (Gevers et al., 2005). Actually, taxonomic identification based on a single gene could be incorrect in case of horizontal gene transfer (HGT) between strains of different species (a strain readily belonging to a given species may have acquired a foreign *recA* allele). Nevertheless, in our hands, MLSA and comparisons of complete genomes reliably showed us that *recA* was not prone to interspecies HGT within the *Agrobacterium tumefaciens* complex (unpublished results). We are thus reasonably confident about the taxonomic assignments of agrobacteria when *recA* was used alone. Moreover, thanks to its stability within species, *recA* is a good candidate to detect the presence of agrobacterial species in composite communities. We assumed that having a positive amplification signal with given primer sets is an easy and efficient way to have a first idea of the taxonomic status of any novel *Agrobacterium* isolate.

Designing of species-specific primers was possible thanks to the availability of numerous *recA* sequences longer than those usually found in databases (mostly from MLSA sequence data). Indeed, MLSA analysis generally delivered ca. 450-500 bp fragments matching the 45-550 region of *recA* in C58, while the present study showed that relevant regions for the designed specific primers were often outside of this region (Table 2). Furthermore, perfect specific amplifications were obtained using species-specific primers, providing a novel way for rapid and accurate identification of *Agrobacterium* species.

Primers designed for superior taxa enabled amplification of most species that were usually classified within *Agrobacterium* spp. or Rhizobiaceae. The so-called *Agrobacterium*-specific primers were found to be able to amplify almost all *Agrobacterium* species except *A. vitis*, while it allowed the amplification of *Allorhizobium undicola*. This illustrates the close relatedness of the later species and might be an argument to include *Allorhizobium undicola* in the genus *Agrobacterium* since this inclusion is still a pending question (Lindström and Young, 2011). As

a consequence, it is likely that primers F8360 and F8361 target a very large subset of *Agrobacterium* spp. and only *Agrobacterium* spp. as shown by results obtained with DNA extracted both from synthetic and environmental community (Table 2-2-4 and 2-2-5). The Rhizobiaceae-specific primers allowed amplifying *recA* sequences from available isolates of *Agrobacterium*, *Ensifer* and *Rhizobium* species (Table 2-2-5). *A. undicola* was not amplified by these primers. However, they amplify all other Rhizobiaceae members tested in this study. Moreover using those primers on metagenomic DNA from an unknown community allowed us to detect sequences of Rhizobiaceae family members and no sequences of other family (Figure 5). This shows that *Agrobacterium* and Rhizobiaceae-specific primers could be useful tools to study more easily *Agrobacterium* and Rhizobiaceae biodiversity in complex environments, i.e. without standard microbiology methods requiring strains isolation and biochemical typing tests. All primer sets were found to be relevant for the identification of the taxonomic status of bacteria at the family (Rhizobiaceae), genus (*Agrobacterium*) or species levels of bacteria in pure culture, but also for simply detecting the presence of agrobacteria in composite environmental DNAs as shown with the known synthetic community or the unknown environmental community (Table 2-2-4 and 2-2-5). Notably, primers designed for superior taxa allowed us to consistently detect the presence of Rhizobiaceae and *Agrobacterium* spp. in all the soils we have tested to date. This was verified in the present study for the French LCSA soil (Table 2-2-5), but also for 66 soils sampled at 16 locations from northern to the middle of Tunisia, within orchards, oasis and uncultivated semi-desert stands (data not shown).

The cloning-sequencing approach revealed that G8 agrobacteria prevailed over all other *Agrobacterium* or Rhizobiaceae species at LCSA, but species specific primers also showed the occurrence of several other species (Table 2-2-5). This was a confirmation of previous microbiological inspections that reliably led to the isolation of G1, G4, G8 and G2 and G7 agrobacteria at LCSA (Vogel et al., 2003; and unpublished results). Thus, species specific primers could be useful to point the presence at low density of particular *Agrobacterium* species in environmental samples. However, the agrobacterial taxonomic diversity assessment was more rapidly obtained by analyzing the allele content of amplicons obtained with primers targeting superior taxa (Tables 2-2-5). Hence, this approach also revealed that *recA*-G8-1 was the dominant allele in this rhizosphere. This was in complete agreement with previous microbiological studies that showed the frequent presence in this soil of agrobacteria such as J-07 that had a typical *recA*-G8-1 allele (Vogel et al., 2003; Cosechareyre et al., 2010). Finally, the

prevalence of G8 noted in the present study suggested that G8 agrobacteria were preferentially recruited in a particular rhizosphere. This kind of analyses used at larger scale could help deciphering the ecology of the different agrobacterial species such as finding factors that control the preferential recruitments of *Agrobacterium* species by particular rhizospheres.

In conclusion, the microbiological approach used so far to identify *Agrobacterium* species is hard to use for exhaustive investigations of the species composition of bacterial communities. Thanks to the *recA* specific primers designed in the present study, it is now possible to achieve an accurate and rapid first identification of all *Agrobacterium* species in both pure cultures and complex environments. Coupled with a cloning-sequencing approach, primers targeting superior taxa will allow investigation of agrobacterial and Rhizobiaceae microdiversity in the different biotopes they inhabit, including bulk and rhizosphere soils, tumors or even legume nodules for the purposes of thorough ecological, agroecological and epidemiological studies.

Acknowledgment

This work was supported by the EcoGenome project of Agence Nationale de la Recherche (grant number ANR-08-BLAN-0090). M.S. was supported by a CNRS contract in the framework of EcoGenome. The authors would like to thank R. Mhamdi and J. Puławska for providing novel strains of novel species, P. Portier at CFBP (<http://www-intranet.angers.inra.fr/cfbp/>) for strain repository, and the translator Dr D. Manley for reading the manuscript and providing suggestions.

References

- Hao X, Xie P, Johnstone L, Miller SJ, Rensing C, Weia G. 2012. Genome sequence and mutational analysis of plant-growth-promoting bacterium *Agrobacterium tumefaciens* CCNWGS0286 isolated from a zinc-lead mine tailing. *Appl. Environ. Microbiol.* 78:5384–5394.
- 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.
- Costechareyre D, Rhouma A, Lavire C, Portier P, Chapulliot D, Bertolla F, Boubaker A, Dessaux Y, Nesme X. 2010. Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis. *Microb. Ecol.* 60:862–872.
- Lindström K, Young JPW. 2011. International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*: Minutes of the

- meeting, 7 September 2010, Geneva, Switzerland. *Int. J. Syst. Evol. Microbiol.* 61:3089–3093.
- Shams M, Campillo T, Lavire C, Muller D, Nesme X, Vial L. 2012. Rapid and Efficient Methods to Isolate, Type Strains and Determine Species of *Agrobacterium* spp. in Pure Culture and Complex Environments. In Dr. Jose C. Jimenez-Lopez (ed.)-Biochemical Testing. InTech. DOI: 10.5772/2250
- Keane PJ, Kerr A, New PB. 1970. Crown gall of stone fruit. II. Identification and nomenclature of *Agrobacterium* isolates. *Aust. J. Biol. Sci.* 23:585–595.
- Puławska J, Willems A, De Meyer SE, Süle S. 2012. *Rhizobium nepotum* sp. nov. isolated from tumors on different plant species. *Syst. Appl. Microbiol.* 35:215–220.
- Panday D, Schumann P, Das SK. 2011. *Rhizobium pusense* sp. nov., isolated from the rhizosphere of chickpea (*Cicer arietinum* L.). *Int. J. Syst. Evol. Microbiol.* 61:2632–2639.
- Lassalle F, Campillo T, Vial L, Baude J, Costechareyre D, Chapulliot D, Shams M, Abrouk D, Lavire C, Oger-Desfeux C, Hommais F, Guéguen L, Daubin V, Muller D, Nesme X. 2011. Genomic species are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. *Genome Biol. Evol.* 3:762–781.
- Pulawska J, Willems A, Sobiczewski P. 2012. *Rhizobium skierniewicense* sp. nov. isolated from tumours on chrysanthemum and cherry plum. *Int. J. Syst. Evol. Microbiol.* 62: 895-899
- Popoff MY, Kersters K, Kiredjian M, Miras I, Coynault C. 1984. Position systématique de souches de *Agrobacterium* d'origine hospitalière. *Ann. Inst. Pasteur Microbiol.* 135:427–442.
- 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.
- Mougel C, Thioulouse J, Perrière G, Nesme X. 2002. A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* 52:573–586.
- Wibberg D, Blom J, Jaenicke S, Kollin F, Rupp O, Scharf B, Schneiker-Bekel S, Szczepanowski R, Goesmann A, Setubal JC, Schmitt R, Pühler A, Schlüter A. 2011. Complete genome sequencing of *Agrobacterium* sp. H13–3, the former *Rhizobium lupini* H13–3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J. Biotechnol.* 155:50–62.
- De Meyer SE, Van Hoorde K, Vekeman, B, Braeckman T, Willems A. 2011. Genetic diversity of rhizobia associated with indigenous legumes in different regions of Flanders (Belgium). *Soil Biol. Biochem.* 43:2384–2396.
- Li A, Geng J, Cui D, Shu C, Zhang S, Yang J, Xing J, Wang J, Ma F, Hu S. 2011. Genome Sequence of *Agrobacterium tumefaciens* Strain F2, a Bioflocculant-Producing Bacterium. *J. Bacteriol.* 193:5531.
- Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF Jr, Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D Sr, Chapman P, Clendenning J, Deatherage G, Gillet W, Grant C, Kuttyavin T, Levy R, Li MJ, 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 ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science.* 294:2317–2323.

- Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Quorollo B, Goldman BS, 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.
- Ruffing AM, Castro-Melchor M, Hu WS, Chen RR. 2011. Genome Sequence of the Curdlan-Producing *Agrobacterium* sp. Strain ATCC 31749. *J. Bacteriol.* 193:4294–4295.
- Martens M, Delaere M, Coopman R, De Vos P, Gillis M, Willems A. 2007. Multilocus sequence analysis of *Ensifer* and related taxa. *Int. J. Syst. Evol. Microbiol.* 57:489–503.
- Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, Burr TJ, Banta L, Dickerman AW, Paulsen I, Otten L, Suen G, Welch R, Almeida NF, Arnold F, Burton OT, Du Z, Ewing A, Godsy E, Heisel S, Houmiel KL, Jhaveri J, Lu J, Miller NM, Norton S, Chen Q, Phoolcharoen W, Ohlin V, Ondrusek D, Pride N, Stricklin SL, Sun J, Wheeler C, Wilson L, Zhu H, Wood DW. 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* 191:2501–2511.
- Kawaguchi A. 2011. Genetic diversity of *Rhizobium vitis* strains in Japan based on multilocus sequence analysis using the sequences of *pyrG*, *recA* and *rpoD*. *J. Gene. Plant. Pathol.* 77:299–303.
- Young JM, Pennycook SR, Watson DR. 2006. Proposal that *Agrobacterium radiobacter* has priority over *Agrobacterium tumefaciens*. Request for an opinion. *Int. J. Syst. Evol. Microbiol.* 56:491–493.
- González V, et al. González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramírez MA, Jiménez-Jacinto V, Collado-Vides J, Dávila G. 2006. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci.* 103: 3834–3839.
- Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668–672.
- 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.
- Ophel K, Kerr A. 1990. *Agrobacterium vitis* sp. nov. for strains of *Agrobacterium* biovar 3 from grapevines. *Int. J. Syst. Bacteriol.* 40:236–241.
- Huber T, Faulkner G, Hugenholtz P. 2004. Bellerophon; a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*. 20:2317–2319.
- Gaunt MW, Turner SL, Rigottier-Gois L, Lloyd-Macgilps SA, Young JPW. 2001. Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int. J. Syst. Evol. Microbiol.* 51:2037–2041

Chapter 2-3

**Assessing the presence of agrobacteria and Ti plasmids in soil
at a country scale.**

Agrobacteria are considered as common soil and rhizosphere inhabitants. However, to our knowledge, there is no exhaustive data about their actual presence in a large variety of soils (Chapter 1-3). The same question could be addressed to the presence of Ti plasmid harboring agrobacteria. Indeed, it is not known whether pathogenic or Ti plasmid-free agrobacteria are present everywhere or if biotic or abiotic factors may influence their presence in particular environments. Answering to this question may have important issues for crown gall control itself because finding factors that limit the survival of agrobacteria and especially of the Ti plasmid harboring ones may provide us new ways to control perennial soil contaminations by pathogenic agrobacteria.

In this chapter we thus addressed the question of the presence of i) agrobacteria and ii) Ti plasmids in soils at the country and local scales. For this purpose, we sampled soils in various locations in Tunisia consisted of agronomic and non-agronomic soils as well as soils collected in extreme condition such as salty and semi-arid areas that had or had not known history of crown gall.

Material and methods

Soil sampling and processing.

To look for the presence of agrobacteria, 16 stations located from northern to the middle of Tunisia have been arbitrarily selected. Sampled stations consisted of orchard, oasis and uncultivated semi-desert plots (Fig. 2-3-1), with 1 to 10 different sampling points per station for a total of 66 (Table 2-3-1). A special focus was given to the governmental fruit tree nursery located at Chbika (Kairouan governorate) in which one plot (Chbika-1) has repeatedly been reported to be infested by crown gall (Rhouma et al., 2006; Costechareyre et al., 2010) (Fig. 2-3-1').

GPS coordinates recorded for each sampling point were transformed into decimal degrees using the on-line service http://www.toutimages.com/google_gps/sis_gps.htm. Soil samples (ca. 40 g per point) were collected in the 5-10 cm superficial soil layers in sterile Falcon tubes, sealed and stored at greenhouse temperature (25°C) up to construction of individual soil microcosmes in the following weeks. Microcosmes consisted of sterile glass tubes containing 1.3 g perlite filled with ca. 20 g soil, and one germinated seed of *Medicago truncatula* ecotype Jemalong line A17. Before sowing, seeds had been surface sterilized with hypochlorite, scalpel scarified, vernalized at 4°C during 48 h then overnight germinated on sterile agar plate at 25°C in the dark.

Microcosmes irrigated with 1 ml of sterile nutrient solution (N/P/K: 18/6/26 , Soluplant, Duclos International, Lunel Viel, France) just after sowing were incubated in a growth chamber for three weeks at 28°C and 70% of humidity and a 16/8 day/night regime, with irrigations once a week with 1 ml of nutrient solution. After 3 weeks, microcosmes containing seedlings at the 2 leaf stage (Fig. 2-3-2) were treated to separate bulk from rhizosphere soils.

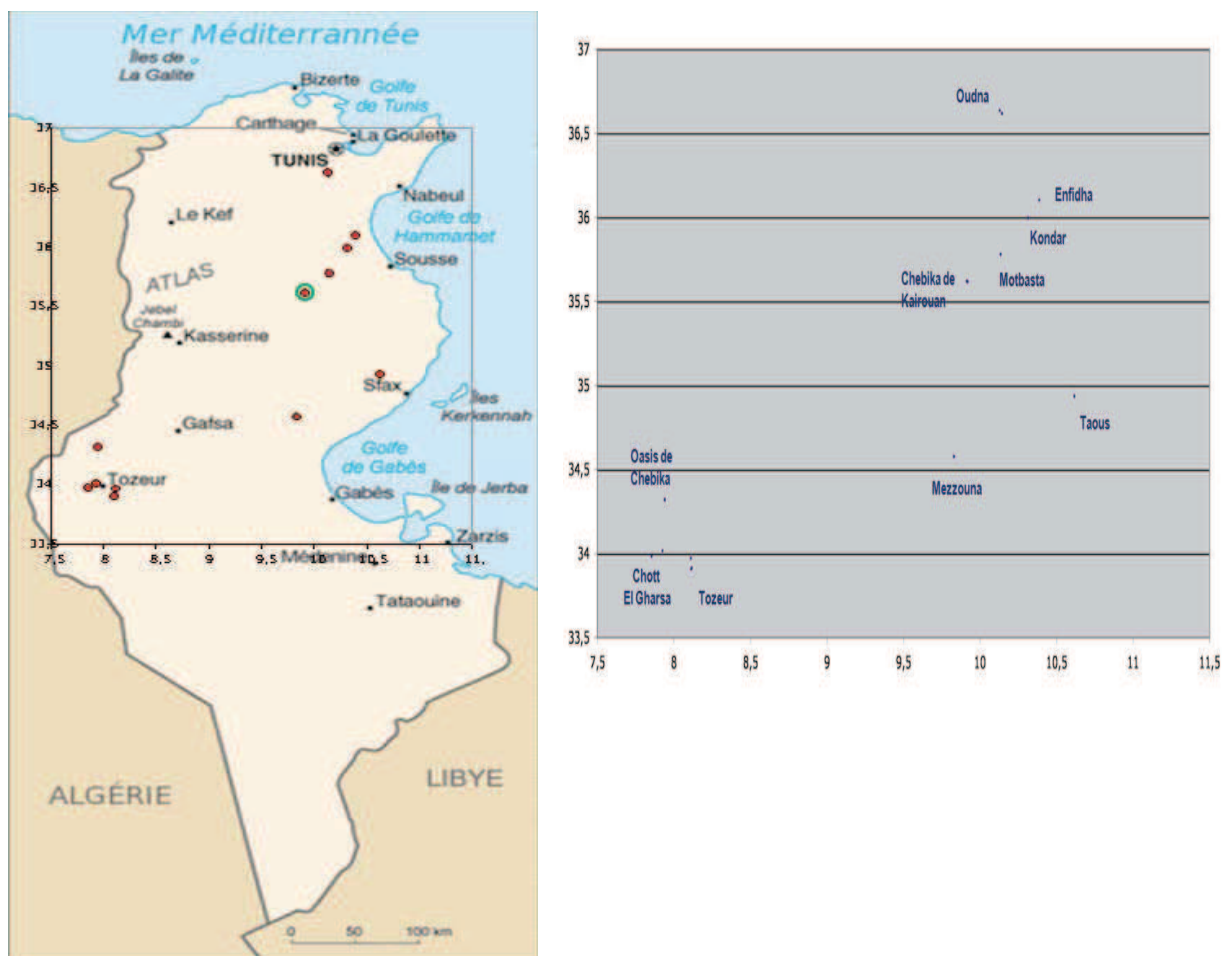


Figure 2-3-1. Location of soils selected for assessing the agrobacterial diversity at a country scale.
A) Location on the Tunisian map. red, selected locations; green, Chbika (Kairouan governorate). B) GPS coordinates (decimal degrees) of locations.

Table 2-3-1. Location of Tunisian samples used in this study.

Name	Station	Latitude	Longitude	Altitude
TA1	LOUDNA-1	36°38'06,0	10°07'41,0	60
TA2	LOUDNA-1	36°38'05,7	10°07'41,5	
TB1	LOUDNA-2	36°37'03,4	10°08'39,1	65
TB2	LOUDNA-2	36°37'03,2	10°08'38,8	
TB3	LOUDNA-2	36°37'02,3	10°08'36,1	64
TB4	LOUDNA-2	36°37'00,5	10°08'37,6	60
TB5	LOUDNA-2	36°37'00,9	10°08'40,7	
TB6	LOUDNA-2	36°37'01,8	10°08'42,5	76
TB7	LOUDNA-2	36°37'03,4	10°08'41,7	73
TB8	LOUDNA-2	36°37'12,2	10°08'36,7	
TC1	ENFIDHA	36°06'06,4	10°23'12,2	21
TC2	ENFIDHA	36°06'06,2	10°23'11,7	14
TC3	ENFIDHA	36°06'06,2	10°23'10,9	
TD1	KONDAR	35°59'48,2	10°18'51,6	26
TD2	KONDAR	35°59'47,2	10°18'51,4	25
TD3	KONDAR	35°59'46,4	10°18'51,1	27
TE1	MOTBASTA	35°46'47,6	10°08'05,6	31
TE2	MOTBASTA	35°46'47,5	10°08'05,0	39
TE3	MOTBASTA	35°46'47,2	10°08'04,1	
TF1	CHBIKA-1	35°37'04,8	09°55'05,3	131
TG1	CHBIKA-1	35°37'04,5	09°55'05,4	130
TG2	CHBIKA-1	35°37'05,0	09°55'05,7	
TG3	CHBIKA-1	35°37'05,6	09°55'06,1	125
TG4	CHBIKA-1	35°37'06,2	09°55'06,6	124
TG5	CHBIKA-1	35°37'06,8	09°55'07,0	127
TG6	CHBIKA-1	35°37'05,4	09°55'08,2	122
TG7	CHBIKA-1	35°37'05,6	09°55'07,5	124
TG8	CHBIKA-1	35°37'05,8	09°55'06,7	125
TG9	CHBIKA-1	35°37'05,9	09°55'06,0	125
TG10	CHBIKA-1	35°37'06,2	09°55'05,1	
TH1	CHBIKA-2	35°37'05,0	09°55'03,6	125
TH2	CHBIKA-2	35°37'05,6	09°55'01,9	126
TH3	CHBIKA-2	35°37'06,6	09°54'59,4	127
TH4	CHBIKA-2	35°37'04,0	09°55'00,0	125
TH5	CHBIKA-2	35°37'07,2	09°55'02,0	124
TI1	CHBIKA-3	35°37'07,3	09°54'46,3	130
TI2	CHBIKA-3	35°37'07,8	09°54'46,1	124
TI3	CHBIKA-3	35°37'08,6	09°54'45,8	131
TI4	CHBIKA-3	35°37'08,4	09°54'45,5	125
TI5	CHBIKA-3	35°37'12,7	09°54'45,3	130
TI6	CHBIKA-3	35°37'12,6	09°54'44,8	127
TJ1	CHBIKA-4	35°37'03,5	09°55'06,0	105
TJ2	CHBIKA-4	35°37'03,1	09°55'07,5	132
TJ3	CHBIKA-4	35°37'02,0	09°55'08,3	125
TJ4	CHBIKA-4	35°37'01,0	09°55'09,5	130
TK1	TAOUS	34°56'08,9	10°37'00,5	124
TK2	TAOUS	34°56'07,7	10°37'00,5	112
TK3	TAOUS	34°56'06,0	10°37'00,0	124
TL1	MEZZOUNA	34°34'33,6	09°49'45,8	157
TL2	MEZZOUNA	34°34'33,7	09°49'47,0	159
TL3	MEZZOUNA	34°34'35,1	09°49'46,4	155
TL4	MEZZOUNA	34°34'35,1	09°49'46,4	
TM1	TOZEUR-1	33°58'16,8	08°06'39,8	32

TM2	CHOTT EL-GHARSA	34°00'59,7	07°55'33,3	-25
TM3	CHOTT EL-GHARSA	34°00'59,7	07°55'32,2	-27
TM4	CHOTT EL-GHARSA	33°58'59,4	07°51'18,1	-9
TN1	TOZEUR-2	33°54'48,7	08°06'58,2	55
TN2	TOZEUR-2	33°54'49,1	08°06'58,5	59
TN3	TOZEUR-2	33°54'49,7	08°07'00,8	46
TO1	CHEBIKA OASIS	34°19'23,9	07°56'23,2	145
TO2	CHEBIKA OASIS	34°19'23,3	07°56'27,0	141
TO3	CHEBIKA OASIS	34°19'18,3	07°56'24,4	168
TO4	CHEBIKA OASIS	34°19'04,3	07°56'23,3	112
TP1	TOZEUR-3	33°54'32,7	08°06'46,1	
TP2	TOZEUR-3	33°54'39,2	08°06'54,8	56
TP3	TOZEUR-3	33°54'37,4	08°06'57,0	60

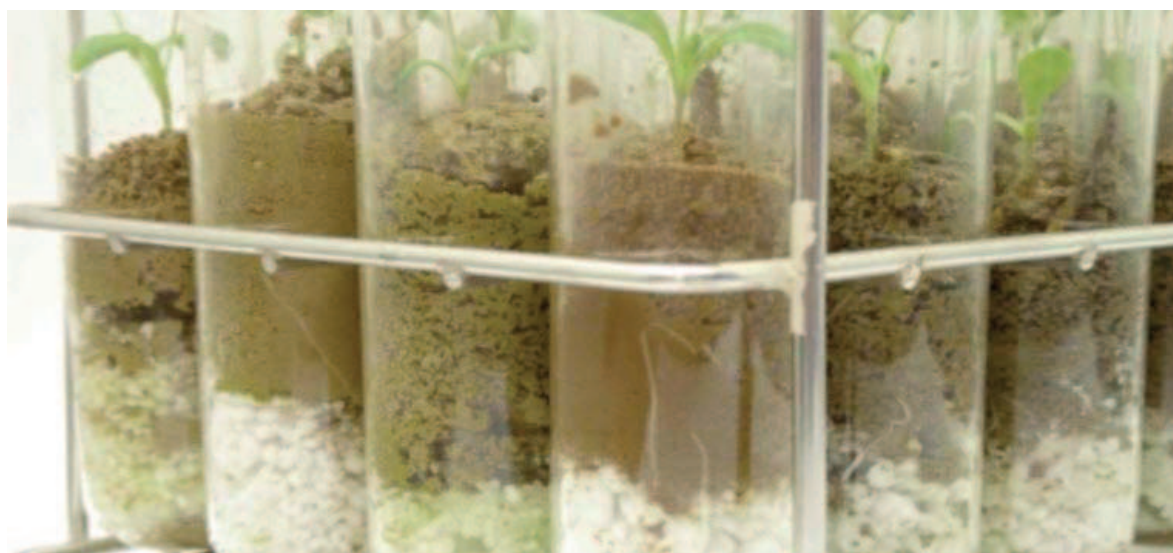


Figure 2-3-2. *Medicago truncatula* genotype Jemalong A17 seedlings after 3 weeks of growth.

***Agrobacterium* isolation and metagenomic DNA extraction.**

For each microcosme, soil was divided into bulk and rhizosphere soil compartments. Bulk soil compartments consisted of visually root free soil fractions and rhizosphere soil compartments consisted of soil fractions tightly adhering to roots. The later compartments were obtained by removing non-adherent soil particules of roots by vigorous manual shaking. Bulk soils (ca. 1 g) as well as shaken root systems were suspended into 1 ml sterile water under agitation for 2 hours. Soil suspension aliquots were used to isolate agrobacteria, while remaining aliquots deprived of root were centrifuged at 4000 g for 10 min to obtain soil pellets for subsequent soil metagenomic

DNA extractions. *Agrobacterium tumefaciens* complex bacteria were isolated using the elective 1A-Te medium as described by Mougél et al. (2001). The bona fide *Agrobacterium* status of isolates was done using biochemical methods described Chapter 1-1. Metagenomic DNAs were extracted from soil samples using the PowerSoil® DNA isolation kit (Mo Bio, Carlsbad, CA, US).

Primers and PCR conditions.

Primers used to detect and identify agrobacteria were defined Chapter 2-2, while primers F14 and F749 used to detect Ti plasmids were described Chapter 1-1. PCRs were carried out as described in these chapters

Statistics

Chi square statistics were calculated on-line at <http://marne.u707.jussieu.fr/biostatgv/?module=tests/chideux>

Results

***Agrobacterium* prevalence in Tunisian soils.**

Presence of agrobacteria in soils was first assessed by PCR using *Agrobacterium* specific primers with sample metagenomic DNAs. Positive amplifications were obtained for all the 132 samples that included equal numbers (66) of bulk and rhizospheric soils. This molecular detection was confirmed by the isolation of *A. tumefaciens* bacteria from all samples. Using species specific primers, we found that isolates gave positive PCR results either with G4 or G7 specific primers, 97/132 or 35/132, respectively, with no cross reactions with other primer sets as expected. There were more G4 than G7 isolates, but the two species were equally distributed in bulk and rhizosphere soil samples (Chi square=0.039; N=132; df=1; p-value=0.843), indicating no effect of biotope compartment upon *A. tumefaciens* species relative frequency (Table 2-3-2).

A more exhaustive analysis of the agrobacterial diversity in soils was performed using species-specific primers with metagenomic DNAs of the Chbika-1 station. All samples yielded easily visible bands of expected sizes using G4 and G7 specific primers as well as faint bands with *A. larrymoorei* and *A. vitis* (Fig. 2-3-3) specific primers. Conversely, no amplifications were obtained using G1, G2, G3, G5, G6, G8, G9, G13, G14 specific primer sets in any samples (data not shown).

Presence of Ti plasmids in different Tunisian soils.

We looked for the presence of Ti plasmids by performing PCRs with primers F14 and F749, that permit the amplification of 432 or 384 bp fragment of the *vir* regions of nopaline or octopine and agrocin-mannopine type Ti plasmids, respectively (Chapter 1-1). No amplifications were obtained with *Agrobacterium* isolates, but bands of ca. 432 bp were obtained with 36/132 metagenomic DNAs. Amplifications occurred equally in bulk and rhizosphere soil samples (Chi square=1.375; N=132; df=1; p-value=0.241), indicating no effect of biotope compartment upon Ti plasmid harboring bacteria distribution (Table 2-3-3). Moreover, Ti plasmids were generally detected in both the bulk and the rhizosphere soils of the same sampling points, indicating a high level of reliability of the detection for a given sampling point corresponding to a single microcosme (data not shown).

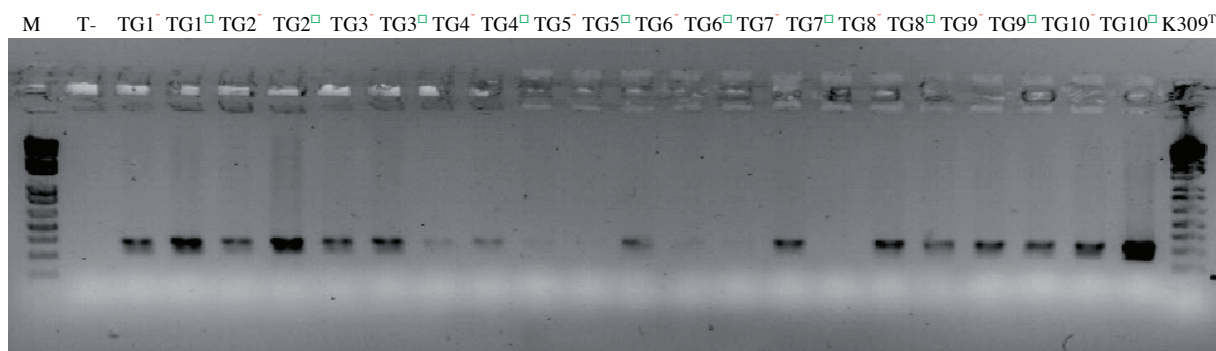


Figure 2-3-3. PCR amplification in metagenomic DNA from Chbika using *A. vitis*-specific primers (G0004-G0005).
 ◻, Bulk soil; ◻, rhizospheric soil; T-, negative control; K309T, *A. vitis*; M, 1kb + Invitrogen DNA marker.

In summary, *vir* ampli were obtained in 12/16 Tunisian locations (Table 2-3-4). The greatest numbers of detection were obtained at Chbika-1, that was known to be severely contaminated by pathogenic agrobacteria. However, there was not detection at Oudna-2 which is a station where large crown gall tumors were easily visible in the close vicinity of sampling points. Conversely, Ti plasmids were detected with soils of Motbasta and Tozeur-1 that were both non-agronomic soils from a salty and a semi-arid area, respectively.

Table 2-3-2. Number of isolates per *Agrobacterium* species Tunisian soils according to biotope compartments.

Biotope	Genomovar		Total
	G4	G7	
BS	48	18	66
RS	49	17	66
Total	97	35	132

BS, bulk soil; RS, rhizopshere soil. Species identifications of bacterial isolates were done according to PCR results using G4 or G7 specific primers defined Chapter 2-2.

Table 2-3-3. Distribution of Ti plasmid in Tunisian soils according to sampled biotopes.

Biotope	Vir+	Vir-	Total
BS	21	45	66
RS	15	51	66
Total	36	96	132

BS, bulk soil; RS, rhizopsheric soil. Ti plasmid presence in metagenomic DNAs were inferred from PCR results using Ti plasmid sspecific primers described Chapter 1-1.

Table 2-3-4. PCR based Ti plasmid detection in Tunisian soils.

Location	Soil characteristic	pTi ratio	sampling point
Oudna-1	non agronomic soil	0/2	
Oudna-2	crown galled cherry orchard	0/9	
Enfidha	ploughed fallow soil (after cereal)	1/3	TC2
Kondar	common olive orchard	1/3	TD2
Motbasta	non agronomic salty soil	1/3	TE1
Chbika-1	ploughed contaminated plot	7/11	TG1,2,3,5,6,7,8,10
Chbika-2	ploughed plot	1/5	TH3
Chbika-3	stone fruit tree nursery plot	1/6	TI5
Chbika-4	olive orchard plot	3/4	TJ1,2,4
Taous	olive orchard	1/3	TL2
Mezzouna	stone fruit orchard	1/4	TK2
Tozeur-1	non agronomic semi-arid soil	1/1	TM1
Chott-El-Gharsa	non agronomic soil in arid zone	0/3	
Tozeur-2	hotel park soil	1/3	TN1
Chebika oasis	oasis soil	2/4	T01,3
Tozeur 3	palm tree orchard	0/3	

pTi ratio, Ti plasmid detection number per sampling point number.

Discussion

In this chapter we addressed the question of the presence of agrobacteria and Ti plasmids in soils at a large scale. Taken together, our result showed that agrobacteria and especially members of the *A. tumefaciens* complex were present in all Tunisian soils we investigated. This result is remarkable and partly unexpected because some sampled soils were non-agronomic ones

collected in extreme environments such as arid and salty areas (Table 2-3-4). It is certainly too early to claim that agrobacteria are present in every soils on earth. Nevertheless, the present study showed us that we now have the suitable microbiological and metagenomic methodologies to perform extensive *Agrobacterium* prevalence studies.

Only G4 and G7 agrobacteria were isolated from Tunisian soils in the present study. This result was surprising because other species were very likely present in numerous soils, at least at Chbika-1 as previously described by Rhouma et al. (2006) and Costechareyre et al. (2010). As an hypothesis, we wonder if the unique presence of G4 and G7 amongst isolates could be an artifact due to faster growth speeds of G4 and G7 over other species. It is thus our further aim to compare growth speeds of *Agrobacterium* species on 1A-Te medium. Nevertheless, the fact that we got so numerous isolates of the same species in very different locations offered us the opportunity to address the novel question of the presence of unique strains of agrobacteria (of G4 or of G7) at a country scale. We unfortunately didn't have time to do this analysis, but this investigation would be easily amenable because we have both the availability of numerous isolates belonging to the same species in various locations and the suitable methodology to answer the question of Tunisian G4 or G7 strain uniqueness. For this purpose, the *recA* sequencing currently in progress would offer a first indication. Nevertheless, MLSA will be preferable in order to know whether G4 and G7 isolates belong to single clonal complex or to markedly different strains. This population genetic information would help us to understand if G4 and G7 agrobacteria have spread over all Tunisia or if their dispersal was limited to sampling points of the same stations. We expect having soon these results because it would be help us in understanding how long agrobacterial strain actually spread out of stations.

We found no Ti plasmid harboring agrobacteria among isolates, confirming that Ti plasmid-free, non-pathogenic agrobacteria are dominant in soils. Conversely, we detected Ti plasmid signature in many soil community DNAs. However, those so-called Ti plasmid-specific primers indeed detected *vir* homologous sequences in non-agronomic soils that had no reported history of crown gall infestations and conversely not in visually crown gall infested orchards (Table 2-3-4). The biological relevance of those results is questionable. In a one hand, if F14-F749 primers did amplify only Ti plasmid sequences, the lack of detection in crown gall infested stations might have been caused either by a low density of Ti plasmid-free agrobacteria that was below the PCR detection threshold, either because primers did not match the particular Ti plasmids that caused the visible crown galls. In any cases, if primers readily detected Ti plasmids, results would

suggest a wide dispersal of pathogenic agrobacteria in numerous soils. This is an unexpected and surprising result that absolutely requires additional investigations for definitive conclusions. To answer this question, we need to isolate pathogenic agrobacteria from stations that provided unexpected results, but we unfortunately didn't have the opportunity to perform these investigations. In the other hand, F14-F749 primers might simply have targeted homologous *vir* regions outside Ti plasmids. Indeed, performing blastn with the F14-F749 amplified *vir* region of C58 in GenBank, showed us that part of this region has a strong homology also with non-Ti plasmids such as a *vir* region of plasmid p42a in *Rhizobium etli* that matches significantly with F749, or as the symbiotic island of *Mesorhizobium loti* that matches F14 (data not shown). It is thus very likely that *vir* homologous regions occurred by chance in soils. Hence, we assume that results obtained with F14-F749 primers just provided a presumptive indication upon the presence of Ti plasmids in soils. Detecting other Ti plasmid-specific regions is therefore required to ascertain such a presence. Conserved T-DNA regions might be used for this purpose, targeting for instance the oncogene region as early proposed by Ponsonnet and Nesme (1994) and Pionnat et al. (1999). Indeed, in our opinion, the PCR amplification of a single region can't be a definitive method to ascertain the presence of Ti plasmid in environmental samples. Moreover, especially when investigation are undertaken for forensic purposes, isolation of viable pathogenic agrobacteria from suspect areas are required to definitively confirm the first indications given by PCR. In fine, pathogenicity testing on indicator plants is the ultimate and definitive clue for the presence of viable and pathogenic agrobacteria. Nonetheless, direct PCRs remain faster. To gain in significance, we propose thus to use multiplex PCRs targeting *vir* plus other Ti plasmid-specific regions to achieve trustable Ti plasmid detection.

As a general feature, the present analysis showed us that bulk and rhizosphere soils of the same sampling points mostly led to the isolation of the same *Agrobacterium* species (Table 2-3-2) or the same detection of Ti plasmid (Table 2-3-3). This could lead us to suggest that there was no selective plant effect upon the recruitment of *Agrobacterium* species or of Ti plasmid harboring bacteria in rhizospheres. This suggestion is however questionable. In our opinion, the absence of bulk vs. rhizosphere effect might be caused by a failure in our sampling procedure. Indeed, we collected bulk and rhizosphere soils in the same microcosmes and not in independent test tubes. It is thus very likely that cross contaminations between the two compartments might occur. Actually, in those relatively small test tubes, root systems were rather developed (Fig. 2-3-2) and leaved us little chance to well select really root free soil fractions. However, if there is no real

compartment difference, the two samples of a given microcosme should be considered as simple replicates of the same soil community. If so, results however showed a high reliability of results obtained with two independent metagenomic DNAs extracted from the same community sampled corresponding to the microcosme scale.

References

- Rhouma A, Boubaker A, Nesme X, Dessaux Y (2006) Plasmid and chromosomal diversity of a Tunisian collection of *Agrobacterium tumefaciens* strains. Tunis J Plant Protect 1:73–84
- Ponsonnet C., Nesme X. 1994. Identification of *Agrobacterium* strains by PCR-RFLP analysis of pTi and chromosomal regions. Arch. Microbiol. 161 : 300-309.
- Pionnat S., Keller H., Héricher D., Bettachini A., Dessaux Y., Nesme X., Poncet C. 1999. Ti plasmids from *Agrobacterium* characterize rootstock clones that initiated an outspread of crown gall disease in Mediterranean countries. Appl. Environ. Microbiol. 65:4197-4206.
- Rhouma A., Boubaker A., Nesme X., Dessaux Y. 2007. Plasmid and chromosomal diversity of a Tunisian collection of *Agrobacterium tumefaciens* strains. Tunisian journal of plant protection 1:73-84.
- Costechareyre D., Rhouma A., Lavire C., Portier P., Chapulliot D., Bertolla F., Boubaker A., Dessaux Y., Nesme X. 2010. Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis. Microbiol. Ecol. 60:862-872.

Chapter 2-4

Assessing the pyrosequencing approach to analyze *Agrobacterium* microdiversity⁵

⁵ We are very indebt to C. Mougel and A. Zancarini who very kindly gave us plant associated microbial DNAs to study plant selective effect upon *Agrobacterium* communities.

In preceding chapters, we showed that analyzing the sequence composition of amplicons obtained from composite communities is a way to determine their taxonomic compositions. In these works, internal amplicon compositions were analyzed using the cloning-sequencing approach. The method was found relevant as it revealed the actual composition of synthetic communities of known taxonomic composition (Chapter 2-3). This methodology is nevertheless limited by the number of clones that can be reasonably sequenced for a given community (i.e. 50-100 at the best), hampering therefore the comparison of multiple communities.

Conversely, the new genome sequencing (NGS) technologies provide thousands of reads in a single run. Coupled with multiplexing NGSs allowed the simultaneous sequencing of numerous pooled samples and noticeably of pooled amplicons. NGSs therefore offer the scope for the development of affordable high throughput analysis of multiple communities. The present limit of NGSs is the maximum length of DNA fragment that can be sequenced. Indeed, sequence features must be long enough to carry enough evolutive informations for subsequent phylogenetic treatments. In this respect, the best methodology to date is the pyrosequencing technology with Titanium chemistry that permits the sequencing of DNA fragments of 450 bp. Hence, we designed *Agrobacterium* specific primers F8360 and F8361 described Chapter 2-2 precisely taking into account this technical constraint.

In the present work, we aimed at setting up and assessing the pyrosequencing approach to analyse agrobacterial microdiversity. For this purpose: i) we analyzed the sequence composition of amplicons obtained with communities of known taxonomic compositions; ii) we then used the approach to deliver an exhaustive investigation of a field agrobacterial community (i.e. Chbika-1) and to test whether plant genotypes had an effect upon agrobacteria recruited in rhizospheres.

Material and methods

Synthetic communities

To construct synthetic communities of known species allelic compositions, suspensions of overnight bacterial culture in standard LPG adjusted to the same OD₆₀₀ of 0.6 were assembled in defined quantities. These quantities were chosen in order to construct two kind of communities distinguished by their patterns of taxon density distribution. A first set, namely communities CS1 and CS2, were constructed with taxon densities distributed according to a "broken stick" model in which a small number of taxa strongly dominated numerous other taxa present at low and equivalent frequencies. A second set, namely communities CS3 and CS4, were constructed with

taxon density distributions that followed an exponential decrease model in which ranked taxa decreased according to a 1/2 ratio (Table 2-4-1).

***Agrobacterium* communities.**

Metagenomic DNAs used to evaluate the *Agrobacterium* species in a standard field plot was those of Chbika-1 described Chapter 2-3. Metagenomic DNAs used to test the plant genotype effect upon *Agrobacterium* species recruitment in rhizospheres were a kind gift of A. Zancarini and C. Mougel (INRA, Dijon, France), who prepared DNAs as explained in Zancarini et al. (2012). Selected genotypes were two Algerian ecotypes DZA 045-6, DZA 315-16 and the standard line Jemalong A17. Plants were grown in a silt-clay loam soil from Mas d'Imbert (France). Investigated compartments were bulk and rhizosphere soils, and root tissues.

Sequencing and data processing.

The Roche 454™ Genome Sequencer FLX™ System, with FLX Titanium chemistry was used to sequence pooled *recA* PCR products with MID labelling according to manufacturer's instructions. For this purpose, samples were 5'-labeled during PCR using forward primers consisted of the FLX adapter A, a sample specific barcode (MID) and the *Agrobacterium* specific *recA* sequence F8360 defined Chapter 2-3-2, and a reverse primer consisted of the FLX adapter B and the *Agrobacterium* specific *recA* sequence F8361 defined Chapter 2-3-2:

forward primer: 5'-CCATCTCACCCCTGCGTGTCTCCGACTCGAAGCTCGGTTCGAATGAAA-3'

reverse primer: 5'-CCTATCCCCCTGTGTGCCTTGGCAGTCTCAGGCTTGCGCAGCGCCTGGCT-3'

Sequencings were done using ¼ FLX slides for pools of 24 or 48 samples by Genoscreen (Lille, France). Raw sequence data were cleaned and denoised using the mothur v.1.25.0 pipeline (<http://www.mothur.org>), authorizing 1 error in barcode (bdiffs=) and 2 errors in primer (pdiffs=2), aligned against *recA* sequences of the fungene data base and cleaned of chimera retrieved with the Uchime module of mothur.

Data treatment and statistics

Cleaned sequences were classified in operational taxonomic units (OTUs) according to their best identity hit (Imax) with an exhaustive Rhizobiaceae *recA* database that essentially corresponded to those given in Table 2-3-1 using an analysis pipeline set up by A. Chosson. If $Imax < 90\%$, sequences were classified in the "unclassified *Rhizobium*" OTU class; if $90\% \leq Imax < 94\%$, in

"unclassified *Agrobacterium*"; if $95\% \leq I_{\max} < 98\%$, in the corresponding species; if $98\% \leq I_{\max} < 100\%$, in the corresponding *recA* allele. I_{\max} data were used to construct a contingency table giving the number of sequences per OTU of each sample.

Percentage confidence intervals were calculated on-line at http://ww3.ac-poitiers.fr/math/prof/resso/cali/ic_phrek.html#sommaire.

Results

Diversity of synthetic communities.

The first aim of the present study was to evaluate whether the pyrosequencing technology applied to analyze the sequence feature composition of *recA* amplicons would provide relevant data to analyze the microdiversity of *Agrobacterium* spp. In a first step, we used synthetic communities to evaluate the methodology abilities and limits. In the second step, the methodology was used to analyze the diversity of field *Agrobacterium* communities and to test whether plant genotypes had a selective effect upon this diversity.

A great majority of sequence features obtained with synthetic communities were readily assigned to species that were used to construct synthetic communities since only 13/9195 (1.3%) displayed less than 95% identity with any introduced species and were thus classified in the "undetermined *Agrobacterium* class" class (Table 2-4-2). This class was not an homogenous OTU, since it grouped unclassified sequences. Most of unclassified sequence features branched indeed at different places of the *Agrobacterium* phylogenetic tree. These sequences were "single singletons" that occurred only once in a single synthetic community. Even if no experimental data supports our assumption, we assumed that "single singletons" resulted by chance from stochastic errors occurring during PCR and/or sequencing and constituted the background noise of the methodology. Due to this background noise, the methodology could lead to false detections of species that were readily absent in communities. For this reason, species occurring as single singletons in an analysis were considered as falsely detected and removed of subsequent community analyses.

At the infra-specific level, 6196 sequence features (67%) were assigned to one of the introduced alleles by using the 0-1% divergence criterion (Table 2-4-2). Conversely, numerous sequences were not assigned to any introduced allele even if they were could be assigned to introduce species. This showed a severe limit of the methodology, which confidently allowed the

assignment of sequence features at the species level but difficultly at the allele level. In further studies, we only investigated the species status revealed by sequence features.

Comparisons of observed and expected occurrence of species in synthetic communities.

As a general feature, there was a rather good agreement between observed and expected species frequency distributions (Fig. 2-4-1). This was particularly verified for the most dominant species that actually dominated in all communities. Discrepancies occurred however with *A. rubi*, *A. larrymoorei*, *A. sp.* 1650 and *A. undicola* that were poorly or not detected at all. Control PCRs using the community DNAs with primers targeting the same *recA* region showed that *A. rubi* and *A. sp.* 1650 were present. *A. larrymoorei* was not amplified as expected. Nevertheless, for unknown reasons, *A. undicola* also failed to amplify with these primers, but was detected by using *A. undicola* specific primers defined in Chapter 2-2. This suggested that the lack of *A. undicola* sequences was caused by an unexpected mutation but was not related to a particular limitation of the method. Except these particular cases, the species distribution shapes were grossly the same between introduced and analyzed community, showing that the approach could be used to give a snapshot of the *Agrobacterium* species distribution in a community. Finally, as a rule of thumb, species of the *A. tumefaciens* complex were always detected if a minimum of 4 sequence features were expected.

***Agrobacterium* diversity assessment in a model field plot.**

Amplicons obtained with metagenomic DNAs of the Chbika-1 station (Cf. Chapter 2-3) delivered sequences that generally matched with known *Agrobacterium* species (Table 2-4-3), but an important set of sequences found in the sample TG9-BS did not match with any known *Agrobacterium* or *Rhizobium* species. These unclassified sequences were almost identical and showed a best identity of 87% with *A. sp.* 1650, suggesting they originate from the same species closely related but different of presently known *Agrobacterium*. As a result, the "unclassified *Agrobacterium*" class corresponded to a bona fide OTU. Similarly, the "undetermined *Rhizobium*" was an OTU for a single *Rhizobium* species related to *R. loessense*. These OTUs were kept in further analyses.

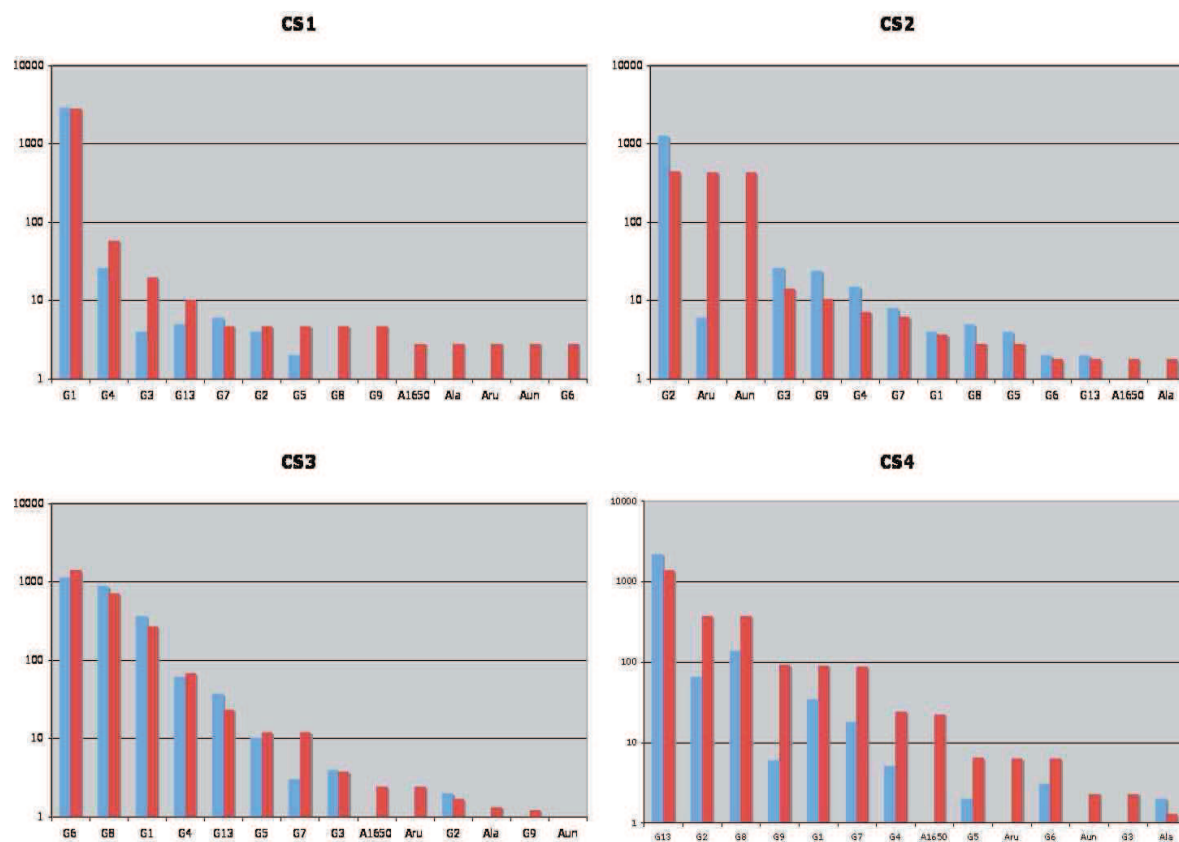


Figure 2-4-1. Observed and expected occurrence of *Agrobacterium* species in synthetic communities as given by multiplex pyrosequencing of a *recA* amplicon. Expected (red). Observed (blue). Axis: *Agrobacterium* species. Ordinate: log of sequence occurrence in amplicons. Bars indicate the 5% confidence interval for expected data.

Statistical analyses including principal component analysis (PCA) performed to compare bulk and rhizosphere soils (i.e. soil compartment), or to analyze effect of sample field geographic positions, showed that these parameters had no significant effect on sample species contents (data not shown). This result indicated indeed that the 19 samples were independent replicates of the Chbika-1 community.

From those analyses, the *Agrobacterium* species richness of Chbika-1 consisted in 13 *Agrobacterium* and 1 *Rhizobium* species. The number of species detected per sample varied from 1 to 6. Conversely, species were detected in only one sample for six species (underdetermined *Rhizobium*, *A. sp.* 1650, *A. undicola*, G3, G6, G9), while G14 was detected in 12/19 samples (Table 2-4-3). The latter species was the most prevalent in the Chbika-1 station. Simulations showed that dominant and sub-dominant species found in more than one samples were obtained by analyzing ca. 10 samples in average (data not shown). As a consequence, a good estimate of

the *Agrobacterium* species diversity for a field such as Chbika-1 would be obtained by analyzing the composition of 9 amplicons that could be obtained by sequencing three independent pools of three amplicons for each.

Table 2-4-1. Relative abundance of *Agrobacterium* strains assembled in to construct synthetic communities.

Species	Allele	Strain	CS1	CS2	CS3	CS4
<i>A. sp.</i> 1650	<i>recA</i> -Asp	NCPPB1650	1	1	64	1024
<i>A. larrymoorei</i>	<i>recA</i> -Ala	AF3-10	1	1	16	16
<i>A. rubi</i>	<i>recA</i> -Aru	LMG17935	1	500	64	256
<i>A. undicola</i>	<i>recA</i> -Aun	ORS922	1	500	1	64
<i>A. tumefaciens</i> G1	<i>recA</i> -G1-1	TT111	500	1	4096	64
<i>A. tumefaciens</i> G1	<i>recA</i> -G1-2	CFBP5771	500	1	4096	4096
<i>A. tumefaciens</i> G1	<i>recA</i> -G1-3	S4	500	1	4096	4
<i>A. tumefaciens</i> G2	<i>recA</i> -G2-2	LMG147	1	10	16	1024
<i>A. tumefaciens</i> G2	<i>recA</i> -G2-5	CIP43-76	1	500	16	16384
<i>A. tumefaciens</i> G3	<i>recA</i> -G3-1	CFBP6623	5	5	64	1
<i>A. tumefaciens</i> G3	<i>recA</i> -G3-2	CFBP6624	5	10	64	64
<i>A. tumefaciens</i> G4	<i>recA</i> -G4-1	B6	10	1	1024	16
<i>A. tumefaciens</i> G4	<i>recA</i> -G4-3	Kerr14	10	1	1024	1024
<i>A. tumefaciens</i> G4	<i>recA</i> -G4-5	CFBP7126	10	5	1024	64
<i>A. tumefaciens</i> G5	<i>recA</i> -G5-1	CFBP6625	1	1	256	4
<i>A. tumefaciens</i> G5	<i>recA</i> -G5-2	CFBP6626	1	1	256	256
<i>A. tumefaciens</i> G6	<i>recA</i> -G6-1	NCPPB925	1	1	65536	256
<i>A. tumefaciens</i> G7	<i>recA</i> -G7-1	RV3	1	5	256	4096
<i>A. tumefaciens</i> G7	<i>recA</i> -G7-3	Zutra3-1	1	1	256	16
<i>A. tumefaciens</i> G8	<i>recA</i> -G8-2	LMG75	1	1	16384	1024
<i>A. tumefaciens</i> G8	<i>recA</i> -G8-4	C58	1	1	16384	16384
<i>A. tumefaciens</i> G9	<i>recA</i> -G9-1	H0362	1	1	4	256
<i>A. tumefaciens</i> G9	<i>recA</i> -G9-2	H0363	1	10	4	4096
<i>A. tumefaciens</i> G13	<i>recA</i> -G13-1	CFBP6927	5	1	1024	65536

A. tumefaciens G1 to G13, genomospecies of the *Agrobacterium tumefaciens* species complex (Lindström and Young, 2012).

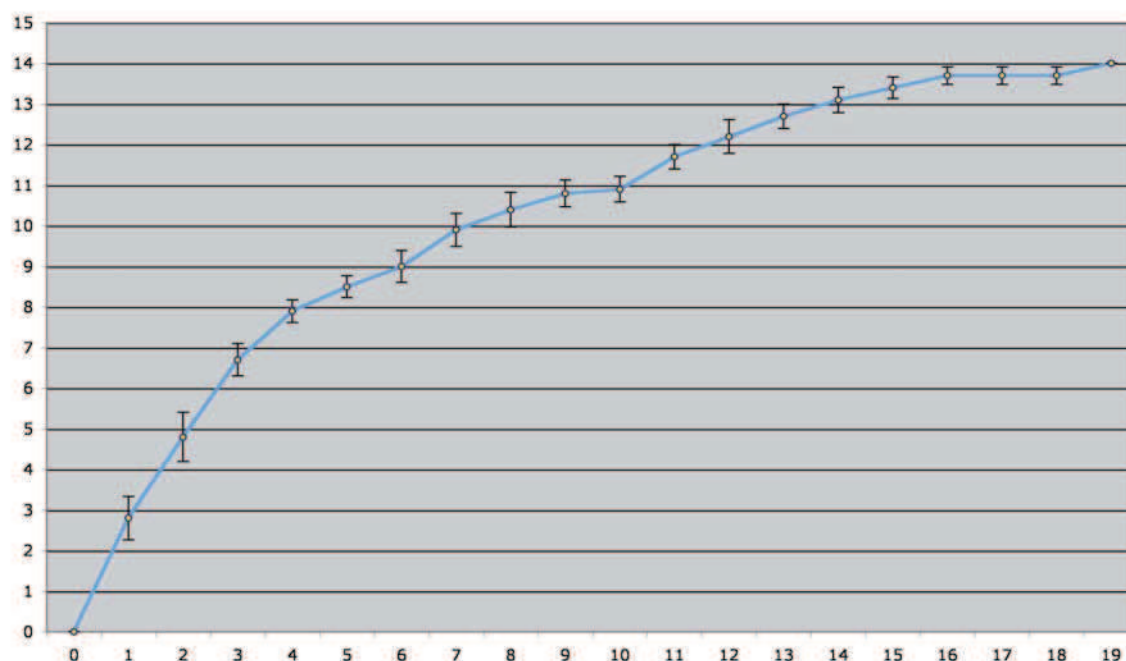


Figure 2-4-2. Average number of novel *Agrobacterium* species detected in the metagenomic DNA of Chbika-1 in successive samples. Axis: successive independent samples. Ordinate: number of novel species detected amongst *recA* sequence in successive sample amplicons. Bars indicate average standard error. Dominant and sub-dominant species found in more than one sample were obtained by sequencing ca. 10 sample amplicons in average.

Plant genotype effect upon the *Agrobacterium* diversity of root inhabiting bacterial communities.

Height species were detected in the soil metagenomics DNA of Mas d'Imbert used to grow *M. truncatula* genotypes (Table 2-4-4). Five species dominated in plant compartments over the bulk soil, but no significant differences were observed between Rhizosphere soil and soil tissue communities. Later data were thus assembled in subsequent analyses. The *Agrobacterium* species relative abundance significantly varied according to plant genotypes (Fig. 2-4-3). G1 was preferentially recruited in the root systems of Algerian ecotypes DZA 045-6 and DZA 315-16, while these genotypes poorly recruited GA members. Conversely, Jemalong A17, recruited equivalent amounts of G1 and G4. This suggested that Jemalong A17 was more tolerant than Algerian ecotypes toward the recruitment of G4 members.

Table 2-4-2. *Agrobacterium* species and allele assignation of sequences detected in synthetic communities.

Species	CS1	CS2	CS3	CS4	Total
A. sp.	0/0	0/5	0/5	0/2	0/12
Ala	0/0	0/0	0/0	0/1	0/1
Aru	0/0	0/5	0/0	0/0	0/5
G1	1750/2875	1/3	209/368	32/33	1992/3279
G2	0/3	0/1250	0/1	0/64	0/1318
G3	2/3	13/25	1/3	0/0	16/31
G4	19/25	11/14	30/59	2/4	52/102
G5	0/1	1/3	6/8	0/1	7/13
G6	0/0	1/1	1023/1133	2/2	1025/1136
G7	3/5	2/7	1/2	9/16	15/30
G8	0/0	4/4	755/886	126/135	885/1025
G9	0/0	15/23	0/0	5/5	20/28
G13	4/4	1/1	34/36	2145/2174	2184/2215
Total	1778/2916	38/1341	2059/2501	2321/2437	6196/9195

Numbers indicates the number of sequence matching with alleles present in the synthetic communities on the total number of detected sequence features. "A. sp." groups sequence features that did not match with introduced species. This group had no taxonomic value.

Table 2-4-3. *Agrobacterium* species occurrence at Chbika-1.

Sample	Rhizo	Agro	1650	Ala	Aun	G1	G2	G3	G4	G6	G7	G8	G9	G14
TG1-BS	0	0	0	63	0	1489	0	0	0	0	0	0	0	176
TG1-RS	0	0	0	0	0	112	0	0	0	0	2	849	0	0
TG2-BS	0	0	0	32	0	0	6	0	64	0	0	117	0	526
TG2-RS	0	0	0	2	0	1	137	0	1	0	0	3	0	1174
TG3-BS	0	0	0	0	0	876	0	0	0	536	0	0	0	0
TG3-RS	0	0	0	0	0	167	0	0	1	0	73	240	0	366
TG4-BS	0	0	0	0	0	0	0	0	0	0	0	0	0	879
TG4-RS	0	2	0	56	0	190	0	0	0	0	0	140	0	0
TG5-BS	0	0	0	129	0	1	0	0	0	0	118	0	0	435
TG5-RS	0	0	0	0	0	0	0	0	0	0	0	290	0	0
TG6-BS	0	0	0	0	0	0	0	0	0	0	0	0	0	323
TG6-SR	0	0	0	0	0	0	0	0	39	0	0	0	0	2249
TG7-BS	0	0	0	0	0	178	0	0	0	0	0	0	0	0
TG7-RS	0	0	0	0	0	0	0	0	204	0	76	0	0	6
TG8-BS	13	13	0	9	0	0	1	0	0	0	7	0	0	388
TG8-RS	0	2	0	26	952	478	0	0	121	1	4	1	0	0
TG9-BS	1	606	0	406	0	111	0	0	0	0	0	0	0	348
TG9-RS	0	0	0	0	0	0	0	20	0	0	0	1	0	1685
TG10-BS	0	0	14	391	0	33	1	0	31	0	0	4477	22	0

Sample plots TG1 to TG10, with bulk (BS) or rhizosphere (RS) soils were as described Chapter 2-4. Agro: "unclassified *Agrobacterium*" sequences detected at TG9-BS had the same sequence with a best identity of 87% with *A. sp.* 1650, forming a single OTU for a single bona fide *Agrobacterium* species. Rhizo: "unclassified *Rhizobium*" sequences detected at TG8-BS had the same sequence with a best identity of 88% with *R. loessense*, forming a single OTU for a single bona fide *Rhizobium* species

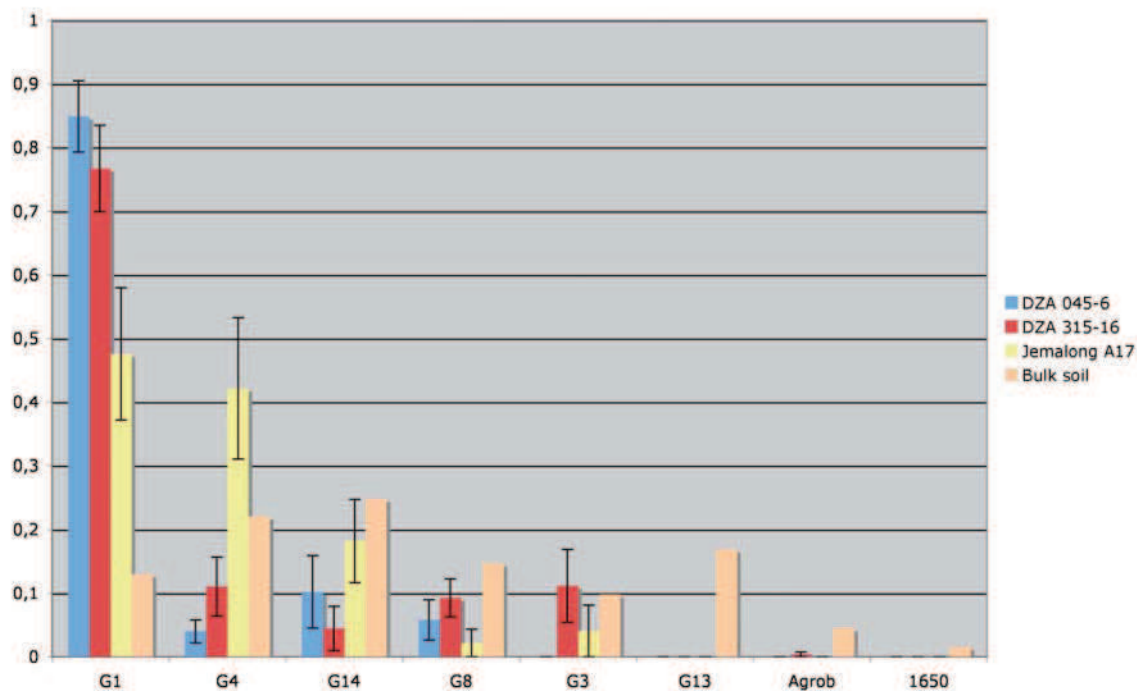


Figure 2-4-3. Relative *Agrobacterium* species *recA* sequences frequencies in different *Medicago truncatula* genotypes. Axis: *Agrobacterium* species. Ordinate: average percentage of *recA* sequence occurrence in amplicons. Bars indicate average standard error.

Discussion

The pyrosequencing technology allowed us to sequence numerous individual PCR fragments of composite PCR products obtained with metagenomic DNAs extracted from complex communities. In this approach each individual pyrosequence was assumed to arise from a single cell and the relative number of fragments with identical sequences is assumed to be a measure of the relative abundance of cells harboring this sequence in the part of the community whose *recA* was amplifiable with the concerned primers. While, the pyrosequencing technology allowed to deliver huge numbers of individual sequences, it is however unnecessary to have huge numbers of sequences to analyze the species and allelic diversity of a taxon of limited extent such as *Agrobacterium* spp. Conversely, we needed to analyze and compare numerous communities. Thus, we pooled and sequenced together numerous samples individually identified by tags (i.e. MIDs).

Table 2-4-4. *Agrobacterium* species occurrence in *Medicago truncatula* genotypes.

Sample	Agrob	1650	G1	G3	G4	G8	G13	G14	Total
211-R	0	0	1029	0	0	0	0	0	1029
211-SR	0	0	1167	0	17	120	0	5	1309
212-R	0	0	1004	0	18	0	0	358	1380
212-SR	0	0	919	0	0	15	0	102	1036
213-R	0	0	709	0	0	0	0	0	709
213-SR	0	0	1462	0	33	21	0	3	1519
217-R	0	0	752	0	18	36	0	0	806
217-SR	0	0	417	34	141	29	0	0	621
218-R	0	0	920	0	0	0	0	0	920
218-SR	0	0	1597	0	2	58	0	0	1657
219-R	0	0	950	428	25	0	0	167	1570
219-SR	1	0	669	57	43	52	0	6	828
235-R	0	0	217	0	170	0	0	202	589
235-SR	0	0	127	45	126	0	0	25	323
236-R	0	0	875	0	84	0	0	0	959
236-SR	0	0	1029	0	339	0	0	153	1521
237-R	0	0	2	0	355	0	0	0	357
237-SR	0	0	426	0	141	31	0	165	763
613S-N	2	0	3	2	490	5	1	4	507
615S-N	111	0	962	0	0	1227	0	0	2300
616S-N	0	0	0	249	0	0	652	0	901
618S-N	0	8	1	0	0	0	0	835	844

211, 212, 213: replicates of DZA 045-6. 217, 218, 219: replicates of DZA 315-16. 235, 236, 237: replicates of Jemalong A17. R: root tissue; SR: rhizosphere soil; N: bulk soil of Mas d'Imbert.

In the present setting up of *recA* amplicon pyrosequencing for the purpose of *Agrobacterium* microdiversity analysis, we verified that the methodology provided a relevant snapshot of the *Agrobacterium* community composition. Inherent limits of the methodology however allowed us to obtain relevant informations at the species but not at the allele levels. The methodology allowed detecting a species when at least 4 *recA* sequences were expected, while sequencing error could provide false sequences occurring as singleton. Thus, we proposed to consider a species as being readily present only at least 4 sequences of this species were detected.

Applied to environmental communities, the methodology revealed us the presence of up to 14 species in the model station of Chbika-1. While most species detected through the amplicon sequencing approach had been already detected by standard microbiology or by the cloning-sequencing approach (Chapter 2-3), the most frequent species revealed by amplicon pyrosequencing was G14, a species never found previously. Investigations using G14-specific primers did not confirm the occurrence of this species in the Chbika-1 metagenomic DNA. We

have no reasonable explanations for this strong discrepancy, leading us to consider cautiously the present Chbika-1 analysis.

Using various approaches, we did not find differences in the *Agrobacterium* species content of between bulk and rhizosphere soils in Chbika-1 (Chapter 2-3 and 2-4). This was very likely due to our compartment separation procedure that was used "bulk soils" that were harvested in the close vicinity of roots. Conversely, the compartment separation procedure used by A. Zancarini with *M. truncatula* was more stringent, and a clear difference between bulk and rhizosphere soils was thus obtained with her DNAs (Fig. 2-4-3). A third compartment, consisted of washed root tissue was also considered by A. Zancarini, but no significant differences were detected between these two plant relate compartment. Thanks two those DNAs, we were able to show that plant genotypes displayed significant different abilities toward the recruitment of *Agrobacterium* species. To our knowledge, this is the first report of such a plant effect upon the recruitment of very closely related bacterial species by very closely related plant genotypes. This finding is in agreement with the result of Zancarini et al. (2012), who showed clear community differences between the same *M. truncatula* genotypes using a taxonomically less resolutive but more general 16S based approach. The effect of different plants on agrobacterial diversity leads to forming different agrobacteria community in each rhizosphere depending on needs, completion, advantages or difficulties for each bacterial species. It conducts to preferential population formations in different rhizospheres, but also to differences according to soils. As showed by P. Portier (2004), the rhizosphere of *Medicago truncatula* (Jemalong A17) was more favorable to G8 and G4 than to G1 members when grown in the French soil of La-Côte-Saint-André (LCSA). Presently, in the soil of Mas d'Imbert, G8 members were found in the Jemalong rhizospher but it was sub-dominant compared to G1 and G4. As a consequence, question remains pending whether the plant selective effect found in the present study is caused by marked *Agrobacterium*-specific traits or by characteristic traits of the sole agrobacterial strains present in Mas d'Imbert.

In conclusion, the *recA* amplicon-pyrosequencing approach is very promising tool to analyze environmental *Agrobacterium* species diversity. It allows the simultaneous and high throughput deliverance of relevant taxonomic data requested to study biotic and abiotic factors that are shaping agrobacterial populations. Amongst, the very promising way of research, the methodology will permit to more clearly show how plant root systems is influencing this community.

References

- Zancarini A., C. Mougel, S. Terrat, C. Salon, N. Munier-Jolain. 2012. Combining ecophysiological and microbial ecological approaches to study the relationship between *Medicago truncatula* genotypes and their associated rhizosphere bacterial communities. Plant Soil. DOI 10.1007/s11104-012-1364-7
- Portier P. 2004. "Sélection d'écotypes bactériens pathogènes et non-pathogènes par la plante en relation avec la différenciation en espèces génomiques chez *Agrobacterium spp.*". Thèse de doctorat, UCB-Lyon 1, pp. 122.

Discussion and Perspectives

The genus *Agrobacterium* encompasses several species and several strains within species that are now clearly named using a novel nomenclature reviewed in chapter 1.1. In each population different techniques have been used to recognize their members. In the present experimental work, we have adapted several methodologies that have been set up by other researchers. We have also amended different techniques or developed new techniques to rapid and accurate identification of agrobacteria. We also studied aspects of agrobacterial life such as pathogenicity at a country scale or the role of plants on the composition of agrobacterial population.

To identify and study agrobacteria diversity, a set of efficient procedures is now available that can be used on pure culture isolates as well as on agrobacteria directly in soil microbiomes. These methodologies are necessary to unmask the ecologies of *Agrobacterium* species that usually live sympatrically in soils (Voget et al., 2003; Costechareyre et al., 2010). We also used this methodology to look for a novel environmental strain of genomovar G8 in order to verify whether such a novel strain actually harbour the species-specific genes of this species. This was my contribution to the paper of Lasalle et al. (2011) given in annexe 3. We verified indeed that strain MKS-01, a novel isolate from LCSA assigned to G8 thanks to our G8-specific primers harbored a novel *recA* allele (*recA*-G8-5), and had indeed the predicted G8-specific genes in its genome. More generally, we are expecting that these knowledge will now allow us to differentially manage agrobacterial species in soils to decrease the crown gall disease impact as well as to stimulate agrobacteria having potential PGPR effects.

Among techniques that could be used for rapid identification of agrobacterial isolates, MALDI-TOF MS was evaluated and found to be suitable to delineate all *Agrobacterium* species. By using this technique, we could identify *Agrobacterium* species undistinguishable by 16S rRNA sequencing. However the methodology still requires improvements, and the construction of a database that could allow the identification of mass peaks specific for each *Agrobacterium* species (i.e. a set of markers present in fingerprints of all strains from a given species). The MALDI-TOF MS technique is however not precise enough to discriminate strains within a species, and is thus not suitable for epidemiological studies. Unfortunately again, we did not

found specific spectra for Ti plasmid even in presence of Ti plasmid protein production stimulator (Ferulic acid). We nevertheless think that this methodology can still be improved for this purpose. Results presented in this work showed that MALDI-TOF MS is a reliable and rapid method for agrobacterial identification at the species level, comparable to what is obtainable with housekeeping gene sequencing. This implies the implementation of a database and of course that laboratories have bought the MALDI-TOF apparatus. However those analyses can be externalized to companies or plate-form facilities owning this technology and expertise. Hence, MALDI-TOF MS opens a new and very useful way for diversity and ecological studies applicable to analysis of large populations of isolates allowing the differentiation of species and genera of agrobacteria.

To achieve our dual objective of rapid agrobacterial species identification and agrobacterial diversity assessment in complex environments, we designed *recA* based-PCR primers with specificities at different taxonomic levels. This was possible thanks to the availability of numerous *recA* sequences longer than those usually found in databases (mostly from MLSA sequence data). Indeed, MLSA analysis generally delivered ca. 450-500 bp fragments which are not enough long to find suitable regions for primer designing. Such designing was also possible because *recA* allele sequences belonging to a given species show enough diversity among and great differences with other agrobacteria. Moreover, primers designed for superior taxa enabled the amplification of most species that were usually classified within *Agrobacterium* spp. or Rhizobiaceae members.

The microbiological approach used so far to identify *Agrobacterium* species is difficult to use for exhaustive investigations of the species composition of bacterial communities. Thanks to *recA* specific primers, it is now possible to achieve accurate and rapid first identifications of all *Agrobacterium* species in both pure cultures and complex environments. In chapter 2-3, we addressed the question of the presence of agrobacteria and Ti plasmids in soils at large scales. Taken together, our results showed that agrobacteria and especially members of the *A. tumefaciens* complex were present in all Tunisian soils we investigated so far. This result is remarkable and partly unexpected because some sampled soils were non-agronomic ones collected in extreme environments such as arid and salty areas. It is certainly too early to claim that agrobacteria are present in every soils on earth. However, as mentioned in chapter 1.3, bacteria closely related to *A. tumefaciens* complex were found in exotic samples such as deeply buried sediments of the eastern Pacific Ocean (D'Hondt *et al.* 2004) or up to 200,000-year-old marine sapropels (Süss *et al.* 2006). The present study showed that we now have the suitable

microbiological and metagenomic methodologies to perform extensive *Agrobacterium* prevalence studies, to determine where and in which extent *Agrobacterium* is present everywhere.

From biodiversity studies on colonies isolated from Tunisian soils, only G4 and G7 agrobacteria were detected. This result was surprising because other species were very likely present in numerous soils, at least at Chbika-1 as previously described by Rhouma et al. (2006) and Costechareyre et al. (2010). The unique presence of G4 and G7 amongst isolates could be an artifact due to faster growth speeds of G4 and G7 over other species. The question of *Agrobacterium* species growth speeds on 1A-Te medium should be addressed for all species as it could also have modified our view of *Agrobacterium* population diversity. This could explain some differences we obtained comparing biodiversity studies that were culture-dependant or culture-independent (chapters 2-2, 2-3 and 2-4 for *A. tumefaciens* G14). Interestingly, the fact that we got numerous isolates of the same species in very different locations is offering us the opportunity to address the novel question of the presence of unique strains of agrobacteria (of G4 or of G7) at a country scale. Information about the presence of a given strain at a country scale has great consequences for epidemiological studies. Of course, reliable taxonomic identifications require further investigations such as the sequencing of other housekeeping genes, as proposed with the multi-locus sequencing analysis (MLSA) approach (Gevers et al., 2005; Costechareyre, 2008). Indeed, taxonomic identification based on a single gene could be incorrect in case of horizontal gene transfer (HGT) between strains of different species (a strain readily belonging to a given species may have acquired a foreign *recA* allele). Nevertheless, *recA* was not prone to interspecies HGT within the *Agrobacterium tumefaciens* complex. We are thus reasonably confident about the taxonomic assignments of agrobacteria when *recA* was used alone. Anyway, MLSA will be added in order to know whether G4 and G7 isolates belong to single clonal complexes or to markedly different strains. This population genetic information would help us to understand how far agrobacterial strain actually spread out of local stations such as tree nurseries.

Most species detected through the amplicon sequencing approach had been already detected by standard microbiology or by the cloning-sequencing approaches. The most frequent species revealed by amplicon pyrosequencing at Chbika-1 was G14, a species never found previously. As mentioned above, we have no reasonable explanations for this strong discrepancy between isolation and metagenomic analysis especially for the present Chbika-1 samples. Concerning the

isolation method, it is extremely important to go back to stations and look for G14 isolates using specific-primers. We propose here to use both isolation and metagenomic methods by keeping selective media at 28°C during the selection of well isolated agrobacteria in order to have metagenomic stocks for subsequent analyses. Metagenomic stocks must be later analysed using species-specific at first, or if needed by *Agrobacterium*-specific primers. This technique can assure us to have all of possible collected and grown agrobacteria on a plate that were selected or no via isolation methods.

The *recA* amplicon-pyrosequencing approach allowed us to show that plant genotypes displayed significant different abilities to recruit different species of *Agrobacterium*. This is the first report of a differential recruitment of *Agrobacterium* species by so closely related closely related plant genotypes. This finding is a complement and in agreement with results of Zancarini et al. (2012), who showed clear community differences between the same *M. truncatula* genotypes using a taxonomically less resolutive —but more general— 16S rDNA based approach. The effect of different plants on agrobacterial diversity leading to forming different agrobacteria community in each rhizosphere depends on ecological needs, advantages or difficulties of each bacterial species. The soil itself is also strongly influencing the interaction between two partners (plant and bacteria), leading to tripartite interaction. Portier (2006) showed that the rhizosphere of *Medicago truncatula* (Jemalong A17) was more favorable to G8 and G4 than to G1 members when grown in the soil of La-Côte-Saint-André. However, from our present study, with the soil of Mas d'Imbert, G8 members were detected in the rhizosphere of Jemalong but this latter species was sub-dominant compared to G1 and G4. Those results may have been influenced by different methods used in the two studies, microbial approaches via isolation and metagenomic approaches. As a consequence, question remains pending whether the plant selective effect found in the present study is caused by marked *Agrobacterium*-specific traits or by characteristic traits of the sole agrobacterial strains present in Mas d'Imbert.

We found no Ti plasmid harboring agrobacteria among isolates, confirming that Ti plasmid-free agrobacteria are prevalent in all tested soils. Conversely, we detected Ti plasmid signature in many soil community DNAs. However, those so-called Ti plasmid-specific primers indeed detected *vir* homologous sequences in non-agronomic soils that had no reported history of crown gall infestations and conversely not in visually crown gall infested orchards. In any cases, if primers readily detected Ti plasmids, results would suggest a wide dispersal of pathogenic agrobacteria in numerous soils. This is an unexpected and surprising result that absolutely

requires additional investigations for definitive conclusions. We need therefore to isolate pathogenic agrobacteria from stations to ascertain this result by testing the pathogenicity of those newly isolated agrobacteria. Nevertheless, this unexpected result could be an indication that *vir* primers do not target only Ti plasmids. Of course, Ti plasmid detection in metagenomic condition is risky and it is shown clearly by Yakabe et al. (2012). Indeed, *vir* genes have homologous in many plasmids especially in Rhizobiaceae. Due to this homology and also to horizontal gene transfer (HGT) between plasmids it is likely that *vir* primers might sometime hybridize non-Ti plasmids. Alternatively, designing new conserved primers targeting other Ti plasmid-specific regions would allow the unambiguously detection of pTi from genomic DNA, as it is made for strain with MLSA technique. Conserved T-DNA regions might be used for this purpose, targeting for instance the oncogene region as early proposed by Ponsonnet and Nesme (1994) and Pionnat et al. (1999). In fine, pathogenicity testing on indicator plants is the ultimate and definitive clue for the presence of viable and pathogenic agrobacteria. Nonetheless, direct PCRs remain faster. To gain in significance, we propose thus to use multiplex PCRs targetting *vir* plus other Ti plasmid-specific regions to achieve trustable Ti plasmid detection.

In conclusion, we have reviewed the *Agrobacterium* taxonomy and the methods used to isolate, type strains and determine species of *Agrobacterium*. This has highlighted the need to improve and develop new tools that would be valuable to perform large scale studies on *Agrobacterium* populations. Tools were thus developped for studing isolates (MALDI-TOF MS and *recA* primers) but also to asses *Agrobacterium* diversity from metagenomic DNAs. Coupled with a cloning-sequencing approach, primers targeting superior taxa allow easy and affordable investigations of *Agrobacterium* and Rhizobiaceae microdiversity. The *recA* amplicon-pyrosequencing approach was found very promising to assess the environmental diversity of *Agrobacterium* populations. Alternatively, we could indeed imagine that sequencing the whole metagenomic DNA is a better and richer alternative providing much more information about the role of abiotic and biotic factors, since this approach is thought to provide the real genomic diversity of an ecosystem. Nevertheless, the complete DNA sequencing of a soil is still expensive and would require a lot of times for analysis... for perhaps lesser results concerning our interested object! Indeed, due to the extreme abundance of metagenomic data it is likely that relevant informations concerning only a target taxon—such as *Agrobacterium* spp.—would be very highly diluted and thus less accessible than with targeted approaches such as those developped in

the present thesis. Indeed, with this limited approach, we were able to evaluate the *Agrobacterium* diversity from a model field soil of Tunisian and from rhizospheres allowing us to draw new hypothesis about *Agrobacterium* prevalence in soils. Those results need to be confirmed by studies at larger scales. The technology for simultaneous and high throughput deliverance of relevant taxonomic data requested to study biotic and abiotic factors that are shaping agrobacterial populations is now available. Amongst, results obtained during our thesis, the plant genotype effect upon the differential recruitment of agrobacteria in root systems appears us a promising way for further researches in *Agrobacterium* agro-ecology.

Assessing the diversity of agrobacterial populations

RESUME

Les bactéries du genre *Agrobacterium* forment un ensemble taxonomiquement diversifié composé de nombreuses espèces, présent dans la plupart des sols et dans la rhizosphère des plantes. Les agrobactéries sont le plus souvent anodines voire stimulatrices de la croissance des plantes mais celles qui hébergent un plasmide Ti induisent la maladie de la galle du collet à de nombreuses plantes d'intérêt agronomique. Les plasmides Ti peuvent se transmettre par conjugaison aux autres agrobactéries qui deviennent à leur tour pathogènes. Ainsi, du point de vue écologique comme du point de vue épidémiologique, il est impératif de connaître la diversité des populations d'agrobactéries du sol, hôtes potentiels des plasmides Ti. Dans ce contexte, nous avons dans notre travail de thèse d'une part repréciser l'état actuel des connaissances sur la taxonomie du genre *Agrobacterium*, et nous avons fait une revue des méthodes d'isolement et de typage de ces bactéries. D'autre part, nous avons cherché à mettre au point des méthodes d'identification rapides et fiables des différentes espèces. Une première méthode est basée sur la technique MALDI-TOF MS. Elle permet d'identifier facilement les espèces de bactéries cultivées mais la méthode n'est pas assez résolutive pour typer des souches et encore moins la présence de plasmides Ti dans les isolats. Nous avons alors développé des amorces de PCR spécifiques de chaque espèce (17), pour le genre *Agrobacterium* et la famille Rhizobiaceae. Ces amorces se sont révélées efficaces pour identifier les bactéries cultivées mais aussi pour détecter leur présence dans des communautés microbiennes. Nous avons ensuite utilisé ces outils pour étudier la répartition des agrobactéries à l'échelle d'un pays, d'une station et entre sols nus et sols rhizosphériques en utilisant soit des isolats soit des ADN métagénomiques extraits des différents environnements. Enfin, nous avons montré que le clonage-séquençage ou le séquençage à haut débit d'amplicons obtenus à partir d'ADN de communautés microbiennes nous permettaient de connaître facilement et avec une grande profondeur la diversité des populations d'agrobactéries. Les outils que nous avons développés sont fiables et faciles d'accès. Ils permettront d'étudier l'effet des facteurs de l'environnement sur la structure génétique des populations d'*Agrobacterium* pour répondre, par exemple, à la question de l'effet sélectif des plantes sur les espèces d'agrobactéries du sol recrutées dans la rhizosphère.

DISCIPLINE

Ecologie Microbienne

MOTS-CLES

Agrobacterium, complexe d'espèces, identification, détection, biodiversité des agrobactéries

INTITULET ET ADRESSE DU LABORATOIRE

Ecologie Microbienne, UMR CNRS 5557 / USC INRA 1364

Université Claude Bernard Lyon 1, 16 rue R. Dubois, F-69622 Villeurbanne cedex, France