Understanding the composition and role of the prokaryotic diversity in the potato rhizosphere for crop improvement in the Andes

Jonas Ghyselinck

Dissertation submitted in fulfilment of the requirements for the degree of

Doctor (Ph.D.) in Sciences, Biotechnology

Promotor - Prof. Dr. Paul De Vos

Co-promotor - Dr. Kim Heylen

Ghyselinck Jonas – Understanding the composition and role of the prokaryotic diversity in the potato rhizosphere for crop improvement in the Andes

ISBN-number: 978-94-6197-119-7

Printed by University Press | www.universitypress.be

Ph.D. thesis, Faculty of Sciences, Ghent University, Ghent, Belgium.
This Ph.D. work was financially supported by European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement N° 227522
Publicly defended in Ghent, Belgium, May 28th 2013

EXAMINATION COMMITTEE

Prof. Dr. Savvas Savvides (chairman)
Faculty of Sciences
Ghent University, Belgium

Prof. Dr. Paul De Vos (promotor)
Faculty of Sciences
Ghent University, Belgium

Dr. Kim Heylen (co-promotor)
Faculty of Sciences
Ghent University, Belgium

Prof. Dr. Anne Willems
Faculty of Sciences
Ghent University, Belgium

Prof. Dr. Peter Dawyndt
Faculty of Sciences
Ghent University, Belgium

Prof. Dr. Stéphane Declerck
Faculty of Biological, Agricultural and Environmental Engineering
Université catholique de Louvain, Louvain-la-Neuve, Belgium

Dr. Angela Sessitsch
Department of Health and Environment, Bioresources Unit
AIT Austrian Institute of Technology GmbH, Tulln, Austria

Dr. Barbara Doyle Prestwich
School of Biological, Earth and Environmental Sciences
University College Cork, Cork, Ireland

DANKWOORD

Doctoreren omvat het zelfstandig uitdenken en uitvoeren van experimenten, het schrijven van publicaties (meestal gevolgd door het schrijven van een rebuttal), het halen van de deadline voor het indienen van een doctoraatsthesis en de uiteindelijke verdediging ervan. Dit alles zou uiteraard niet mogelijk geweest zijn zonder de hulp van vele mensen.

Vooreerst is er mijn promotor Prof. Paul De Vos. Paul, u heeft me de vrijheid gegeven om telkens de richting in te slaan die ik wilde uitgaan met mijn onderzoek, zonder veel op mijn vingers te kijken. Op geregelde tijdstippen vroeg u hoe de huidige stand van zaken was; de daaruit voortvloeiende reactie maakte wel duidelijk of ik al dan niet in de goede richting aan het gaan was. Ondanks het feit dat die vrijheid mij toeliet onderzoek te doen in wat mij echt interesseerde, werd op die manier ook een zekere zelfstandigheid en verantwoordelijkheid opgelegd; iets waar al uw doctoraatstudenten ongetwijfeld sterk uitkomen. Paul, bedankt dat ik onder uw toezicht aan een doctoraat mocht beginnen, bedankt voor het geven van feedback wanneer nodig, voor het nalezen van geschreven papers en het uiteindelijke doctoraat, en niet te vergeten voor de lekkere truffels van de volley die regelmatig in het koffiekot terug waren te vinden!

Een aantal maanden na de start van mijn doctoraat werd een nieuwe post-doc aangenomen. Dr. Kim Heylen kwam op het juiste moment om mijn onderzoek bij te sturen in de richting van een eerste paper. Met jouw ervaring en kennis binnen het domein van de microbiologie was het mogelijk om tegenwerk te bieden waar nodig; iets wat resulteerde in stimulerende en richtinggevende discussies. Kim, dank je wel voor de vele discussies bij de opzet van nieuwe experimenten, voor het grondig nalezen van meerdere tussenversies van papers en voor de leuke babbels over reizen, films en Deense reeksen in het koffiekot. Bedankt voor je grote bijdrage aan dit doctoraat; je rol als begeleider heb je uitstekend ingevuld!

I would like to thank the people with whom I cooperated frequently, and whom I met for the first time at our annual meeting in Lima, Peru. Dr. Angela Sessitsch, Dr. Barbara Doyle Prestwich, Stefan Pfeiffer and Siva Velivelli, thank you very much for sharing results and expertise which made joint publication possible and which also allowed me to gain new insights within the broad field of microbiology. You were not only excellent researchers to work with, but also very pleasant people to travel with! I would also like to thank all the people that were part of the Jury of my PhD defense (Prof. Paul De Vos, Dr. Kim

Heylen, Prof. Stéphane Declerck, Prof. Savvas Savvides, Prof. Anne Willems, Dr. Angela Sessitsch, Dr. Barbara Doyle Prestwich and Prof. Peter Dawyndt) for carefully proofreading my PhD thesis, providing comments that led to improvement of the manuscript and of course for allowing me to defend my PhD.

Uiteraard wil ik mijn collega's van de PDV groep bedanken voor de zeer leuke werksfeer; Ines, Bram, Joanna en – door velen aanzien als mijn tweelingbroer – Sven. Ik kon mij geen betere collega's indenken! De vele frietkotbezoeken op vrijdagmiddag, de etentjes aan de vijver bij mooi weer, de lachmomenten en de koffiepauzes in het koffiekot ga ik niet snel vergeten. Ook de meeting ter ere van de 100-jarige verjaardag van NvvM in Nederland is er één om niet snel te vergeten. Kort gezegd; 't was wijs met jullie, ik heb mij enorm geamuseerd!

My colleague PhD student and good friend Stefan Pfeiffer; thank you, not only for the excellent cooperation we had during our joint research but also for the pleasant (but often intense) travelling we did during the weeks following annual meetings. I must say that during our trips in Peru and Bolivia, together with our companion Prof. Arthur Schüssler, we probably had several of the best laughs ever; one of them including Stefan's struggle with the word 'Ollantaytambo' and the fact that some people tend to 'eat coffee'. I'm really happy to have met you guys during the VALORAM project; and to have travelled with you from the tropical regions near the Amason basin to the isolated high altitudes of the Altiplano of South-America. This was an adventure! I sincerely hope we meet again in the very near future!

Door de input van vele collega's is dit werk uiteindelijk geworden tot wat het is. Bedankt Liesbeth, Emly en Evi voor de vele sequenties die jullie hebben gelopen; vooral tijdens de laatste maanden van mijn doctoraat. Liesbeth, bedankt voor de hulp in het labo, en voor de grappige momenten. An Coorevits, heel erg bedankt voor de uren uitleg wat betreft de complexe taxonomie van *Pseudomonas*; zelfs in de periode dat je zelf druk bezig was met het schrijven van jouw doctoraat. Bedankt voor de uitleg rond ARB, en de bijdrage tot de laatste paper! Isabel, de discussies gaande van het uitdenken van pyrosequencing experimenten tot hoe interessant het eigenlijk wel niet is om een huis te kopen en te verbouwen hebben mij nieuwe inzichten gegeven. Timo en Anne Becker, met jullie was het steeds leuk om muziekervaringen te delen, gaande van persoonlijke dj-sets, tot de concerten van Anne en de optredens van Timo (waarvoor ik mijn schade nog moet inhalen). Koenraad, heel erg bedankt voor je toewijding en het uitdenken van de statistiek wat betreft de eerste paper. Wim De Smet en Bart Verheyde, bedankt voor het schrijven van scripts die mijn werk aanzienlijk sneller deden gaan. Bart Hoste en Margo, heel erg bedankt voor de weelderige uitleg als er ergens problemen of

onduidelijkheden waren. Nadine en Claudine, bedankt voor het verzenden van stammen en het verwerken van de vele permits die daarmee gepaard gingen.

Mijn ouders en zus! Zonder jullie was ik nooit aan een doctoraat begonnen; doctoreren stond ver van mijn bed. Dankzij jullie motiverende en stimulerende woorden ben ik er uiteindelijk toch aan begonnen en ik moet zeggen, het was de juiste beslissing; een 23-jarige weet blijkbaar toch niet alles beter. Ook jullie interesse in mijn werk heb ik enorm geapprecieerd. Bedankt voor het aanbieden van een warm nest tijdens de vier jaren van m'n doctoraat, maar ook voor de vele jaren daarvoor en ongetwijfeld voor de vele jaren die nog moeten komen!

Mijn vriendin Julie! Doorheen de vele jaren dat we elkaar kennen heb ik me altijd gesteund gevoeld in mijn werk en in het feit dat ik mijn doctoraat zou behalen. Je hebt me bijgestaan waar nodig; het laatste jaar was druk en ik was in de laatste maanden misschien niet altijd de meest attente partner, maar toch hebben we het samen heel goed doorgemaakt. Heel binnenkort is het tijd voor de vakantie waar we zo lang naar hebben uitgekeken. We gaan het er eens goed van pakken! Samen op naar al het nieuws dat nog komen zal!

LIST OF ABBREVIATIONS

AA	Amino Acid
ACC	1-aminocyclopropane-1-carboxylate
AG	Anastomosis Group
AMF	Arbuscular Mycorrhizal Fungi
ANI	Average Nucleotide Identity
CHCA	cyano-4-hydroxycinnamic acid
dNMS	diluted Nitrate Mineral Salts
FAME	Fatty Acid Methyl Ester
FAO	Food and Agriculture Organization
IAA	Indola Acetic Acid
ISR	Induced Systemic Resistance
LGT	Lateral Gene Transfer
LTP	Living Tree Project
MALDI-TOF	Matrix Assisted Laser Desorption Ionisation Time Of Flight
MS	Mass Spectrometry
MDS	Multi Dimensional Scaling
ML	Maximum Likelihood
MLSA	Multi Locus Sequence Analysis
MPCA	Microbial Pest Control Agent
MSDS	Material Safety Data Sheet
NBRIP	National Botanical Research Institute's Phosphate
NFL	Nearly Full-Length
NGS	Next Generation Sequencing
OECD	Organization for Economic Co-operation and Development
OTU	Operational Taxonomic Unit
PBS	Phosphate Buffered Saline
PC	Pearson Correlation
PCR	Polymerase Chain Reaction
PDA	Potato Dextrose Agar
PGP	Plant Growth-Promotion
PIA	Pseudomonas Isolation Agar
PPMC	Pearson Product Moment Correlation
QS	Quorum Sensing
RAPD	Randomly Amplified Polymorphic DNA
Rep	Repetitive Element
RF	Robinson-Foulds
SDS	Sodium Dodecyl Sulphate

SR	Short Read
TSA	Trypticase Soy Agar
TSB	Trypticase Soy Broth
UPGMA	Unweighted Pair Group Method with Arithmetic averages
VALORAM	Valorizing Andean Microbial diversity
vCEED	verboonian Comparison of Embedded Evolutionary Distances
WRF	Weighted Robinson-Foulds
wRMSD	weighted Root Mean Square Deviation

CONTENTS

PART I

Background, Aims & Summary of Work

I. OUTLINE

The research leading to this work was performed within the frame of the VALORAM project. This project results from the joint collaboration between five European and three Latin American institutions. VALORAM stands for Valorizing Andean Microbial diversity through sustainable intensification of potato-based farming systems. The project was set up to analyze the microbial diversity associated with *Solanum tuberosum* (potato) in the Central Andean Highlands, which is the region of origin of the plant. The aim of the project was to get insight into the microbial diversity that is associated with potato plants, both cultivation dependent and cultivation independently. Subsequently, microorganisms obtained would be checked for plant growth-promotion activities based on a series of *in vitro*, greenhouse and local field trials. The research may lead to the discovery of microorganisms with interesting plant growth-promotion properties, hence allowing a reduction in the amounts of chemical fertilizers and pesticides used.

The dissertation starts with a description of the background and objectives of the performed research, and a brief description of the experimental work (**Part I**). The introduction (**Part II**) consists of two chapters. Chapter 1 discusses the benefit and obstacles which involve plant growth-promotion by means of microorganisms. Chapter 2 describes the importance of and pitfalls associated with cultivation dependent and independent analyses of bacterial diversity. The experimental work (**Part III**) has been arranged into four chapters (Chapters 3-6). Each chapter addresses a different aspect involving microbial diversity. Each of these chapters concludes with a reflection on the work performed. The dissertation ends with a final conclusion (**Part IV**).

II. BACKGROUND AND OBJECTIVES

The rhizosphere, i.e. the environment directly surrounding the plant roots, is a nutrient rich environment. These increased nutrient levels relative to the surrounding bulk soil result from compounds that are excreted by plant roots, a process referred to as root exudation. It explains why the rhizosphere soil is much richer in bacteria than the surrounding bulk soil. In addition, many bacteria reside inside the plant roots. Such bacteria are referred to as bacterial endophytes. Most interestingly, many of these rhizo- and endophytic bacteria have shown to exert plant growth-promotion properties, and thus have the potential to replace harmful agrochemicals.

Potato plants have their origin in the Central Andean Highlands. Due to the long term association between potato plants and bacteria in this region, it is likely that bacteria historically developed interesting plant growth-promotion properties for the potato plant. This hypothesis motivated the investigation and discovery of the microbial potential that resides in these potato fields. However, setting up cultivation dependent and independent diversity studies remains a challenging task, and researchers are confronted with difficulties at each step of the analysis.

The objectives of this study were dual and aimed at isolating bacteria to investigate their plant growth-promotion properties on the one hand, while on the other hand the goal was to specifically highlight and clarify some of the issues that involve bacterial diversity assessments (both cultivation dependent and independent). Hence, the experimental work performed within the frame of this PhD dissertation has been implemented into different chapters, each of which discusses a different aspect that is related to bacterial diversity.

Chapter 3 describes an investigation of the applicability of MALDI-TOF MS to dereplicate a group of bacterial isolates at a taxonomic level similar to that of rep-PCR. The evaluation is based on a number of criteria, including taxonomic resolution, reproducibility, suitability for high-throughput automation and time and cost effectiveness.

Chapter 4 describes an analysis of the plant growth-promoting properties of a collection of bacterial strains that were isolated from potato plants in the Central Andean Highlands. The approach consists of an initial *in vitro* screening, followed by trials on potato microplants. All antagonistic isolates were identified, with extra focus on *Pseudomonas* and *Bacillus*.

Chapter 5 describes the effect of primer choice on the outcome of high-throughput amplicon sequencing based diversity studies that use the 16S rRNA gene. A global understanding of the impact was obtained by analyzing primer coverage rates, short read phylogeny, OTU-richness and taxonomic assignment for ten well established primers targeting dispersed regions of the bacterial 16S rRNA gene.

Chapter 6 describes an evaluation of the diversity of *Pseudomonas* isolates obtained on three cultivation media to identify the one yielding the largest diversity. A taxonomic marker was used that was selected based on the results of a foregoing investigation of the taxonomic resolution of the 16S rRNA, *rpoD*, *gyrB* and *rpoB* genes.

Each of these chapters ends with hindsight reflections on the work performed, and highlights a number of aspects that – with current knowledge and understanding – would have improved the research performed.

III. SUMMARY OF WORK

Potato fields in the Central Andean Highlands are mostly owned by small farming communities that are settled in remote areas, free of intense anthropogenic influences. These communities grow potatoes because they provide cheap but nutritious foods to the farmers. The biggest share of grown potatoes is used for consumption by the communities; only a small proportion is meant for export. These farming communities are often poor, and largely depend on crop yields. Crop diseases are disastrous and as a result, local farmers are willing to put their fields at microbiologists' disposal for studying disease protection programs.

The Central Andean Highlands are the center of origin of the potato plant. The long-term cohabitation between potato plants and bacteria in this region leads to the hypothesis that local potato fields contain bacteria with interesting plant growth-promotion properties. Years of cohabitation may have induced a mutualistic relationship between plant and bacteria; making these fields interesting targets for microbial research. However, almost no literature is available on bacterial diversity studies in the Central Andean Highlands, implying that there is yet much to be discovered.

Within the frame of this PhD research, bacteria residing in potato fields in the Central Andean Highlands were cultivated and screened for plant growth-promotion properties. In addition, the broad diversity of bacteria obtained presented an ideal target for evaluating the dereplication potential of MALDI-TOF MS for a broad range of bacterial species. Since many of the plant growth-promoting bacteria were identified as members of the genus *Pseudomonas*, and because previous studies also demonstrated the important role of this group of bacteria with respect to plant growth-promotion, the size and nature of *Pseudomonas* populations obtained with different cultivation media was investigated. The research concludes with a thorough investigation on the effect of primer choice on the outcome of cultivation independent diversity studies.

A total of 585 bacterial isolates were isolated from the rhizosphere of potato plants in the Central Andean Highlands. Identification of a large number of bacterial isolates is often preceded by a dereplication step. Dereplication involves the process of recognizing identical isolates at a specific taxonomic level and grouping them accordingly. This has the advantage that further analyses in the identification process can be restricted to representatives of each group, thus avoiding unnecessary screening effort. Dereplication can significantly reduce time and financial costs, especially in large-scale studies. The first study within the frame of this PhD was an evaluation of MALDI-TOF MS for

dereplication of bacterial isolates. The suitability of MALDI-TOF MS was evaluated relative to rep-PCR, a technique which is frequently used for this purpose. A number of criteria were taken into account for comparison, including taxonomic resolution, reproducibility, suitability for high-throughput automation and time and cost effectiveness. MALDI-TOF MS proved to have higher reproducibility than rep-PCR and seemed to be more promising with respect to high-throughput analyses, automation, and time and cost efficiency. Its taxonomic resolution was situated at the species-to-strain level. MALDI-TOF MS was considered a powerful tool for dereplication and a promising alternative for rep-PCR.

All isolated bacteria were screened for antagonistic activities against the severe plant pathogenic fungus *Rhizoctonia solani* and the oomycete*Phytophthora infestans*. Isolates which tested positive against at least one of both pathogens in *in vitro* assays, were screened for the production of compounds likely to induce promotion of plant growth. After dereplication with MALDI-TOF MS, all of the antagonistic strains were identified. Identification showed that most isolates were members of the genera *Pseudomonas* and *Bacillus*, but also *Paenibacillus*, *Flavobacterium*, *Curtobacterium*, *Pedobacter* and *Enterobacter* species were obtained. Potato microplant trials were set up to test the effect of bacterial isolates on plant growth itself, and suppression of diseases caused by *Rhizoctonia solani*. A total of 23 antagonistic isolates were associated with plant growth-promotion and/or disease suppression activities. A number of isolates even outperformed the commercial strain *Bacillus subtilis* FZB24® WG.

The third study describes the impact of primer choice on the outcome of next generation sequencing efforts. The approach used consists of an elaborate series of analyses, which allow the assessment of primer coverage rate, short read phylogeny, OTU richness and taxonomic assignment performance of sequenced reads. These analyses allow a thorough evaluation of the information obtained from sequencing with different 16S rRNA gene targeting primers. With the obtained results, it was possible to provide a global view on the outcome that is to be expected with sequencing different regions of the bacterial 16S rRNA gene.

Since many of the plant growth-promoting isolates were identified as *Pseudomonas* species, three growth media were evaluated for their individual capacities to retrieve a high diversity of *Pseudomonas* isolates. The rationale was that an increased *Pseudomonas* diversity would increase chances of isolating plant growth-promoting *Pseudomonas* strains. The media in question were the general media Trypticase Soy Agar (TSA) and Potato Dextrose Agar (PDA), and the *Pseudomonas* specific *Pseudomonas* Isolation Agar (PIA). The *Pseudomonas* diversity on each of the growth media was expressed in terms of *Pseudomonas rpoD* sequence diversity. The choice to use the *rpoD* gene was motivated by an introductory study in

which the taxonomic resolution of the gene was investigated. TSA and PDA were found to generate the highest *Pseudomonas* diversity, while PIA generated the smallest. However, communities obtained with TSA and PDA overlapped, while those obtained with PIA were unique.

The thesis illustrates that the Central Andean Highlands harbor interesting plant growth-promoting strains, thus fulfilling the expectations. However, their efficiency in the field remains to be evaluated. The experimental design of both cultivation dependent and independent diversity studies has an enormous impact on the outcome of the experiment; this PhD thesis specifically highlighted and clarified some of the issues that involve bacterial diversity assessments. These insights can be extrapolated to other studies and guide researchers in the design of new experiments.

IV. SAMENVATTING

Aardappelvelden in de Andische Hoogvlakten zijn in hoofdzaak eigendom van kleine boerengemeenschappen. Deze boerengemeenschappen zijn zeer vaak gevestigd in afgelegen gebieden, vrij van intensieve antropogene invloeden. Deze gemeenschappen telen aardappelen omdat dit een goedkope, maar tezelfdertijd zeer voedzame voedingsbron vormt voor de lokale boeren. Het grootste aandeel van de verkregen opbrengsten is dan ook bedoeld voor eigen consumptie; slechts een klein deel gaat naar export. De boerengemeenschappen zijn vaak arm en bijgevolg zeer afhankelijk van hun gewasopbrengsten. Gewasziektes zijn catastrofaal voor hen. Omwille hiervan stellen deze boeren graag hun velden ter beschikking voor microbiologisch onderzoek dat doelt naar microbiële gewasbescherming.

De aardappel vindt zijn oorsprong in de Andische Hoogvlakte. De eeuwenlange cohabitatie tussen aardappelplant en bacteriën in deze regio doet veronderstellen dat in deze aardappelvelden bacteriën voorkomen met interessante plantengroei bevorderende eigenschappen. Mogelijks heeft deze jarenlange associatie een wederzijdse samenwerking tussen plant en bacterie teweeggebracht; wat van deze velden een interessant doelwit maakt voor microbiologisch onderzoek. Toch is er bijna geen literatuur beschikbaar over bacteriële diversiteitstudies in de Hooglanden van de Andes.

Binnen het kader van dit doctoraatsonderzoek werden bacteriën met herkomst in de Andische Hoogvlaktes gecultiveerd, en vervolgens getest op plantengroei bevorderende eigenschappen. De brede diversiteit aan verkregen micro-organismen vormde bovendien een ideaal doelwit om de toepassing van MALDI-TOF MS voor dereplicatie van bacteriële isolaten te onderzoeken. Een groot deel van de bacteriën werd geïdentificeerd als zijnde *Pseudomonas.* Voorgaande studies toonden ook al de belangrijke rol van deze groep van bacteriën aan binnen het domein van de plantengroeibevordering. Omwille van deze redenen, werd de verkregen *Pseudomonas* diversiteit op drie verschillende groeimedia dieper onderzocht. Het doctoraatsonderzoek besluit met een grondige studie naar het effect van primerkeuze op de uitkomst van cultuuronafhankelijke diversiteitstudies.

Een totaal van 585 bacteriële isolaten werden geïsoleerd uit de rhizosfeer van aardappelplanten in de Andische Hoogvlaktes. Identificatie van een groot aantal isolaten wordt veelal voorafgegaan van een dereplicatiestap. Dereplicatie slaat op het herkennen van identieke isolaten op een bepaald taxonomisch niveau, en het dienovereenkomstig groeperen ervan. Dit biedt het voordeel dat verdere analyses in het identificatieproces beperkt kunnen worden tot representatieve stammen van elke groep.

Dereplicatie kan de tijdskost en financiële kosten aanzienlijk drukken, zeker in studies op grote schaal. Wij evalueerden de toepasbaarheid van MALDI-TOF MS voor de dereplicatie van bacteriële isolaten door een vergelijking te maken met rep-PCR, een techniek die frequent wordt gebruikt voor deze toepassing. De evaluatie was gebaseerd op een aantal criteria, inclusief taxonomische resolutie, reproduceerbaarheid, geschiktheid voor high-throughput automatisatie en efficiëntie met betrekking tot tijd en kost. MALDI TOF MS bleek een hogere reproduceerbaarheid te hebben dan rep-PCR, en was veelbelovender wat betreft mogelijkheid tot high-throughput analyse, automatisatie en tijd- en kostenefficiëntie. Zijn taxonomische resolutie situeerde zich op het species-tot-stamniveau. MALDI-TOF MS werd beschouwd als zijnde een krachtig instrument voor dereplicatie, en een veelbelovend alternatief voor rep-PCR.

Alle bacteriële isolaten werden gescreend voor antagonistische activiteit tegen de plant pathogene fungus *Rhizoctonia solani* en de oomyceet *Phytophthora infestans*. Isolaten die antagonistische activiteit vertoonden tegen minstens één van beide pathogenen in *in vitro* testen, werden verder gescreend voor de productie van plantengroei bevorderende componenten. Na dereplicatie met MALDI-TOF MS werden alle antagonistische isolaten geïdentificeerd. De resultaten van deze identificatie toonden aan dat de meeste isolaten behoorden tot de genera *Pseudomonas* en *Bacillus*. De overige isolaten werden geïdentificeerd als *Paenibacillus*, *Flavobacterium*, *Curtobacterium*, *Pedobacter* en *Enterobacter* species. Vervolgens werden aardappel microplant testen opgezet om het effect van de bacteriën op gebied van groeibevordering enerzijds, en bescherming tegen ziekte veroorzaakt door *R. solani* anderzijds te bestuderen. In totaal werden 23 isolaten bevonden plantengroei te bevorderen, en/of bescherming te bieden tegen *R. solani* ziekte. Een aantal isolaten presteerden zelfs beter dan het commercieel verkrijgbare product gebaseerd op *Bacillus subtilis* FZB24® WG.

Een derde studie beschrijft de impact van primerkeuze op de uitkomst van Next Generation Sequencing experimenten. Het onderzoek bestaat uit een uitgebreide reeks analyses die toelaten inzicht te verkrijgen in de primer coverage rates enerzijds, en de fylogenetische informatie, OTU richness en identificaties verkregen met korte sequenties (reads) anderzijds. Deze analyses laten een diepgaande evaluatie toe van de informatie die bevat zit in korte sequenties gegenereerd met verschillende 16S rRNA gen primers.

Gezien een groot deel van de antagonistische isolaten behoorde tot het genus *Pseudomonas*, werden in een daaropvolgende studie drie groeimedia geëvalueerd op basis van hun capaciteiten wat betreft het verkrijgen van een grote diversiteit aan *Pseudomonas* isolaten. De redenering hierachter was dat een

verhoogde diversiteit aan *Pseudomonas* isolaten gelijktijdig een verhoogde kans biedt op het isoleren van plantengroei bevorderaars. De gebruikte groeimedia waren de twee universele media Trypticase Soy Agar (TSA) en Potato Dextrose Agar (PDA), en het *Pseudomonas* specifieke *Pseudomonas* Isolation Agar (PIA). *rpoD* sequentie divergentie van de gecultiveerde *Pseudomonas* isolaten diende als maatstaf voor het meten van de diversiteit. De keuze voor het *rpoD* gen werd bepaald door een inleidende studie waarin de taxonomische resolutie van het gen werd geëvalueerd. TSA en PDA genereerden de hoogste *Pseudomonas* diversiteit, terwijl PIA de kleinste diversiteit genereerde. De verkregen diversiteit met PDA en TSA bleek echter te overlappen, terwijl de diversiteit verkregen met PIA eerder uniek was.

Deze doctoraatsthesis illustreert dat onze veronderstelling dat aardappelvelden in de Andische Hoogvlakten rijk zijn aan interessante plantengroei bevorderende bacteriën blijkt te kloppen. De efficiëntie van de verkregen organismen moet echter nog worden aangetoond in het veld. Het experimentele ontwerp van cultuurafhankelijke en cultuuronafhankelijke diversiteitstudies blijkt een significante impact te hebben op de uitkomst ervan. Deze thesis beklemtoont en bestudeert enkele van deze problemen; de verkregen inzichten kunnen geëxtrapoleerd worden naar andere studies.

PART II

Introduction

CHAPTER 1

PLANT GROWTH-PROMOTION BY BACTERIA

The discovery of natural suppressive soils [1], i.e. soils exerting naturally protective effects on plants, created excitement amongst researchers, as this suggested that a huge potential for sustainable agriculture resided in soils. The naturally protective effect seemed to be attributed to (micro)organisms with plant beneficial properties. Plant growth-promotion (PGP) by (micro)organisms has indirectly been used for many years by farmers to maintain soil fertility and protect their crops against plant pathogenic organisms (biopesticides). Crop rotation, green manure, soil solarization and biofumigation all rely on the PGP properties of (micro)organisms [2]. While crop rotation and green manure are widely recognized techniques, soil solarization and biofumigation are less generally known. Soil solarization consists of covering the soil with a transparent plastic cover to trap solar energy. Most plant pathogenic organisms are unable to grow at temperatures exceeding 31 to 32°C. As a result, the high temperatures underneath the plastic cover kill many pathogens either directly or indirectly due to their weakened state and increased vulnerability to thermophylic bacteria. Beneficial microorganisms usually survive the soil solarization process or recolonize the soil quickly after heating. Biofumigation is based on the cultivation of plants containing high levels of glucosinolates (mainly *Brassica* species). Glucosinolates are naturally occurring plant sulphur compounds that can enzymatically be degraded into isothiocyanates. This enzymatic degradation process is mediated by soil microflora [3]. Isothiocyanates are compounds which provide protection to the plant. This chapter digs deeper into the PGP effects exerted by bacteria, and the possibilities for sustainable agriculture associated herewith.

1.1 IMPORTANCE

It is justified to state that plant PGP by micro-organisms is a 'hot topic'. The amount of literature that is available in this field of research is enormous. In 2009 the global agrochemicals market was reported to be $37.9 billion, of which biopesticides represented about 3.5% [4]. Glar and colleagues (2012) [4] reported that the biopesticide market is growing more rapidly than the agrochemicals market and is expected to contribute for about 7.7% to the global pesticide market by 2014. Consumers are becoming increasingly aware of the environmental dangers that are associated with the use of agrochemicals, as

very often they involve risks for both the consuming and operating individual. Agrochemicals also persist in the environment and may cause alterations in indigenous microbial or animal populations. It is thus not surprisingly that regulatory actions have been undertaken such as the withdrawal of agrochemicals from the market [5]. There is increasing awareness of growing resistance development against chemical pesticides within pathogenic populations. Simultaneously, the discovery of new agrochemicals has become increasingly difficult and costly [4]. Last but not least, an increase in severity of diseases, which is attributed to the use of specific chemical products (iatrogenic diseases), has already been recorded. Hence, plenty of arguments to stimulate the search for alternative pest control strategies.

Breeding of pathogen resistant plants is an environmentally-friendly alternative. However, it is a long-term activity and has been found to be insufficient to prevent disease of important crops [1]. The genetic modification of plants is significantly faster compared to breeding and offers promising perspectives. However, it has not yet been widely accepted amongst consumers. Biological disease control and fertilizing agents offer a promising environmentally-friendly alternative to agrochemicals, with potentially higher efficiency as they are active on or near the plant surface. Plant roots are relatively inaccessible to agrochemicals. Because biopesticides involve low risks to non-target organisms, crops can be sprayed up to harvest; pre-harvest intervals are not necessary [6]. Moreover, secondary metabolites produced by the biocontrol agent are biodegradable, and consequently don't result in any toxic residues. Contrary to agrochemicals, the action of biopesticides is often complex and doesn't rely on a single target site for efficacy. This prevents, or at least delays, the development of resistance in the pathogenic populations [6]. Development of new biopesticides is faster (three years versus five years for agrochemicals) and significantly cheaper (five million dollars versus 200 million dollars) than the development of agrochemicals in the USA [6]. Last but not least, biopesticides may additionally act as plant growth stimulators via direct plant growth promotion (§1.2).

The best known and widely used biopesticides are formulations based on *Bacillus thuringiensis.* This biopesticide produces insecticidal proteins and has been in use for over 50 years (http://www.biopesticideindustryalliance.org). Its long application history breaks any record of synthetic pesticides, and thus illustrates the lack of resistance development in the target populations. Other examples of commercially available biopesticides include formulations of *Trichoderma harzianum* (Bioworks, Inc.) for ornamentals, *Bacillus thuringiensis* (Valent Bio.) for applications on vegetables, vines and fruits, *Myrothecium verrucaria* (Valent Bio.) for grapes and vegetables, *Bacillus subtilis* and *Bacillus pumilus* (AgraQuest) for protection of wine grapes, lettuce and tomatoes, *Agrobacterium radiobacter* (AgBioChem) for ornamentals, fruits and nuts, *Pseudomonas fluorescens* (BlighBan) for fruits, potato,

almond and tomato and *Pseudomonas syringae* for post-harvest protection of apples, pears, lemons, oranges and grapefruit [2,6]. Recently, Bayer Cropscience developed a biopesticide product Votivo® which is based on formulations of *Bacillus firmus*. It is used in combination with the synthetic insecticide Poncho®. An extensive list of biopesticides has been published by Fravel in 2005 [2]. A complete overview of commercially available biocontrol agents can be found in 'The Manual of Biocontrol Agents, Fourth Edition', edited by Copping in 2004 [7].

1.2 MECHANISMS OF PLANT GROWTH-PROMOTION BY BACTERIA

Several bacteria living in close proximity to plants have the ability to promote plant growth. The rhizosphere is defined as the soil directly surrounding the plant roots and is known to be rich in nutrients due to the exudation by the plant roots. Root exudates are mainly composed of organic components which include amino acids, fatty acids, nucleotides, phenolics, organic acids, plant growth regulators, putrescine, sterols, sugars and vitamins [8]. The actual composition depends on a number of factors such as plant species, plant growth stage [9], presence of microbes [10,11], presence of products from rhizobacteria [12], stress conditions [13] and growth substrate [9]. As such, the rhizosphere composition is not constant for different plant species and may alter in time. Specific root exudates have been found to attract bacteria by triggering genes that are involved in chemotaxis [14,15]. They have also been found to play an important role in root colonization [15]. Hence, root exudates are one of the reasons that rhizosphere soil is much richer in bacteria than the surrounding bulk soil. Although this work focuses on PGP by bacteria occurring in the rhizosphere, PGP may as well be practiced by bacteria inside the plant roots (endophytic bacteria) or in the phyllosphere. Understanding the mechanisms leading to the promotion of plant growth may help to design the process of developing a PGP organism into a commercial product on the one hand, and selecting the best formulation and application methods (§1.4.3) on the other hand. This, for its part, may increase the efficacy and consistency of the biological PGP agent in the field.

Two types of PGP are known to occur and are referred to as direct and indirect PGP. Classification into one of these categories depends on whether or not the bacteria offer protection against plant pathogenic organisms. Direct PGP refers to the direct effect of an organism on plant growth; for instance by the production of plant hormones or through the delivery of nutrients via phosphate solubilization or the fixation of nitrogen. Indirect mechanisms, however, refer to the protective effect of an organism against a plant pathogenic organism. Indirect PGP can be established through a variety of mechanisms of which the production of secondary metabolites and the induction of systemic resistance

in plants are two examples. Depending on the mechanism of action, direct and indirect PGP properties are divided into a number of subclasses [8].

The first subclass that contributes to the mechanism of direct PGP is biofertilization. Biofertilization refers to the supply of nutrients to the plant and can be performed through fixation of nitrogen [16,17] or the enzymatically induced solubilization of phosphate [17]. A second mechanism of direct PGP, which is referred to as rhizoremediation, results from the degradation of pollutants [18]. Rhizoremediation allows the plant to grow normally in polluted grounds where it would not be able to grow without the interference of the PGP bacteria. In some cases, however, pollutant degradation is insufficient to accommodate the nutrient requirements of the bacteria. In such cases, the microorganisms are supported by the plant which provides root exudates that guarantee the organism's primary metabolism [18]. Other bacteria stimulate plant growth through the production of components such as phytohormones (auxins, cytokinins) [19] and volatiles (2,3-butanediol, acetoin) [20]. This is referred to as phytostimulation. However, adverse effects on plant growth as a result of excessive auxin levels have also been reported [19,21]. The last mechanism of direct PGP is stress control. Several bacteria produce the enzyme 1-aminocyclopropane-1-carboxylate deaminase (ACC deaminase) [22], which interferes with the production of ethylene by converting its precursor ACC into 2-oxobutanoate and NH_3. Ethylene is activated under stress conditions, and reduces plant growth. Preventing ethylene production thus indirectly increases plant growth.

Indirect PGP can be the result of secondary metabolite production, production of lipopeptides harmful to the pathogenic organism, signal interference, predation and parasitism, induction of systemic resistance in plants, siderophore production and competition for nutrients and binding sites on the plant roots [8,22]. Initially it was assumed that competition for nutrients and niches was a property specific for fungi, but more recently Kamilova and colleagues [23] showed that this can also be used by bacteria. Signal interference refers to the enzymatical degradation of signaling molecules that are involved in Quorum Sensing (QS). QS molecules are typically expressed at high bacterial density-levels and thus indirectly under conditions of limited nutrient availability. Signaling molecules can trigger pathogenicity in organisms that are not pathogenic at low cell densities. Degradation of QS molecules prevents these organisms from becoming pathogenic. Predation and parasitism occur through the production of fungal cell wall degrading enzymes (i.e. chitinases, proteases, β-glucanases and cellulases) [24]. The joint action of different lytic enzymes can have a synergistic effect, resulting in an increased antifungal effect [25]. Similarly, an increased biocontrol efficiency has been observed from synergism between lytic enzymes and antibiotics [26]. This indicates a huge potential for microorganisms that exert biocontrol through a

variety of mechanisms. Alternatively, mixed biocontrol formulations or mixtures of biocontrol agents with lytic enzyme preparations may be more effective than the individual agents. Biocontrol can also be achieved by not directly acting against the pathogen, but through stimulation of defense mechanisms in the plant, which renders it less susceptible to the attack of pathogenic organisms. This mechanism is referred to as Induced Systemic Resistance (ISR). It is triggered by bacterial components which amongst others include antibiotics, lipopolysaccharides, flagella, siderophores, N-acylhomoserine lactones and volatiles such as 2,3-butanediol [27]. One major advantage of ISR is that it is not a prerequisite that the ISR inducing bacterium is an efficient root colonizer. Moreover, the acquired resistance remains active over longer periods of time and offers the plant protection against a wider range of plant pathogens [27]. Hence ISR can be a very effective biocontrol mechanism. Ultimately, biocontrol can be mediated by competition for iron ions. Under conditions of iron depletion, several bacteria excrete iron chelating molecules with high affinity for Fe^{3+} ions. As a result, the pathogen cannot dispose of iron and is inhibited by the siderophore producing organism. In addition, Van Loon (2007) [27] showed that siderophores may also be involved in ISR.

1.3 STATUS OF CURRENT RESEARCH

A complete overview of commercially available biocontrol agents can be found in 'The Manual of Biocontrol Agents, Fourth Edition', edited by Copping in 2004 [7]. However, within the frame of this dissertation I mainly focused on indirect PGP against two severe potato pathogens *Rhizoctonia solani* and *Phytophthora infestans*. Although the focus within this thesis was mainly on PGP bacteria, arbuscular mycorrhizal fungi (AMF) are also known to suppress *R. solani* and *P. infestans* induced diseases in plants. *R. solani* is the causing agent of scurf disease and stem canker in potato plants [28]. *P. infestans* causes potato late blight disease, one of the most devastating diseases of potato worldwide [29,30]. *R. solani* is difficult to control because it has the ability to survive as sclerotia under adverse environmental conditions for many years, is capable of surviving as a saprophyte and has a very wide host range [31]. Disease transmission occurs via infected seed tubers. *P. infestans* infection, on the other hand, mainly occurs from airborne contamination by sporangia [32]. These sporangia can spread over wide distances during the potato growing season. *P. infestans* infections are very aggressive and are often associated with complete field destruction. Moreover, the time required for the pathogen to complete its life-cycle can be as short as three days, and as such thousands of spores can be formed in a very short period of time [33], contributing to the large scale often associated with infection.

1.3.1 *Rhizoctonia solani* disease suppression

R. solani disease occurs in potato production throughout the world [28]. Symptoms manifest on below- and aboveground parts of the plant as black scurf and stem canker respectively [28] (Fig. 1.1).

Figure 1.1 Black scurf disease in potato, caused by *Rhizoctonia solani* infection. Extracted from [34].

Black scurf develops later in the growing season and can be recognized from the appearance of black, irregular sclerotia on the tuber. Although differences in susceptibility amongst potato cultivars have been observed, no resistant cultivars have been identified nor developed [28]. The species *R. solani* consists of a number of anastomosis groups (AGs) [35] that are not equally infective to potato. Currently, disease control occurs by chemical fungicides. However, the different AGs are not equally susceptible to these agents. Efficacy of disease control depends on the stage of infection, and whether the infection was soil borne or tuber borne. Tuber borne *R. solani* infections are relatively easy to control compared to soil borne infections [28]. Chemical fungicide treatments may not always be effective against soil borne infections. Treatments seem to perform well in the early stages of disease development; however, in the presence of high inoculum levels higher doses are needed to be effective. As a result, *R. solani* disease is a complex disease to manage. Various biological agents, however, have proven to have promising effects with respect to control of the pathogen. Literature study shows the

effectiveness of fungi belonging to the genera *Trichoderma* [36,37,38,39], *Verticillium* [40,41], *Cladorrhinum* [42], binucleate *Rhizoctonia* [43], *Streptomyces* [44], and *Gliocladium* [45]. Likewise, several members of bacterial genera have proven to be effective suppressors of *R. solani* disease. These include *Bacillus* [46,47], *Burkholderia* [47,48], and *Pseudomonas* [46,47,48]. Van den Boogert and Luttikholt (2004) [41] found that the biocontrol fungus *Verticillium biguttatum* had a synergistic effect on *Rhizoctonia*-specific (pencycuron, flutanolil) fungicides. They also found that *V. biguttatum* extended the control spectrum of oomycete-specific chemical fungicides (cymoxanil and propamocarb). Grosch et al. (2005) [49] found two *Pseudomonas* strains and one *Serratia* strain, all of which were isolated from potato roots, that were able to suppress *R. solani* disease during field trials with potato. Ikeda and colleagues (2012) [50] performed field tests with infected potato seed tubers to test the biocontrol efficacy of *Pythium oligandrum* and obtained disease suppression at a level similar to that achieved by chemical control. Moreover, their study showed the expression of defense-related genes in the potato plant, which reduced tuber disease severity upon challenge with *R. solani*. Their observations indicated that *P. oligandrum* induced resistance in the potato plant. Wilson et al. (2008) [36] performed field trials to test the efficacy of *Trichoderma harzianum* in controlling soil borne potato infection and found that *T. harzianum* was capable of suppressing disease both in combination with the chemical fungicide flutolanil and when applied alone. Disease suppression in both cases was higher than if flutolanil was used alone.

1.3.2 *Phytophthora infestans* disease suppression

P. infestans disease causes enormous economic damage, which is estimated at $5.2 billion worldwide annually [51]. Late blight disease (Fig. 1.2) was responsible for the Great Famine in Ireland around 1845.

Figure 1.2 *Phytophthora infestans* infection in potato plants (Late blight disease). (A) Aboveground symptoms, (B) belowground symptoms. Extracted from [52].

The worldwide breeding for resistant potato varieties only had little effect so far [51]. Frequently a genetic variety was obtained which seemed promising with respect to *P. infestans* resistance. However, whenever the variety was grown for a few years and at a larger scale, the resistance was repeatedly lost. Current disease control measurements consist of an array of tactics [32]. These include planting healthy seed tubers, eliminating on-farm sources that may be or become contaminated with *P. infestans* (e.g. destruction of potatoes in waste heaps), applying chemical fungicides for disease control [53] and elongating the time between potato planting cycles by means of crop rotation, which is necessary since *P. infestans* survives in the soil after the growing season has ended [32]. Sexual reproduction of the pathogen has created more aggressive *P. infestans* strains with increased virulence [52,53,54], thus increasing the need for pesticide application. However, increasing fungicide resistance in the pathogen populations simultaneously renders agrochemicals less effective. Moreover, fungicides to control late blight disease are based on copper, which is known to have a negative environmental impact [32]. Excess amounts of copper in the environment are harmful for aquatic and soil organisms [55], and may cause adverse health effects in humans [56]. In Belgium, over 1000 tons of active agents are applied annually to ensure control of *P. infestans*. In Flanders, an average of about 17 kg of active component is applied per hectare per year [52]. Therefore, public concern puts further pressure on the use of copper based fungicides. It is clear that there is a great need for alternative treatments. Genetic modification of potato varieties [52] or biopesticide applications may be valuable alternatives. However, due to public concern about genetically modified organisms, biopesticides may be the preferred approach.

Axel et al. (2012) [32] made an interesting overview of all published studies in which the biocontrol efficiency of microorganisms against *P. infestans* was tested. The authors concluded that so far the application of microorganisms as biological control agents did not result in any consistent field performance. However, more recently Wharton et al. (2012) [57] tested the efficiency of formulations of *Trichoderma harzianum* and *Bacillus subtilis* in suppressing *P. infestans* in field trials with potato plants and found that the *B. subtilis* formulations were able to control seed piece decay caused by *P. infestans* with 57% in trials performed in 2006 and with 98% in 2007. *T. harzianum* was able to suppress seed piece decay with 81.5% in 2006 and 77% in 2007. This was similar to the level of suppression obtained with a commercially available mixture of fludioxonil and mancozeb. The authors found that pre-storage conditions of treated tubers played a significant role in disease suppression activity, as sub-optimal storage conditions did not result in disease suppression. Similarly, disease emergence was higher with the chemical fungicide mixture after sub-optimal pre-storage conditions. The authors also noted the fact that due to the effective root colonization of *T. harzianum*, biocontrol applications were less prone to being washed away during the growing season, resulting in longer efficiency compared to chemical fungicides. Field trials that were performed by Dorn et al. (2007) [58] with a selection of commercially available biocontrol agents were less promising. None of the agents reached the same level of control as copper based fungicides. The failure to suppress late blight disease was mainly attributed to detrimental environmental conditions. The copper based fungicides were more stable. Although fungicide stability is a desired trait, it simultaneously raises concern about the persistence of copper-based preparations in the environment. Dorn et al. (2007) [58] did not perform pre-storage of treated tubers but sprayed the biopesticides onto already planted tubers which were subsequently infected with *P. infestans* a few weeks later. As Wharton et al. (2012) [57] demonstrated the importance of pre-storage, the results may have been better if Dorn et al. (2012) [58] had accounted for this. Axel and colleagues (2012) [32] postulated that direct application of metabolites responsible for *P. infestans* inhibition may be a valuable alternative for the application of the producing organisms. However as this is beyond the scope of this dissertation, I refer to their review for further information.

1.4 DIFFICULTIES

Numerous publications illustrate the extent of ongoing research in the field of biological PGP. Still, the number of biopesticides that are currently available on the market is in relation rather low. Hence, an important question that researchers in this specialized area should wonder about is ‘What is causing this

discrepancy?'. This chapter describes some of the issues that are associated with the commercialization of biological plant growth-promoters.

1.4.1 Lack of communication

According to Lidert (2001) [59] one of the main reasons for the low number of biopesticides on the market is that many PGP studies are performed by academic institutions, which often do not focus on commercialization of PGP formulations. No universal standard protocol is available that describes the screening process required for bringing PGP bacteria on the market. Consequently, consistency between approaches followed by different research groups is lacking. Very often the approaches followed by academic institutions do not meet the industry's requirements. The focus of academic research differs from that of the industry in that academic research aims at publishing important results. Lidert (2001) [59] states that academic researchers often (1) overestimate the importance of sustainability as an economic driver, (2) lack knowledge of grower's needs, registration strategy and competitive forces, (3) have naïve ideas about market strategy, (4) underestimate registration costs and difficulties, and (5) pay less attention to cost-performance and shelf-life. Better adjusting academic research towards industrial needs could thus aid the commercialization of biological PGP agents. According to Glare et al. (2012) [4], there is a need for public and private organizations to combine and educate grower, retailer and public on the use and merits of biological disease control agents.

1.4.2 Legal restrictions

Restrictions on the international distribution of products containing living organisms continue to limit market access for biopesticides [4]. Guidelines are available on the export, shipment, import and release of biological control agents and other beneficial organisms [60]. The Organization for Economic Co-operation and Development (OECD) is an intergovernmental organization which consists of 30 countries in North America, Europe and the Pacific. The goal of the OECD is the co-ordination and harmonization of government policies. With their pesticide program they aim at harmonizing pesticide review procedures and reduce risks that are associated with the use of biopesticides. Regulatory authorities require detailed information on the microbial agent that is considered for commercialization as direct or indirect plant growth-promoter. A profound characterization of the microbial agent is required to assess its potential risks to the people and to the environment, and also to confirm its effectiveness with respect to pest suppression and/or plant growth-promotion. Their rationale is that microbial plant growth-promoters may infect or cause disease in other living creatures and result in displacement of non-target organisms and microorganisms. The risk assessment is based on the biological and ecological profile of the microorganism, and a set of short term pathogenicity and toxicity tests [61].

The requirements provided in this section refer to the active component of a formulated product. A detailed description of the microbial agent is based on nine points. The first point concerns the identity of the microbial pest control agent and includes the scientific name of the organism at species level, or a level sufficient to show taxonomic relation to known microorganisms, especially pathogens. If the microbial pest control agent (MPCA) during production was exposed to microbial impurities, these have to be identified to a taxonomic level that is required to support the hygienic state of the product. The maximum content of such impurities in the end product must also be reported. Similarly, non-microbial impurities such as additives, metabolites and fermentation residues, have to be reported. The technical grade of the MPCA, and if not pure its concentration, are also required. Furthermore, the producer should provide quality criteria that were applied for the production of the MPCA. In this context the OECD mainly focuses on the possible exposure to toxins or pathogens during any stages of production. Quality control data have to be reported. Last but not least, a theoretical discussion is required on the formation of unintentional ingredients, mainly from a toxicological concern.

The second point refers to the biological properties of the MPCA. All historical information that concerns isolation and preservation has to be provided. This includes amongst others isolation source, geographic distribution and the ecological niche which it was isolated from. Furthermore, the producer should provide a detailed description with respect to its mode of action, host specificity and the possible effects on species that are closely related to the target organism. The MPCA's life-cycle and more specifically the differences in pathogenic or toxigenic character of the various forms that may occur are important. The manufacturer is also required to provide specific information on closely related species, mainly with respect to pathogenicity and formation of toxic metabolites, but also about their physiological properties, presence of plasmids encoding genes involved in pesticidal/pathogenic/toxic activity, and genetic stability. In conclusion, the manufacturer has to report any relationship of the MPCA to any known human dermatophyte or any resistance to antimicrobial agents used in human or veterinary medicine.

Point three includes any information on functioning, mode of action and handling of the MPCA. The manufacturer has to provide details on the antagonistic nature of the target organism (e.g. fungi, bacteria), details of the crops for which it is intended, details of the harmful organisms against which it is active, the effects achieved and a statement on the mode of action of the MPCA (i.e. the biochemical and physiological pathways involved). Further details on the nomenclature and the mode of action of both active metabolites and degradation products are required, with special focus on toxic ones. If

resistance or cross-resistance may occur, this also has to be specified. Further requirements refer to the safe handling of the MPCA agent. A Material Safety Data Sheet (MSDS) for the microbial agent and recommendations in case of an accident with the MPCA agent have to be provided.

The fourth point concerns all information on preservation and the production process of the MPCA product. The manufacturer has to describe the techniques that were used during the production process that guarantee a uniform end product which has the same characteristics as the original strains. He should also be able to guarantee the absence (or low level) of microbial and toxicological impurities. The quality control measures and monitoring methods that are taken to ensure that the obtained end product is the original bacterial strain have to be specified. Concerning the preservation of the microbial agent, the shelf life of the agent has to be specified and stability tests have to be performed.

Point five encompasses information on risks that are associated with exposure to the MPCA. This involves both the workers in the production environment, as well as the users of the end-product. A number of pathogenicity and toxicological studies have to be performed within the context of this paragraph, such as allergenic responses, hypersensitivity, and oral, tracheal or intravenous infectivity. Furthermore, the possible occurrence of viruses, bacteria and protozoa that replicate intracellular, and the production of toxic metabolites and genotoxins has to be accounted for.

Point six concerns residues that remain after treating crops with the MPCA. The manufacturer has to guarantee that MPCA residues are not hazardous to mammals, and that it is unlikely that the MPCA will occur on treated food in concentrations that are considerably higher than under natural conditions.

Points seven, eight and nine are directed towards the effect of the MPCA on the environment. This encompasses information on amongst others survival and residual metabolites of the microorganism in the environment, mobility and multiplication. It also tries to assess the effects on non-target organisms.

1.4.3 Formulation and delivery

Formulation of a biocontrol agent plays a crucial role in its preservation and shelf-life [43,44], and can either improve or diminish PGP performance [62]. Poor shelf-life is one of the reasons that promising microorganisms do not become commercial products [2]. A variety of factors such as water activity of the formulated product [43,63], pH [64], the matrix used for product stability improvement [44,64,65] and preservation temperature [43,44] are known to affect product shelf-lives. Obviously, the nature of the biocontrol agent also has a significant impact. For instance, *Pseudomonas* based formulations

generally have shorter shelf-lives than *Bacillus* based formulations [1], making them less practical to work with.

Products with high water contents are susceptible to growth of contaminant microorganisms [64]. These may interfere with the PGP activity of the active component and carry health risks for the operator. In addition, if the contaminant produces gas as a result of microbial activity, explosive release of the active component may occur prior to, or on opening of the recipient [64].

Whether or not the moment of application of the biocontrol agent, and its concentration in the formulated product, impact biocontrol efficiency will likely depend on the nature of the biocontrol agent, the pathosystem and the cropping system [2]. For instance, Honeycutt and Benson (2001) [43] found that both the moment of application of the biocontrol agent on the seedlings, and increasing the concentration of the active component from 0.47% to 0.9% (vol:vol) in the formulated product did not influence biocontrol efficiency. Landa et al. (2004) [66], however, noted a significant effect of sowing date (i.e. spring versus winter) on biopesticide performance. Similarly, the mode of application may influence biopesticide performance. The agent should be formulated in a way that allows it to be easily applied in the crop production system and simultaneously leads to high efficiency with an adequate number of cells [67]. Trivial details in application mode can make a huge difference. For instance, advances in spray technology which involve optimizations of droplet size and formulations in granules that slowly release the active component may significantly increase biocontrol efficiency. Biopesticides have been applied as dusts, granules and briquettes [64], oil (for low volumes) and water suspensions , but also through insect vectors like bees [68] and ants [69]. However, the application of biopesticides on the plant by insect vectors may be prone to significant variation in treatment efficiency.

Seed treatment with the biocontrol agent and pre-storage for a certain period of time may result in longer disease suppressing effects compared to in-furrow applications. Seed treatment and pre-storage allow the biocontrol agent to colonize the seeds prior to planting, which makes them more resistant to washing away by water in the soil. For the same reason they are likely to have a longer lasting effect compared to chemical fungicides. Of course, much also depends on the colonization efficiency of the biocontrol agent.

1.4.4 Field Performance and practical issues

The efficiency of biocontrol agents may be greater in the greenhouse than under field conditions. The failure rate when moving from greenhouse experiments to field experiments is relatively high. As the greenhouse environment is a controlled environment, the interaction between PGP organisms and pathogens is often not disturbed by third parties. In the field, the PGP agents may be exposed to competition or antagonism by other (micro)organisms. In addition, edapho-climatic conditions may hamper PGP performance, as the optimal conditions for crop production are not necessarily optimal conditions for PGP activity [70].

Huang et al. [71] reported inconsistent performance of biopesticides after performing field experiments from 1992 to 1994. A number of factors may have accounted for the inconsistencies observed, such as physical and chemical soil composition, moisture levels and incidence of light, microbial community composition, soil pH and differences in temperature between the different field trials. Therefore, it may be advisable to determine to what extent these variables affect biocontrol performance [2]. A number of studies, which involve degradation of pollutants in the soil, have investigated the potential of introducing microorganisms with plasmid encoded bioremediation properties [72,73,74]. For instance, Zhang et al. (2012) [72] introduced bacteria into the soil which encoded genes involved in synthetic xenobiotic degradation on broad host-range plasmids. The authors found that these bacteria horizontally transferred plasmids to soil bacteria, which acquired the ability to degrade toxic components within five days after introduction. This approach results in higher efficiency due to guaranteed survival of the soil organisms and copes with the problem of poor survival following the introduction of biopesticides into the field. Genes involved in plant colonization, PGP and biocontrol have been found to be encoded on large plasmids [75,76]. Introduction of bacterial strains with plasmid encoded PGP genes into the soil may thus lead to the horizontal dissemination of the plasmid to soil microbial communities and may result in increased PGP efficiency. Post-harvest treatments are less prone to inconsistency in performance of biopesticides, and therefore may be more suited for their application. Storage conditions can be controlled, i.e. parameters such as temperature, humidity, incidence of light and gas composition can be kept constant [62]. Similarly, there is a huge potential in the application of endophytic PGP organisms, since endophytic microorganisms are encapsulated by the plant matrix, which offers some degree of protection against adverse environmental conditions.

Many commercially available biopesticides target a single pest [4]. This small spectrum of activity is often not desired by end-users, although it inherently implies that the active component contributes to environmental and non-target safety. Therefore, to encourage the use of biopesticides, there is a need

for biological control agents with a broader spectrum of disease control. Similarly, microorganism biodegradability, which is responsible for the alleviation of environmental and non-target concerns, inherently implies that the active components of biopesticides do not persist longtime in the environment. Consequently, the application of biopesticides rarely results in long term activity on plant or soil surfaces compared to agrochemicals [4]. However, the induction of systemic resistance in plants was described above (§1.2) as a biocontrol mechanism offering the plant protection for longer periods of time. Considering this, ISR involving biocontrol formulations may have greater potential on the pesticide market.

Research and field trials show that the most effective control strategy is through the combined use of biopesticides with traditional synthetic pesticides (http://www.biopesticideindustryalliance.org). However, a prerequisite for succeeding is that the chemical fungicide does not inhibit the biocontrol agent. Combined use offers several advantages; it allows the successful managing of pests through a combination of different control mechanisms, which reduces the development of pest resistance in the pathogenic populations, and it reduces the environmental impact compared to agrochemical-only applications.

Unfortunately, very often the farmer's perception on biopesticides hampers the expansion of this market [6]. Farmers generally have the wrongful perception that biopesticides are less efficient compared to agrochemicals. Moreover, once accustomed to an agrochemical product and having experienced positive results, they will not take the risk of switching to a biopesticide product which they have no experience with. To lower the threshold of switching from traditional agrochemicals to biopesticides, it is desirable that biopesticide application can be performed with the equipment used for the application of chemical pesticides, since it is unlikely that the farmer will purchase specialized equipment [64]. Very often, however, farmers lack knowledge of the harmful nature of agrochemicals and underestimate their impact on health of the consuming individual and the environment. Therefore, they not always recognize the importance of switching to alternatively solutions.

CHAPTER 2

INTRODUCTION TO BACTERIAL DIVERSITY

Bacterial diversity is everywhere in the environment, and mapping it has proven to be a very complex task. Initially, diversity assessments were mainly cultivation based. Cultivation based approaches encompass the cultivation of organisms in a sample, after which the cultivated members are identified and/or characterized. However, it is well-known that only a fraction of the bacterial diversity within a sample is able to grow on currently used cultivation media. Therefore, the diversity obtained is often not representative for the natural diversity of the sample. To cope with this major limitation, cultivation independent approaches were developed, in which the intermediate step of cultivation is skipped and the diversity is directly assessed from the sample. Cultivation independent diversity assessments consist of extracting the bacterial DNA directly from the sample, and subsequently amplifying and sequencing specifically chosen target genes. Although the accuracy of obtaining the true diversity is likely to be determined by DNA extraction efficiency, PCR artifacts and sequencing errors, it does provide a more representative picture of the diversity than currently used cultivation dependent approaches.

2.1 MEASURES TO STUDY BACTERIAL DIVERSITY

2.1.1 The important role of the 16S rRNA gene

The application of 16S rRNA gene sequences to study prokaryotic phylogeny was first introduced by Carl Woese. It was raised as an alternative for classical approaches that were based on phenotypic properties of organisms. Although phenotype based reconstructions of organism phylogenies seemed to work well for multicellular eukaryotes, it was more difficult for prokaryotic organisms due to the limited information that could be obtained from a prokaryotic phenotype. The 16S rRNA gene based phylogeny led to the discovery of a third kingdom, next to the Bacteria and Eukarya [77]; the so-called ‘Archaea’. Its discovery seemed to correspond with observations in organism phenotypes, as archaea differed from bacteria with respect to the composition of their cell membranes and some essential proteins that are involved in gene transcription and translation [78]. This consistency between data led to the wide assumption that the bacterial 16S rRNA gene was a useful marker for determining evolutionary

relationships between organisms. The highly and less conserved regions of the 16S rRNA gene were considered to function as a molecular clock and document the history of microbial evolution [79].

The gene meets the principal requirements of a taxonomic marker [80,81]. It is universally present in all prokaryotes and its length is large enough to contain considerable phylogenetic information. In addition, the 16S rRNA gene is functionally stable, and contains highly conserved regions that guarantee gene homology, guide sequence alignments and allow the design of universal primers. It also contains variable regions that provide meaningful phylogenetic information. The different variable regions within the 16S rRNA gene have evolved at different evolutionary rates. This results from different evolutionary pressures acting upon the independent structural elements. However, regardless of the degree of evolution that is observed within the gene, its vital role in prokaryotes dictates the evolutionary preservation of its secondary structure amongst all prokaryotes [80]. This beneficial 16S rRNA gene specific property is used to guide sequence alignments. Sequence alignments are performed to compare homologous bases in a pair of sequences. However, identification of homologous bases is often difficult, especially in gene regions that are characterized by a high mutation rate. The secondary structure, which is determined by the conserved fragments, provides guidance during the alignment in that sense that homologous bases are easily recognizable in those conserved regions.

It is worth mentioning that, due to its larger size, the 23S rRNA gene is considered to be a more informative chronometer than the 16S rRNA gene. However, for technical and economical reasons, historically the 16S rRNA gene was preferred [81]. Still, with current sequencing technology, sequencing of the 23S rRNA gene would no longer present an obstacle. However, due to the long history of using the 16S rRNA gene, switching back to the 23S rRNA gene would no longer be attractive due to the extent of currently available 16S rRNA gene databases.

The importance of the 16S rRNA gene in prokaryotic taxonomy is reflected by its incorporation in the current definition of prokaryotic species, and the size of current 16S rRNA gene databases. The current species definition states that organisms sharing less than 97% 16S rRNA gene sequence similarity represent different species. However, this definition is not reversible, as organisms sharing more than 97% sequence similarity do not necessarily belong to the same species. Due to its limited taxonomic resolution, extra analyses including DNA-DNA hybridizations remain necessary to confirm. The current species definition states that two bacterial strains are considered to be the same species, if their DNA-DNA relatedness is approximately 70% or greater [82]. Phenotypic characteristics should agree with this observation. Recently, Stackebrandt and Ebers (2006) [83] re-evaluated this species definition and

suggested to increase the current 97% threshold to 98.7-99% 16S rRNA gene sequence similarity. However, this renewed definition has not yet been widely accepted amongst taxonomists.

2.1.2 Housekeeping genes as alternative biomarkers for bacterial diversity

Several studies have investigated the use of alternative phylogenetic markers to deduce organism phylogeny. While a number of studies focused on specific prokaryotic lineages [84,85,86,87], other studies took into account a broader range of prokaryotic lineages [88,89,90,91].

Ludwig and Klenk (2001) [80] put forward a number of requirements for a gene to be considered a valid phylogenetic marker for prokaryotic evolution. These requirements state that the marker should be universally distributed amongst prokaryotes and is functionally constant. The gene should be long enough to contain sufficient phylogenetic information, and there should be a sufficient amount of sequence variation amongst different prokaryotic lineages to provide adequate resolution. A sequence database should be available containing at least representatives of the major taxonomic groups. Not mentioned by Ludwig and Klenk (2001) [80], but equally important for technical reasons, is that the taxonomic marker preferably contains conserved regions that allow the design and application of universal primers on the one hand, and offer guidance during sequence alignment on the other hand.

Comparative analyses of complete genomes suggest that only a limited amount of taxonomic markers meet these requirements. Only a small proportion of genes are universally present amongst prokaryotes that share sufficient sequence similarity to be recognized as ortho- or paralogs [80]. Estimates show that only about 40-100 genes fulfill the requirements mentioned above [80].

However, several drawbacks are associated with alternative phylogenetic markers [80]. Not every gene contains sufficient information to reconstruct organism phylogeny. In addition, due to the widely accepted role of the 16S rRNA gene in studying prokaryotic diversity, no other gene database is as extensive as the SSU rRNA gene database. Certain genes are prone to Lateral Gene Transfer (LGT), and it is clear that in order to reconstruct vertical descent in prokaryotic organisms no horizontally acquired genes are informative. Still, even for vertically acquired genes, a frequently encountered problem is the conflicting tree topologies obtained with independently evolving phylogenetic markers. The authors also mention the importance of comparing orthologous genes rather than paralogous genes. Paralogous genes result from historical gene duplication events. However, as gene duplications are often followed by changes of function in one or both paralogs [92], comparisons of paralogous genes may bias the

reconstruction of phylogenetic relationships between organisms. However, differentiating orthologous from paralogous genes seems to be a challenging task.

2.1.3 Phylogeny under attack

The idea that the phylogeny of one gene represents the evolution of complete genomes has received major criticism the last few years; all the more with the discovery of LGT. Current insights indicate that the 16S rRNA model of prokaryotic phylogeny is an oversimplification of the complexity of prokaryotic evolution [93]. It was longtime assumed that 16S rRNA gene based phylogeny represented the evolutionary history of a prokaryotic organism (and thus the genome). However, current understandings suggest otherwise. The initial idea of a universal tree of life which unites the kingdoms Bacteria, Archaea and Eukarya seems to be a misconception. Representation of evolutionary history of organisms by a tree-like structure seems to work well for multicellular eukarya. However, the situation is more complex for prokaryotes. This especially became apparent the last few years, with an increasing amount of genomes being sequenced. In prokaryotes, the mechanism of evolution seems to be different from the mechanism in multicellular eukaryotes, with the occurrence of LGT events. LGT can be induced by intercellular movement of DNA [94,95], mediated by transformation, transduction and conjugation processes, but also by gene transfer agents [96] and integrons [95]. Adding more complexity to the system, LGT has been found to cross taxonomic boundaries [97] and to show different rates of occurrence between bacterial lineages [94,98]. Moreover, LGT events were found not to occur randomly, but to be driven by selective processes operating in environments of residence [97]. The event of LGT has been considered to be too important to be regarded as a secondary mechanism of prokaryotic evolution. Therefore, it cannot be ignored for construction of prokaryotic phylogeny. Since LGT implies that not all genes within a genome have the same history, it raises questions on our current tree-based representation methods, and the species concept [99]. Bapteste and colleagues (2009) [97] stated that the belief in the existence of a universal tree of life is stronger than the evidence from genomes to support it. Alternative solutions have been proposed, ranging from the construction of supertrees [100] and phylogenetic trees based on a set of core genes [101] to averaging the phylogenetic signal across a set of genes [98] and constructing trees in which taxa appear several times according to their positions in trees derived from each of the contributing genes [98]. However, each of these tree-based representations fails to represent the true prokaryotic relationships. Therefore, it is most likely that in order to represent true prokaryotic evolution, microbiologists will have to step aside from traditional tree-based representations, and switch to reticulate networks instead [97,102] (Fig. 2.1). Reticulate networks represent evolutionary histories and reflect reticulate events such as hybridization, LGT or recombination between taxa [102]. As current genome sequencing projects

expand, it is likely that in the near future we will be in a much better position to evaluate prokaryotic relationships and deal with classification and taxonomy more effectively [103].

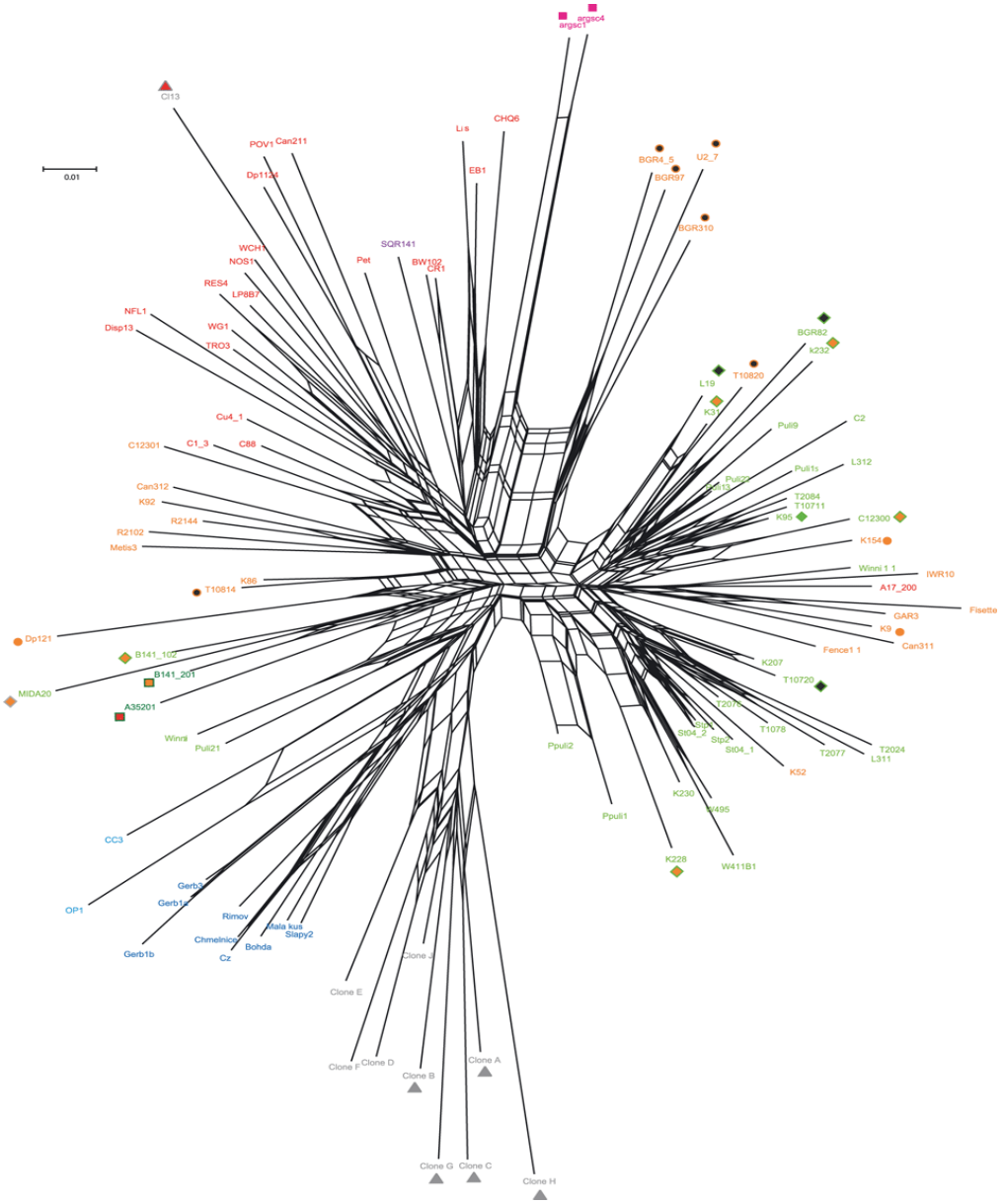


Figure 2.1 Example of a reticulate network representing the evolution of the *Daphnia pulex* complex as revealed by nuclear markers. Extracted from [104].

Still, it should be stressed that, regardless of the weaknesses that are inherent to the tree-building process, current tree-building methods remain useful for deducing single-gene phylogenies. However, it is clear that a distinction has to be made between gene-based phylogeny and organism based phylogeny in prokaryotes, and that single-gene phylogeny does not represent organism based phylogeny as once was assumed. The extent of differences between single-gene and organism based phylogenies is likely to depend on the size of the genetic pool to which the organism is exposed, and thus the environmental niche in which the organism resides.

This literature survey shows that the biggest concern is not a matter of which technique can be used to delineate species, as the ongoing discussion on Average Nucleotide Identity (ANI) versus DNA-DNA hybridization illustrates [105], but rather a matter of whether it is scientifically possible to delineate bacterial species. Current insights in prokaryotic evolution raise the possibility that a Darwinian way of thinking is not transferrable from multicellular eukaryotic organisms to prokaryotes. Nonetheless it is difficult to step aside from a theorem and a way of thinking that has been established longtime and is widely accepted amongst scientists, taxonomists should focus on fully understanding the mechanisms with which prokaryotes evolve and use this information to answer the question on whether a definition of prokaryotic species is justified. It is likely that, with the advent of increasing amounts of prokaryotic genomes being sequenced, more knowledge will be obtained on this matter. Clearly, the last words in this scientifically and philosophically intriguing discussion have not yet been spoken.

2.2 CULTIVATION DEPENDENT PROKARYOTIC DIVERSITY

2.2.1 Great plate-count anomaly

For decades, researchers have struggled with the issue known as ‘the great plate count anomaly’. This implies that only a small fraction of the existing prokaryotic diversity can be cultivated. It is a misconception that the cultivable fraction encompasses the numerically dominant and/or functionally important organisms in their original environments [106]. Instead, the cultivable fraction represents those bacteria that were able to grow on the nutrients, temperature, pH and atmosphere provided during cultivation. Current estimations argue that as many as 31 of an estimated total of 61 bacterial phyla have no cultivable representatives [107]. Explanations for this immense gap can be found in the organism’s growth requirements, which are often not met by artificial growth media and incubation conditions used. Alternatively, competition for nutrients between different organisms cultured together, production of antimicrobial components, as well as presence of growth-inhibiting substances in the growth medium [108] may also be responsible for bacterial growth inhibition.

In some cases, bacteria live in obligate interdependent relationships and rely on the cross-feeding or metabolic cooperation by their symbiotic partners [109]. This cooperation often consists in the production of growth-determining compounds such as siderophores [110,111], vitamins [112,113], specific carbon sources [112,114,115] and other essential nutrients [116,117]. In some cases, signaling molecules are required for bacterial growth. Nichols and colleagues (2008) [118] showed that short peptides were essential factors in initiating growth of non-growing cells.

Cultivation of yet uncultured bacteria does not always require complex modifications to traditional approaches. Increased incubation times [119] and lowered nutrient concentrations have been efficient strategies to induce bacterial growth. Oligotrophic bacteria may be inhibited as a result of the exposure to high nutrient concentrations, by a mechanism referred to as substrate accelerated death [120,121]. Hence, dilution of traditional rich media and the incorporation of polymers as substrates have been reported to result in the cultivation of yet uncultured bacteria. Since polymers must be hydrolyzed before serving as nutrients for bacteria, they prevent a sudden exposure to high nutrient concentrations [122].

It is clear that successful cultivation of yet unculturables is not a straightforward process. Besides creativity, it requires in-depth insight into the biochemistry that is responsible for mediating growth. In theory, any micro-organism can be retrieved on artificial growth media, provided that the right selective conditions are met [123]. Thorough physicochemical analysis of a sample prior to cultivation is thus encouraged and is likely to increase chances of success.

2.2.2 The importance of cultivation

Cultivation of bacteria remains to be an important aspect of microbiology. The great plate count anomaly inherently implies that our understanding of microbiology is largely biased towards the cultivated fraction. Not much is known about the uncultivable fraction beyond their geographical distribution, as obtained from culture-independent approaches. Cultivation is still useful in understanding the metabolism and function of bacteria. As cultivation-independent diversity assessments are frequently based on the amplification of well-chosen target genes, such as the 16S rRNA gene, a thorough characterization of the organisms is impossible. Metagenomic approaches are different in that they don't rely on prior amplification of target genes, but instead aim at sequencing the complete DNA pool within a sample. Although tracing back DNA fragments to the original organisms may be possible for environments that are characterized by a low diversity (such as the well-known acid

main drainage site [124]), it is a nearly impossible task to perform on samples characterized by a high diversity, such as soil and marine ecosystems. It is clear that in order to characterize bacteria, and thus extend our knowledge about microbiology, cultivation remains indispensable. Moreover, applications such as plant growth-promotion for sustainable agriculture and wastewater treatment by activated sludge rely on the application of bacteria, and thus cultivation.

2.3 CULTIVATION INDEPENDENT PROKARYOTIC DIVERSITY

The last few years have known an exponential increase regarding the number of next generation sequencing (NGS) experiments being performed. Many of these studies aim at mapping prokaryotic communities in environments of interest. In order to do so, a taxonomic marker is selected which allows to reveal the identities of all members of a bacterial community. Due to its phylogenetic and taxonomic value, the 16S rRNA gene is often the gene of interest. As a result, public 16S rRNA gene databases have grown exponentially within a very short period of time.

2.3.1 General issues

The major issue with cultivation independent NGS is the limited control over the whole process from DNA-extraction to data analysis and interpretation. It is of utmost importance for a researcher to be aware of the weaknesses that involve these types of analyses. Pitfalls may occur at all stages of the process.

First of all, it is well-known that DNA extraction does not always result in extraction of the complete prokaryotic gene pool due to different susceptibilities of prokaryotic lineages to the DNA extraction protocol. Secondly, a number of amplification artifacts, which include limited primer universality [125,126,127], preferential amplification [128], chimera and heteroduplex formation [129,130,131,132], lowered reproducibility due to barcode-tagged primers [133], error rates of *Taq*-polymerases [128,131] and unwanted co-amplification of host-organelle DNA, may influence the outcome of the experiment. Thirdly, sequencing errors resulting from monomer regions in sequences, multiple template binding on beads, or incomplete removal of nucleotide solution between subsequent flows as in pyrosequencing, may occur and have an effect on sequencing accuracy. Ultimately, a series of decisions made during data processing may directly affect the end results and conclusions drawn from the experiment [134]. Decisions refer to the stringency of settings used for quality filtering, settings applied for chimera detection, in- or excluding singleton (or doubleton) sequences for further analyses and choices of algorithms used for sequence clustering and taxonomic assignments. An even more important limitation

of this relatively new technique is that current NGS instruments only allow sequencing parts of the complete 16S rRNA gene. Hence, only part of the phylogenetic and taxonomic information can be recovered and relied upon to draw final conclusions.

Due to this dependency of cultivation independent studies on this multitude of variables, it can be expected that NGS experiments will not generate reproducible outcomes. This was an observation made by Zhou and colleagues (2011) [135], who ventured to reveal the weaknesses involving NGS. However, it should be clear that we're only at the beginning of an era regarding NGS. Increasing insight in the weaknesses of this technique will certainly allow us to make better supported decisions at each step of the analysis in the future.

2.3.2 Reality of the rare biosphere

A consistent observation that was made with intensive sequencing of different environments was the appearance of very low-abundant sequences from species that had not been characterized previously [136,137,138]. Rarefaction curves constructed with stringent OTU (Operational Taxonomic Unit) cutoff levels frequently illustrated that with increased sampling intensity more sequences would be obtained; even for already intense sequencing efforts. These sequences, which constituted the 'rare biosphere', had systematically been masked by dominant populations in traditional molecular techniques. However, the last few years, increasing concern has arisen about the origin of these sequences. Do these sequences truly lurk in every environment in nature, or are they simply the result of unwanted and uncontrollable processes that occur during PCR and sequencing?

Factors that may account for overestimations of prokaryotic diversity, and thus indirectly contribute to the idea of the existence of a rare biosphere, are chimera formation and PCR- and sequencing errors (§2.3.1). Chimera formation is the process that occurs during PCR, in which original sequences recombine to form new sequences [130] (Fig. 2.2), and consequently new OTUs. Chimera formation was observed to be formed reproducibly among independent amplifications [139].

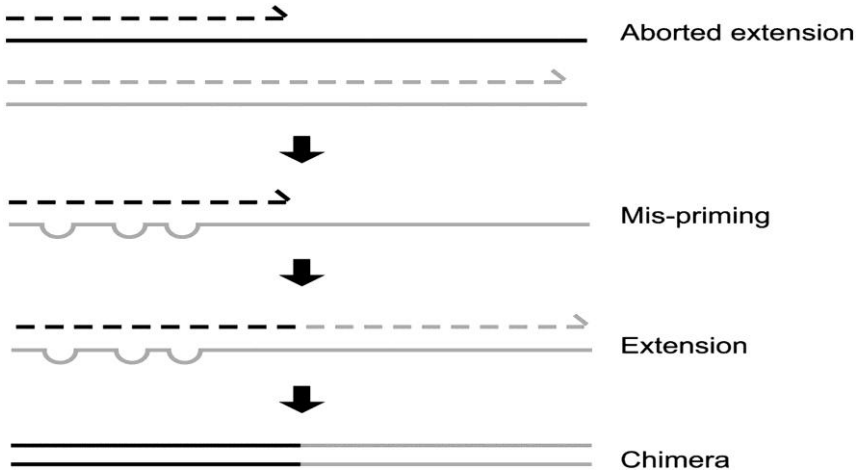


Figure 2.2 The process of chimera formation. Extracted from [139].

Huse and coworkers (2010) [140] mentioned that highly diverse amplicon libraries, free of any conserved regions, may reduce chances of chimera formation. Breakpoint curves (Fig. 2.3C), constructed by Ashelford and colleagues [141], confirmed these findings. Fig. 2.3A was constructed by allocating to each base position in the *E. coli* reference sequence the frequency with which the most common base (A, G, C or T/U) occurred in a collection of 4383 type strain sequences. For instance, a position that was occupied by an adenine in all 4383 sequences was designated frequency 1, while a position on which all bases were equally distributed was designated frequency 0.25. This curve was smoothened by calculating the mean frequency of the most common residues in a 50 base window to present Fig. 2.3B. The smoothened curve shows the known variable and conserved regions of the 16S rRNA gene. Fig. 2.3C was constructed based on observations made with the PinTail software [141] which is used for chimera detection. It shows the locations within the 16S rRNA gene where chimera formation is likely to occur. It is clear that chances of chimera formation alter with the gene region sequenced. Haas and colleagues (2011) [139] were able to assess the effect of the gene region sequenced on chimera formation experimentally, and found that the V6-V9 region showed a higher chimera rate than the V1-V3 region and the V3-V5 region. The highest numbers of chimera breakpoints seem to occur in the conserved regions, which is not surprisingly due to the ease of recombination. A strong positive correlation was observed between sequence similarities of the sequences constituting a chimera, and the amount of chimeras observed [139], indicating that similar sequences are more likely to form chimeras. As breakpoints are located in dispersed regions of the 16S rRNA gene (Fig. 2.3C), splicing may occur at multiple locations. As such, chimera formation is not limited to the recombination of two fragments; chimeric sequences may as well consist of three or more parent sequences.

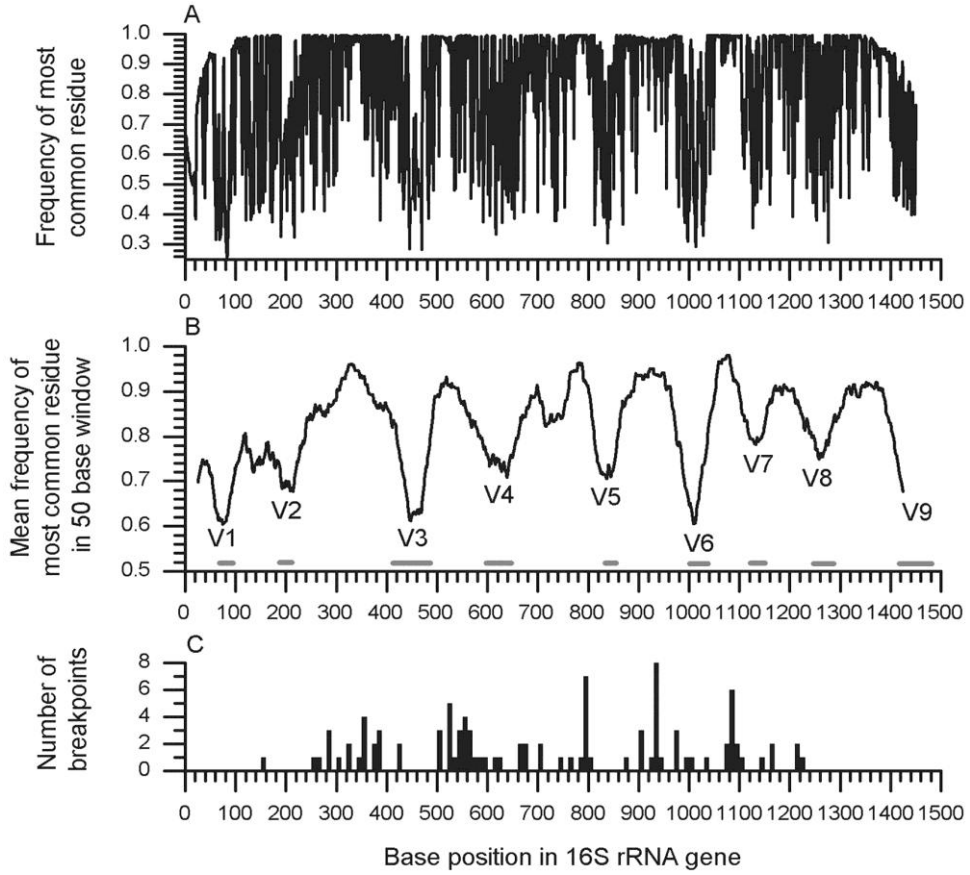


Figure 2.3 The variable regions V1-V9 in the 16S rRNA gene and the positions of chimeric breakpoints. (A) The curve was constructed by allocating to each base position in the *E. coli* reference sequence the frequency with which the most common base (A, G, C or T/U) occurred in a collection of 4383 type strain sequences. (B) The curve was obtained from smoothening the data in (A), by calculating the mean frequency of the most common residues in a 50 base window. It shows the known variable and conserved regions of the 16S rRNA gene. (C) This plot shows the positions in the 16S rRNA gene where chimera formation is likely to occur. Hence, these locations are referred to as chimeric breakpoints. Extracted from [141].

Huse and colleagues (2010) [140] demonstrated that the number of error containing sequences increased with the intensity of the sequencing effort. Therefore, for deep-sequencing experiments, identification of erroneous sequences during data pre-processing is advisable. Filtering of chimeric sequences was shown to reduce the amount of incorrect OTU assignments, and thus constituted to a diversity assessment that better reflected reality [136]. Several programs such as PinTail [141], Bellerophon [142], Chimera Slayer [139] and Perseus [143] were specifically designed to detect chimeras. Quince et al. (2009) [144] demonstrated that, due to sequencing errors, species diversity estimates (i.e. the total number of species extrapolated from a finite sample) were at least one order of

magnitude higher than the actual diversity in the sample (Fig. 2.4). This underlines the need for noise-removal methods. However, Reeder and Knight (2009) [136] argued that even after noise removal the total number of sequences is still heavily inflated. They postulated that a large proportion of sequencing errors probably constitute the rare biosphere.

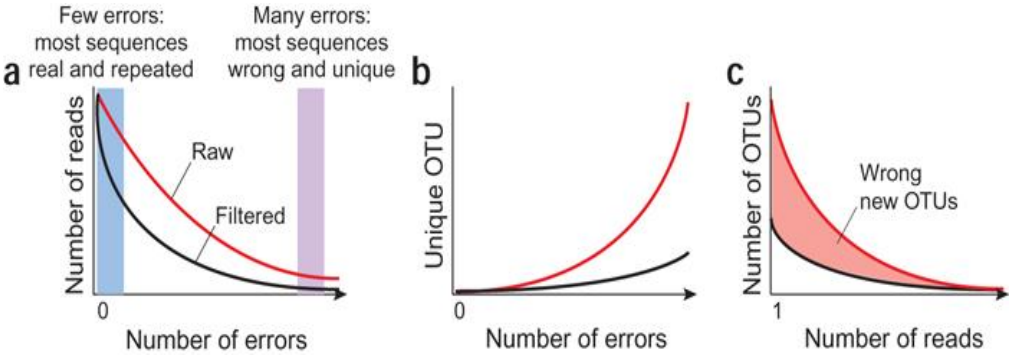


Figure 2.4 The effect of noise removal on diversity estimations. (a) High number of errors gives rise to unique sequences, while repeated sequences indicate few errors and thus real sequences. The graph shows that raw and filtered reads differ mainly in the small number of reads that contain many errors, indicating that filtering does not reject the real sequences. (b) Due to rejecting the erroneous unique reads, a lower number of OTUs is obtained after filtering. (c) Removal of single reads that constitute the rare biosphere has a significant effect on OTU-richness. Extracted from [136].

Even more alarming was the observation that a significant proportion of sequences available in public repositories were found to contain anomalies, with chimeras accounting for the majority of problematic sequences [141]. This illustrates that the problem is not limited to individual studies, but any researcher consulting these databases will indirectly include these anomalies in their work. Moreover, chimera detection programs such as Chimera Slayer [139] rely on the detection of parent reference sequences from public databases. These considerations stress the need for raised awareness amongst researchers and database curators.

Template dilution [128,131], PCR cycle number [128,131], elongation time [131], sample species diversity [131] and the type of *Taq*-polymerase used for PCR amplification [128,132] all affect PCR accuracy, and consequently diversity. Low annealing temperatures may lead to non-specific primer binding due to reduced stringency, and result in amplification of non-target PCR products. If not checked for, it may lead to increased OTU richness. Huse et al. (2007) [145] showed that sequencing errors may result from the presence of homopolymer sequences (i.e. a series of the same bases in a sequence), insufficient flushing between successive flows, nonsynchronized read extensions and multiple templates binding to a single bead on the picotiterplate. However, these errors only account for pyrosequencing.

Due to different operating procedures in other NGS platforms, different error types may occur. Presence of ambiguous bases in the sequences and aberrant read lengths relative to the expected read length were considered indicators of low-quality sequences [145]. Removal of such sequences will presumably improve the quality of the remaining dataset.

Lynch et al. (2012) [138] showed that not all constituents of the rare biosphere result from PCR and sequencing errors. In their approach, they selected low-abundance OTUs (less than 1.10^{-4}% of all sequences in an environmental DNA extract) with only weak similarities to known organisms in a dataset consisting of approximately 6.5 million assembled paired-end Illumina reads from the 16S rRNA gene. Based on these sequences, the authors designed specific forward primers from the highly variable 3' end of the V3 region, and used these primers in combination with the universal 1492r primer to amplify the corresponding genes from the same environmental sample as which the 6.5 million read library was constructed from. The appearance of an amplification product, and sequencing of the obtained amplicons, showed that at least part of the sequences that were identified as belonging to the rare biosphere were genuine organisms residing in nature in very low abundances. The nearly complete 16S rRNA gene amplicons allowed a thorough phylogenetic study of the sequences obtained. The phylogenetic tree containing the members of the rare biosphere is presented in Fig. 2.5.

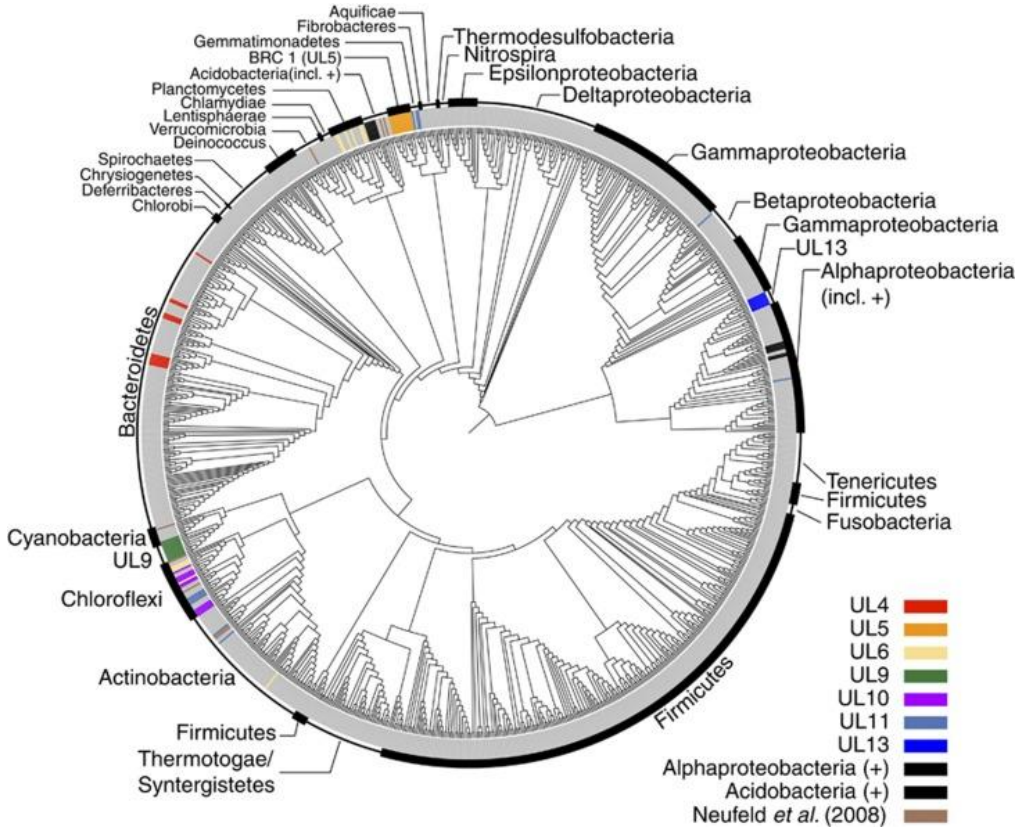


Figure 2.5 Maximum Likelihood tree constructed from the nearly full-length 16S rRNA gene sequences (UL4-UL13) that constitute the rare biosphere in the study by Lynch et al. (2012) [138].

It is clear that sequences that constitute the rare biosphere partly originate from sequencing and PCR artifacts and partly originate from genuine low abundance organisms residing in specific environments. However, this raises the problem of how to differentiate genuine rare biosphere sequences from erroneous sequences. Some researchers argue that the best way of recognizing a genuine rare biosphere species would be by detecting its appearance across many biological samples. This may indicate that the sequence was not obtained from sequencing error. The construction of large databases containing sequences from numerous different environments thus aids the differentiation of sequencing errors from genuine sequences. However, as chimera formation was found to occur reproducibly [139], the proposed method may not always be effective. Therefore, the best option to prevent chimera formation is by applying experimental conditions that prevent their formation [129,132,139,146,147]. Still, as a small number of chimeras may slip through, we heavily rely on bioinformatics programs for

their detection. While the effect of chimeras on OTU richness will not be filtered by lowering OTU cutoff values, the latter may aid to disregard the effect of PCR or sequencing mutations [128,132].

2.3.3 Other factors biasing prokaryotic diversity estimations

Huse et al. (2010) [140] and Sun and colleagues (2009) [148] showed that increased OTU richness was not only due to PCR and sequencing errors. They found that the alignment strategy used (i.e. pairwise versus multiple alignment) and the clustering algorithm (i.e. single linkage, average linkage or complete linkage) had a significant effect on OTU richness. Publications reported that the average linkage clustering method was least sensitive to sequencing noise and was the most robust amongst the different methods [144,149]. Including a single linkage preclustering step was found to reduce the number of spurious OTUs in data sets of known composition by approximately 90% [140]. Efficiency in spurious OTU removal was comparable to PyroNoise [144]. However, the computational expense of running PyroNoise was significantly higher compared to single linkage preclustering.

Experiments indicated that sequence length had an effect on OTU richness [140]. For instance, at the 3% OTU level, 400 bp reads are allowed to contain 12 erroneous nucleotides (relative to the original template) to be included in the same OTU, whereas 100 bp reads are allowed to only contain 3 erroneous nucleotides. Although error rates increase with increasing read lengths, the relation may not always be linear. Consequently, shorter reads are more likely to show increased OTU richness compared to longer reads (due to a reduced error buffering effect).

Similarly, sampling depth was identified as yet another factor influencing the number of spurious OTUs. As indicated in Fig. 2.6, OTU richness increased with increasing sampling depth, although not linear.

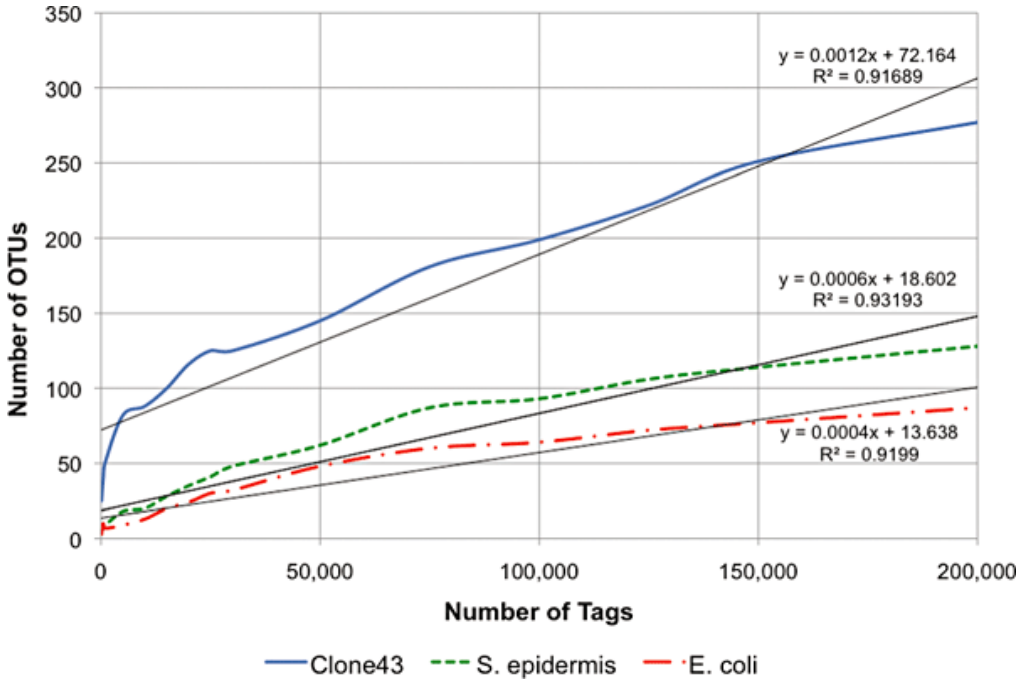


Figure 2.6 The effect of sampling depth on OTU richness. The graph shows the number of spurious OTUs (in three samples Clone43, *S. epidermidis* and *E. coli*) as a function of sampling depth. Extracted from [140].

Non-specific primer binding to non-rRNA sequences within the DNA pool may contribute to the creation of spurious OTUs [140]. These non-target sequences, however, can be identified by their anomalously poor alignment against target sequences. Consequently, by disregarding all sequences that show less than a predetermined alignment length with known target sequences, a filtered dataset can be obtained in which only target sequences will be retained. This approach may also help to detect and disregard chimeric sequences.

2.3.4 How culture independent techniques aid the cultivation of organisms

Culture independent techniques can assist the isolation of yet uncultivated organisms. For instance, cultivation independent sequencing techniques may direct the development of specific probes, which can be used to screen a series of samples for presence of the target organism [107]. The target organism then needs to be selectively enriched or physically isolated from non-target organisms in the sample [106,150,151]. However, often information on metabolic pathways is lacking and physical isolation of the target organism is preferred. Several methods have been developed to physically isolate microorganisms. These include sample dilution, single cell encapsulation combined with flow cytometry, micromanipulators and optical tweezers, filtration, cell sorting by flow cytometry and density-gradient

centrifugation [106,152]. Once isolated, organism growth can be established through trial and error by applying high-throughput culturing methods, diffusion growth chambers [153], cell-free extracts, extended incubation times [154] and use of gellan gum as a solidifying agent instead of agar [155]. Alternatively, the isolated organism's genome can be sequenced and direct the isolation of the organism. Indeed, the genome sequence of a given organism allows insight into the growth requirements of the organism. Identification and analysis of genes involved in the organism's metabolism may reveal its requirements with respect to carbon source, energy source and electron acceptors.

BIBLIOGRAPHY

1. Haas D, Defago G (2005) Biological control of soil-borne pathogens by fluorescent pseudomonads. Nature Reviews Microbiology 3: 307-319.
2. Fravel DR (2005) Commercialization and implementation of biocontrol. Annual Review of Phytopathology 43: 337-359.
3. Motisi N, Dore T, Lucas P, Montfort F (2010) Dealing with the variability in biofumigation efficacy through an epidemiological framework. Soil Biology & Biochemistry 42: 2044-2057.
4. Glare T, Caradus J, Gelernter W, Jackson T, Keyhani N, et al. (2012) Have biopesticides come of age? Trends in Biotechnology 30: 250-258.
5. Bailey A, Chandler D, Grant WP, Greaves J, Prince G, editors (2011) Biopesticides: Pest Management and Regulation: CABI.
6. Marrone P (2011) Biopesticides - Legitimate products for integration into IPM and certified organic production.
7. Copping LG (2004) The Manual of Biocontrol Agents, Fourth Edition. Alton, Hampshire, UK: CABI. pp. 896.
8. Lugtenberg B, Kamilova F (2009) Plant-Growth-Promoting Rhizobacteria. Annual Review of Microbiology 63: 541-556.
9. Kamilova F, Kravchenko LV, Shaposhnikov AI, Azarova T, Makarova N, et al. (2006) Organic acids, sugars, and L-tryptophane in exudates of vegetables growing on stonewool and their effects on activities of rhizosphere bacteria. Molecular Plant-Microbe Interactions 19: 250-256.
10. Kamilova F, Kravchenko LV, Shaposhnikov AI, Makarova N, Lugtenberg B (2006) Effects of the tomato pathogen *Fusarium oxysporum* f. sp *radicis-lycopersici* and of the biocontrol bacterium *Pseudomonas fluorescens* WCS365 on the composition of organic acids and sugars in tomato root exudate. Molecular Plant-Microbe Interactions 19: 1121-1126.
11. Meharg AA, Killham K (1995) Loss of exudates from the roots of perennial ryegrass inoculated with a range of microorganisms. Plant and Soil 170: 345-349.
12. Phillips DA, Fox TC, King MD, Bhuvaneswari TV, Teuber LR (2004) Microbial products trigger amino acid exudation from plant roots. Plant Physiology 136: 2887-2894.
13. Song FB, Han XY, Zhu XC, Herbert SJ (2012) Response to water stress of soil enzymes and root exudates from drought and non-drought tolerant corn hybrids at different growth stages. Canadian Journal of Soil Science 92: 501-507.

14. Neal AL, Ahmad S, Gordon-Weeks R, Ton J (2012) Benzoxazinoids in Root Exudates of Maize Attract *Pseudomonas putida* to the Rhizosphere. Plos One 7: 10.

15. de Weert S, Vermeiren H, Mulders IHM, Kuiper I, Hendrickx N, et al. (2002) Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens*. Molecular Plant-Microbe Interactions 15: 1173-1180.

16. Mulas D, Garcia-Fraile P, Carro L, Ramirez-Bahena MH, Casquero P, et al. (2011) Distribution and efficiency of *Rhizobium leguminosarum* strains nodulating *Phaseolus vulgaris* in Northern Spanish soils: Selection of native strains that replace conventional N fertilization. Soil Biology & Biochemistry 43: 2283-2293.

17. Marra LM, Soares C, de Oliveira SM, Ferreira PAA, Soares BL, et al. (2012) Biological nitrogen fixation and phosphate solubilization by bacteria isolated from tropical soils. Plant and Soil 357: 289-307.

18. Kuiper I, Lagendijk EL, Bloemberg GV, Lugtenberg BJJ (2004) Rhizoremediation: A beneficial plant-microbe interaction. Molecular Plant-Microbe Interactions 17: 6-15.

19. Hussain A, Hasnain S (2011) Phytostimulation and biofertilization in wheat by cyanobacteria. Journal of Industrial Microbiology & Biotechnology 38: 85-92.

20. Farag MA, Ryu CM, Sumner LW, Pare PW (2006) GC-MS SPME profiling of rhizobacterial volatiles reveals prospective inducers of growth promotion and induced systemic resistance in plants. Phytochemistry 67: 2262-2268.

21. Barazani O, Friedman J (2000) Effect of exogenously applied L-tryptophan on allelochemical activity of plant-growth-promoting rhizobacteria (PGPR). Journal of Chemical Ecology 26: 343-349.

22. Glick BR, Todorovic B, Czarny J, Cheng ZY, Duan J, et al. (2007) Promotion of plant growth by bacterial ACC deaminase. Critical Reviews in Plant Sciences 26: 227-242.

23. Kamilova F, Validov S, Azarova T, Mulders I, Lugtenberg B (2005) Enrichment for enhanced competitive plant root tip colonizers selects for a new class of biocontrol bacteria. Environmental Microbiology 7: 1809-1817.

24. Haran S, Schickler H, Chet I (1996) Molecular mechanisms of lytic enzymes involved in the biocontrol activity of *Trichoderma harzianum*. Microbiology-Uk 142: 2321-2331.

25. Lorito M, Harman GE, Hayes CK, Broadway RM, Tronsmo A, et al. (1993) Chitinolytic enzymes produced by *Trichoderma harzianum* - Antifungal activity of purified endochitinase and chitobiosidase. Phytopathology 83: 302-307.

26. Schirmbock M, Lorito M, Wang YL, Hayes CK, Arisanatac I, et al. (1994) Parallel formation and synergism of hydrolytic enzymes and peptaibol antibiotics, molecular mechanisms involved in

the antagonistic action of *Trichoderma harzianum* against phytopathogenic fungi. Applied and Environmental Microbiology 60: 4364-4370.

27. van Loon LC (2007) Plant responses to plant growth-promoting bacteria. European Journal of Plant Pathology 119: 243-254.
28. Tsror L (2010) Biology, Epidemiology and Management of *Rhizoctonia solani* on Potato. Journal of Phytopathology 158: 649-658.
29. Runno-Paurson E, Remmel T, Ojarand A, Aav A, Mand M (2010) The structure of *Phytophthora infestans* populations from organic and conventional crops. European Journal of Plant Pathology 128: 373-383.
30. Costanzo S, Christ BJ, Haynes KG (2004) Late blight resistance in a diploid full-sib potato family. Plant Breeding 123: 377-381.
31. Ghyselinck J, Velivelli SL, Heylen K, O'Herlihy E, Franco J, et al. (2013) Bioprospecting in potato fields in the Central Andean Highlands: Screening of rhizobacteria for plant growth-promoting properties. Systematic and Applied Microbiology.
32. Axel C, Zannini E, Coffey A, Guo JH, Waters DM, et al. (2012) Ecofriendly control of potato late blight causative agent and the potential role of lactic acid bacteria: a review. Applied Microbiology and Biotechnology 96: 37-48.
33. Mizubuti ESG, Júnior VL, Forbes GA (2007) Management of Late Blight with Alternative Products. Pest Tech 1: 106-116.
34. El Bakali AM, Martin MP (2006) Black scurf of potato. Mycologist 20: 130-132.
35. Sneh B, Burpee L, Ogoshi A, editors (1991) Identification of *Rhizoconia* species. St. Paul, USA: American Phytopathological Society.
36. Wilson P, Ahvenniemi P, Lehtonen M, Kukkonen M, Rita H, et al. (2008) Biological and chemical control and their combined use to control different stages of the *Rhizoctonia* disease complex on potato through the growing season. Annals of Applied Biology 153: 307-320.
37. Lahlali R, Hijri M (2010) Screening, identification and evaluation of potential biocontrol fungal endophytes against *Rhizoctonia solani* AG3 on potato plants. FEMS Microbiology Letters 311: 152-159.
38. Grosch R, Lottmann J, Rehn VNC, Rehn KG, Mendonca-Hagler L, et al. (2007) Analysis of antagonistic interactions between *Trichoderma* isolates from Brazilian weeds and the soil-borne pathogen *Rhizoctonia solani*. Journal of Plant Diseases and Protection 114: 167-175.
39. Brewer MT, Larkin RP (2005) Efficacy of several potential biocontrol organisms against *Rhizoctonia solani* on potato. Crop Protection 24: 939-950.

40. Jager G, Velvis H, Lamers JG, Mulder A, Roosjen J (1991) Control of *Rhizoctonia solani* in potato by biological, chemical and integrated measures. Potato Research 34: 269-284.

41. van den Boogert P, Luttikholt AJG (2004) Compatible biological and chemical control systems for *Rhizoctonia solani* in potato. European Journal of Plant Pathology 110: 111-118.

42. Lewis JA, Fravel DR, Papavizas GC (1995) *Cladorrhinum foecundissimum* - A potential biological control agent for the reduction of *Rhizoctonia solani*. Soil Biology & Biochemistry 27: 863-869.

43. Honeycutt EW, Benson DM (2001) Formulation of binucleate *Rhizoctonia* spp. and biocontrol of *Rhizoctonia solani* on impatiens. Plant Disease 85: 1241-1248.

44. Sabaratnam S, Traquair JA (2002) Formulation of a *Streptomyces* biocontrol agent for the suppression of *Rhizoctonia* damping-off in tomato transplants. Biological Control 23: 245-253.

45. Lewis JA, Fravel DR, Lumsden RD, Shasha BS (1995) Application of biocontrol fungi in granular formulations of pregelatinized starch-flour to control damping-off diseases caused by *Rhizoctonia solani*. Biological Control 5: 397-404.

46. Gonzalez-Sanchez MA, Perez-Jimenez RM, Pliego C, Ramos C, de Vicente A, et al. (2010) Biocontrol bacteria selected by a direct plant protection strategy against avocado white root rot show antagonism as a prevalent trait. Journal of Applied Microbiology 109: 65-78.

47. Yangui T, Rhouma A, Triki MA, Gargouri K, Bouzid J (2008) Control of damping-off caused by *Rhizoctonia solani* and *Fusarium solani* using olive mill waste water and some of its indigenous bacterial strains. Crop Protection 27: 189-197.

48. De Curtis F, Lima G, Vitullo D, De Cicco V (2010) Biocontrol of *Rhizoctonia solani* and *Sclerotium rolfsii* on tomato by delivering antagonistic bacteria through a drip irrigation system. Crop Protection 29: 663-670.

49. Grosch R, Faltin F, Lottmann J, Kofoet A, Berg G (2005) Effectiveness of 3 antagonistic bacterial isolates to control *Rhizoctonia solani* Kuhn on lettuce and potato. Canadian Journal of Microbiology 51: 345-353.

50. Ikeda S, Shimizu A, Shimizu M, Takahashi H, Takenaka S (2012) Biocontrol of black scurf on potato by seed tuber treatment with *Pythium oligandrum*. Biological Control 60: 297-304.

51. Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009) Applied Biotechnology to Combat Late Blight in Potato Caused by *Phytophthora Infestans*. Potato Research 52: 249-264.

52. VIB (2010) Scientific background report *Phytophthora*-resistant potatoes.

53. Cooke LR, Schepers H, Hermansen A, Bain RA, Bradshaw NJ, et al. (2011) Epidemiology and Integrated Control of Potato Late Blight in Europe. Potato Research 54: 183-222.

54. Fry WE, Goodwin SB (1997) Resurgence of the Irish potato famine fungus. Bioscience 47: 363-371.

55. Flemming CA, Trevors JT (1989) Copper toxicity and chemistry in the environment - A review. Water Air and Soil Pollution 44: 143-158.

56. Stern BR, Solioz M, Krewski D, Aggett P, Aw TC, et al. (2007) Copper and human health: Biochemistry, genetics, and strategies for modeling dose-response relationships. Journal of Toxicology and Environmental Health-Part B-Critical Reviews 10: 157-222.

57. Wharton PS, Kirk WW, Schafer RL, Tumbalam P (2012) Evaluation of biological seed treatments in combination with management practices for the control of seed-borne late blight in potato. Biological Control 63: 326-332.

58. Dorn B, Musa T, Krebs H, Fried P, Forrer H (2007) Control of late blight in organic potato production: evaluation of copper-free preparations under field, growth chamber and laboratory conditions. European Journal of Plant Pathology 119: 217-240.

59. Lidert Z (2001) Biopesticides: Is there a path to commercial success? In: Vurro M, Gressel J, Butt T, Harman GE, Pilgeram A et al., editors. Enhancing Biocontrol Agents and Handling Risks. Amsterdam: IOS Press. pp. 295.

60. IPPC, FAO (2005) Guidelines for the export, shipment, import and release of biological control agents and other beneficial organisms. Rome.

61. OECD ED (2003) Guidance for Registration Requirements for Microbial Pesticides. Paris Cedex, France.

62. Mari M, Bertolini P, Pratella GC (2003) Non-conventional methods for the control of post-harvest pear diseases. Journal of Applied Microbiology 94: 761-766.

63. Connick WJ, Daigle DJ, Boyette CD, Williams KS, Vinyard BT, et al. (1996) Water activity and other factors that affect the viability of *Colletotrichum truncatum* Conidia in wheat flour-kaolin granules ('Pesta'). Biocontrol Science and Technology 6: 277-284.

64. Burges HD, editor (1998) Formulation of Microbial Biopesticides - Beneficial Microorganisms, Nematodes and Seed Treatments. Dordrecht: Kluwer. 412 p.

65. Minuto A, Migheli Q, Garibaldi A (1995) Evaluation of antagonistic strains of *Fusarium* spp in the biologial and integrated control of Fusarium wilt of cyclamen Crop Protection 14: 221-226.

66. Landa BB, Navas-Cortes JA, Jimenez-Diaz RM (2004) Integrated management of fusarium wilt of chickpea with sowing date, host resistance, and biological control. Phytopathology 94: 946-960.

67. Lumsden RD, Lewis JA, Fravel DR (1995) Formulation and delivery of biocontrol agents for use against soilborne plant pathogens In: Hall FR, Barry JW, editors. Biorational Pest Control Agents: Formulation and Delivery. Washington: Amer Chemical Soc. pp. 166-182.

68. Thomson SV, Hansen DR, Flint KM (1992) Dissemination of bacteria antagonistic to *Erwinia amylovora* by honey bees. Plant Disease 76: 1052-1056.

69. Gracia-Garza JA, Fravel DR, Bailey BA, Hebbar PK (1998) Dispersal of formulations of *Fusarium oxysporum* f. sp. *erythroxyli* and *F. oxysporum* f. sp. *melonis* by ants. Phytopathology 88: 185-189.

70. Gullino ML, Garibaldi A (2007) Critical aspects in management of fungal diseases of ornamental plants and directions in research. Phytopathologia Mediterranea 46: 135-149.

71. Huang HC, Bremer E, Hynes RK, Erickson RS (2000) Foliar application of fungal biocontrol agents for the control of white mold of dry bean caused by *Sclerotinia sclerotiorum*. Biological Control 18: 270-276.

72. Zhang Q, Wang BC, Cao ZY, Yu YL (2012) Plasmid-mediated bioaugmentation for the degradation of chlorpyrifos in soil. Journal of Hazardous Materials 221: 178-184.

73. Wang YJ, Li H, Zhao W, He XL, Chen J, et al. (2010) Induction of toluene degradation and growth promotion in corn and wheat by horizontal gene transfer within endophytic bacteria. Soil Biology & Biochemistry 42: 1051-1057.

74. Dejonghe W, Goris J, El Fantroussi S, Hofte M, De Vos P, et al. (2000) Effect of dissemination of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation plasmids on 2,4-D degradation and on bacterial community structure in two different soil horizons. Applied and Environmental Microbiology 66: 3297-3304.

75. De Maayer P, Chan WY, Blom J, Venter SN, Duffy B, et al. (2012) The large universal *Pantoea* plasmid LPP-1 plays a major role in biological and ecological diversification. BMC Genomics 13: 12.

76. Salman M (2010) Determination of antibiotic activity on plasmids from fluorescent pseudomonads isolates CW2, WB15 and WB52 against pre-emergence damping-off caused by *Pythium ultimum* and *Rhizoctonia solani* in cucumber. Biological Control 53: 161-167.

77. Balch WE, Magrum LJ, Fox GE, Wolfe RS, Woese CR (1977) An ancient divergence among the bacteria. Journal of Molecular Evolution 9: 305-311.

78. Doolittle WF (2000) Uprooting the tree of life. Scientific American 282: 90-95.

79. Ludwig W, Schleifer KH (1994) Bacterial phylogeny based on 16S and 23S ribosomal RNA sequence-analysis. FEMS Microbiology Reviews 15: 155-173.

80. Ludwig W, Klenk H-P (2001) Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In: Boone DR, Castenholz RW, Garrity GM, editors. Bergey's Manual of Systematic Bacteriology. 2 ed. New York: Springer. pp. 49-65.

81. Yarza P, Ludwig W, Euzeby J, Amann R, Schleifer KH, et al. (2010) Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. Systematic and Applied Microbiology 33: 291-299.

82. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, et al. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic Bacteriology 37: 463-464.

83. Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiology Today 33: 152-155.

84. Thompson CC, Thompson FL, Vandemeulebroecke K, Hoste B, Dawyndt P, et al. (2004) Use of *recA* as an alternative phylogenetic marker in the family *Vibrionaceae*. International Journal of Systematic and Evolutionary Microbiology 54: 919-924.

85. Kim KS, Ko KS, Chang MW, Hahn TW, Hong SK, et al. (2003) Use of *rpoB* sequences for phylogenetic study of *Mycoplasma* species. FEMS Microbiology Letters 226: 299-305.

86. Farfan M, Minana-Galbis D, Garreta A, Loren JG, Fuste MC (2010) Malate dehydrogenase: A useful phylogenetic marker for the genus *Aeromonas*. Systematic and Applied Microbiology 33: 427-435.

87. Scholz HC, Al Dahouk S, Tomaso H, Neubauer H, Witte A, et al. (2008) Genetic diversity and phylogenetic relationships of bacteria belonging to the *Ochrobactrum-Brucella* group by *recA* and 16S rRNA gene-based comparative sequence analysis. Systematic and Applied Microbiology 31: 1-16.

88. Eisen JA (1995) The *RecA* protein as a model molecule for molecular systematic studies of bacteria: Comparison of trees of *RecA*s and 16S rRNAs from the same species. Journal of Molecular Evolution 41: 1105-1123.

89. Gruber TM, Bryant DA (1997) Molecular systematic studies of eubacteria, using sigma(70)-type sigma factors of group 1 and group 2. Journal of Bacteriology 179: 1734-1747.

90. Karlin S, Weinstock GM, Brendel V (1995) Bacterial classifications derived from *recA* protein sequence comparisons. Journal of Bacteriology 177: 6881-6893.

91. Arahal DR, Sanchez E, Macian MC, Garay E (2008) Value of *recN* sequences for species identification and as a phylogenetic marker within the family "*Leuconostocaceae*". International Microbiology 11: 33-39.

92. Lynch M, O'Hely M, Walsh B, Force A (2001) The probability of preservation of a newly arisen gene duplicate. Genetics 159: 1789-1804.

93. Gest H (2003) Names of bacteria and their evolutionary relationships. Microbiology-Sgm 149: 1956-1958.

94. Boucher Y, Bapteste E (2009) Revisiting the concept of lineage in prokaryotes: a phylogenetic perspective. BioEssays 31: 526-536.

95. Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: The agents of open source evolution. Nature Reviews Microbiology 3: 722-732.

96. Lang AS, Beatty JT (2007) Importance of widespread gene transfer agent genes in alpha-proteobacteria. Trends in Microbiology 15: 54-62.

97. Bapteste E, O'Malley MA, Beiko RG, Ereshefsky M, Gogarten JP, et al. (2009) Prokaryotic evolution and the tree of life are two different things. Biology Direct 4: 20.

98. Galtier N, Daubin V (2008) Dealing with incongruence in phylogenomic analyses. Philosophical Transactions of the Royal Society of London B Biological Sciences 363: 4023-4029.

99. Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. Genome Biology 7: 7.

100. Bininda-Emonds OR (2004) The evolution of supertrees. Trends in Ecology & Evolution 19: 315-322.

101. Daubin V, Gouy M, Perriere G (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome Research 12: 1080-1090.

102. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267.

103. Gest H (1999) Bacterial classification and taxonomy: a 'primer' for the new millenium. Microbiology Today 26: 70-72.

104. Vergilino R, Markova S, Ventura M, Manca M, Dufresne F (2011) Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. Molecular Ecology 20: 1191-1207.

105. Richter M, Rossello-Mora R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences of the United States of America 106: 19126-19131.

106. Hugenholtz P (2002) Exploring prokaryotic diversity in the genomic era. Genome Biology 3: 8.

107. Vartoukian SR, Palmer RM, Wade WG (2010) Strategies for culture of 'unculturable' bacteria. FEMS Microbiology Letters 309: 1-7.

108. Tamaki H, Sekiguchi Y, Hanada S, Nakamura K, Nomura N, et al. (2005) Comparative analysis of bacterial diversity in freshwater sediment of a shallow eutrophic lake by molecular and improved cultivation-based techniques. Applied and Environmental Microbiology 71: 2162-2169.

109. Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, et al. (2006) Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. Applied and Environmental Microbiology 72: 3593-3599.

110. Guan LL, Onuki H, Kamino K (2000) Bacterial growth stimulation with exogenous siderophore and synthetic N-acyl homoserine lactone autoinducers under iron-limited and low-nutrient conditions. Applied and Environmental Microbiology 66: 2797-2803.

111. Thanh VN, Van Dyk MS, Wingfield MJ (2002) *Debaryomyces mycophilus* sp nov., a siderophore-dependent yeast isolated from woodlice. FEMS Yeast Research 2: 415-427.

112. Cavedon K, Canaleparola E (1992) Physiological interactions between a mesophilic cellulolytic *Clostridium* and a non-cellulolytic bacterium. FEMS Microbiology Ecology 86: 237-245.

113. Graber JR, Breznak JA (2005) Folate cross-feeding supports symbiotic homoacetogenic spirochetes. Applied and Environmental Microbiology 71: 1883-1889.

114. Brown SA, Whiteley M (2007) A novel Exclusion mechanism for carbon resource partitioning in *Aggregatibacter actinomycetemcomitans*. Journal of Bacteriology 189: 6407-6414.

115. van der Meer MTJ, Schouten S, Damste JSS, Ward DM (2007) Impact of carbon metabolism on C-13 signatures of cyanobacteria and green non-sulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). Environmental Microbiology 9: 482-491.

116. Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, et al. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature 452: 741-744.

117. Dangmann E, Stolz A, Kuhm AE, Hammer A, Feigel B, et al. (1996) Degradation of 4-aminobenzenesulfonate by a two-species bacterial coculture - Physiological interactions between *Hydrogenophaga palleronii* S1 and *Agrobacterium radiobacter* S2. Biodegradation 7: 223-229.

118. Nichols D, Lewis K, Orjala J, Mo S, Ortenberg R, et al. (2008) Short peptide induces an "uncultivable" microorganism to grow in vitro. Applied and Environmental Microbiology 74: 4889-4897.

119. Sait M, Hugenholtz P, Janssen PH (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environmental Microbiology 4: 654-666.

120. Straskrabova V (1983) The effect of substrate shock on populations of starving aquatic bacteria. Journal of Applied Bacteriology 54: 217-224.

121. Postgate JR, Hunter JR (1964) Accelerated death of *Aerobacter aerogenes* starved in the presence of growth-limiting substrates. J Gen Microbiol 34: 459-473.

122. Chin KJ, Hahn D, Hengstmann U, Liesack W, Janssen PH (1999) Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. Applied and Environmental Microbiology 65: 5042-5049.

123. Pedros-Alio C (2007) Dipping into the rare biosphere. Science 315: 192-193.

124. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428: 37-43.

125. Mao D-P, Zhou Q, Chen C-Y, Quan Z-X (2012) Coverage evaluation of universal bacterial primers using the metagenomic datasets. BMC Microbiology 12.

126. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. (2012) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research.

127. Wang Y, Qian P-Y (2009) Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. Plos One 4.

128. Wu JY, Jiang XT, Jiang YX, Lu SY, Zou F, et al. (2010) Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiology 10: 7.

129. Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Research 30: 2083-2088.

130. Shuldiner AR, Nirula A, Roth J (1989) Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Research 17: 4409-4409.

131. Qiu XY, Wu LY, Huang HS, McDonel PE, Palumbo AV, et al. (2001) Evaluation of PCR-generated chimeras: Mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology 67: 880-887.

132. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology 71: 8966-8969.

133. Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification Applied and Environmental Microbiology 78: 612-612.

134. Schloss PD (2010) The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. Plos Computational Biology 6.

135. Zhou J, Wu L, Deng Y, Zhi X, Jiang Y-H, et al. (2011) Reproducibility and quantitation of amplicon sequencing-based detection. Isme Journal 5: 1303-1313.

136. Reeder J, Knight R (2009) The 'rare biosphere': a reality check. Nature Methods 6: 636-637.

137. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the United States of America 103: 12115-12120.

138. Lynch MDJ, Bartram AK, Neufeld JD (2012) Targeted recovery of novel phylogenetic diversity from next-generation sequence data. Isme Journal 6: 2067-2077.

139. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research 21: 494-504.

140. Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology 12: 1889-1898.

141. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Applied and Environmental Microbiology 71: 7724-7736.

142. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology 72: 5069-5072.

143. Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing Noise From Pyrosequenced Amplicons. BMC Bioinformatics 12: 18.

144. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nature Methods 6: 639-U627.

145. Huse SM, Huber JA, Morrison HG, Sogin ML, Mark Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology 8.

146. Lahr DJG, Katz LA (2009) Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. BioTechniques 47: 857-863.

147. Wang GCY, Wang Y (1996) The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology-Uk 142: 1107-1114.

148. Sun Y, Cai Y, Liu L, Yu F, Farrell ML, et al. (2009) ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. Nucleic Acids Research 37.

149. Schloss PD, Westcott SL (2011) Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. Applied and Environmental Microbiology 77: 3219-3226.

150. Huber R, Burggraf S, Mayer T, Barns SM, Rossnagel P, et al. (1995) Isolation of a hyperthermophilic archaeum predicted by in situ RNA analysis. Nature 376: 57-58.

151. Pham VHT, Kim J (2012) Cultivation of unculturable soil bacteria. Trends in Biotechnology 30: 475-484.

152. Liesack W, Janssen PH, Rainey FA, Ward-Rainey NL, Stackebrandt E (1997) Microbial diversity in soil: the need for a combined approach using molecular and cultivation techniques. In: Van Elsas JD, Trevors JT, Wellington EMH, editors. Modern Soil Microbiology. New York: Marcel Dekker. pp. 375-439.

153. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science 296: 1127-1129.

154. Davis KER, Joseph SJ, Janssen PH (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. Applied and Environmental Microbiology 71: 826-834.

155. Tamaki H, Hanada S, Sekiguchi Y, Tanaka Y, Kamagata Y (2009) Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. Environmental Microbiology 11: 1827-1834.

PART III

Experimental Work

Author Contributions :

JG performed the experiments and wrote the manuscript. KVH gave advice and made suggestions concerning statistic analyses. KH and JG designed the experiment. KVH, BH, KH & PDV proofread the manuscript.

CHAPTER 3

EVALUATION OF MALDI-TOF MS AS A TOOL FOR HIGH-THROUGHPUT DEREPLICATION

*Redrafted from: Ghyselinck, J., Van Hoorde, K., Hoste, B., Heylen, K. & De Vos, P. (2011). Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. J Microbiol Meth **86**(3), 327-336*

SUMMARY

The present study examined the suitability of matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) for the rapid grouping of bacterial isolates, i.e. dereplication. Dereplication is important in large-scale isolation campaigns and screening programs since it can significantly reduce labor intensity, time and costs in further downstream analyses. Still, current dereplication techniques are time consuming and costly. MALDI-TOF MS is an attractive tool since it performs fast and cheap analyses with the potential of automation. However, its taxonomic resolution for a broad diversity of bacteria remains largely unknown. To verify the suitability of MALDI-TOF MS for dereplication, a total of 249 unidentified bacterial isolates retrieved from the rhizosphere of potato plants, were analyzed with both MALDI-TOF MS and repetitive element sequence based polymerase chain reaction (rep-PCR). The latter technique was used as a benchmark. Cluster analysis and inspection of the profiles showed that for 204 isolates (82%) the taxonomic resolution of both techniques was comparable, while for 45 isolates (18%) one of both techniques had a higher taxonomic resolution. Additionally, 16S rRNA gene sequence analysis was performed on all members of each delineated cluster to gain insight in the identity and sequence similarity between members in each cluster. MALDI-TOF MS had higher reproducibility than rep-PCR, was found to be suited for high-throughput analyses, offered possibilities for automation, and was more time and cost efficient than rep-PCR. Its taxonomic resolution was situated at the species to strain level. The present study demonstrated that MALDI-TOF MS is a powerful tool for dereplication.

3.1 INTRODUCTION

The last decade has seen a renewed interest in bacterial cultivation. New approaches for isolating bacteria have been developed, either through adjusting growth and incubation conditions using alternative gelling agents such as gellan gum [1] and prolonging incubation times [2], or through elaborating new technologies such as high-throughput culturing methods [3] and diffusion growth chambers [4]. To significantly reduce the work load once bacteria are obtained in culture, most studies perform a dereplication step. Conventionally, dereplication refers to the process of eliminating knowns from unknowns. However, the term is often used in an unconventional way, referring to the process of recognizing identical isolates at a specific taxonomic level and grouping them accordingly. Subsequent selection of representatives of each group reduces the number of isolates to be analyzed in further downstream analyses, and thus prevents unnecessary screening efforts. Dereplication in this meaning originally referred to the grouping of bacterial isolates at the lowest taxonomic level, the strain level [5]. However, nowadays the term is somewhat ambiguous and often used in a broader sense, also indicating grouping at subspecies [6], species or any higher taxonomic level [7,8]. A broad range of techniques has been used in the context of dereplication such as repetitive element sequence based Polymerase Chain Reaction (rep-PCR) [6], randomly amplified polymorphic DNA (RAPD) [9] and fatty acid methyl ester (FAME) analyses [8]. In some studies, techniques have been used of which the taxonomic resolution was not validated, e.g. Boroczky and coworkers (2006) used a specific type of GC-analysis for dereplication of a set of bacteria isolated from marine environments. A suitable dereplication technique should comply with the following criteria: (i) hold a universal character, i.e. applicability to all bacterial strains; (ii) robustness; (iii) produce easy to interpret data; (iv) have a high taxonomic resolution and (v) provide the possibility of high-throughput application/automation with low operational costs and labor intensity.

Rep-PCR has proven to be a powerful tool in microbial ecology and environmental microbiology [10]. It is a widely applied DNA fingerprinting technique targeting repetitive sequences interspersed throughout the bacterial genome [11], and largely fulfills above-mentioned criteria that make a technique suitable for dereplication. PCR amplification of the DNA between these repetitive elements and subsequent electrophoresis results in easy to interpret bacterial fingerprints that allow differentiation at the subspecies to strain level for a wide range of bacterial species [12,13,14,15]. Furthermore, rep-PCR is a robust technique since factors like e.g. culture age and the number of subcultures prior to DNA extraction have shown not to influence the genomic fingerprint [11]. Nevertheless, this technique also has drawbacks. Trials with different primer sets may be required to produce good quality fingerprints

[16] and both intra- [17] and interlaboratory reproducibility [18] can sometimes be lacking. These shortcomings hamper (semi-) automation of the technique, making rep-PCR quite laborious to be used as dereplication tool in extensive isolation campaigns.

Matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) is a more recent technique in microbiology and has become an important tool with promising applications, especially in diagnostics [19,20,21]. MALDI-TOF MS generates protein mass spectra which can be used to group and identify bacteria. These mass spectra contain mainly peaks corresponding to ribosomal proteins because of their very high abundance in the bacterial cell [22]. It could be an interesting tool for dereplication since it has the same advantages as rep-PCR (i.e. applicability for a wide range of bacterial species, generation of easy interpretable data and robustness) with the additional plus-point that it could be automated, resulting in time and cost reduction. Furthermore, the reagents required to prepare bacterial cell extracts and to do the analysis are cost effective. Numerous studies have already explored whether MALDI-TOF MS has the ability to discriminate at the strain level [23,24,25,26]. However, these studies were systematically limited to a specific taxon (particular genus or species), and therefore insufficient to evaluate the taxonomic resolution of MALDI-TOF MS as a broad range dereplication tool.

In this study, the applicability of MALDI-TOF MS for high-throughput dereplication of a large and unidentified variety of bacterial isolates that were isolated from the potato rhizosphere in Peru and Bolivia was evaluated. Rep-PCR was performed in parallel as a benchmark and both techniques were compared based on grouping of isolates, taxonomic resolution, reproducibility, suitability for high-throughput automation and time and cost effectiveness.

3.2 MATERIALS AND METHODS

Bacterial isolates

Bacteria used in this study were isolated from the rhizosphere of potato plants from the Central Andean Highlands. In short, 5 ml phosphate-buffered saline and 10 sterile glass beads (6 mm) were added to 1 g rhizosphere soil, and vortexed for 2 min. Serial dilutions (10^0-10^{-6}) were plated (100 µl) on ten-fold diluted Trypticase Soy Agar (TSA), supplemented with 0.005% (w/v) cycloheximide to inhibit fungal growth. After incubation for 48 hours at 28°C, isolates were picked and subcultured to purity.

Rep-fingerprinting

Genomic DNA was released from the bacterial cells through alkaline lysis. Therefore, a small amount of cells was lysed in 20 µl alkaline lysis buffer (0.25% (w/v) SDS and 0.05 M NaOH) for 15 min at 95°C. Subsequently, 180 µl sterile milliQ-water was added and lysates were immediately used for PCR. Rep-PCR was performed with the $(GTG)_5$-primer because in-house experience showed this primer targeted the largest bacterial diversity (unpublished data). The PCR-mixture was prepared as described previously [27]. Amplification was performed in a GeneAmp PCR System 9600 (Applied Biosystems) with the following temperature-time profile: 7 min 95°C, 30 cycles of 94°C for 1 min, 40°C for 1 min and 65°C for 8 min, and a final step of 16 min at 65°C. Electrophoresis was performed in 1 x TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA) on a 1.5% agarose gel (w/v) under highly standardized conditions (55 V, 400 mA, 960 min, 4°C). Fourteen samples were loaded per gel. Four reference markers, 6 µl each composed of 1.10 µl Molecular Ruler 500 bp (Bio-Rad), 1.40 µl Molecular Ruler 100 bp (Bio-Rad), 2 µl TE buffer (1 mM EDTA, 10 mM Tris-HCl (pH 8.0)) and 1.50 µl loading dye, were included on every gel. Profiles were visualized under ultraviolet light after staining with ethidium bromide. Digitized images of gels were normalized and analyzed with the BioNumerics 5.1 software (Applied Maths, Belgium). Similarity matrices of densitometric curves of the gel tracks were calculated with Pearson's product-moment correlation coefficient. Cluster analyses of similarity matrices were performed by unweighted pair group method with arithmetic averages (UPGMA). Reproducibility was assessed by analyzing a random subset comprising 10% of all isolates (24 out of 249) in triplicate (starting from growth and DNA extraction to analysis of fingerprint).

MALDI-TOF MS

Preparation of cell extracts

Isolates were grown from stock on tenfold diluted TSA for 48h at 28°C and subcultured twice prior to analysis to ensure all isolates were in the same physiological state. For preparation of the extracts, a small amount of bacterial cells was picked up and suspended in 300 µl milliQ water. Next, 900 µl of absolute ethanol was added and the bacterial suspension was centrifuged for 3 min at 18,000 x *g*. After removing the supernatant, the bacterial pellet was resuspended in 50 µl formic acid (70%). Finally, 50 µl of acetonitrile was added and mixed until complete suspension. The extract was centrifuged for 3 min at 18,000 x *g* and the supernatant was used for MALDI-TOF MS analysis or was preserved at -20°C for later use.

MALDI-TOF MS analysis

Bacterial cell extracts (1 µl) were spotted on a 384 Opti-TOF 123mm x 81mm stainless steel MALDI-TOF MS target plate (AB Sciex) and dried at room temperature. Subsequently, the sample spot was overlaid with 1 µl of a 0.5% (w/v) α-cyano-4-hydroxycinnamic acid (α-CHCA) in 50:48:2 acetonitrile:water:trifluoroacetic acid solution. The spotted plate was analyzed with the 4800 *Plus* MALDI TOF/TOF Analyzer (AB Sciex) which was used in linear, positive-ion mode. Ions were generated by a 200 Hz tripled UV Nd:YAG laser, accelerated at 20 kV through a grid at 19.2 kV and separated according to their m/z ratio in a 1.5 m long linear, field-free drift region. Each generated spectrum resulted from 40 laser shots at 50 random positions within the measuring spot. MALDI-TOF mass spectra were generated in the mass range 2-20 kDa. Calibration was performed with the Protein Calibration Standard I (Bruker) (composition: insulin ($[M+H]^+$, *m/z* 5734.6), ubiquitin I ($[M+H]^+$, *m/z* 8565.9), cytochrome C ($[M+H]^+$, *m/z* 12361.5), myoglobin ($[M+H]^+$, *m/z* 16952.3)) to which ACTH Fragment 18-39 MALDI-MS Standard ($[M+H]^+$, m/z 2465.7) (Sigma-Aldrich) was added. With every set of measurements, the Bruker Bacterial Test Standard (Bruker) was included as a positive control.

Analysis of spectral data

Mass spectra were obtained in t2d format and were converted to txt files using the Data Explorer 4.9 software (AB Sciex). The txt files were imported in BioNumerics 5.1 software (Applied Maths, Belgium) and converted to fingerprints for further analyses. To obtain reliable data analysis, the spectra with extensive noise and/or insufficient signal intensities were excluded. The similarity between the spectra was expressed using Pearson's product moment correlation coefficient and the spectra were clustered using the UPGMA clustering algorithm. Reproducibility was assessed as for rep-PCR, by analyzing the

same 10% of all isolates (24 out of 249) in triplicate (starting from growth and cell extract preparation to analysis of fingerprint).

16S rRNA gene sequence analysis

16S rRNA gene amplification and sequencing was performed as described by Heyrman and Swings (2001) and Heylen et al. (2006) respectively. Partial 16S rRNA gene sequences (first 300-500 bp) were assembled using the BioNumerics 5.1 software. Identification was obtained in two steps: (i) query in the "Classifier" program of Ribosomal Database Project II [28] of the partial 16S rRNA gene sequence of an isolate, (ii) the type strains of all species of all genera mentioned in the Classifier report were compared in an exhaustive pair wise manner with the query sequence of each isolate in BioNumerics 5.1. The isolates were assigned to a genus based on the obtained pairwise 16S rRNA gene sequence similarities.

Nucleotide sequence accession numbers

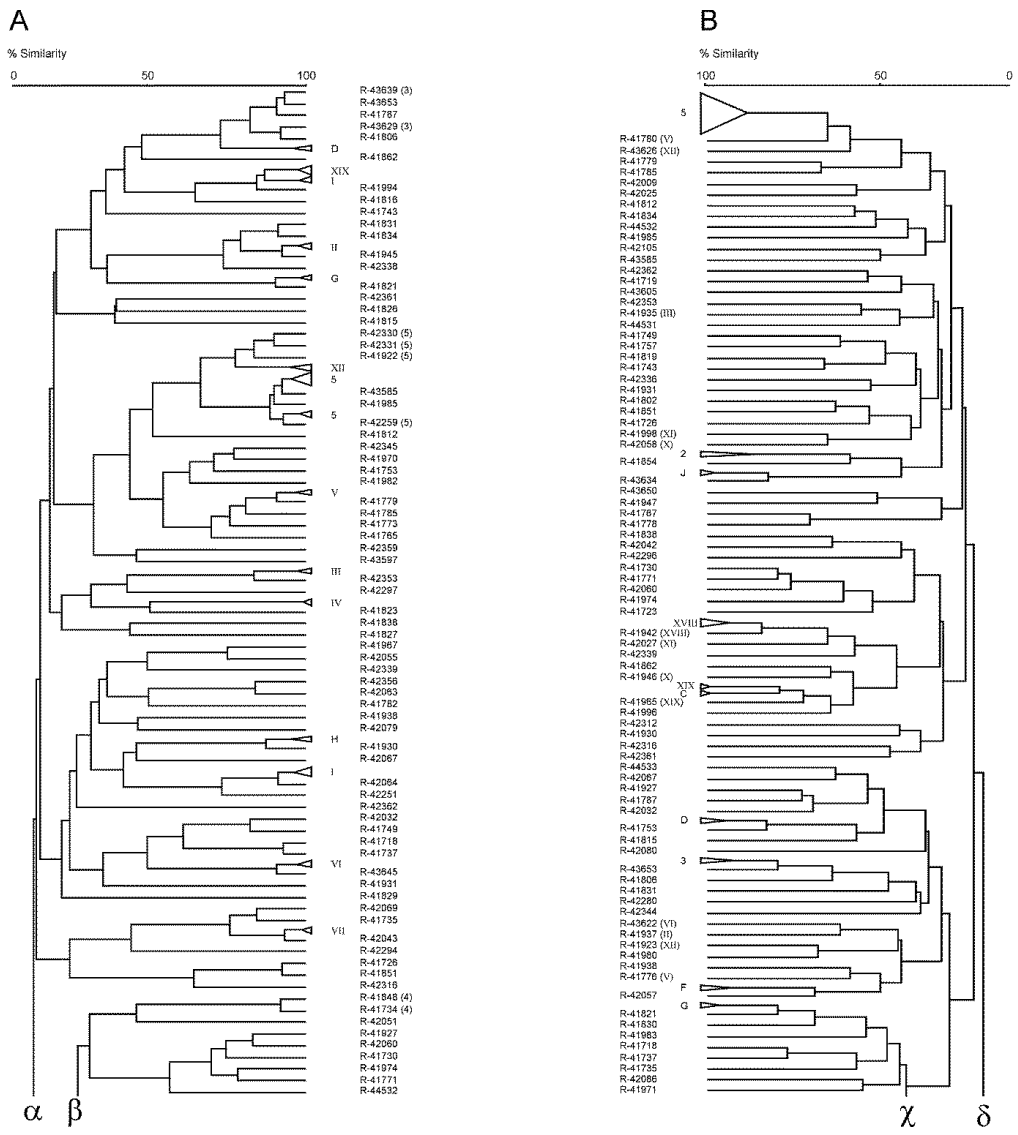
The 16S rRNA gene sequence data generated in this study has been deposited in GenBank/EMBL/DDBJ with accession numbers FR727740 to FR727837 and FR773155.

3.3 RESULTS

Rep-fingerprinting

$(GTG)_5$-PCR fingerprints were generated from 249 unidentified bacterial isolates and cluster analysis was performed. To objectively delineate clusters, a cut-off value was calculated based on triplicate analysis of 24 arbitrarily chosen isolates. For this purpose, the mean similarity and standard deviation were calculated for each of the 24 sets. Mean similarities ranged from 71.20% – 99.01%. From these 24 mean similarities, the overall mean similarity and its standard deviation were determined (93.82 $\pm$ 7.43%). Through subtracting the standard deviation from the mean, the cut-off was obtained (86.39%). A lower similarity value between two fingerprints than this cut-off value was assumed to be caused by genetic variation among the isolates and not by methodological variations. Applying this cut-off level on the dendrogram, 22 $(GTG)_5$ clusters, composed of two to twelve isolates (Fig. 3.1) were delineated. In addition, 210 unique $(GTG)_5$ -patterns were observed (Fig. 3.1), suggesting a high genetic diversity among the isolates.

The reproducibility of $(GTG)_5$-PCR was calculated from two parameters. The first parameter was the overall mean similarity (93.82%, see above). The second parameter, 3.55%, was calculated by dividing the sum of the standard deviations of the mean similarities of each replicate set by the number of triplicate sets. Good reproducibility is reflected by a high value of the former parameter, while the latter, being a measure for the general experimental variation of the technique, should be low.



Continued on the next page.

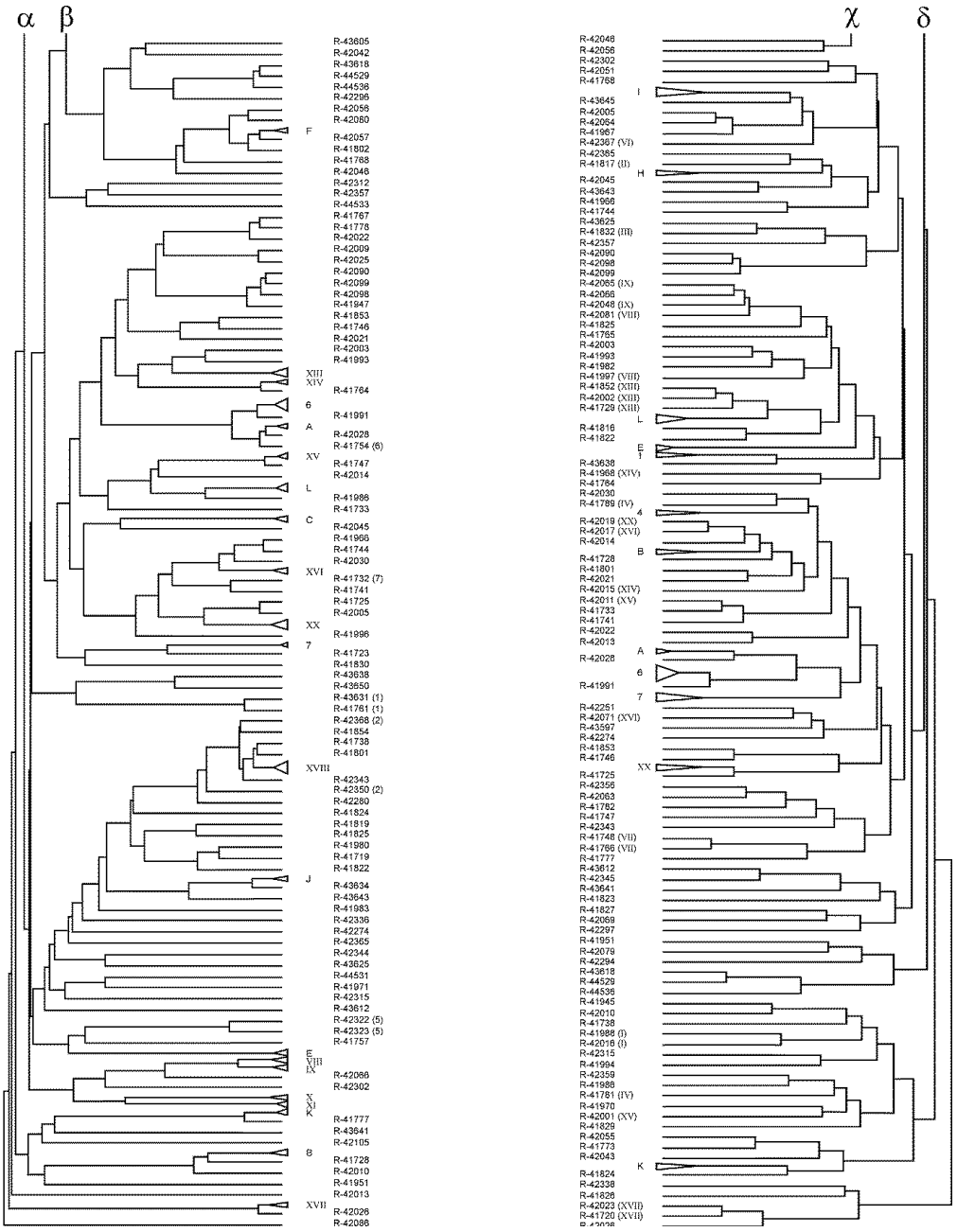


Figure 3.1 Comparison between dendrograms derived from MALDI-TOF MS (A) and rep-PCR (B) fingerprinting. Capital letters (A-L) were used to indicate identical grouping of isolates by both techniques, thus suggesting both techniques had a similar taxonomic resolution. Roman numbers (I-XX) were used to indicate clusters which suggested a higher taxonomic resolution for rep-PCR. Clusters which suggested a higher taxonomic resolution for MALDI-TOF MS were assigned Arabic numbers (1-7). Symbols between brackets indicate which cluster the isolate belongs to in the dendrogram generated by the other technique. Greek symbols mark the orientation of the figure.

MALDI-TOF MS fingerprinting

MALDI-TOF MS spectra were generated from the same 249 isolates and cluster analysis was performed. Spectra were checked for high background in the lower m/z ranges (slopes) and limited number or low intensity peaks. Good quality spectra were obtained for all isolates. As for $(GTG)_5$-PCR, a cut-off was calculated to delineate clusters based on triplicate analysis of the same 24 isolates. Mean similarities ranged from 90.63% – 99.80%. From these 24 values, the overall mean similarity and its standard deviation were determined (96.85 $\pm$ 2.10%). The cut-off was thus set at 94.75%. In total, 36 MALDI-TOF MS clusters, composed of two to four isolates, were delineated (Fig. 3.1). In addition, cluster analysis showed the presence of 202 unique MALDI-TOF MS spectra, confirming the large diversity that was observed with $(GTG)_5$-PCR.

The reproducibility of MALDI-TOF MS was again deduced from two parameters. For the first parameter (the overall mean similarity), a value of 96.85% was calculated (see above), being higher than for $(GTG)_5$-PCR. The second parameter was lower with a value of 1.83%, again demonstrating the higher reproducibility of MALDI-TOF MS.

Dereplication and taxonomic resolution

Suitability of MALDI-TOF MS for dereplication was evaluated using $(GTG)_5$-PCR, which has discriminatory power at subspecies to strain level, as a benchmark. Although the nature of the obtained data differs, i.e. genetic data versus protein profiles, cluster analyses of both data sets were performed identically (UPGMA and Pearson product-moment correlation coefficient). Isolates from all clusters, were also analyzed with 16S rRNA gene sequence analysis (Table 3.1).

In theory, comparison of two dereplication techniques can result in the following: (i) delineated groups are identical for both techniques, suggesting a similar taxonomic resolution of both techniques, (ii) groups delineated by $(GTG)_5$-PCR are subdivided by MALDI-TOF MS, suggesting higher resolution of MALDI-TOF MS, (iii) groups delineated by MALDI-TOF MS are subdivided by $(GTG)_5$-PCR, suggesting higher resolution of $(GTG)_5$-PCR, and (iv) isolates are grouped differently by both techniques (the latter possibility was not observed in this study).

Cluster[a]	# Isolates in cluster[c]	16S rRNA gene sequence similarity within cluster (%)	Type strain with highest 16S rRNA gene sequence similarity to query sequence			
			Species name	Sequence similarity (%) with the query sequences	Strain number	Accession number
A	2	99.6	*Pseudomonas azotoformans*	98.3-99.4	IAM 1603^T	D84009
B	2	99.5	*Pseudomonas taiwanensis*	96.5-97.9	BCRC 17751	EU103629
C	2	100	*Microbacterium foliorum*	99.5	DSM 12966^T	AJ249780
D	2	100	*Bacillus weihenstephanensis*	99.4-100	DSM 11821^T	AB021199
E	2	100	*Pseudomonas nitroreducens*	99.7	IAM 1439^T	AM088473
F	2	100	*Arthrobacter psychrolactophilus*	99.5	B7^T	AF134179
G	2	100	*Pseudosphingobacterium domesticum*	97.9	DC-186^T	AM407725
H	2	97.1	*Paenibacillus odorifer*	95.5	LMG 19079^T	AJ223990
I	3	99.8 - 100	*Paenibacillus xylanexedens*	99.6-99.8	B22a^T	EU558281
J	2	100	*Stenotrophomonas chelatiphaga*	98.2	LPM-5^T	EU573216
K	2	100	*Pseudomonas brassicacearum* subsp. *brassicacearum*	97.4	CFBP 11706^T	AF100321
L	3	98.7 - 100	*Pseudomonas jessenii*	96.1-98.7	CIP 105274^T	AF068259
1	2	99.9	*Pseudomonas veronii*	97.1-100	CIP 104663^T	AF064460
2[b]	2	/	*Stenotrophomonas humi*	98.9	R-32729^T	AM403587
3	2	99.9	*Bacillus weihenstephanensis*	99.9	DSM 11821^T	AB021199
4	2	100	*Pseudomonas azotoformans*	98.3	IAM 1603^T	D84009
5	12	99.9 - 100	*Rhodococcus erythropolis*	99.7	DSM 43066^T	X79289
6	5	99.8 - 99.9	*Pseudomonas azotoformans*	98.1-99.6	IAM 1603^T	D84009
7	3	100	*Pseudomonas azotoformans*	98.3	IAM 1603^T	D84009
I	2	99.9	*Flavobacterium resistens*	95.1-96.7	BD-b365^T	EF575563
II	2	99.2	*Pedobacter panaciterrae*	99.4-99.8	Gsoil 042^T	AB245368
III	2	100	*Rhizobium radiobacter*	96.9	IAM 12048^T	AB247615

Continued on the next page.

Cluster[a]	# Isolates in cluster[c]	16S rRNA gene sequence similarity within cluster (%)	Type strain with highest 16S rRNA gene sequence similarity to query sequence			
			Species name	Sequence similarity (%) with the query sequences	Strain number	Accession number
IV	2	99.3	*Streptomyces anulatus*	99.8-100	NRRL B-2000^T	DQ026637
V	2	94.3	R-41776 *Rhodococcus koreensis*; R-41780 *Rhodococcus erythropolis*	96.9; 99.8	DNP 505^T; DSM 43066^T	AF124342; X79289
VI	2	99.7	*Paenibacillus xylanexedens*	99.5-99.6	B22a^T	EU558281
VII	2	99.5	*Ochrobactrum pseudogrignonense*	97.0-98.6	CCUG 30717^T	AM422371
VIII[b]	2	/	*Arthrobacter psychrolactophilus*	99.5	B7^T	AF134179
IX	2	100	*Arthrobacter psychrolactophilus*	99.5	B7^T	AF134179
X	2	99.5	*Pseudomonas marginalis*	97.0-97.3	LMG 2210^T	Z76663
XI	2	100	*Pseudomonas marginalis*	98.5	LMG 2210^T	Z76663
XII[b]	2	/	*Rhodococcus qingshengii/ R. jialingiae/R. baikonurensis*	98.9	djl-6^T/djl-6-2^T/GTC 1041^T	DQ090961/ DQ185597/ AB071951
XIII	3	97.5 - 100	*Pseudomonas agarici*	96.2-98.9	LMG 2112^T	Z76652
XIV	2	99.9	*Pseudomonas cedrina* subsp. *fulgida*	99.9-100	LMG 21467^T	AJ492830
XV	2	97.6	*Pseudomonas alcaligenes*	97.4-98.6	IAM 12411^T	D84006
XVI	2	96.1	R-42071 *Pseudomonas taiwanensis*; R-42017 *Pseudomonas extremaustralis*	97.6; 99.5	BCRC 17751; CT14-3T	EU103629; AJ583501
XVII	2	99.5	*Chryseobacterium soli*	97.9-98.5	JS6-6^T	EF591302
XVIII	4	99.9 - 100	*Stenotrophomonas humi*	99.1-99.4	R-32729^T	AM403587
XIX	3	100	*Flavobacterium psychrolimnae*	95.6	LMG 22018^T	AJ585428
XX	3	99 - 100	*Pseudomonas migulae*	96.5-99.3	CIP 105470^T	AF074383

Table 3.1 Overview of the identity of isolates from each cluster and their mutual 16S rRNA gene sequence similarities. For species identification, all type strains of all species of all genera mentioned in the RDP Classifier report [28] were compared in an exhaustive pair wise manner with the query sequence of each isolate in BioNumerics 5.1 (Applied Maths, Belgium).

[a], cluster names were taken from Fig. 3.1.

[b], no sequence could be obtained for one isolate of this cluster.

[c], the number given in the table refers to the maximum number of isolates in each cluster, i.e. refers to the cluster in the dendrogram of the technique with the lowest discriminatory power for this cluster.

The first possibility, i.e. both techniques having a similar taxonomic resolution, was reflected by a total of 150 isolates which were found to occupy unique positions in both the MALDI-TOF MS and rep-PCR dendrogram. In addition, 26 isolates grouping into twelve $(GTG)_5$ clusters (clusters A-L) (Fig. 3.1; Table 3.1), each representing the same subspecies or strain, also formed identical groups based on their respective mass spectra. Visual inspection of the $(GTG)_5$-fingerprints indicated that the isolates from MALDI-TOF MS clusters I, VII, XIII and XX did not cluster in the $(GTG)_5$-dendrogram although they had identical rep-profiles. As such, rep-PCR for these isolates in fact agreed with the respective clusters formed by MALDI-TOF MS. The reason for the separation was that these rep-fingerprints in fact had the exact same patterns but the bands had slightly shifted positions in the same direction. As a result, Pearson product-moment correlation coefficient separated these patterns in the dendrogram.

Differences in clustering suggesting a higher resolution with MALDI-TOF MS (the second possibility) also occurred (clusters 1-7) (Fig. 3.1; Table 3.1). Isolates from $(GTG)_5$ clusters 1 (*Pseudomonas*), 2 (*Stenotrophomonas*), 3 (*Bacillus*), 4 (*Pseudomonas*), 5 (*Rhodococcus*), 6 (*Pseudomonas*) and 7 (*Pseudomonas*), were subdivided by MALDI-TOF MS analysis (Fig. 3.1). Mass spectra generated from these isolates were analyzed more into detail in order to find peaks justifying their separation in the MALDI-TOF MS dendrogram (Fig. 3.2). Detailed visual inspection of the spectral data confirmed the observed. For all isolates, except isolates from cluster 1, presence or absence of specific peaks explaining their separate grouping could be demonstrated (Fig. 3.2). These observations suggest that, within the genera *Pseudomonas*, *Stenotrophomonas*, *Rhodococcus* and *Bacillus,* MALDI-TOF MS might be able to differentiate strains where $(GTG)_5$-PCR fails to do so. In contrast, isolates from cluster 1 (Fig. 3.1) were separated in the MALDI-TOF MS dendrogram while no discriminatory peaks were found responsible for their separation. This, however, could be explained by looking at the peak intensities of the respective profiles which are expressed relative to the most intense peak. One of the spectra contained a very intense peak with the outcome that other peaks in the profile had very low relative intensities and could hardly be distinguished from the background. As a consequence, similarity calculations using Pearson's product moment correlation coefficient separated both profiles in the MALDI-TOF MS dendrogram. Thus, based on the raw data and the use of adjusted software, isolates from cluster 1 would group together.

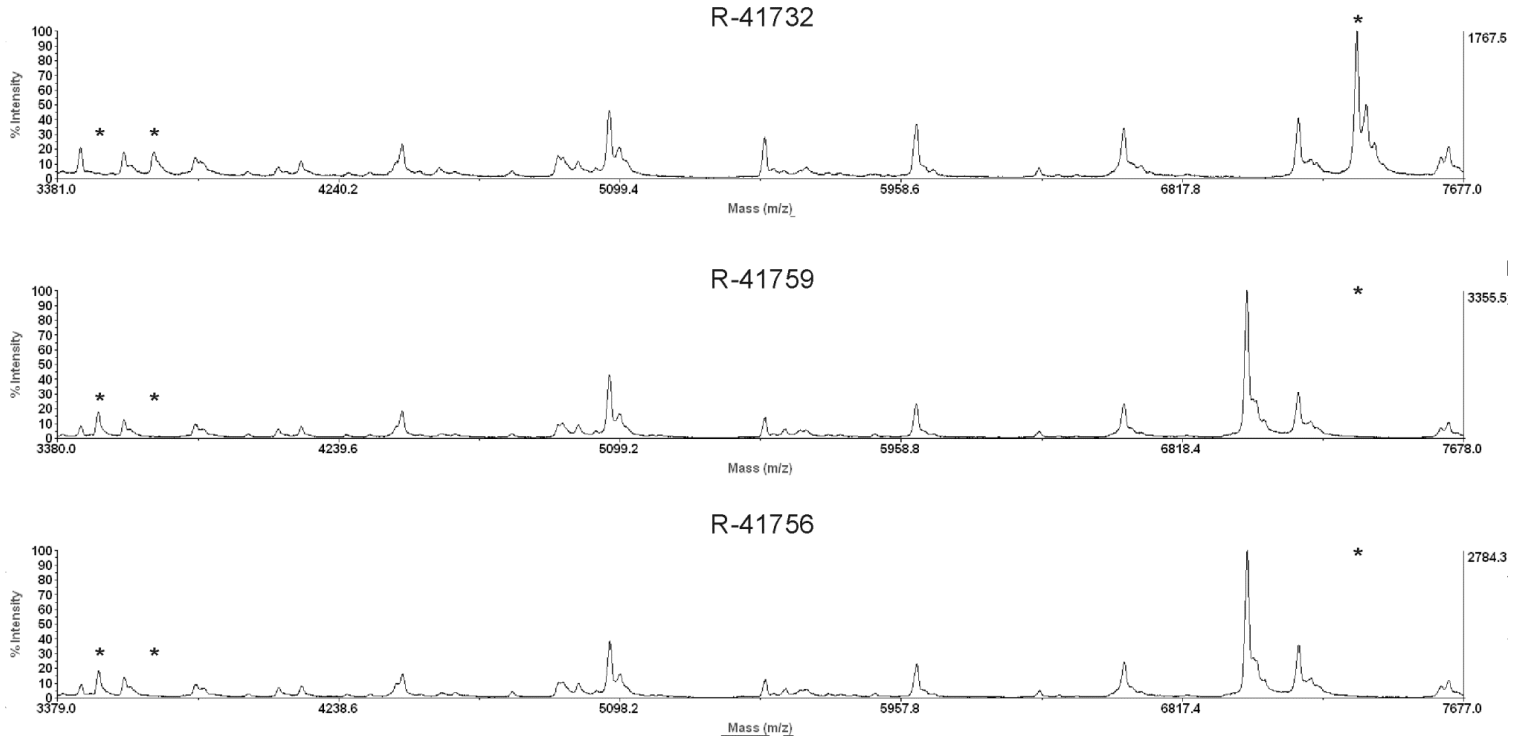


Figure 3.2 A detailed view of the 3380-7670 Da region of the mass spectral profiles of the isolates from cluster 7. The mass spectrum of isolate R-41732 clearly differs from the mass spectra of isolates R-41759 and R-41756 at positions 3505 Da, 3677 Da, 7351 Da and 7378 Da (indicated by the stars). It is clear that these differences support the separation of isolate R-41732 in the MALDI-TOF MS dendrogram.

Clusters suggesting a lower discriminatory power of MALDI-TOF MS (third possibility) were encountered as well (clusters I-XX) (Fig. 3.1; Table 3.1). Clusters I-XVII, with the exception of I, VII, XIII (see above), contained isolates with unique $(GTG)_5$ fingerprints that did cluster together based on their respective mass spectra. Furthermore, isolates from 3 $(GTG)_5$ clusters (XVIII (*Stenotrophomonas*), XIX (*Flavobacterium*), and XX (*Pseudomonas*)), formed similar clusters based on their mass spectra, but extra isolates were included in these clusters as well (Fig. 3.1). Upon detailed visual inspection of the spectral profiles, it was clear that in clusters III, IV, VI, VIII, IX, XI, XVIII and XIX mass spectra were indeed identical. These clusters contained representatives of *Rhizobium*, *Streptomyces*, *Paenibacillus*, *Arthrobacter*, *Pseudomonas*, *Stenotrophomonas* and *Flavobacterium* respectively. Thus, it was concluded that for these clusters the resolution of MALDI-TOF MS was lower than that of $(GTG)_5$-PCR. However, clusters II, V, X, XII, XIV, XV, XVI and XVII including representatives of genera *Pedobacter*, *Rhodococcus*, *Pseudomonas* and *Chryseobacterium* respectively, contained non-identical mass spectra per cluster, and thus should have grouped separately, as with $(GTG)_5$-PCR. These results demonstrate that MALDI-TOF MS analysis itself was able to differentiate the isolates within these clusters, but subsequent data analysis was not. However, further software developments will allow differentiating these isolates.

In this study, cluster analysis and inspection of the profiles showed that for 204 isolates (82%) grouping was similar with both techniques. For 26 isolates (10.4%), however, MALDI-TOF MS had higher discriminatory power than rep-PCR, while 19 isolates (7.6%) were better differentiated by rep-PCR. Pairwise comparison of 16S rRNA gene sequences of members of each unique MALDI-TOF MS cluster, i.e. clusters of which the high degree of similarity between the spectra was confirmed by visual inspection (clusters III, IV, VI, VIII, IX, XI), resulted in similarities ranging from 99 to 100%, (Table 3.1) suggesting resolution at least at the species level (mostly set at 97-98% 16S rRNA gene sequence similarity [29]).

Cost-benefit analysis

The suitability of a dereplication technique is not only dependent on the taxonomic resolution of the technique, but also on technological and other aspects, i.e. cost and time efficiency. Therefore, a cost-benefit analysis was performed for both rep-PCR and MALDI-TOF MS (Table 3.2).

Our data demonstrated a higher reproducibility for MALDI-TOF MS than for $(GTG)_5$-PCR. The large number of samples that can be analyzed per batch with MALDI-TOF MS further benefits the analysis. Rep-profiles, however, are usually generated in high numbers of batches and, as a consequence,

differences in experimental conditions are more likely to occur than for MALDI-TOF MS. Unlike MALDI-TOF MS analysis, which allowed 384 samples to be analyzed per batch, rep-PCR analysis, as performed in this study, allowed only 14 samples per rep-gel, taking into account the necessary molecular ladders for normalization, positive controls and blanks. Notwithstanding $(GTG)_5$-PCR was performed under standardized conditions, experimental variations during electrophoresis, staining and digitalizing will have occurred.

Rep-PCR is a DNA-based technique. High-quality DNA, which improves reproducibility (unpublished data), can be extracted with time-consuming protocols and costly commercial kits. For this study, we opted for a cheap and quick alternative DNA extraction method which, however, resulted in a crude extract of lower quality and which of course may have affected reproducibility. MALDI-TOF MS is a chemotaxonomic technique and can be performed on whole bacterial cells, reducing sample preparation to almost zero but also negatively effecting reproducibility. To minimize the latter, our protocol included preparation of cell extracts, taking approximately a day's work for 100 samples. This time cost could be significantly decreased through process automation, e.g. using a colony picker and liquid handling robot. Further automation of MALDI-TOF MS would also be possible by automated spotting of the target plate.

	Rep-PCR	**MALDI-TOF MS**
Sample material	DNA	Cells or cell extracts
Cost	Low capital cost High consumable cost	High capital cost Low consumable cost
Interlaboratory comparison	Limited	High if a minimum standard is followed[a]
Time cost	Possible short sample prep (alkaline lysis, lower quality) (Very) long sample prep (Pitcher, high quality) Time-consuming PCR & gel electrophoresis	Possible short sample prep (intact cell spotting, lower quality) Long sample prep (cell extracts, high quality) Quick spectral profile generation
Reproducibility	High Overall Mean Similarity 93.82% Mean Standard Deviation 3.55%	Higher Overall Mean Similarity 96.85% Mean Standard Deviation 1.83%
Taxonomic resolution	Subspecies-to-strain	Species-to-strain
Influence of growth conditions	No influence	Limited influence (if following minimum standard, mass range of 2 to 20 kDa) [b]
Possibility for automation	Sample prep PCR	Sample prep Spotting target plate Analysis
High-throughput	Dependent on electrophoresis capacity in lab (14 samples/gel)	384 samples/target plate
Storage sample material without quality loss (-20°C)	Years (if high-quality DNA) Months (alkaline lysates)	Months (cells in ethanol) [c] Weeks (cell extracts) [c]
General applicability	Yes, but trials for suitable primers necessary	Yes

Table 3.2 General overview comparing the advantages, disadvantages and possibilities of rep-PCR and MALDI-TOF MS.
[a], not tested in this study, but taken from [30] and [31]
[b], not tested in this study, but taken from [30] and [32]
[c], not tested in this study but taken from [30]

3.4 DISCUSSION

Since dereplication is a vital step in diversity studies and screening programs, it is of utmost importance to select a technique from which maximum benefit can be gained. Our research demonstrated that time-saving and cost effectiveness make MALDI-TOF MS the preferred tool for dereplication. However, whether it is a worthy alternative for rep-PCR (a technique often used for dereplication) with respect to taxonomic resolution remained unclear until now. Promising results were obtained by Siegrist et al. (2007) who demonstrated that MALDI-TOF MS more effectively grouped environmental isolates of *E. coli* according to their respective sources than rep-PCR and that it was able to differentiate strains. Because MALDI-TOF MS detects a large spectrum of proteins, theoretically the technique should be able to discriminate between closely related species and to classify organisms at the subspecies level [21]. However, since the large majority of detected proteins and fragments have a ribosomal origin, skepticism may arise concerning its claimed taxonomic resolution. Yet, despite the highly conserved nature of ribosomal proteins, slight sequence variations can occur even at the subspecies and strain level [33]. Our work demonstrated that for some isolates belonging to the genera *Stenotrophomonas*, *Bacillus*, *Rhodococcus*, and *Pseudomonas*, MALDI-TOF MS was able to discriminate where rep-PCR failed, while, on the contrary, MALDI-TOF MS was unable to differentiate other isolates belonging to the genera *Rhizobium*, *Streptomyces*, *Paenibacillus*, *Arthrobacter* and *Pseudomonas* which were clearly distinguished by rep-PCR. Still, isolates within these latter clusters showed 99 to 100% 16S rRNA gene sequence similarities (Table 3.1), suggesting MALDI-TOF MS can differentiate at least onto species level, or possibly lower. MALDI-TOF MS analysis could correctly group serotypes of *Listeria monocytogenes* in three lineages, corresponding to results obtained with pulsed-field gel electrophoresis [30], discriminated subspecies of *Francisella tularensis* [32] and successfully differentiated epidemiologically related *Legionella* strains [34]. However, MALDI-TOF MS was unable to differentiate between the subspecies *Lactococcus lactis* subsp. *cremoris* and *L. lactis* subsp. *lactis* based on their mass spectra [35]. These reports confirm our observations that the taxonomic resolution of MALDI-TOF MS is taxon dependent.

Data were analyzed with the BioNumerics 5.1 software (Applied Maths, Belgium). This software allowed to objectively compare results generated from different experiment types. Pearson product moment correlation coefficient was used to calculate similarities between both MALDI-TOF MS [26,36,37] and rep fingerprints. This resulted in a higher reproducibility than with the binary Dice coefficient (data not shown), confirming results from another study [38]. It was shown previously that pearson product

moment correlation coefficient was better suited for identification of DNA fingerprinting profiles than band matching algorithms [39]. Moreover, pearson product moment correlation coefficient takes into account the overall pattern of the fingerprint and is therefore less sensitive to small variations arising from faint bands or small shifts in the overall pattern [40]. Finally it can be stated that similarity coefficients based on binary data require intense visual inspection of the fingerprint to verify band allocation, as such not only reducing the reliability and reproducibility of the band-calling process but also interfering with the process of automation [40].

An overall mean similarity between replicate MALDI-TOF MS profiles of 96.85 $\pm$ 2.10% was observed. These results imply good reproducibility and are in agreement with results from a previous study [41]. Since MALDI-TOF MS is a chemotaxonomic technique, several parameters, e.g. medium [42,43] and cell age [43], can affect its reproducibility. Although these experimental differences do not significantly hamper identification at the species level, distinction at the subspecies level might become difficult [33]. However, other studies showed that medium, time period of cultivation and preparation protocols did not interfere in species and subspecies differentiation [30,32]. If standardized, and measured in a mass range of 2 to 20 kDa, the technique has proven to have high interlaboratory reproducibility [30,31] and robustness under different culture conditions [41]. However, apart from these considerations, one could question the use of MALDI-TOF-MS for dereplication. Since organisms are best analyzed from the media on which they grow best, cultivation on alternative media might initiate the production of stress-induced compounds and generate extra peaks in the MALDI-TOF MS profile. This is an important aspect in dereplication since in many diversity studies the identity of the isolated population is unknown. However, it is known that stress response systems show a high degree of similarity in prokaryotes [44]. It is more a matter of which conditions provoke the reaction. This means that at the lower taxonomic levels (species to strain level), we expect bacteria to react in a similar way and thus generate spectral profiles that will not be differentiated based on stress response. For rep-PCR, an overall mean similarity of 93.82 $\pm$ 7.43% was obtained, confirming results from a previous study [40] and demonstrating that MALDI-TOF MS had better reproducibility than rep-PCR. However, the opposite was found in a study by Siegrist et al. (2007). Reasons for this could be the authors' choice to directly deposit bacterial cells from agar plates on the MALDI-TOF MS target plate instead of cell extracts, the different mass ranges used and the different matrix.

MALDI-TOF MS is a fast, accurate and inexpensive tool for identification of bacteria [33,45] and does not require a high level of staff training [32]. A general overview comparing the advantages, disadvantages and possibilities of rep-PCR and MALDI-TOF MS is given in Table 3.2. It is possible to prepare 100

bacterial cell extracts in an 8h working day, a number which could significantly increase with process automation. Inactivated cells can be stored in ethanol for months (in advance of cell extraction) and bacterial cell extracts for weeks without significant loss in spectrum quality [30]. This allows for the analysis of large numbers of samples per batch, limiting the occurrence of experimental variation. Generated mass spectra can efficiently be evaluated and analyzed in high-throughput [33]. Nevertheless, if alkaline lysis is used for DNA extraction, rep-PCR sample preparation is less time consuming. Still, the rep-PCR experiment itself is significantly longer and the number of samples to be processed per batch is limited by the size and number of gel electrophoresis tanks available. For some isolates (other than the 249 analyzed in this study (unpublished data)), no rep-fingerprints could be obtained whereas all isolates generated mass spectra. As a consequence, total time between sample preparation and accessibility of the results was substantially longer with rep-PCR. However, recently a high speed semi-automated rep-PCR kit has been developed for bacterial strain typing [46]. This kit, however, is not applicable for dereplication of an unknown diversity since prior knowledge of the identity of the organisms is required. Although current advances for amplicon detection in rep-PCR using microfluidics make it possible to significantly shorten the time between the PCR reaction and accessibility to results, the technique has its limitations which increase both cost and turnaround time, and the number of isolates that can be analyzed per batch is restricted [47].

Based on the results of this study, we propose MALDI-TOF MS as the best dereplication technique currently available. It complies with the criteria stated in the introduction: (i) MALDI-TOF MS spectra were obtained for all isolates, whereas this was not the case for rep-PCR; it has shown to (ii) be robust, (iii) produce easy to interpret mass spectra, (iv) discriminate a broad bacterial diversity at the species-to-strain level, and (v) offer possibilities for use in high-throughput with low operational costs and labor intensity. As such, MALDI-TOF MS is the recommended dereplication tool for next generation cultivation studies.

3.5 ACKNOWLEDGEMENTS

Our special thanks to Prof. Dr. Peter Vandamme for critically reading the manuscript and providing new insights. The Research Fund University College Ghent is acknowledged for the Post Doctoral fellowship of Dr. Koenraad Van Hoorde. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement N° 227522.

3.6 REFLECTING ON THE WORK PERFORMED

Brief summary of work

Motivated by the intensive isolation campaign performed at the start of the project, the suitability of MALDI-TOF MS for dereplication was studied. Although this technique had proven to have high taxonomic resolution for a number of bacterial lineages, its resolution for a wide range of bacterial species remained unknown. Since rhizosphere soils are known to harbor a wide range of taxonomically divergent microorganisms, the isolates obtained from the sampling campaign were considered to be appropriate targets for this study. Different aspects of this technique were studied relative to rep-PCR. Aspects included taxonomic resolution, reproducibility, suitability for high-throughput automation and time and cost effectiveness. The results obtained allowed us to conclude that MALDI-TOF MS was a promising alternative for rep-PCR. However, in hindsight, a number of aspects could have been improved or worked out better. In this chapter I will reflect on the research performed.

In hindsight

The weakness of the clustering method

The research was based on the comparison of clusters obtained from rep-PCR and MALDI-TOF MS data. However, the clustering algorithm used was the weakest point of the analysis. Still, decisions made were the best options at that time, as no valuable alternatives were available. The problem is that only presence or absence of bands is the criterion that should be used for profile clustering, and not band intensity. Therefore, the Dice algorithm was preferred above the Pearson Product Moment Correlation (PPMC) method. However, as the number of bands in MALDI-TOF MS profiles is enormous relative to rep-PCR fingerprints, the automatic band assignment tool in BioNumerics v5.1, which is designed for rep-data, was not adjusted to MALDI-TOF MS fingerprints. Therefore, automatic band assignment would have required manual checking and correction. However, this operation would have been less objective and optimal compared to clustering with PPMC. Although PPMC clustered profiles were checked visually, software specifically implemented for processing MALDI-TOF MS data may have been better for the purpose of this study. It would be interesting to check whether, with the optimal tools available, results would have been improved. Still, no major deviations from the conclusions drawn are expected, considering the thorough (and time consuming) visual inspections of the clustered profiles.

Mean similarities between rep-patterns obtained from replicate analyses indicated a lower limit of 71.20%. However, the lowest mean similarity between replicate MALDI-TOF MS profiles was much higher with a value of 90.63%. This was rather unexpected, considering the fact that rep-PCR is a

genomic technique while MALDI-TOF MS is a chemotaxonomic technique, and thus more prone to variation. Re-analyzing the rep-profiles for which only 71.20% mean similarity was obtained (Fig. 3.3), illustrated that all band patterns were of good quality. Rep-patterns A and B showed 97.1% similarity. However, profiles A and B showed only 50.7 and 63.8% similarity with pattern C. Fig. 3.3 shows that a number of large DNA fragments occurring in profile C are missing in profiles A and B. Although difficult to say, this may have resulted from sheared genomic DNA, which occurred during DNA extraction, or from DNA degradation by DNA degrading enzymes. As DNA extraction was performed by alkaline lysis, unlike Pitcher DNA extraction DNA degrading enzymes were not removed. Still, as PCR was performed shortly after DNA extraction, the latter explanation seems unlikely. Another explanation may be that bands are present, but not visible to the naked eye due to processes occurring during PCR. Low intensities relative to the bands in C may have caused the low similarity percentages observed, as clustering was based on PPMC. My personal opinion favors the last option, since it is unlikely that both DNA extracts in A and B would have experienced the same DNA shearing. If the latter is true, this is another confirmation of the weakness of the PPMC algorithm for clustering rep-fingerprints. However, this observation for replicate analyses was rather an exception than the rule.

Figure 3.3 GTG_5-patterns obtained from replicate analysis on strain R-41784.

Another unexpected observation was that the reproducibility of MALDI-TOF MS was found to be higher compared to rep-PCR. It is clear that the low 71.20% value mentioned above lowered the reproducibility of rep-PCR relative to MALDI-TOF MS. However, as the profiles obtained were of good quality and not due to experimental error, they cannot be excluded for the purpose of assessing reproducibility. Moreover, excluding this replicate similarity value did not result in higher reproducibility of rep-PCR. As MALDI-TOF MS is a chemotaxonomic technique, and thus prone to differences in cultivation conditions, and rep-PCR is a genomic technique, implying its higher robustness, the observation questions the rep-PCR approach. One would expect the reproducibility of the rep-PCR method to be superior instead. Many factors or a combination of factors may have accounted for this, including the use of the PPMC

method for pattern clustering, alkaline lysis instead of Pitcher DNA extraction, PCR artifacts, differences in gel densities for electrophoresis or the process of gel imaging.

Resolution at the species-to-strain level

In the final conclusion, the statement was made that the taxonomic resolution of MALDI-TOF MS is situated at the species-to-strain level. However, in hindsight, this may be a bit too simplistic. Further analyses are required to be accurate. Several isolates that clustered together based on their rep-profiles were differentiated by MALDI-TOF MS. Although this suggests a higher resolution of MALDI-TOF MS, we cannot be sure; no more than it would allow strain differentiation. Additional experiments are necessary. For a start, it would be interesting to analyze the isolates in question with Pulsed Field Gel Electrophoresis or Amplified Fragment Length Polymorphism, which are known to harbor a higher resolution compared to rep-PCR. If the isolates can be differentiated with one of both techniques, our question is answered and we prove that MALDI-TOF MS has a higher resolution than rep-PCR. However, if not the case, the differences observed may have been due to experimental variation (which exceeded the cut-off value postulated from replicate analyses). In this case, the following reasoning is applicable. A technique is only good at distinguishing groups of objects if the variation (i.e. distance between the fingerprints) within the group is less than the variation between the groups. Although positive results were obtained from replicate analyses, it remained unclear whether the organisms involved were closely related or distantly related. The question thus remains to which degree of relatedness between organisms MALDI-TOF MS is able to differentiate them from replicates. As mentioned above we cannot be sure that MALDI-TOF MS differentiates at the strain level. At a certain degree of relatedness, the variation inherent to the experiment may become higher than the difference between the organisms. At that point, the technique reaches its maximum differentiating power. Indeed, differentiating power does not only depend on the taxonomic resolution of the technique, but is also restricted by the amount of variation inherent to the experiment. Apart from their 16S rRNA gene sequences, no further information was available on the isolates. As such, the question concerning organism relatedness could not be answered and requires further investigation. However, making the situation even more complex, similarities between replicate profiles may be taxon dependent (this could e.g. be related to the number of peaks generated in the profile), and therefore the maximum variation between replicate profiles may be different per taxon.

As a summary, the ideal approach would be to first verify whether the higher resolution of MALDI-TOF MS relative to rep-PCR was genuinely due to higher resolving power. This should be assessed from comparisons with other techniques. If the answer remains unclear, experimental variation should be

checked for. This may occur by calculating the maximum variation between replicates per taxon, and relating the value obtained to the distance to the closest neighbor (the choice of closest neighbor should be determined by the taxonomic level one is interested in). This can be achieved with the TaxonGap software [48].

Coming back to the case of the taxonomic resolution of MALDI-TOF MS relative to rep-PCR, this reasoning implies that if the variation within the replicate group is higher than the distance to the closest neighbor (which is defined at the resolution level of rep-PCR), the 'higher resolution' of MALDI-TOF MS relative to rep-PCR was due to experimental variation which exceeded the proposed cut-off value. In that case, the isolates could in fact not be differentiated by MALDI-TOF MS, although clustering suggested otherwise. However, as the results obtained in this study do not allow the approach proposed here, this may be material for further research.

Conversely, we concluded from partial 16S rRNA gene sequences that the resolution of the technique suggested at least species level differentiation. This conclusion may be an oversimplification, as there is no guarantee that 16S rRNA gene sequence similarities higher than 97% guarantee that the corresponding organisms belong to the same species. Moreover, the conclusions were drawn from partial 16S rRNA gene sequences containing only the V3 region of the bacterial 16S rRNA gene. Still, our study on the phylogenetic information content of short read sequences (Chapter 5) indicated that the slope in the pairwise distance correlation plot for the given primer (Fig. 3.4) was 1.07, thus well-representative for full length 16S rRNA gene sequences. As such, the short sequences do not significantly over- or underestimate pairwise distances between full length sequences and may well be representative.

.

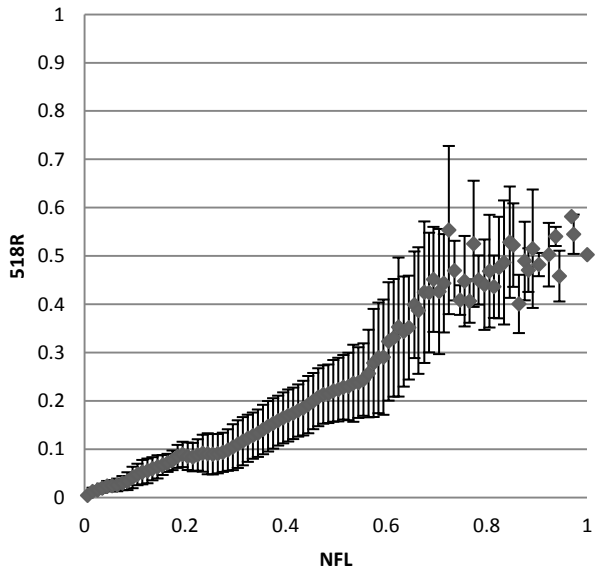


Figure 3.4 Pairwise distance correlation plot between reads generated from the 518r primer that was used for identification in this study, and corresponding nearly full-length (NFL) sequences.

Combined use with a colony picker

If MALDI-TOF MS could be automated in combination with a colony picker, this would mean tremendous potential for high-throughput isolation campaigns. However, at this moment, no colony picker is commercially available that has been developed specifically for use in combination with MALDI-TOF MS. Nevertheless, certain colony picker types possess functionalities that could allow their use for this purpose. As mentioned in the manuscript, there are two sample preparation methods for MALDI-TOF MS: cell smears and cell extracts. Technically, the cell smear method might be easiest to automate. However, as this method negatively affects reproducibility, it is preferable to analyze bacterial cell extracts. Full automation of cell extract preparation is difficult, since the protocol requires centrifugation and homogenization steps. Still, semi-automation and the combined use of a colony picker and liquid handling robot could be an option that would significantly decrease time and personnel cost for sample preparation.

3.7 REFERENCES

1. Tamaki H, Hanada S, Sekiguchi Y, Tanaka Y, Kamagata Y (2009) Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. Environmental Microbiology 11: 1827-1834.
2. Davis KER, Joseph SJ, Janssen PH (2005) Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. Applied and Environmental Microbiology 71: 826-834.
3. Connon SA, Giovannoni SJ (2002) High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Applied and Environmental Microbiology 68: 3878-3885.
4. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science 296: 1127-1129.
5. Bull AT, Goodfellow M, Slater JH (1992) Biodiversity as a source of innovation in biotechnology. Annual Review of Microbiology 46: 219-252.
6. De Clerck E, De Vos P (2002) Study of the bacterial load in a gelatine production process focussed on *Bacillus* and related endosporeforming genera. Systematic and Applied Microbiology 25: 611-617.
7. Heyrman J, Swings J (2001) 16S rDNA sequence analysis of bacterial isolates from biodeteriorated mural paintings in the Servilia tomb (necropolis of Carmona, Seville, Spain). Systematic and Applied Microbiology 24: 417-422.
8. Coorevits A, De Jonghe V, Vandroemme J, Reekmans R, Heyrman J, et al. (2008) Comparative analysis of the diversity of aerobic spore-forming bacteria in raw milk from organic and conventional dairy farms. Systematic and Applied Microbiology 31: 126-140.
9. Martin-Platero AM, Maqueda M, Valdivia E, Purswani J, Martinez-Bueno M (2009) Polyphasic study of microbial communities of two Spanish farmhouse goats' milk cheeses from Sierra de Aracena. Food Microbiology 26: 294-304.
10. Ishii S, Sadowsky MJ (2009) Applications of the rep-PCR DNA fingerprinting technique to study microbial diversity, ecology and evolution. Environmental Microbiology 11: 733-740.
11. Kang HP, Dunne WM (2003) Stability of repetitive-sequence PCR patterns with respect to culture age and subculture frequency. Journal of Clinical Microbiology 41: 2694-2696.

12. Souza RA, Pitondo-Silva A, Falcao DP, Falcao JP (2010) Evaluation of four molecular typing Methodologies as tools for determining taxonomy relations and for identifying. species among *Yersinia* isolates. Journal of Microbiological Methods 82: 141-150.

13. Olive DM, Bean P (1999) Principles and applications of methods for DNA-based typing of microbial organisms. J Clin Microbiol 37: 1661-1669.

14. Dombek PE, Johnson LK, Zimmerley ST, Sadowsky MJ (2000) Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Applied and Environmental Microbiology 66: 2572-2577.

15. Rademaker JLW, Louws FJ, Versalovic JV, de Bruijn FJ (2008) Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting. In: Kowalchuck GA, de Bruijn FJ, Head IM, Akkermans ADL, van Elsas JD, editors. Molecular Microbial Ecology Manual. Dordrecht, the Netherlands: Springer. pp. 611-644.

16. Masco L, Huys G, Gevers D, Verbrugghen L, Swings J (2003) Identification of *Bifidobacterium* species using rep-PCR fingerprinting. Systematic and Applied Microbiology 26: 557-563.

17. Rasschaert G, Houf K, Imberechts H, Grijspeerdt K, De Zutter L, et al. (2005) Comparison of five repetitive-sequence-based PCR typing methods for molecular discrimination of *Salmonella enterica* isolates. Journal of Clinical Microbiology 43: 3615-3623.

18. Deplano A, Schuermans A, van Eldere J, Witte W, Meugnier H, et al. (2000) Multicenter evaluation of epidemiological typing of methicillin-resistant *Staphylococcus aureus* strains by repetitive-element PCR analysis. Journal of Clinical Microbiology 38: 3527-3533.

19. Bizzini A, Greub G (2010) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. Clinical Microbiology and Infection 16: 1614-1619.

20. Eigner U, Holfelder M, Oberdorfer K, Betz-Wild U, Bertsch D, et al. (2009) Performance of a Matrix-Assisted Laser Desorption Ionization-Time-of-Flight Mass Spectrometry System for the Identification of Bacterial Isolates in the Clinical Routine Laboratory. Clinical Laboratory 55: 289-296.

21. Murray PR (2010) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: usefulness for taxonomy and epidemiology. Clinical Microbiology and Infection 16: 1626-1630.

22. Ryzhov V, Fenselau C (2001) Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Analytical Chemistry 73: 746-750.

23. Siegrist TJ, Anderson PD, Huen WH, Kleinheinz GT, McDermott CM, et al. (2007) Discrimination and characterization of environmental strains of *Escherichia coli* by matrix-assisted laser

desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). Journal of Microbiological Methods 68: 554-562.

24. Arnold RJ, Reilly JP (1998) Fingerprint matching of *E. coli* strains with matrix-assisted laser desorption ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. Rapid Communications in Mass Spectrometry 12: 630-636.
25. Dickinson DN, La Duc MT, Haskins WE, Gornushkin I, Winefordner JD, et al. (2004) Species differentiation of a diverse suite of *Bacillus* spores by mass spectrometry-based protein profiling. Applied and Environmental Microbiology 70: 475-482.
26. Vargha M, Takats Z, Konopka A, Nakatsu CH (2006) Optimization of MALDI-TOF MS for strain level differentiation of *Arthrobacter* isolates. Journal of Microbiological Methods 66: 399-409.
27. Rademaker JLWdB, F. J. (1997) Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer assisted pattern analysis. DNA Markers: Protocols, Applications and Overviews: 151-171.
28. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Research 33: D294-D296.
29. Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiology Today 33: 152-155.
30. Barbuddhe SB, Maier T, Schwarz G, Kostrzewa M, Hof H, et al. (2008) Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Applied and Environmental Microbiology 74: 5402-5407.
31. Mellmann A, Bimet F, Bizet C, Borovskaya AD, Drake RR, et al. (2009) High Interlaboratory Reproducibility of Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry-Based Species Identification of Nonfermenting Bacteria. Journal of Clinical Microbiology 47: 3732-3734.
32. Seibold E, Maier T, Kostrzewa M, Zeman E, Splettstoesser W (2010) Identification of *Francisella tularensis* by Whole-Cell Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: Fast, Reliable, Robust, and Cost-Effective Differentiation on Species and Subspecies Levels. Journal of Clinical Microbiology 48: 1061-1069.
33. Freiwald A, Sauer S (2009) Phylogenetic classification and identification of bacteria by mass spectrometry. Nature Protocols 4: 732-742.
34. Fujinami Y, Kikkawa HS, Kurosaki Y, Sakurada K, Yoshino M, et al. (2009) Rapid discrimination of *Legionella* by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Microbiol Res.

35. Tanigawa K, Kawabata H, Watanabe K (2010) Identification and Typing of *Lactococcus lactis* by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. Applied and Environmental Microbiology 76: 4055-4062.

36. Vanlaere E, Sergeant K, Dawyndt P, Kallow W, Erhard M, et al. (2008) Matrix-assisted laser desorption ionisation-time-of of-flight mass spectrometry of intact cells allows rapid identification of *Burkholderia cepacia* complex. Journal of Microbiological Methods 75: 279-286.

37. Dieckmann R, Strauch E, Alter T (2010) Rapid identification and characterization of *Vibrio* species using whole-cell MALDI-TOF mass spectrometry. Journal of Applied Microbiology 109: 199-211.

38. Giebel RA, Fredenberg W, Sandrin TR (2008) Characterization of environmental isolates of *Enterococcus* spp. by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Water Research 42: 931-940.

39. Hane BG, Jager K, Drexler HG (1993) The Pearson Product-Moment Correlation-Coefficient is better suited for identification of DNA fingerprint profiles than band matching algorithms. Electrophoresis 14: 967-972.

40. Chokesajjawatee N, Zo YG, Colwell RR (2008) Determination of clonality and relatedness of *Vibrio cholerae* isolates by genomic fingerprinting, using long-range repetitive element sequence-based PCR. Applied and Environmental Microbiology 74: 5392-5401.

41. Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, et al. (2008) Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. Journal of Clinical Microbiology 46: 1946-1954.

42. Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K (2005) Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. Applied and Environmental Microbiology 71: 58-64.

43. Ruelle V, El Moualij B, Zorzi W, Ledent P, De Pauw E (2004) Rapid identification of environmental bacterial strains by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry 18: 2013-2019.

44. Ron EZ (2006) Bacterial Stress Response. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes: Springer. pp. 1012-1027.

45. Hartmeyer GN, Jensen AK, Bocher S, Bartels MD, Pedersen M, et al. (2010) Mass spectrometry: Pneumococcal meningitis verified and *Brucella* species identified in less than half an hour. Scandinavian Journal of Infectious Diseases 42: 716-718.

46. Healy M, Huong J, Bittner T, Lising M, Frye S, et al. (2005) Microbial DNA typing by automated repetitive-sequence-based PCR. Journal of Clinical Microbiology 43: 199-207.

47. Shutt CK, Pounder JI, Page SR, Schaecher BJ, Woods GL (2005) Clinical evaluation of the DiversiLab microbial typing system using repetitive-sequence-based PCR for characterization of *Staphylococcus aureus* strains. Journal of Clinical Microbiology 43: 1187-1192.

48. Slabbinck B, Dawyndt P, Martens M, De Vos P, De Baets B (2008) TaxonGap: a visualization tool for intra- and inter-species variation among individual biomarkers. Bioinformatics 24: 866-867.

Author Contributions:

JG performed the bacterial isolations, identifications, lytic enzyme and siderophore production assays, statistics and co-wrote the manuscript with SV. Antagonism tests, direct plant growth-promotion tests and microplant trials were performed by SV under supervision of BDP and EOH. KH helped designing the experiment. JF and MR provided the samples for the experiment. KH, EOH, PDV & BDP proofread the manuscript.

CHAPTER 4

BIOPROSPECTING IN POTATO FIELDS IN THE CENTRAL ANDEAN HIGHLANDS: SCREENING OF RHIZOBACTERIA FOR PLANT GROWTH-PROMOTION PROPERTIES

*Redrafted from: Ghyselinck, J., Velivelli, S. L. S., Heylen, K., O'Herlihy, E., Franco, J., Rojas, M., De Vos, P. & Prestwich, B. D. (2013). Bioprospecting in potato fields in the Central Andean Highlands: Screening of rhizobacteria for plant growth-promoting properties. Syst Appl Microbiol **36**(2), 116-127*

SUMMARY

The Central Andean Highlands are the center of origin of the potato plant (*Solanum tuberosum*). Ages of mutualism between potato plants and soil bacteria in this region support the hypothesis that Andean soils harbor interesting plant growth-promoting (PGP) bacteria. The aim of this study was to isolate rhizobacteria from Andean ecosystems, and to identify those with PGP properties. A total of 585 bacterial isolates were obtained from eight potato fields in the Andes and were screened for suppression of *Phytophthora infestans* and *Rhizoctonia solani*. Antagonistic mechanisms were determined and antagonistic isolates were further tested for phosphate solubilization, 1-aminocyclopropane-1-carboxylate (ACC) deaminase activity, and production of NH_3- and indole-3-acetic acid (IAA). PGP was studied in healthy and *R. solani* diseased plantlets under growth room conditions. Performance was compared to the commercial strain *B. subtilis* FZB24® WG. Isolates were dereplicated with Matrix Assisted Laser Desorption/Ionisation Time Of Flight Mass Spectrometry (MALDI-TOF MS), and identified with 16S rRNA gene sequencing and Multi Locus Sequence Analysis (MLSA). Ten percent of the isolates were effective antagonists, of which many were able to to solubilize phosphate, and produce IAA, ACC deaminase, NH_3 and hydrogen cyanide (HCN). During growth room experiments, 23 antagonistic isolates were associated with plant growth-promotion and/or disease suppression. Ten isolates had a statistically significant impact on test parameters compared to the uninoculated control. Three isolates significantly promoted plant growth in healthy plantlets compared to the commercial strain, and seven isolates outperformed the commercial strain in *in vitro R. solani* diseased plantlets.

4.1 INTRODUCTION

The potato plant (*Solanum tuberosum*) is a valuable crop worldwide with a key role in the world's global food system. In order to emphasize its importance, the Food and Agriculture Organization (FAO) acclaimed 2008 as the International Year of the Potato [1]. Indeed, the plant possesses a number of interesting properties such as a high nutritional value, a high growth rate, the ability to grow worldwide and a high yield to soil occupation ratio. However, due to a growing demand for potato, crop rotation and fallow times are often reduced, making soil less fertile and increasingly more infested with soil-borne diseases [2]. This often leads to an increased use of chemical fertilizers and pesticides. In many cases, however, chemical pesticides have a harmful non-target environmental impact [3,4], and lose their efficiency over time due to the development of resistance in the pathogen populations [5,6]. Alternative strategies are therefore in great demand. The last decade has seen an increased interest in the application of microorganisms with plant growth-promotion (PGP) properties in agriculture. PGP organisms can be applied either solely [7,8,9,10], or in combination with chemical control agents [4,11]. Different mechanisms of PGP are known, and can be classified as biofertilization, stress control, rhizoremediation, phytostimulation (i.e. direct mechanisms) and pathogen suppression (i.e. indirect mechanisms) [12]. The latter, also referred to as biocontrol, is an environmentally-friendly approach in which a microbial natural antagonist of the plant pathogen is used to prevent plant disease. Unlike chemical pesticides, biocontrol agents use a number of mechanisms that generally do not harm the environment. Certain PGP bacteria work through multiple of the above mentioned PGP mechanisms, and thus are beneficial for plant growth both directly and indirectly. As a result, the use of microbial inoculants to control plant diseases is becoming more and more popular, with an annual increase of approximately 10% [7]. For instance, fungi belonging to the genera *Trichoderma* [13] and *Gliocladium*, and bacteria from the genera *Bacillus*, *Pseudomonas* and *Streptomyces* are commercially available for their application against several plant pathogens such as *Rhizoctonia*, *Fusarium*, *Alternaria*, *Pythium* and *Sclerotinia*.

Rhizoctonia solani and *Phytophthora infestans* are known as two of the most important potato pathogens. The former causes stem canker and scurf diseases, the latter causes potato late blight disease, one of the most devastating diseases of potato worldwide [14,15]. Since control of *R. solani* by chemical fungicides has met with limited success, many studies explored the efficiency of biocontrol agents against the pathogen. Various antagonistic fungi including *Trichoderma* spp. [4,7,16,17], *Verticillium* spp. [11,18] and *Gliocladium* spp. [19], as well as bacteria from the genera *Bacillus* [20,21],

Pseudomonas [20,21,22] and *Burkholderia* [21,22] have been examined for their potential as biocontrol agents against *R. solani*; often with promising results. However, for *P. infestans* the search for fungal or bacterial antagonists appears to be more difficult. Taking into account the destructive nature of this pathogen and the economical losses associated herewith, it is clear that further studies on *P. infestans* disease control are urgently needed.

The potato plant is a crop that is indigenous to the Central Andean Highlands and has been cultivated locally by farming communities for centuries. Plants are often grown at high altitudes under harsh climatic conditions, and in fields where nutrients are often unavailable due to high soil acidity [23]. As such, it is very likely that a strong mutualistic relationship between potato plants and rhizosphere bacteria has evolved over time, which leads one to suspect that in this region plant-associated bacteria play a crucial role for the potato plant. However, to date, only limited data is available on PGP potential of rhizosphere microorganisms from this region [24]. Taking into account the key role of the potato in the world's global food system, exploration of the microbial potential present in these soils may be of great economical value. Therefore, this study aimed at isolating and screening rhizosphere bacteria for antagonistic activities against *R. solani* and *P. infestans* and for direct and indirect PGP properties *in vitro*. In order to do so, eight different potato fields from the Central Andean Highlands of Peru and Bolivia were investigated, varying in altitude, soil composition, climatic conditions and pesticide as well as fertilizer use. If, in an initial screening step, antagonistic activity against these specific pathogens was observed *in vitro*, bacteria were further tested for the ability to solubilize phosphate, and for the production of indole-3-acetic acid (IAA), ammonia (NH_3), hydrogen cyanide (HCN), and 1-aminocyclopropane-1-carboxylate (ACC) deaminase. All antagonistic isolates were then tested for direct PGP effects and biocontrol activity against *R. solani* on potato plantlets *in vitro*. In addition, bacterial isolates were identified with 16S rRNA gene sequencing and Multi Locus Sequence Analysis, and pathogenicity was discussed. The *in vitro* PGP property screening approach presented here, allowed us to select the most promising isolates for field trials.

4.2 MATERIALS AND METHODS

Bacterial isolates

A total of 585 bacteria were isolated from the rhizosphere of *Solanum tuberosum* (potato) plants from eight fields located in the Central Andean Highlands of Peru and Bolivia (Figs. 4.1 and 4.2, Table 4.1). Ten plants were sampled per field. An amount of 5 g of rhizosphere soil adhering directly to the potato roots was collected per plant. Rhizosphere samples were then pooled per field and stored at 4°C until further processing. For bacterial isolations, 5 ml phosphate-buffered saline and 10 sterile glass beads (6 mm) were added to 1 g of each of the pooled rhizosphere soil samples, which were then vortexed for 2 min. Serial dilutions were made (10^{0}-10^{-6}) and plated (100 µl) on ten-fold diluted Trypticase Soy Agar (TSA) and a standard mineral base [25] supplemented with γ-caprolactone as the sole carbon source. Gamma-caprolactone is a compound structurally related to N-acylhomoserine lactones (which are involved in quorum sensing) [26], and was introduced in the medium in order to benefit the isolation of signal molecule degrading bacteria. Both media were supplemented with 0.005% (w/v) cycloheximide to inhibit fungal growth. Incubation temperatures were 15°C and 28°C. At random time intervals, isolates with visually different colony morphologies were picked and subcultured to purity.

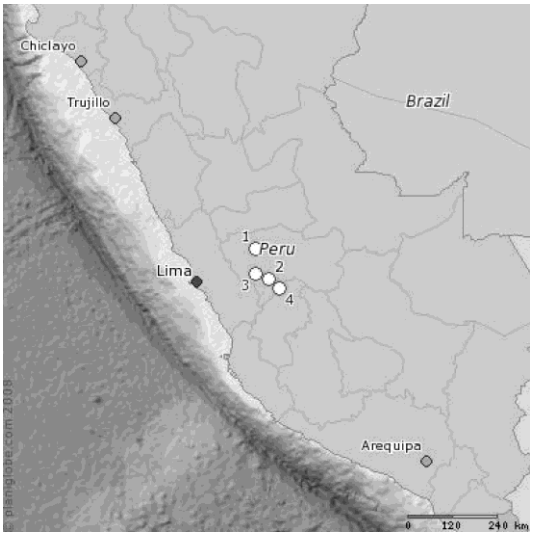


Figure 4.1 Locations of the four sampling sites in Peru (P1-P4) (http://www.planiglobe.com/). GPS coordinates of the sampling sites are the following: P1: 11° 15’ 52.72” latitude, 75° 37’ 17.10” longitude; P2: 12° 00’ 44.4.” latitude, 75° 17’ 29.70” longitude; P3: 11° 52’ 26.16” latitude, 75° 37’ 6.24” longitude; P4: 12° 13’ 27.36” latitude, 75° 02’ 48.6” longitude.

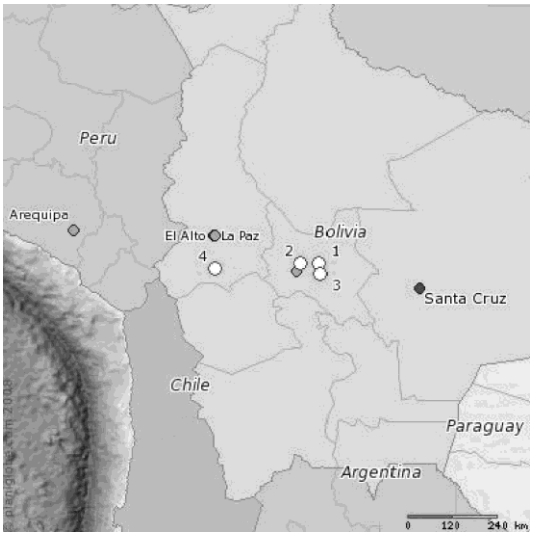


Figure 4.2 Locations of the four sampling sites in Bolivia (B1-B4) (http://www.planiglobe.com/). GPS coordinates of the sampling sites are the following: B1: 17° 9’ 57.6” latitude, 65° 38’ 36.9” longitude; B2: 17° 9’ 57.6” latitude, 66° 5’ 3.7” longitude; B3: 17° 26’ 57.5” latitude, 65° 36’ 57.1” longitude; B4 17° 18’ 46” latitude, 68° 7’ 30” longitude.

	Bolivia				Peru			
	Field 1	**Field 2**	**Field 3**	**Field 4**	**Field 1**	**Field 2**	**Field 3**	**Field 4**
Plant growth stage	Senescence	Senescence	Senescence	Senescence	Elongation	Flowering	Senescence	Senescence
Sampling location	Tiraque, Cochabamba	Chapare, Cochabamba	Tiraque, Cochabamba	Aroma, La Paz	Huasahuasi, Tarma	Sicaya, Huancayo	Huancani, Jauja	Pazos, Tayacaja
Altitude (m)	3 560	3 700	4 010	4 070	2 780	3 280	3 920	4 150
Potato variety	Cultivar Waycha	Cultivar Waycha	Cultivar Waycha	Cultivar Waycha	Unica	Unica	Yungay	Muru Huayro
Field history	Barley – Potato (sampling)	Barley – Potato (sampling)	Fallow for many years - Potato (sampling)	Fallow - Barley - Potato (sampling)	Flowers (Gladiolus) - Pea - Potato - Potato (sampling)	Potato (variety Canchan) - Potato (sampling)	Oat - Oat - Oat - Potato (sampling)	Potato – Potato- Incorporation of fresh manure to soil – Potato (sampling)
Fertilizer	Chemical Fertilizer (low input) + Chicken manure	Chicken Manure	Chemical Fertilizer (low input) + Chicken manure	Sheep Manure	Manure + Chemical Fertilizer	Manure + Chemical Fertilizer	Manure + Chemical Fertilizer	Manure + Chemical Fertilizer
Pesticide application	Insecticide	Insecticide	Insecticide	No	No	No	No	No
pH	4.6	4.7	4.2	6.2	4.8	7.7	4.3	4.4
Soil Organic Matter (%)	5.3	22.3	11.3	5.3	3	2.3	3.7	14.8
N Content (%)	0.06	0.03	0.02	0.06	0.16	0.19	0.19	0.74
P (ppm)	5.7	13.3	6.9	6.1	101.7	42.7	17.5	6.3
Sand (%)	55.4	21.5	50.1	68	60	24	50	Organic
Clay (%)	11	30	14.8	11	10	30	10	Organic
Silt (%)	33.7	48.5	35.2	21	30	46	40	Organic

Table 4.1 Specifications of the sampling locations.

Screening for bioprotectant properties using plate assays (indirect PGP)

Dual-culture assays

The two plant pathogenic strains, namely *R. solani* EC-1 and *P. infestans* EC-1, were grown for 2 weeks on petri dishes of potato dextrose agar and green pea agar respectively, at their optimum temperature of 25°C. *In vitro* antagonistic activity of bacterial isolates against the pathogens was evaluated according to the method described by Dikin and Sijam [27]. Isolated bacteria were streaked on Tryptone Soya Agar (TSA) and grown at 28°C overnight. Control plates with just the mycelial plug were set up and when the pathogen had grown across these control plates, the diameter of growth in the challenge plates was measured. The performance of the isolates was compared to the commercially available rhizobacterial strain, *B. subtilis* FZB24® WG.

The percentage suppression was calculated from the following formula:

$$\frac{\text{(Total growth of the control - measured growth with bacteria)}}{\text{Total growth}} \times 100\%$$

Dual culture assays were repeated five times per isolate and for each pathogen. The average values were calculated and presented.

Enzyme production

Certain bacteria inhibit fungal growth by the production of fungal cell wall degrading enzymes. Bacterial isolates were tested for the production of chitinase, cellulase and β-glucanase in semiminimal medium, i.e. a mixture of minimal dNMS (diluted Nitrate Mineral Salts) medium [28] and nutrient broth (3:1), supplemented with 1.5% agar and 0.2% colloidal chitin, 0.1% AZCL-HE-cellulose (Megazyme, Ireland) and 0.1% AZCL-Pachyman (Megazyme, Ireland) respectively. Plates were incubated at 28°C for 7 days. Cellulose and β-glucan degradation were detected by the formation of blue haloes around the colonies. Chitin degradation was detected by the appearance of clearing zones around the colonies after flooding the plates with a 0.1% Congo Red in distilled water solution. Protease activity was determined from the appearance of clearing zones around bacterial colonies on skimmed milk agar (equal volumes of skimmed milk and 1/5 Trypticase Soy Broth (TSB), mixed at 60°C and solidified with 1.5% agar) after 7 days of incubation at 28°C. Enzyme production assays were performed once per isolate.

Siderophore production

Siderophore production was detected on solid medium according to Schwyn & Neilands [29]. Prior to medium preparation, all glass vials and vessels were deferrated by rinsing with 6M HCl. The casamino

acid solution was deferrated by extraction with a 3% (w/w) 8-hydroxyquinoline in chloroform solution before being introduced into the medium. Bacteria were incubated at 28°C for a period of 7 days to check for siderophore production. Siderophore production assays were performed once per isolate.

HCN Production

Bacterial isolates were grown on nutrient agar amended with glycine (4.4 g/l). A Whatman filter paper no.1 soaked in a 2% sodium carbonate in 0.5% picric acid solution was placed on top of each plate. Plates were sealed with parafilm and incubated at 28°C for 4 days. Development of orange to red color indicated HCN production [30].HCN production assays were performed twice per isolate.

Screening for direct PGP properties in selected bacterial isolates

Phosphate-solubilization

The phosphate-solubilization assay was carried out by growing bacteria in TSB to a concentration of 10^8 cfu/ml. A volume of 5 µl of bacterial cultures was then inoculated onto plates of the National Botanical Research Institute's Phosphate medium (NBRIP) [31] containing insoluble tricalcium phosphate, giving it an opaque appearance. Plates were incubated at 28°C and clearing zones appeared around the colonies if the isolate was positive for phosphate-solubilization. Clearing zones were measured every 48 hours for 30 days to record the solubilization process. Phosphate solubilization assays were repeated four times per isolate. The average values were calculated and presented.

IAA production

Indole acetic acid production was determined quantitatively through a colorimetric microplate assay using the method described by Bano and Musarrat [32]. IAA production assays were repeated twice for each isolate. The average values were calculated and presented.

ACC deaminase activity

Bacterial isolates were grown in TSB medium at 28°C for 2-4 days and the ACC deaminase activity was determined by the method described by Penrose & Glick [33]. ACC deaminase assays were repeated three times for each isolate. The average values were calculated and presented.

NH_3 Production

Freshly grown bacterial cultures were inoculated in 10 ml peptone water (Sigma) and incubated for 48-72 h at 28°C. Nessler's reagent (0.5 ml) was added to each tube and the development of yellow to brown

colour indicated a positive result for ammonia production [34]. NH_3 production assays were performed twice per isolate.

Growth room experiments

Effect of bacterial isolates on plant growth in vitro

Potato (*Solanum tuberosum*) microplants of cultivar 'Unica' were taken from stock cultures (School of Biological, Earth and Environmental Sciences, University College Cork, Ireland) and were grown on heterotrophic medium (1/2 strength M&S basal medium (Sigma, Cat no. M-5519) containing 0.1 mg/l kinetin, 0.2 mg/l gibberellic acid, 15 g/l sucrose, 6 g/l agar, adjusted to pH 5.8). Microplants were grown under growth room conditions of 23°C, 16 h photoperiod for four weeks. Subsequently, the four week old rooted microplants were grown in Magenta culture vessels (6 plants per vessel × 5 replicates) in polyurethane foams imbibed with 50 ml autotrophic medium (1/2 strength M&S salts, pH 5.8). After two weeks the microplants were inoculated with 1 ml of 10^6 CFU/ml of each bacterial isolate [35]. The experiments were repeated five times per isolate. Plants were carefully removed using forceps from the Magenta culture vessel. Excess moisture was removed by blotting the plant material dry prior to weighing. The effect of each isolate was determined by measuring plant growth (expressed as plant weight) after four weeks. Uninoculated control microplants were also set up. The performance of the isolates was compared to the commercially available rhizobacterial strain, *B. subtilis* FZB24® WG.

Effect of bacterial isolates on plant growth in vitro of plantlets inoculated with Rhizoctonia solani

Potato microplants (cultivar 'Unica') after one week bacterization with the bacterial isolates (as described previously) were challenged with *R. solani* by placing an 8 mm plug of a two week old culture in the centre of the growth vessel. The experiments were repeated five times per isolate. Uninoculated control microplants challenged with *R. solani* were also set up by inoculating with a disc of the fungal culture. Plants were carefully removed using forceps from the majenta culture vessel. Excess moisture was removed by blotting the plant material dry prior to weighing. The effect of each isolate was determined by measuring plant growth (expressed as plant weight) three weeks after the pathogen challenge. The performance of the isolates was compared to the commercially available rhizobacterial strains *B. subtilis* FZB24® WG.

Identification of the bacterial isolates

Dereplication with MALDI-TOF MS

Preparation of cell extracts, MALDI-TOF MS analysis and analysis of spectral data was performed according to Ghyselinck et al. [36]. Mass spectra with extensive noise and/or insufficient signal

intensities were excluded from the analysis. Dendrograms were created using the BioNumerics 5.1 software (Applied Maths, Belgium). Similarities between spectra were expressed using Pearson's product moment correlation coefficient. Spectra were clustered using the unweighted pair group method with arithmetic averages (UPGMA) clustering algorithm. For cluster delineation, a cluster cut-off value of 94.75% was used, which was calculated based on triplicate analysis of randomly chosen bacterial strains (data not shown) according to Ghyselinck et al. [36]. This allowed to dereplicate at the species-to-strain level [36]. Representatives per cluster and separately clustered isolates were considered to be different strains after dereplication.

16S rRNA, gyrB, rpoB and rpoD gene sequence and phylogenetic analysis

Genomic DNA was released from the bacterial cells using alkaline lysis. A small amount of cells were lysed in 20 µl alkaline lysis buffer (0.25% (w/v) SDS and 0.05 M NaOH) for 15 min at 95°C. Subsequently, 180 µl sterile milliQ-water was added and lysates were used for PCR. Genus identification of all *in vitro* PGP strains was performed by sequence analysis of the 16S rRNA gene. Amplification of the 16S rRNA gene was performed as described by Heyrman & Swings [37]. Amplicons were purified with the Nucleofast® 96 PCR system (Millipore, Belgium) and sequenced according to Heyrman & Swings [37]. Sequencing products were purified with the BigDye XTerminator® Purification Kit (Applied Biosystems, USA) and gene sequences were analyzed using a 3130 XL Genetic Analyzer (Applied Biosystems, USA). Sequences were assembled with the BioNumerics 5.1 software (Applied Maths, Belgium). Preliminary genus identification was obtained by query in the "Classifier" program of Ribosomal Database Project II [38] of the 16S rRNA gene sequence of each strain. Analyses for further identification were dependent on genus affiliation.

For strains assigned to the genus *Bacillus*, *gyrB* gene sequence analysis was performed in order to obtain an in-depth identification [39,40]. However, no *gyrB* amplicons could be obtained with the primers used [41], even after several attempts and with different PCR conditions tested.

Strains that were assigned to the genus *Pseudomonas* were identified in-depth by additionally sequencing of the *gyrB*, *rpoB* and *rpoD* genes. The genes in question were amplified according to Yamamoto & Harayama [41], Tayeb et al. [42] and Yamamoto & Harayama [43] respectively. For sequencing of the *rpoB* gene, the same primers were used as for gene amplification. The *rpoD* gene was sequenced using the 70FS and 70RS primers [43] and the *gyrB* gene using the degenerate primers UP-1S and UP-2Sr [41]. The sequencing protocol was identical to that of the 16S rRNA gene. The 16S rRNA gene sequences of the *Pseudomonas* strains were aligned using the integrated aligner of ARB [44].

Highly variable positions were filtered out using the positional variability filter from the ARB software, which is based on all sequences of the domain Bacteria within the complete SSU SILVA datasets. *gyrB*, *rpoB* and *rpoD* gene sequences were checked by nucleotide to amino acid, and subsequent amino acid to protein translation using Transeq (http://www.ebi.ac.uk/Tools/emboss/transeq/) and pBLAST [45] respectively. Subsequently, these genes were aligned using the Molecular Evolutionary Genetics Analysis (MEGA 5) software [46]. Concatenated sequences were constructed as suggested by Mulet et al. [47], using the SeaView 4.3.0 software [48]. Similarly, concatenated genes were constructed of the *Pseudomonas* type strains representing the different groups and subgroups within the genus *Pseudomonas* (Table S4.1). Sequences of the four genes in question were obtained from international databases, through query in StrainInfo [49]. The lengths of the 16S rRNA, *gyrB*, *rpoD*, and *rpoB* genes in the concatenated sequences were 1137, 734, 501 and 678 bp respectively. The jModelTest 0.1.1 program [50] was then applied on the concatenated gene data set, which included both the strains under research and the *Pseudomonas* type strains, and the HKY evolutionary model was determined as the best available substitution model. The MEGA 5 software was then used to construct Maximum Likelihood (ML) phylogenetic trees using the HKY, Gamma distributed with Invariant sites evolutionary model. Bootstrap analysis [51], based on 1000 replicates, was used to calculate the statistical significance of the branches of the phylogenetic tree.

For the remaining strains, the 16S rRNA gene sequences of all type strains of all species of all genera mentioned in the Classifier report were compared in an exhaustive pair wise manner with the query sequence of each strain in BioNumerics 5.1 (Applied Maths, Belgium). The strains were assigned to a genus based on the obtained pairwise 16S rRNA gene sequence similarities.

Statistical analyses

Data normality, which was checked in SPSS, and the relation between the variables (i.e. linearity or monotonicity) were the main criteria to determine the coefficient best fit for calculating correlations. To detect the statistical significance of differences between means obtained from the *in vitro* tests on plantlets, the Kruskal-Wallis test was performed. The student t-test was performed to deduce the statistical significance of the relation between the production of lytic enzymes, siderophores, NH_3- and HCN on one hand, and both dual-culture assays and *in vitro* disease suppression tests on plantlets on the other hand.

Nucleotide sequence accession numbers

The 16S rRNA, *rpoB*, *rpoD* and *gyrB* gene sequence data generated in this study has been deposited in GenBank/EMBL/DDBJ with accession numbers HE603489-HE603535, HE603536-HE603560, HE603586-HE603610 and HE603561-HE603585 respectively.

4.3 RESULTS

Isolation campaign

Rhizosphere bacteria were isolated from potato plants from eight different fields in the Central Andean Highlands of Peru (Fig. 4.1) and Bolivia (Fig. 4.2); four fields were sampled per country differing in altitude, soil composition, climatic conditions and pesticide and fertilizer use (Table 4.1).

The isolation procedure, in which three different isolation conditions were used and only morphologically distinct isolates were selected, aimed at the retrieval of a broad bacterial diversity to increase chances of encountering bacteria with PGP properties. A total of 585 bacteria were isolated; 365 isolates from Bolivia and 220 from Peru. Results of the isolation campaign (Table 4.2) demonstrated that retrieval of a large variety of isolates was not correlated with the presence of antagonistic bacteria.

	Field	1.10^{-1} diluted TSA 15°C		1.10^{-1} diluted TSA 28°C		γ-caprolactone	
		# isolates	# isolates with PGP potential[a]	# isolates	# isolates with PGP potential[a]	# isolates	# isolates with PGP potential[a]
Bolivia	1	43	5	6	3	1	0
	2	41	4	58	7	17	0
	3	42	2	48	5	0	0
	4	58	12	36	2	15	1
Peru	1	/	/	42	6	0	0
	2	/	/	40	2	0	0
	3	/	/	39	3	0	0
	4	/	/	84	6	15	0
Total		184	23	353	34	48	1

Table 4.2 Overview of the numbers of morphologically different isolates isolated from each growth medium used, and arranged per country and per field. Note that these retrieved isolates do not reflect the cultivable richness of bacteria in these soils; they rather indicate their 'colony morphological variation of the cultivable fraction'.
[a], PGP= Plant Growth Promotion. Isolates were considered to have PGP potential if at least one of the indirect plant growth promotion assays was positive.

For example, from field 1 in Bolivia six isolates were retrieved on tenfold diluted TSA at 28°C of which three (or 50%) showed antagonistic activities in dual-culture assays, while from field 2 in Peru 40 isolates were isolated under the same conditions of which only two (or 5%) showed antagonistic activities. Field 4 from Bolivia – with a relatively high pH of 6.2, not treated with pesticide and fertilized with sheep manure – rendered most *in vitro* PGP isolates and the highest ratio *in vitro* PGP bacteria to total bacteria investigated. Surprisingly, only one out of 48 isolates that were isolated from the medium with γ-caprolactone (a compound structurally related to N-acylhomoserine lactones) was found to have

antagonistic properties (Table 4.2). Also no bacteria could be retrieved from fields 1, 2 and 3 from Peru, and field 3 from Bolivia on that specific medium.

Antagonistic activity in dual-culture assays

All isolates were screened for antagonistic activity against *R. solani* and *P. infestans* in dual-culture assays. A total of 58 bacterial isolates, corresponding to 9.9% of the isolates set, inhibited growth of either one or both pathogens (Table S4.2). All of the 58 isolates were effective against *R. solani*. However, two isolates failed to inhibit growth of *P. infestans* and as such, 56 isolates were effective against both pathogens. Antagonistic activity against *R. solani* ranged from 24.90% to 53.41% inhibition, while antagonistic activity against *P. infestans* ranged from 0% to 100%.Where there was no evidence of any growth of *P. infestans* it was assumed that there was 100% growth inhibition by the bacterial strain. The overall inhibition, for which the size of the inhibition zones on the plates is taken into account, was largest for *P. infestans*, which could mean that the isolates *in vitro* were more effective in controlling *P. infestans* than *R. solani*. The commercial rhizobacterial strain *B. subtilis* FZB24® WG inhibited growth of *R. solani* and *P. infestans* by 21.32% and 65.11% respectively.

Dereplication, identification and pathogenicity assessment of bacterial isolates

Isolates that had shown antagonistic activity in dual-culture assays, were dereplicated at the species-to-strain level with MALDI-TOF MS. Eight clusters were delineated, each containing two to five isolates, and 39 isolates occupied unique positions in the dendrogram (Fig. 4.3). The unique isolates and one representative of each remaining cluster were considered to be separate strains (47 in total) and were further identified onto genus level (Table 4.3). It should be noted that screening for PGP properties (see below) was performed on the complete set of isolates, and was not restricted to the 47 strains mentioned above, since it is known that PGP properties can be strain-specific in some taxa.

Bacterial isolates belonged to the genera *Pseudomonas* (29), *Bacillus* (22), *Paenibacillus* (1), *Flavobacterium* (1), *Curtobacterium* (2), *Pedobacter* (1) and *Enterobacter* (2) (Table 4.3). Since most of the antagonistic isolates belonged to the genera *Pseudomonas* and *Bacillus*, and because these genera also harbor human- and plant-pathogenic species, strains of both genera were identified more in-depth with Multilocus Sequence Analysis (MLSA). Strains of the genus *Pseudomonas* were further identified through phylogenetic analysis of the concatenated sequences of the 16S rRNA, *rpoB*, *rpoD* and *gyrB* genes (Fig. 4.4). Strains mainly belonged to the *P. fluorescens* group (11), but also members of the *P. jessenii* (2), *P. corrugata* (5), *P. koreensis* (6), *P. syringae* (3) and *P. aeruginosa* (2) groups were found.

Deeper identification of *Bacillus* strains would have been obtained from sequencing of the 16S rRNA and *gyrB* gene [39,40], but amplification of *gyrB* genes failed.

Human and plant pathogenicity of the antagonistic isolates was evaluated through *in silico* query in the BCCM/LMG (http://bccm.belspo.be/) and National Collection of Plant Pathogenic Bacteria (http://www.ncppb.com/) catalogues respectively. Eleven isolates belonging to the *P. fluorescens*, *P. corrugata* and *P. syringae* groups, and two isolates identified as *Curtobacterium* species, were indicated as potentially plant pathogenic (Table 4.3). All antagonistic *Pseudomonas* isolates from this study were classified under risk group 1 (according to the Belgian Regional Decrees [52]) and as such could be considered as non-pathogenic to humans. Deeper identification of *Bacillus* strains was unsuccessful, and therefore pathogenicity had to be evaluated based on identifications obtained with 16S rRNA gene sequencing. The latter showed that sixteen *Bacillus* strains showed high sequence similarities with the type strains of *B. mycoides* and *B. weihenstephanensis* (Table 4.3). Query in the BCCM/LMG Catalogue showed that *B. mycoides* is classified as a species non-pathogenic to humans, while *B. weihenstephanensis* was indicated as an opportunistic human pathogen. The remaining antagonistic isolates were classified under risk group 1 (according to the Belgian Regional Decrees [52]), with the exception of two isolates that showed high 16S rRNA gene sequence similarities with the type strain of *Enterobacter amnigenus*, which is a pathogenic species classified under risk group 2 (according to the Belgian Regional Decrees [52]).

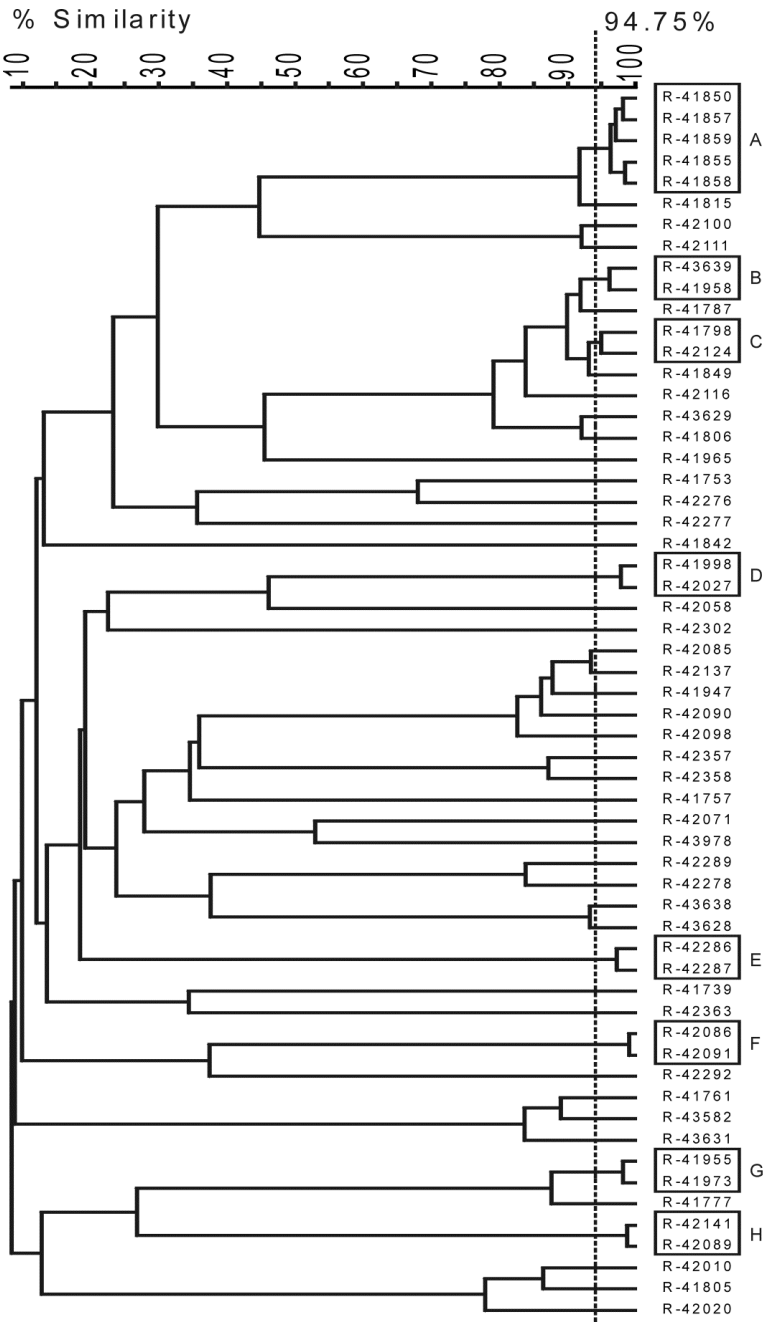
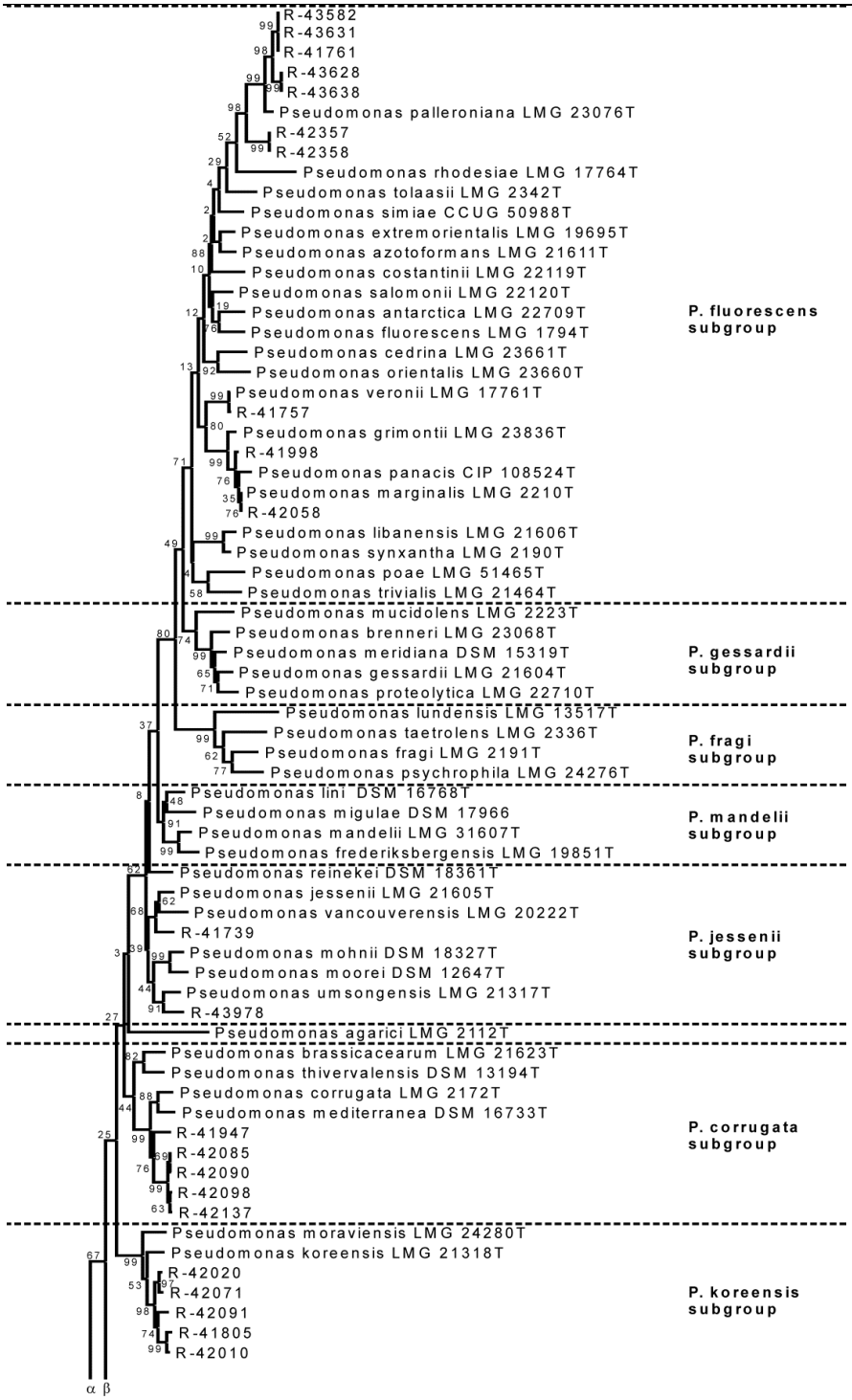


Figure 4.3 Dendrogram derived from MALDI-TOF MS fingerprinting, for dereplication of the biocontrol isolates at the species-to-strain level. The dendrogram was created using Pearson product moment correlation coefficient and UPGMA. Because the taxonomic resolution of MALDI-TOF MS is taxon dependent, clusters A-H group isolates that systematically belong to the same species/subspecies/strain, depending on the taxon. The cluster cut-off was set at 94.75%.



Continued on the next page.

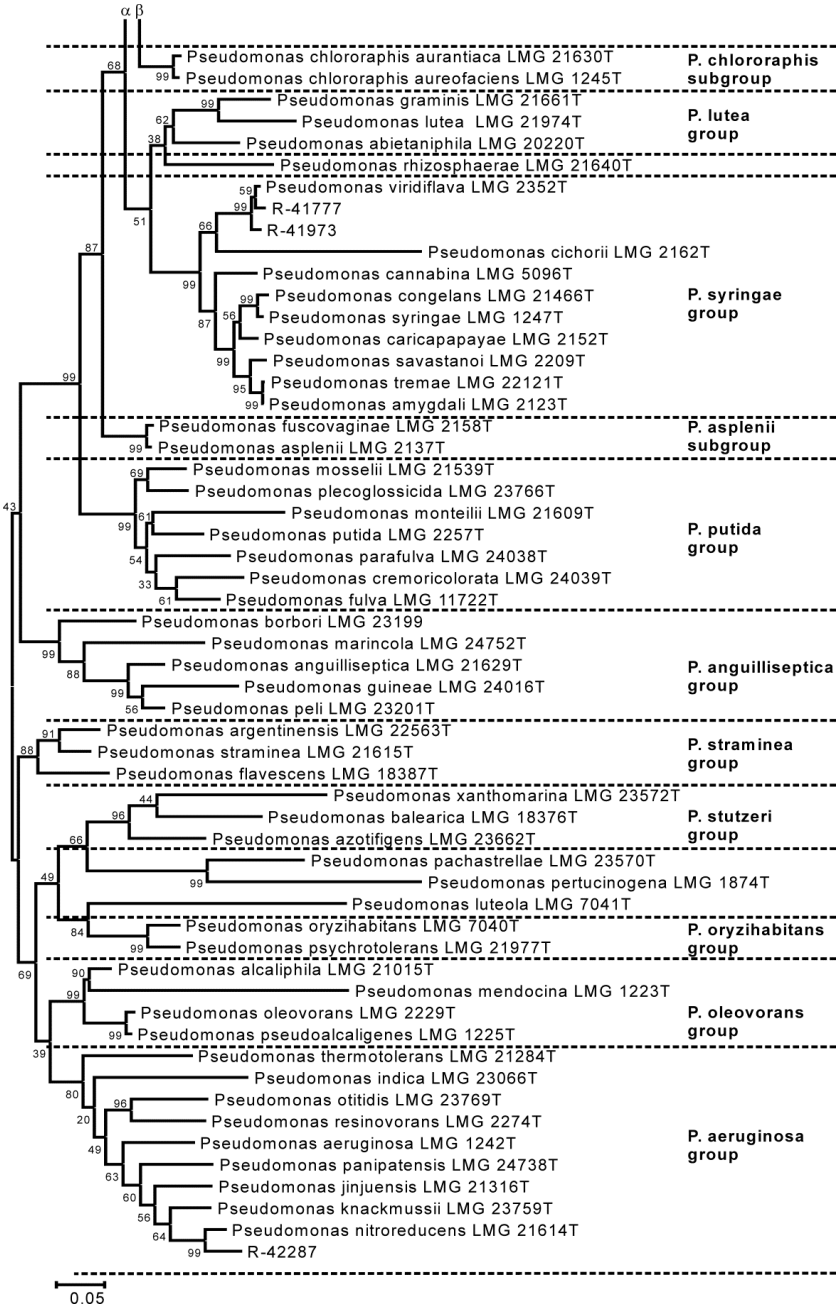


Figure 4.4 Maximum Likelihood tree constructed from a dataset of concatenated gene sequences of the 16S rRNA, *gyrB, rpoD* and *rpoB* genes of (1) the *in vitro* PGP *Pseudomonas* strains and (2) the *Pseudomonas* type strains representing the different groups and subgroups within the genus *Pseudomonas* [47]. The lengths of the concatenated genes were 3050 bp. Bootstrap values were obtained after bootstrap analyses based on 1000 replicates. Greek symbols mark the orientation of the figure.

Identification	**Type strain with highest 16S rRNA/concatenated gene sequence similarity to query sequence**		**Human pathogenicity**[a]	**Plant pathogenicity**[b]	**Isolate**
	Closest related type strain	**Sequence similarity (%)**			
Pseudomonas jessenii subgroup	*P. jessenii* LMG 21605T	97.7	Biohazard group 1	Non pathogenic	R-41739
	P. umsongensis LMG 21317T	97.5	Biohazard group 1	Non pathogenic	R-43978
Pseudomonas fluorescens subgroup	*P. veronii* LMG 17761T	99.9	Biohazard group 1	Non pathogenic	R-41757
	P. palleroniana LMG 23076T	96.8-98.6	Biohazard group 1	Non pathogenic	R-41761, R-42358, R-43582, R-43638, R-42357, R-43628, R-43631
	P. marginalis LMG 2210T	99.4-100	Biohazard group 1	Pathogenic	R-41998, R-42027, R-42058
Pseudomonas syringae group	*P. viridiflava* LMG 2352T	98.7-98.8	Biohazard group 1	Pathogenic	R-41777, R-41955, R-41973
Pseudomonas koreensis subgroup	*P. koreensis* LMG 21318T	97.5-97.9	Biohazard group 1	Non pathogenic	R-41805, R-42010, R-42091, R-42020, R-42071, R-42086
Pseudomonas corrugata subgroup	*P. corrugata* LMG 2172T	97.3	Biohazard group 1	Pathogenic	R-41947
	P. mediterranea DSM 16733^{T}		Biohazard group 1	Non pathogenic	
	P. corrugata LMG 2172T	97.3-97.5	Biohazard group 1	Pathogenic	R-42085, R-42090, R-42098, R-42137
Pseudomonas aeruginosa group	*P. nitroreducens* LMG 21614T	96.1	Biohazard group 1	Non pathogenic	R-42287, R-42286

Continued on the next page.

Identification	Type strain with highest 16S rRNA/concatenated gene sequence similarity to query sequence		Human pathogenicity[a]	Plant pathogenicity[b]	Isolate
	Closest related type strain	Sequence similarity (%)			
Bacillus sp.	*B. weihenstephanensis* LMG 18989T	99.8-100	Opportunistic pathogen	Non pathogenic	R-41753, R-41787, R-41806, R-41798, R-42124, R-41815, R-41849, R-41850, R-41855, R-41857, R-41858, R-41859, R-43629, R-43639, R-41958, R-42116
	B. mycoides LMG 7128T		Biohazard group 1	Non pathogenic	
	B. simplex LMG 11160T	100	Biohazard group 1	Non pathogenic	R-42276, R-42277
	B. aryabhattai LMG 24407T	99.9-100	Biohazard group 1	Non pathogenic	R-42278, R-42289
	B. vallismortis LMG 18725T	100	Biohazard group 1	Non pathogenic	R-42292
	B. subtilis LMG 7135T		Biohazard group 1	Non pathogenic	
	B. amyloliquefaciens LMG 12234T	99.7	Biohazard group 1	Non pathogenic	R-42363
Curtobacterium sp.	*C. flaccumfaciens pv. Flaccumfaciens* LMG 3645T	100	Biohazard group 1	Pathogenic	R-42100, R-42111
Paenibacillus sp.	*P. peoriae* LMG 14832T	99.2	Biohazard group 1	Non pathogenic	R-42302
Enterobacter sp.	*E. amnigenus* LMG 2784T	100	Biohazard group 2	Non pathogenic	R-42089, R-42141
Flavobacterium sp.	*F. psychrophilum* LMG 13179T	97.2	Biohazard group 1	Non pathogenic	R-41965
Pedobacter sp.	*P. ginsengisoli*	98.5	Biohazard group 1	Non pathogenic	R-41842

Table 4.3 Overview of the identity and pathogenic characteristics of the putative biocontrol isolates. Identification of *Pseudomonas* strains was based on phylogenetic analysis of concatenated gene sequences [47]; identification of the other strains was based on phylogenetic analysis of 16S rRNA gene sequences.
[a], Taken from http://bccm.belspo.be/db/lmg_search_form.php
[b], Isolates were considered plant pathogenic if present in the catalogue at http://www.ncppb.com/

***In vitro* characterization of biocontrol mechanisms**

All antagonistic isolates were tested for the production of siderophores, HCN and fungal cell wall degrading enzymes (Table S4.2). A high number of the antagonistic isolates produced siderophores in plate assays, i.e. 54 isolates or 93.1%. No isolates produced all four fungal cell wall degrading enzymes. Yet, one isolate belonging to the genus *Paenibacillus* and isolated from field 2 in Peru was able to produce three, i.e. proteases, cellulases and β-glucanases. Almost all bacterial isolates (86.2%) produced proteases, while only a minority produced chitinase (10.3%) and β-glucanase (5.2%). One isolate (R-42302) produced cellulase. Eight bacterial isolates belonging to the genera *Pseudomonas* (5), *Pedobacter* (1) and *Enterobacter* (2) did not produce any lytic enzymes. Of 58 isolates tested for the production of HCN, 6 *Pseudomonas* isolates (10%) were positive for HCN production.

Direct plant growth-promoting properties in selected bacterial isolates

All antagonistic isolates were screened for phosphate solubilization activity, as well as for IAA, NH_3 and ACC deaminase production. Results are summarized in Table S4.2 and show that many of the antagonistic isolates were also able to directly promote plant growth. Three isolates, all belonging to the genus *Pseudomonas* and originating from Bolivia, inhibited growth of both *R. solani* and *P. infestans in vitro*, and showed activity for the four direct PGP properties tested. Five isolates belonging to the genera *Bacillus* and *Pedobacter*, with antagonistic properties against both *R. solani* and *P. infestans*, did not perform direct PGP, i.e. not for the activities tested. The remaining 45 isolates performed at least one, two or three of the direct PGP activities. Most of the isolates solubilized phosphate; 48 isolates (82.8%) generated clearing zones on plate assays with 2 isolates creating haloes of more than 20 mm. ACC deaminase production ranged from 20 to 310 nmol (α-ketobutyrate)/mg.h. Still, ACC deaminase production was observed less frequently than phosphate solubilization with only 32 isolates (55.2%) producing the enzyme. IAA production ranged from 60.5 to 232.64 mg/ml and was observed least frequently with merely 12 isolates (20.7%) producing the plant growth factor. None of the isolates isolated from the potato fields in Peru produced IAA. Out of 58 isolates tested for the production of NH_3, 21 (36%) were positive. The commercial rhizobacterial strain, *B. subtilis* FZB24® WG was positive for NH_3-production, and produced 26.93 mg/ml IAA and 100 nmol (α-ketobutyrate)/mg.h of ACC. The averaged size of the clearing zone for phosphate solubilization was 1.5 mm.

Direct and indirect plant growth-promotion effects on *in vitro* plantlets

Effect of bacterial isolates on plant growth in vitro

All 58 antagonistic isolates were tested in *in vitro* experiments on plantlets to evaluate their plant growth-promoting abilities. Of 58 bacterial isolates, twelve isolates significantly increased plant growth and development over the uninoculated control. An increase in plant weight was also observed in six other isolates, though not significantly. Plant weight ranged from 1.41 g to 2.04 g. *B. subtilis* FZB24® WG showed an effect on plant growth *in vitro* with a total plant fresh weight of 1.62 g. The uninoculated control microplants had a fresh weight of 1.34 g (Fig. 4.5).

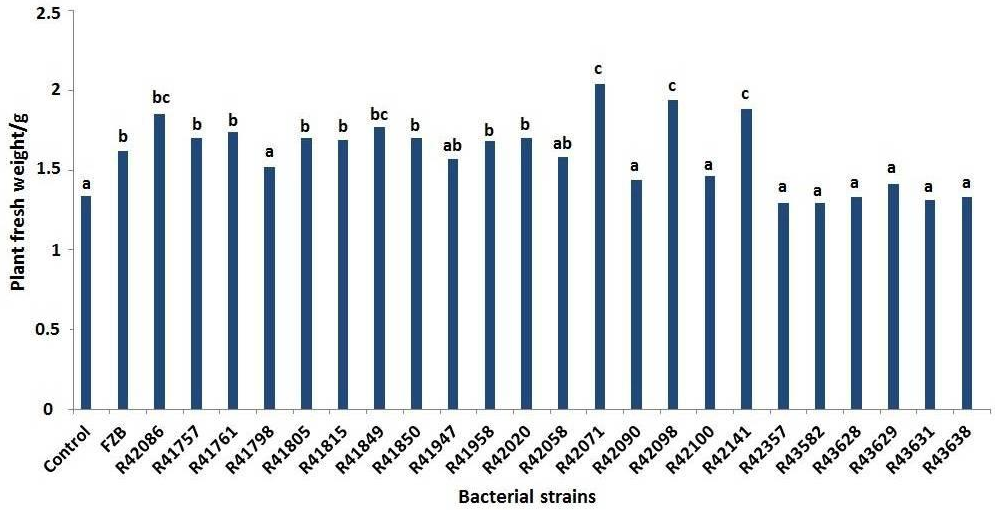


Figure 4.5 Effect of bacterial isolates on growth of potato microplants *in vitro*, four weeks after bacterization. Any two treatments sharing a common letter are not significantly different (p>0.05) using the Kruskal-Wallis test. Only the 23 best performing strains are shown. FZB refers to the commercially available bioprotectant *B. subtilis* FZB24® WG.

Effect of bacterial isolates on plant growth in vitro of plantlets inoculated with Rhizoctonia solani

The plant growth-promoting effect of bacterial isolates on plantlets infected with *R. solani* was evaluated and measured in terms of plant weight. Plant weight ranged from 1.18 g to 1.76 g. Of 58 bacterial isolates, fourteen isolates showed significant effect and improved plant growth over the uninoculated control (Fig. 4.6). An increase in plant weight was also observed in nine other isolates, though not significantly. *B. subtilis* FZB24® WG showed an effect on plant growth *in vitro* with a total plant fresh weight of 1.33 g. The uninoculated control microplants challenged with *R. solani* had a fresh weight of 1.15 g.

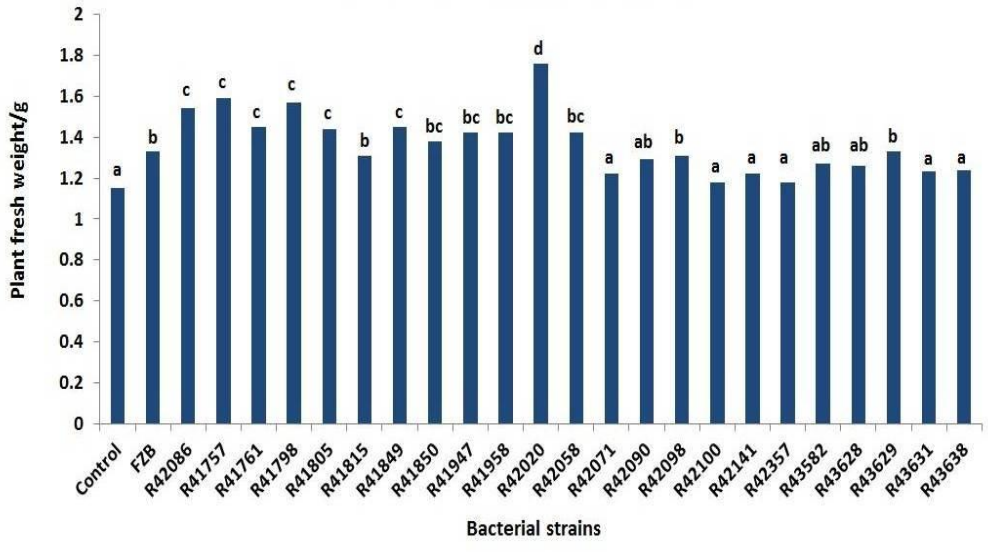


Figure 4.6 Effect of bacterial isolates on disease suppression in *in vitro* potato microplants that were inoculated with *Rhizoctonia solani*. Any two treatments sharing a common letter are not significantly different (P>0.05) using the Kruskal-Wallis test. Only the 23 best performing strains are shown. FZB refers to the commercially available bioprotectant *B. subtilis* FZB24® WG.

4.4 DISCUSSION

Biocontrol can be considered as an environmentally-friendly strategy to protect and promote plant growth. Unlike chemical pesticides, the introduction of bacterial strains as biocontrol agents into the field has shown to have only a limited impact on local microbial communities [53,54,55]. This study focused on the search for bacterial agents expected to suppress diseases caused by *R. solani* and *P. infestans* in the potato plant. Since field tests are time consuming and often associated with legal restrictions and high financial costs, a screening program was developed and tested to evaluate bacterial isolates for their PGP potential *in vitro*. The approach allowed a purposeful selection of isolates for field trials, in order to restrict the problems outlined above.

Two isolates belonging to the *Pseudomonas koreensis* and *Pseudomonas corrugata* subgroups and one isolate identified as *Enterobacter* were found to perform significantly better than the commercial control with respect to growth-promotion in potato plantlets under growth room conditions. Seven isolates, (three belonging to the *P. koreensis* subgroup, two belonging to the *P. fluorescens* subgroup and two *Bacillus* species) were found to perform significantly better than the commercial strain on *R. solani* diseased plants. These observations are promising and indicate that the isolates in question may be interesting targets for the development of new commercial biocontrol agents. However, it is known that there may be a large discrepancy between performance in the greenhouse and performance in the field due to differences in nutrient availability, exposure to secondary metabolites, and absence of competition or predation by other microorganisms under greenhouse conditions. Moreover, it is impossible to mimic edapho-climatic field conditions during growth room experiments. Therefore, field tests are essential, so isolate performance in the field can be monitored, and development of isolates into commercial biocontrol agents can be considered.

P. infestans is known for the complete field destruction that is commonly associated with infection. To date, (bio)control of this oomycete has proven to be difficult, due to its aggressive nature. The time required to complete its life cycle can be as short as three days, and as such, a single lesion may give rise to thousands of spores [56]. Currently, the best control strategy is the use of chemical fungicides in susceptible cultivars. However, toxic residues from chemical fungicides persist in the environment and may affect human health and soil organisms. *R. solani*, on the other hand, is difficult to control because of its ability to survive as sclerotia under adverse environmental conditions for many years, its capability to survive as a saprophyte, and its wide host range. In this study, we observed that the average

suppression of *P. infestans* in dual-culture assays *in vitro* was better with *Bacillus* isolates than with *Pseudomonas* isolates (66.14% vs. 56.81%). The average i*n vitro* suppression of *R. solani*, on the other hand, was better with *Pseudomonas* isolates (38.49% vs. 36.57%), but inhibition values were very similar. *R. solani* and *P. infestans* are two very different types of pathogens, and therefore we hypothesized that bacteria with antagonistic activities against both may potentially protect plants against a broader spectrum of diseases. A total of 56 isolates were found to be active in controlling both *R. solani* and *P. infestans* in dual-culture assays. The two isolates that failed to inhibit growth of *P. infestans* were both part of the same MALDI-TOF MS cluster and were both closely related to *P. nitroreducens* from the *P. aeruginosa* group (Fig. 4.4). In the growth-room experiments, no linear or monotonic relation was found between plant growth-promotion of healthy plants and plant growth of diseased plants, implying that *R. solani* affected the growth stimulation of the PGP bacteria in a non-consistent way, probably strain-dependent.

Growth inhibition of *P. infestans* in dual-culture assays was found to be significantly higher for protease producing strains than for non-producing strains ($p<0.001$). This was not the case for *R. solani*. There were no indications for a statistically significant relation between the production of cellulase, glucanase and chitinase and inhibition of *P. infestans* or *R. solani* in dual-culture assays. Likewise, no statistically significant relation was found between the production of chitinase and protease on one hand, and *in vitro* disease suppression in plantlets on the other hand (hypothesis testing for cellulase- and glucanase-production was not possible). Possible explanations could be a joined effect of direct and indirect PGP on *in vitro* potato plantlets, or that lytic enzyme production was not directly involved in pathogen suppression. Since coevolution of defense and counterdefense strategies in plants and plant pathogens respectively has already been reported, it seems plausible that something similar can occur with antagonistic bacteria, i.e. that pathogens have developed similar counterdefense strategies against harmful components excreted by bacteria. Pathogenic organisms have shown to produce enzyme inhibitors, or develop mechanisms that render them less susceptible to defense systems. *Phytophthora sojae*, for example, developed a counterdefense mechanism by secreting glucanase inhibitor proteins to overcome the action of β-1,3-endoglucanases excreted by a soybean plant [57]. Similarly, the application of different protease-inhibitors [58,59,60] and mechanisms to by-pass the activity of chitinases has been demonstrated in pathogenic organisms [61]. In this context, it should be mentioned that *P. infestans* is able to overcome the action of chitinases because of the structure of its cell-wall which consists mostly of cellulose and only little chitin [62]. Notwithstanding the fact that pathogen inhibition could not be linked with activity of lytic enzymes in general, the mechanism might still have contributed to pathogen control by acting synergistically with other mechanisms involved in biocontrol.

Surprisingly, suppression of *R. solani* in dual-culture assays was significantly higher for non-HCN producing isolates ($p<0.001$) than for HCN-producing isolates. This was confirmed with the *in vitro* disease suppression tests in microplants ($p<0.05$). Antagonistic activity against *P. infestans*, however, was significantly higher for HCN-producing isolates ($p<0.001$); but due to the absence of plant tests with *P. infestans*, we could not confirm the latter.

Many of the antagonistic isolates tested in this study were able to produce siderophores (93.1%). The high percentage of siderophore producing isolates might be explained by the low pH of the sampling locations (Table 4.1), and the herewith frequently associated low Fe^{3+}-ion concentrations in the soil [12]. Siderophore-producing bacteria not only inhibit fungal growth through competition for iron, they can also cause Induced Systemic Resistance (ISR) in plants and as such promote plant growth indirectly. ISR is a biocontrol mechanism frequently encountered in control of *P. infestans* [63,64,65]. It stimulates an immune response of the plant, rendering it less susceptible to pathogen attack. Priming of protease- [66], chitinase- [67] and glucanase-inhibitor production [68] by the plant has been reported, as well as fortification of plant cell wall strength at the sites of pathogen attack [69]. Since 93.1% of the *in vitro* biocontrol population produced siderophores, our results seem promising in this respect. As ISR is the result of plant-pathogen-antagonist interactions, plant testing was the next step in the evaluation of the siderophore-producing isolates from this study.

In this study we found that plant growth was significantly higher for NH_3-producing isolates ($p<0.01$) than for non-producing isolates. For the other PGP factors, i.e. phosphate solubilization, and IAA- and ACC deaminase production, we could not find a linear or monotonic relation with plant biomass increase.

Dereplication of the antagonistic isolates using MALDI-TOF MS (Fig. 4.3) showed a large diversity amongst the antagonistic bacteria at the species-to-strain level. Still, diversity at the genus level was mostly restricted to *Pseudomonas* and *Bacillus*. Previous studies also demonstrated the importance of both genera for biocontrol applications, as strains from *Bacillus subtilis* [63,70], *Pseudomonas putida* [63,65], *Bacillus pumilus* [64] and *Pseudomonas fluorescens* [64] have shown antagonistic activity against *P. infestans*. In a study by Brewer & Larkin [17], a *Bacillus subtilis* strain was found that proved to be consistent and effective against *R. solani* disease. Other experiments showed that *Pseudomonas* strains were able to suppress *R. solani* to levels similar to or better than fungicide treatments during field experiments [22,53]. Both genera are known to harbor plant- and human-pathogenic species, and

thus strain pathogenicity should be evaluated before applying them in agriculture. Biocontrol applications with animal pathogenic organisms have to be prevented as they pose a danger to human health, not only because of their pathogenicity, but also because they can carry multiple resistances on mobile genetic elements due to the exposure to diverse secondary metabolites produced by rhizobacteria [71]. In this study, identity of the bacteria was used to discuss strain pathogenicity. All antagonistic *Pseudomonas* isolates were classified under risk group 1 (according to the Belgian Regional Decrees [52]), and could be considered safe for application. However, amongst the *Pseudomonas* population, a number of isolates were identified as closely related to the plant pathogenic species *P. viridiflava*, *P. corrugata* and *P. marginalis.* Isolate R-41947 was ambiguously identified (Table 4.3), which is why its pathogenicity to plants could not be deduced. The remaining antagonistic strains were identified with 16S rRNA gene sequence analysis. Still, 16S rRNA gene sequencing does not go beyond genus identification, leaving the pathogenic status of these isolates unresolved. We should note that identification is a poor criterion to deduce both human- and plant-pathogenicity since pathogenicity has already shown to be strain-specific. Therefore, further research is recommended.

Sampling and isolation campaign were only performed once. Therefore, statistical correlation between field properties and yield of *in vitro* PGP bacteria was not possible. However, it would be interesting to link presence of PGP bacteria to soil properties. Therefore, for future studies it would be advisable to perform multiple isolation campaigns per sample, since this would provide information on the significance of the proportion of PGP bacteria per sample, which could in turn be linked to soil composition or treatment. Alternatively, one could consider to sample and isolate from different locations in the same area (i.e. from soils with roughly the same composition) that were treated differently (e.g. with respect to manure application).

Since field trials are time consuming and often associated with legal restrictions, a well-founded *in vitro* approach was necessary to narrow the initial population size of 585 bacterial isolates down to a manageable number. Based on the results of the *in vitro* antagonism tests, a total of 58 bacterial isolates were tested for phosphate solubilization and for the production of IAA, ACC deaminase, siderophores, fungal cell wall degrading enzymes, HCN and NH_3. All these isolates were additionally tested *in vitro* for plant growth-promotion and disease suppression on potato plantlets. Of the 58 isolates tested, 23 showed an improved response under disease pressure and/or showed an increase in plant weight over the uninoculated control. These preliminary results based on limited field sampling indicate that the Andean potato rhizosphere is a rich source of biodiversity, harboring bacterial isolates with various plant growth-promoting properties, which may have the potential to be used in the future as biocontrol

inoculants. However, field studies should be undertaken in order to confirm the effectiveness of these 23 strains under field conditions as it is known that *in vitro* and *in vivo* results can differ. Our approach should be seen in the context of the development of an efficient strategy for the large-scale screening of bacterial isolates *in vitro* that – if proven to be equally effective under field conditions – could lead to a commercialisable product in due course.

4.5 ACKNOWLEDGEMENTS

Our special thanks to Gladis Main and Evelyn Urquieta from PROINPA (Bolivia) for providing rhizosphere samples from Bolivian soils, and Andreas Oswald and Juan Sánchez Ccance from CIP (Peru) for providing and analyzing rhizosphere samples from Peruvian soils. The authors would also like to thank An Coorevits for her expertise and contribution with respect to the taxonomy of *Pseudomonas*, and Sven Hoefman for proof reading the manuscript. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement N° 227522.

4.6 REFLECTING ON THE WORK PERFORMED

Brief summary of work

The intention of this research was to obtain a collection of bacterial strains with plant growth-promoting (PGP) properties. As mentioned in the introduction (§1.1), bacteria acting as plant growth-promoters exert a number of benefits relative to agrochemicals. As the potato plant originates from the Central Andean Highlands, it was postulated that local potato plants may benefit from a mutualistic relationship with rhizosphere and/or endophytic bacteria. Therefore, bacteria were isolated from potato plants in this region, and subsequently screened for direct and indirect PGP properties. Antagonistic activity against *Rhizoctonia solani* and *Phytophthora infestans*, two pathogens being responsible for major economical crop losses annually, was investigated. The approach consisted of an initial *in vitro* screening, followed by trials on potato microplants. A number of isolates were found to exert better PGP activity relative to a commercially available *Bacillus* strain. All antagonistic isolates were identified, with extra focus on *Pseudomonas* and *Bacillus*.

In hindsight

How the research could have been improved

The focus of this research was on bacteria that were isolated from the potato rhizosphere. However, many endophytic bacteria and bacteria residing in the phyllosphere are known to exert plant beneficial properties as well. Moreover, endophytic bacteria are less prone to adverse environmental conditions compared to rhizosphere microbiota. This decreased susceptibility to edapho-climatic conditions results from their encapsulation by plant material, which acts as a protective barrier. Hence, it may be interesting to investigate their PGP properties in future research. In the introduction (§1.3.2), the importance of pre-storage of biopesticide treated seeds was discussed. Therefore, investigating the relation between pre-storage of inoculated seed tubers and increased uptake into the plant material (which is likely to improve PGP efficiency) may also be interesting to assess.

gyrB sequencing was initially performed to obtain a deeper identification of *Bacillus* isolates. However, no amplicons could be obtained with different PCR kits and programs tested. Therefore, it is most likely that the primers used for *gyrB* amplification did not target the *gyrB* genes of the *Bacillus* strains isolated in this study. The experiments performed indicate that new primers have to be designed to further characterize these strains.

In this work, biomass increase was measured by weighing the potato plants (plant fresh weight) relative to an uninoculated control. However, measuring dry weight instead might reveal more robust data. The weight increase in dry plants implies an increase in plant biomass, while the fresh weight might be due to increased water content.

In most cases, the individual PGP activities could not be correlated with promotion of plant growth in microplants. Hence, the mechanisms responsible for PGP remained unresolved. However, by generating bacterial mutants which no longer exhibit the phenotype of a given PGP property, and testing their effects on plants relative to the original strain, the mechanism of PGP could be deduced. This implies directly knocking out a gene responsible for a given PGP property and examining concomitant changes in biocontrol or direct PGP activity.

Plant pathogenicity testing

As a first indication to determine whether the PGP strains were plant pathogenic, the hypersensitivity test could have been performed on tobacco leaves (*Nicotiana tobaccum* L.) for Gram negative bacteria and in four o'clock plants (*Mirabilis jalapa* L.) for Gram positive bacteria [72,73]. Phytopathogenic bacteria produce a hypersensitive reaction in leaf mesophyll tissue (Fig. 4.7) [73]. Inoculation is performed by injecting bacterial suspensions into the leaves. However, as the syringe itself may sometimes provoke a hypersensitive reaction, an alternative strategy is to inoculate the bacterial suspension into the plant tissue under pressure. It is an easy, fast and clear assay to perform, and gives evidence of possible plant pathogenicity.

Figure 4.7 The hypersensitivity test on tobacco leaves. Isolate 1 and 2 respectively showed a negative and positive response to the test. Extracted from http://www.npdn.org/webfm_send/1230.

Future research

Shift from fundamental to applied research

Many PGP studies start with an isolation campaign, which is followed by a PGP screening step. However, as discussed during the introduction (§1.4) only few of these agents make their way to the industrial market. This is mostly caused by a lack of common interest between academic institutes and the industry. Although numerous bacterial strains have been discovered that have the potential of becoming interesting commercial agents, in most cases the research ends after establishing their efficacy in the field. Therefore, it may be interesting for future research to adjust the screening approach more towards industrial needs. This would imply a shift from fundamental research to applied research, which could be accomplished by shifting the focus from the search and selection of new isolates towards deep characterization of strain identity and properties, assessments of enhanced field efficacy (e.g. by the optimization of product formulation) and optimization of both product application onto the crop and product persistence. Optimization of product persistence may for instance involve triggering systemic resistance in plants. Information on the mode of action of biocontrol agents could be used to compose formulations and ways of delivery to maximally exploit disease suppression.

Strain characterization, as mentioned above, could include:

- a thorough identification,
- an assessment of crop specificity,
- a study on the plant growth-promoter's life-cycle,
- an evaluation of animal and plant pathogenicity,
- an analysis of produced metabolites and an assessment on the toxicity of their residues,
- a study on the impact on indigenous microbial populations,
- an analysis of plasmid encoded genes involved in pesticidal/pathogenic/toxic activity,
- an evaluation of the organism's genetic stability,
- an evaluation of antimicrobial resistance,
- an assessment of the organism's spectrum of antagonism,
- an analysis of the organism's mode of action (with respect to PGP),
- a study on optimal strain preservation,
- an evaluation of the organism's mobility and multiplication.

A thorough analysis of candidate biopesticides is required due to the ethics that involve global distribution and application (which is inherent to commercialization). This is important, as lacking knowledge on the organism's characteristics might initiate a chain reaction of uncontrollable events

after crop application. Academic institutes can thus, with their knowledge and expertise in the field of microbiology, significantly contribute to risk assessments involving biopesticides.

However, a restricted group of plant pathogenic organisms are still difficult to control by means of biopesticides. An example, which was discussed during the introduction, is *Phytophthora infestans* (§1.3.2). It speaks for itself that such cases strongly encourage the continued quest for new biocontrol agents.

Potential of plasmid based PGP

It was mentioned during the introduction (§1.4.4) that the application of bacteria with plasmid encoded PGP properties on crops may be an interesting topic to investigate in future research. Broad host-range plasmids may be transferred to indigenous bacterial populations and present a way to circumvent the inefficiency of introduced bacteria due to their susceptibilities to biotic and abiotic stresses. Indigenous populations may upon uptake of these plasmids and expression of genes involved in PGP become beneficial to plants. The major advantage is that the indigenous populations are less prone to stresses caused by edapho-climatic conditions. The approach has already been successfully applied within the field of bioremediation [74,75,76], and may be promising within the field of PGP as well.

Remark

This work is the result of a joint effort between University College Cork (UCC) and the Laboratory of Microbiology UGent; it was agreed in advance that the work would be limited to showing the biocontrol activity against *R. solani*. The biocontrol activity against *P. infestans* will be presented by UCC.

4.7 REFERENCES

1. FAO (2008) International Year of the Potato 2008, New Light on a Hidden Treasure. Food and Agriculture Organization of the United Nations. Rome, Italy.
2. Cook RJ (2000) Advances in plant health management in the twentieth century. Annual Review of Phytopathology 38: 95-116.
3. de Weert S, Bloemberg G (2006) Rhizosphere competence and the role of root colonization in biocontrol. Plant-Associated Bacteria: 317-333.
4. Wilson P, Ahvenniemi P, Lehtonen M, Kukkonen M, Rita H, et al. (2008) Biological and chemical control and their combined use to control different stages of the Rhizoctonia disease complex on potato through the growing season. Annals of Applied Biology 153: 307-320.
5. Stehmann C, De Waard MA (1996) Sensitivity of populations of Botrytis cinerea to triazoles, benomyl and vinclozolin. European Journal of Plant Pathology 102: 171-180.
6. Leroch M, Kretschmer M, Hahn M (2010) Fungicide Resistance Phenotypes of Botrytis cinerea Isolates from Commercial Vineyards in South West Germany. Journal of Phytopathology 159: 63-65.
7. Lahlali R, Hijri M (2010) Screening, identification and evaluation of potential biocontrol fungal endophytes against Rhizoctonia solani AG3 on potato plants. FEMS Microbiology Letters 311: 152-159.
8. Zamanizadeh HR, Hatami N, Aminaee MM, Rakhshandehroo F (2011) Application of biofungicides in control of damping disease off in greenhouse crops as a possible substitute to synthetic fungicides. International Journal of Environmental Science and Technology 8: 129-136.
9. Dandurishvili N, Toklikishvili N, Ovadis M, Eliashvili P, Giorgobiani N, et al. (2011) Broad-range antagonistic rhizobacteria Pseudomonas fluorescens and Serratia plymuthica suppress Agrobacterium crown gall tumours on tomato plants. Journal of Applied Microbiology 110: 341-352.
10. Hammami I, Rhouma A, Jaouadi B, Rebai A, Nesme X (2009) Optimization and biochemical characterization of a bacteriocin from a newly isolated Bacillus subtilis strain 14B for biocontrol of Agrobacterium spp. strains. Letters in Applied Microbiology 48: 253-260.
11. Jager G, Velvis H, Lamers JG, Mulder A, Roosjen J (1991) Control of Rhizoctonia solani in potato by biological, chemical and integrated measures. Potato Research 34: 269-284.
12. Lugtenberg B, Kamilova F (2009) Plant-Growth-Promoting Rhizobacteria. Annual Review of Microbiology 63: 541-556.

13. Samuels GJ (1996) Trichoderma: A review of biology and systematics of the genus. Mycological Research 100: 923-935.

14. Runno-Paurson E, Remmel T, Ojarand A, Aav A, Mand M (2010) The structure of Phytophthora infestans populations from organic and conventional crops. European Journal of Plant Pathology 128: 373-383.

15. Costanzo S, Christ BJ, Haynes KG (2004) Late blight resistance in a diploid full-sib potato family. Plant Breeding 123: 377-381.

16. Grosch R, Lottmann J, Rehn VNC, Rehn KG, Mendonca-Hagler L, et al. (2007) Analysis of antagonistic interactions between Trichoderma isolates from Brazilian weeds and the soil-borne pathogen Rhizoctonia solani. Journal of Plant Diseases and Protection 114: 167-175.

17. Brewer MT, Larkin RP (2005) Efficacy of several potential biocontrol organisms against Rhizoctonia solani on potato. Crop Protection 24: 939-950.

18. van den Boogert P, Luttikholt AJG (2004) Compatible biological and chemical control systems for Rhizoctonia solani in potato. European Journal of Plant Pathology 110: 111-118.

19. Lewis JA, Fravel DR, Lumsden RD, Shasha BS (1995) Application of biocontrol fungi in granular formulations of pregelatinized starch-flour to control damping-off diseases caused by Rhizoctonia solani. Biological Control 5: 397-404.

20. Gonzalez-Sanchez MA, Perez-Jimenez RM, Pliego C, Ramos C, de Vicente A, et al. (2010) Biocontrol bacteria selected by a direct plant protection strategy against avocado white root rot show antagonism as a prevalent trait. Journal of Applied Microbiology 109: 65-78.

21. Yangui T, Rhouma A, Triki MA, Gargouri K, Bouzid J (2008) Control of damping-off caused by Rhizoctonia solani and Fusarium solani using olive mill waste water and some of its indigenous bacterial strains. Crop Protection 27: 189-197.

22. De Curtis F, Lima G, Vitullo D, De Cicco V (2010) Biocontrol of Rhizoctonia solani and Sclerotium rolfsii on tomato by delivering antagonistic bacteria through a drip irrigation system. Crop Protection 29: 663-670.

23. Yadav S, Kaushik R, Saxena AK, Arora DK (2011) Diversity and phylogeny of plant growth-promoting bacilli from moderately acidic soil. Journal of Basic Microbiology 51: 98-106.

24. Calvo P, Ormeno-Orrillo E, Martinez-Romero E, Zuniga D (2010) Characterization of Bacillus isolates of potato rhizosphere from andean soils of Peru and their potential PGPR characteristics. Brazilian Journal of Microbiology 41: 899-906.

25. Stanier RY, Palleroni NJ, Doudorof M (1966) Aerobic Pseudomonads - A taxonomic study. J Gen Microbiol 43: 159-271.

26. Cirou A, Diallo S, Kurt C, Latour X, Faure D (2007) Growth promotion of quorum-quenching bacteria in the rhizosphere of Solanum tuberosum. Environ Microbiol 9: 1511-1522.

27. Dikin A, Sijam K (2005) Biochemical and physiological characterization of *Burkholderia cepacia* as biological control agent. Int J Agric Biol 7: 385-388.

28. Hoefman S, van der Ha D, De Vos P, Boon N, Heylen K (2012) Miniaturized extinction culturing is the preferred strategy for rapid isolation of fast-growing methane-oxidizing bacteria. Microb Biotechnol 5: 368-378.

29. Schwyn B, Neilands JB (1987) Universal chemical assay for the detection and determination of siderophores. Analytical Biochemistry 160: 47-56.

30. Ahmad F, Ahmad I, Khan MS (2008) Screening of free-living rhizospheric bacteria for their multiple plant growth promoting activities. Microbiological Research 163: 173-181.

31. Nautiyal CS (1999) An efficient microbiological growth medium for screening phosphate solubilizing microorganisms. FEMS Microbiology Letters 170: 265-270.

32. Bano N, Musarrat J (2003) Characterization of a new Pseudomonas aeruginosa strain NJ-15 as a potential biocontrol agent. Current Microbiology 46: 324-328.

33. Penrose DM, Glick BR (2003) Methods for isolating and characterizing ACC deaminase-containing plant growth-promoting rhizobacteria. Physiologia Plantarum 118: 10-15.

34. Cappuccino JC, Sherman N (1992) Negative staining. Microbiology: A Laboratory Manual. Third Edition ed. New York: Benjamin Cummings. pp. 125-179.

35. Janczura B (2007) The interaction between plant pathogenic bacteria, biocontrol bacteria, human pathogenic bacteria and plants [PhD Thesis]. Cork, Ireland: University College Cork.

36. Ghyselinck J, Van Hoorde K, Hoste B, Heylen K, De Vos P (2011) Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. Journal of Microbiological Methods 86: 327-336.

37. Heyrman J, Swings J (2001) 16S rDNA sequence analysis of bacterial isolates from biodeteriorated mural paintings in the Servilia tomb (necropolis of Carmona, Seville, Spain). Systematic and Applied Microbiology 24: 417-422.

38. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Research 33: D294-D296.

39. Wang LT, Lee FL, Tai CJ, Kasai H (2007) Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the Bacillus subtilis group. International Journal of Systematic and Evolutionary Microbiology 57: 1846-1850.

40. La Duc MT, Satomi M, Agata N, Venkateswaran K (2004) GyrB as a phylogenetic discriminator for members of the Bacillus anthracis-cereus-thuringiensis group. Journal of Microbiological Methods 56: 383-394.

41. Yamamoto S, Harayama S (1995) PCR Amplification and direct sequencing of gyrB genes with universal primers and their application to the detection and taxonomic analysis of Pseudomonas putida strains. Applied and Environmental Microbiology 61: 1104-1109.

42. Tayeb LA, Ageron E, Grimont F, Grimont PAD (2005) Molecular phylogeny of the genus Pseudomonas based on rpoB sequences and application for the identification of isolates. Research in Microbiology 156: 763-773.

43. Yamamoto S, Harayama S (1998) Phylogenetic relationships of Pseudomonas putida strains deduced from the nucleotide sequences of gyrB, ropD and 16S rRNA genes. Int J Syst Bacteriol 48: 813-819.

44. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Research 32: 1363-1371.

45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.

46. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

47. Mulet M, Lalucat J, Garcia-Valdes E (2010) DNA sequence-based analysis of the Pseudomonas species. Environ Microbiol 12: 1513-1530.

48. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27: 221-224.

49. Dawyndt P, Vancanneyt M, De Meyer H, Swings J (2005) Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources. Ieee Transactions on Knowledge and Data Engineering 17: 1111-1126.

50. Posada D (2008) jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25: 1253-1256.

51. Felsenstein J (1985) Confidence limits on phylogenies - an approach using the bootstrap. Evolution 39: 783-791.

52. (WIV-ISP-SBB) BaBU (2005) Belgian classifications for micro-organisms based on their biological risks - Definitions. Brussels.

53. Scherwinski K, Grosch R, Berg G (2008) Effect of bacterial antagonists on lettuce: active biocontrol of Rhizoctonia solani and negligible, short-term effects on nontarget microorganisms. FEMS Microbiology Ecology 64: 106-116.

54. Bankhead SB, Landa BB, Lutton E, Weller DM, Gardener BBM (2004) Minimal changes in rhizobacterial population structure following root colonization by wild type and transgenic biocontrol strains. FEMS Microbiology Ecology 49: 307-318.

55. Viebahn M, Glandorf DCM, Ouwens TWM, Smit E, Leeflang P, et al. (2003) Repeated introduction of genetically modified Pseudomonas putida WCS358r without intensified effects on the indigenous microflora of field-grown wheat. Applied and Environmental Microbiology 69: 3110-3118.

56. Mizubuti ESG, Júnior VL, Forbes GA (2007) Management of Late Blight with Alternative Products. Pest Tech 1: 106-116.

57. Rose JKC, Ham KS, Darvill AG, Albersheim P (2002) Molecular cloning and characterization of glucanase inhibitor proteins: Coevolution of a counterdefense mechanism by plant pathogens. Plant Cell 14: 1329-1345.

58. Tian MY, Benedetti B, Kamoun S (2005) A second kazal-like protease inhibitor from Phytophthora infestans inhibits and interacts with the apoplastic pathogenesis-related protease P69B of tomato. Plant Physiology 138: 1785-1793.

59. Tian MY, Win J, Song J, van der Hoorn R, van der Knaap E, et al. (2007) A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease. Plant Physiology 143: 364-377.

60. Tian MY, Huitema E, da Cunha L, Torto-Alalibo T, Kamoun S (2004) A Kazal-like extracellular serine protease inhibitor from Phytophthora infestans targets the tomato pathogenesis-related protease P69B. Journal of Biological Chemistry 279: 26370-26377.

61. Siegrist J, Kauss H (1990) Chitin deacetylase in cucumber leaves infected by Colletotrichum lagenarium. Physiological and Molecular Plant Pathology 36: 267-275.

62. Kamoun S (2003) Molecular genetics of pathogenic oomycetes. Eukaryotic Cell 2: 191-199.

63. Daayf F, Adam L, Fernando WGD (2003) Comparative screening of bacteria for biological control of potato late blight (strain US-8), using in-vitro, detached-leaves, and whole-plant testing systems. Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie 25: 276-284.

64. Yan ZN, Reddy MS, Ryu CM, McInroy JA, Wilson M, et al. (2002) Induced systemic protection against tomato late blight elicited by plant growth-promoting rhizobacteria. Phytopathology 92: 1329-1333.

65. Kim H-J, Jeun Y-C (2006) Resistance induction and enhanced tuber production by pre-inoculation with bacterial strains in potato plants against *Phytophthora infestans*. Mycobiology 34: 67-72.

66. Petti C, Khan M, Doohan F (2010) Lipid transfer proteins and protease inhibitors as key factors in the priming of barley responses to Fusarium head blight disease by a biocontrol strain of Pseudomonas fluorescens. Functional & Integrative Genomics 10: 619-627.

67. Compant S, Duffy B, Nowak J, Clement C, Barka EA (2005) Use of plant growth-promoting bacteria for biocontrol of plant diseases: Principles, mechanisms of action, and future prospects. Applied and Environmental Microbiology 71: 4951-4959.

68. Tonelli ML, Furlan A, Taurian T, Castro S, Fabra A (2011) Peanut priming induced by biocontrol agents. Physiological and Molecular Plant Pathology 75: 100-105.

69. Benhamou N, Kloepper JW, Tuzun S (1998) Induction of resistance against Fusarium wilt of tomato by combination of chitosan with an endophytic bacterial strain: ultrastructure and cytochemistry of the host response. Planta 204: 153-168.

70. Stephan D, Schmitt A, Carvalho SM, Seddon B, Koch E (2005) Evaluation of biocontrol preparations and plant extracts for the control of Phytophthora infestans on potato leaves. European Journal of Plant Pathology 112: 235-246.

71. Berg G, Eberl L, Hartmann A (2005) The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. Environ Microbiol 7: 1673-1685.

72. Braun-Kiewnick A, Sands DC (2001) Gram negative bacteria. In: Schaad NW, Jones JB, Chun W, editors. Laboratory guide for identification of plant pathogenic bacteria. 3 ed. St. Paul, MN: APS Press. pp. 84-120.

73. Urnesha S, Richardson PA, Kong P, Hong CX (2008) A novel indicator plant to test the hypersensitivity of phytopathogenic bacteria. Journal of Microbiological Methods 72: 95-97.

74. Zhang Q, Wang BC, Cao ZY, Yu YL (2012) Plasmid-mediated bioaugmentation for the degradation of chlorpyrifos in soil. Journal of Hazardous Materials 221: 178-184.

75. Wang YJ, Li H, Zhao W, He XL, Chen J, et al. (2010) Induction of toluene degradation and growth promotion in corn and wheat by horizontal gene transfer within endophytic bacteria. Soil Biology & Biochemistry 42: 1051-1057.

76. Dejonghe W, Goris J, El Fantroussi S, Hofte M, De Vos P, et al. (2000) Effect of dissemination of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation plasmids on 2,4-D degradation and on bacterial community structure in two different soil horizons. Applied and Environmental Microbiology 66: 3297-3304.

Author Contributions:

JG and SP designed the experiment and wrote the manuscript. SP constructed the representative subtree, made Robinson-Foulds calculations and calculated primer coverage rates. The remainder of the experiments, i.e. sequence library and ML tree constructions, branch length based comparisons of phylogenetic trees (PC, vCEED), pairwise distance based comparisons, OTU richness calculations and taxonomic assignments, were performed by JG. KH, AS and PDV proofread the manuscript.

CHAPTER 5

THE EFFECT OF PRIMER CHOICE ON THE OUTCOME OF NEXT GENERATION SEQUENCING STUDIES

Redrafted from: Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A. & De Vos, P. (2013). The effect of primer choice on the outcome of Next Generation Sequencing studies. PLOS ONE, accepted with minor revisions

SUMMARY

Different regions of the bacterial 16S rRNA gene evolve at different evolutionary rates. The scientific outcome of next generation sequencing studies therefore alters with the gene region sequenced. We wanted to gain insight in the impact of primer choice on the outcome of next generation sequencing efforts. All the unknowns associated with sequencing data, i.e. primer coverage rate, phylogeny, OTU-richness and taxonomic assignment, were therefore implemented in one study for ten well established universal primers (338f/r, 518f/r, 799f/r, 926f/r and 1062f/r) targeting dispersed regions of the bacterial 16S rRNA gene. All analyses were performed on nearly full-length and *in silico* generated short read sequence libraries containing 1175 sequences that were carefully chosen as to present a representative substitute of the SILVA SSU database. The 518f and 799r primers, targeting the V4 region of the 16S rRNA gene, were found to be particularly suited for next generation sequencing studies, while the primer 1062r, targeting V6, seemed to be least reliable. Our results will assist scientists in considering whether the best option for their study is to select the most informative primer, or the primer that excludes interferences by host-organelle DNA. The methodology followed can be extrapolated to other primers, allowing their evaluation prior to the experiment.

5.1 INTRODUCTION

Next generation sequencing (NGS) platforms have allowed microbiologists to gain new insights in microbial ecology [1]. Through high-throughput amplicon sequencing of specific target genes such as the 16S rRNA gene, researchers have been enabled to get a glimpse of microbial communities in environments of interest [2]. However, a number of steps, which include sampling, DNA extraction and PCR, may hamper the objective of obtaining results truly representing the environment studied [3]. One essential aspect demanding careful consideration is primer choice. Particular genes, such as the 16S rRNA gene in bacteria, contain regions that have evolved at different evolutionary rates, and as such the scientific outcome may vary with the gene region sequenced [4,5,6,7]. The 16S rRNA gene consists of fast evolving, structural parts that are defined as variable regions V1 – V9, and that allow the identification of bacteria. The term 'hypervariable region' was designated to those regions of the 16S rRNA gene of which the evolutionary rate exceeds the mean evolutionary rate of all nucleotides in the molecule [8]. However, there are clear differences in base heterogeneity and phylogenetic discriminatory power between the different regions [9,10]. The important issue of primer universality has been discussed previously [11,12,13]. The 16S rRNA gene contains several conserved stretches that are shared amongst almost all known bacteria [13,14], and that are used to develop universal primers. However, the coverage rates of such primers differ with the location of their target in the 16S rRNA gene. Online matching tools such as SILVA Test Probe [12] and RDP probe match [15] have been specifically developed to address this problem. Furthermore, Berry et al. [16] have reported biases introduced with barcode-tagging of primers that translate into less reproducible data sets, while Wu and colleagues [17] extensively mentioned the problems of preferential amplification.

The analysis of bacterial communities associated with hosts, such as plants and weeds, may be hampered by the interference of host organelles. In order to efficiently extract the bacterial DNA pool from a host matrix, bacteria ought to be released from the host matrix prior to, or during DNA extraction. This often requires a vigorous DNA extraction, which will also release organelle DNA. As a consequence, microbial community studies that are based on next generation amplicon sequencing of the 16S rRNA gene may experience problems due to the undesired co-amplification of mitochondrial 18S and chloroplast 16S rRNA. As plant organelles sometimes outnumber bacterial cells, it is desirable to specifically amplify prokaryotic genes. The 799 primer [18] could be of special interest for studying microbial communities obtained from host matrices. The 799 primer is known to allow the exclusion of host derived chloroplast sequences by targeting the bacterial 16S rRNA gene, while failing to target the

gene in chloroplasts [18]. Moreover, if used in the forward direction, and in combination with a well-chosen reverse primer, a mitochondrial amplicon will be generated that is larger than the corresponding bacterial amplicon [18], which allows their separation by gel electrophoresis.

Several studies have focused on coverage rates of primers targeting different regions of the bacterial 16S rRNA gene [12,13], while others have analyzed the phylogenetic information that is contained within short reads [10]. Schloss et al. [4] analyzed the effects of different data processing approaches on alpha- and beta-diversity for different regions of the bacterial 16S rRNA gene, while others studied the results of taxonomic assignments with reads generated from different 16S rRNA gene targeting primers [5,6,19,20]. However, uniformity between each of these studies, which provided very useful insights into the advantages and limitations of the NGS approach, is lacking. Therefore, it can be difficult to e.g. be aware of the phylogenetic information that is contained within reads that were generated from a primer with a well documented coverage rate, and what the effect of its use will be on OTU richness and taxonomic assignment. To account for this shortcoming, we implemented the unknowns that are associated with primer choice, i.e. primer coverage rate, OTU-richness, taxonomic assignment, and phylogeny, in one study for ten different primers, including the 799 primer, targeting dispersed regions of the bacterial 16S rRNA gene. Our motivation was to get a clear picture of the intrinsic information loss that is associated with sequencing of short reads compared to their parent nearly full-length sequences covering the V1 – V9 variable regions. The results of this study will allow researchers to select primers based on the objectives of their research, and will assist them with the interpretation of their results. Moreover, the approach followed will allow scientists to evaluate new primers before using them in NGS based experiments.

5.2 MATERIALS AND METHODS

Primer selection and coverage rate

For this study, we chose well established universal 16S rRNA gene primers (Table 5.1), each of which target conserved stretches between the hypervariable regions V1 – V9 of the 16S rRNA gene that were described by Van de Peer et al. [8]. Primer coverage rates were calculated both at the domain and phylum level by using the tool "SILVA Test Probe" [12]. SILVA [21] provides chimera checked, aligned sequences which form today´s standard SSU rRNA database. The primers and their reverse complements were matched against the non redundant SILVA SSU Ref dataset 113 [22], allowing no mismatches.

Selection of sequences and generation of the nearly full-length library

To obtain a practicable but representative subset of the complete SILVA SSU reference dataset, nearly full-length (NFL) sequences were selected from the non redundant SILVA SSU reference database 102 [21]. The database in question contains ~ 262 000 sequences that were chimera and quality checked, and redundancy filtered with the UCLUST tool [23]. In the frame of 'The All Species Living Tree Project' (LTP) [24,25], a maximum-likelihood tree was constructed with RaxML [26] containing all UCLUST quality checked sequences. This allowed the display of the whole database in a tree format in the ARB software package [27]. We used this tree as a baseline for sequence selection, and thus for the construction of the practicable but representative sequence subset. In ARB, all eukaryotic and archaeal entries were removed, and the remaining bacterial tree was screened for phylogenetically distinct bacterial clades up to the species level. Within each clade all except the entry containing the longest sequence were removed. For clades that only contained sequences from uncharacterized cultivation-independent sequence data, one full length, high quality 16S rRNA sequence entry was kept. The resulting tree contained 1186 16S rRNA gene sequences instead of the initial 262 000, while the original SSU Ref 102 LTP tree's branching pattern and phylogenetic distances were conserved. All 1186 sequences were exported in a fasta file. The end-points of all sequences were trimmed with the MEGA 5 software [28] as to obtain maximum overlap between the sequences. Subsequently, the library was analyzed in RAxML v7.3.2 to exclude identical sequences and gap-only characters in the alignment. As a consequence, the dataset was further reduced to 1175 sequences. All sequences of the NFL library contained the V1-V9 variable regions of the bacterial 16S rRNA gene.

Generation of short read libraries

Ten short read (SR) libraries were constructed in MEGA 5 [28]; one library for each of the primers analyzed (Table 5.1). To do so, the NFL library was used as a seed by first locating the respective primers in the NFL library, and then trimming the sequences 280 bp upstream and downstream of the start of each primer (conform to unidirectional sequencing). The length of 280 bp for our SR libraries was based on suggestions made by Schloss and Quince. Although 454 amplicon sequencers generate reads with an average length of 400-700 bp, most quality checked sequences don´t exceed 280 bp due to quality assignments by leading packages Mothur [29] and QIIME [30]. Conversely, other NGS platforms, such as the Illumina sequencers, are now capable of generating longer reads. Therefore, the length of 280 bp, which was applied in this study, makes the results obtained applicable for a variety of NGS sequencers. After trimming primer sequences, libraries were ready for downstream analyses.

Generation of short read and full length 16S rRNA gene trees

Each of the libraries was imported in RAxML v7.3.5 and a Maximum Likelihood (ML) search was performed with the gamma parameter [31], in combination with rapid bootstrapping, which uses the CAT approximation [32]. The substitution model used was GTR. Bootstrapping was performed with 500 replicates. The command line used for the tree search was the following: raxmlHPC-PHTREADS-SSE3 –T <number of processors> -fa -m GTRGAMMA -N <replicates> -x <seed1> -p <seed2> -s <filename> -n <outputfile>. The best scoring ML tree was exported in newick format. Patristic distances, which are defined as the sum of the branch-lengths in the shortest path connecting a pair of taxa in a phylogenetic tree, were calculated for all pairs of taxa within the tree [10].

Branch length based comparison of phylogenetic trees

The Pearson correlation between branch lengths of a pair of phylogenetic trees

In order to calculate the correlation between two ML trees, patristic distances between corresponding pairs of sequences in each of the trees were made into a tuple, which formed the coordinate of a point in a plot. This was performed for all pairs of sequences in each of the trees being compared. For each plot, the Pearson correlation was calculated and used as one measure to study the phylogenetic relation between two regions of the bacterial 16S rRNA gene. In order to present the data in a graph, branch-length distances were normalized to a maximum value of one and were ordered for the NFL tree. For each NFL distance interval of 0.01 we calculated the averages and standard deviations of corresponding patristic distances in the SR tree. Averaged NFL distances (over a 0.01 distance interval) and corresponding averaged SR distances were then plotted in a graph, and the standard deviations on the averaged SR distances were superimposed (as error bars) on the chart.

The degree of fit between a pair of phylogenetic trees using the vCEED approach

Patristic distance matrices were generated from the ML trees by using the PHYLOCOM software [33]. Distance matrices for each of the trees under comparison were used as inputs for the vCEED script that was written in Matlab by Choi and colleagues [34]. Using a distance matrix as an input, each sequence is mapped to a Euclidean space via metric multidimensional scaling (MDS). This produces a multidimensional plot in which each point represents one sequence (or taxon) within the phylogenetic tree (e.g. the NFL tree). The same procedure is then repeated for a second distance matrix, representing the phylogenetic tree we want to compare to the first one (e.g. the SR tree). Subsequently, one embedded point pattern is superimposed on the other and the degree of fit is calculated. The degree of fit is expressed by the weighted Root Mean Square Deviation (wRMSD). A decreasing wRMSD indicates an increasing degree of fit, and thus a higher similarity between trees. In addition, regions of high similarity as well as incongruent regions between the trees can be identified.

Topology based comparison of phylogenetic trees

The Robinson Foulds (RF) metric [35] was used to compare topologies of a pair of unrooted phylogenetic trees. It counts the number of clades that occur in one tree but not in the other. The lower the RF value, the more similar both trees are with respect to tree topology. The weighted Robinson Foulds (WRF) metric, however, takes into account the bootstrap support values of the clades instead of looking at their presence or absence only [36]. A clade with a bootstrap value of 0.6 counts 0.6 instead of 1, and as such the WRF metric penalizes less for lower supported bifurcations. Similarly, another metric was calculated that was derived from the WRF metric, and which we will refer to as WRF2. WRF2 not only includes the support value on each unique bipartition, but additionally includes the differing bootstrap support values of shared bipartitions. This provides additional information on the cladistic distance between a pair of trees. For this study, the RF and both WRF distances were calculated using RAxML v7.4.2.Gui [26,37].

The Pearson correlation between pairwise distances in a pair of sequence libraries and the effect on OTU richness

Pairwise distances were calculated between all pairs of sequences in each sequence library with RAxML v7.3.2. Pairwise distances between corresponding pairs of sequences in each of two libraries under comparison were made into a tuple, which then formed the coordinate of a point in a plot. For each plot, the Pearson Correlation was calculated. To present the data graphically, the same binning step was followed as for the branch-length distance correlation plots. To study the effect of pairwise distances altering with the region of the 16S rRNA gene sequenced on α-diversity, OTU richness was calculated for each SR library and for the NFL library. OTU richness was calculated using the Mothur v1.27.0 software [29] with the average neighbor clustering algorithm (i.e. UPGMA) and a hard cutoff [38]. Results obtained from the SR libraries were compared with results obtained from the NFL library by calculating the ratio of the number of OTU's obtained with each SR library to the number of OTU's obtained with the NFL library.

Taxonomic assignment of sequences

In silico generated reads and the NFL sequences were assigned taxonomically using the Mothur v1.27.0 software, using the classify.seqs() tool. The RDP v9 training set [39] was used as a reference database. The bootstrap cutoff for assigning a sequence to a specific taxon was set at 80%.

5.3 RESULTS

Primer Coverage Rate

With a total coverage rate of 82.2%, primer 518f/r showed the highest coverage amongst all primers investigated. The high value obtained was not only due to a high coverage within the domain Bacteria, but also due to a high coverage of eukaryotic 16S rRNA sequences (Table 5.1). This non-specificity, however, should be taken into consideration for bacterial community sequencing in many habitats, as it could cause contamination with eukaryotic 16S rRNA gene sequences. Primer 799f/r covered 78.5% of bacterial and 71.7% of archaeal sequences in the database. Primers 338f/r, 926f/r and 1062f/r showed almost no homology with sequences within the domains Eukarya and Archaea, which makes them almost exclusive for Bacteria.

Primer[a]	Sequence (5´->3´)	*E. coli* Position	Coverage (%)[b]				Reference
			Eukarya	Bacteria	Archaea	Total	
338r	GCTGCCTCCCGTAGGAGT	355-338	-	88,4	-	75,6	Suzuki (1996) [40]
518r	ATTACCGCGGCTGCTGG	542-518	88,3	85,1	0,4	82,2	Muyzer (1993) [41]
799f	AACMGGATTAGATACCCKG	781-799	-	78,5	71,7	69,4	Chelius (2001) [18]
926f	AACTCAAAGGAATTGACGG	908-926	-	77,4	-	65,7	Lane (1991) [42]
1062r	CTCACRRCACGAGCTGAC	1081-1064	-	89,5	2,4	77,1	Allen (2005) [43]

Table 5.1 Primer sequences and their domain specific coverage rates.
[a], Primer names according to first description; primer names indicate both position and direction
[b], According to SILVA SSU Ref 113 non redundant database

Because total coverage rates bias towards large bacterial phyla such as the *Proteobacteria* and *Firmicutes*, non-coverage rates were calculated per phylum (Fig. 5.1). Non-coverage rates reflect the percentage of sequences that will not be covered by the primer investigated. Of the better represented phyla in the database, primer 799f/r was found to discriminate against almost all sequences of *Cyanobacteria*, against about 80% of *Planctomycete*s and *Verrucomicrobia* and against more than 50% of *Acidobacteria*. As chloroplasts are classified within the phylum *Cyanobacteria*, primer 799f/r can be considered to be of special interest for host-associated bacterial community studies. The lowest total coverage rate that was observed for the 926f/r primer (Table 5.1) seemed to be attributed to a low coverage of proteobacterial 16S rRNA gene sequences (Fig. 5.1). The highest total coverage rate in Bacteria was attributed to primer 1062f/r; its non-coverage rate did not exceed 40% in any of the phyla studied (Fig. 5.1). The non-coverage rates of primers 338f/r and 518f/r were generally low for the best

represented phyla in the database. However, they were found to discriminate against specific taxonomic groups such as the *Verrucomicrobia* (Fig. 5.1).

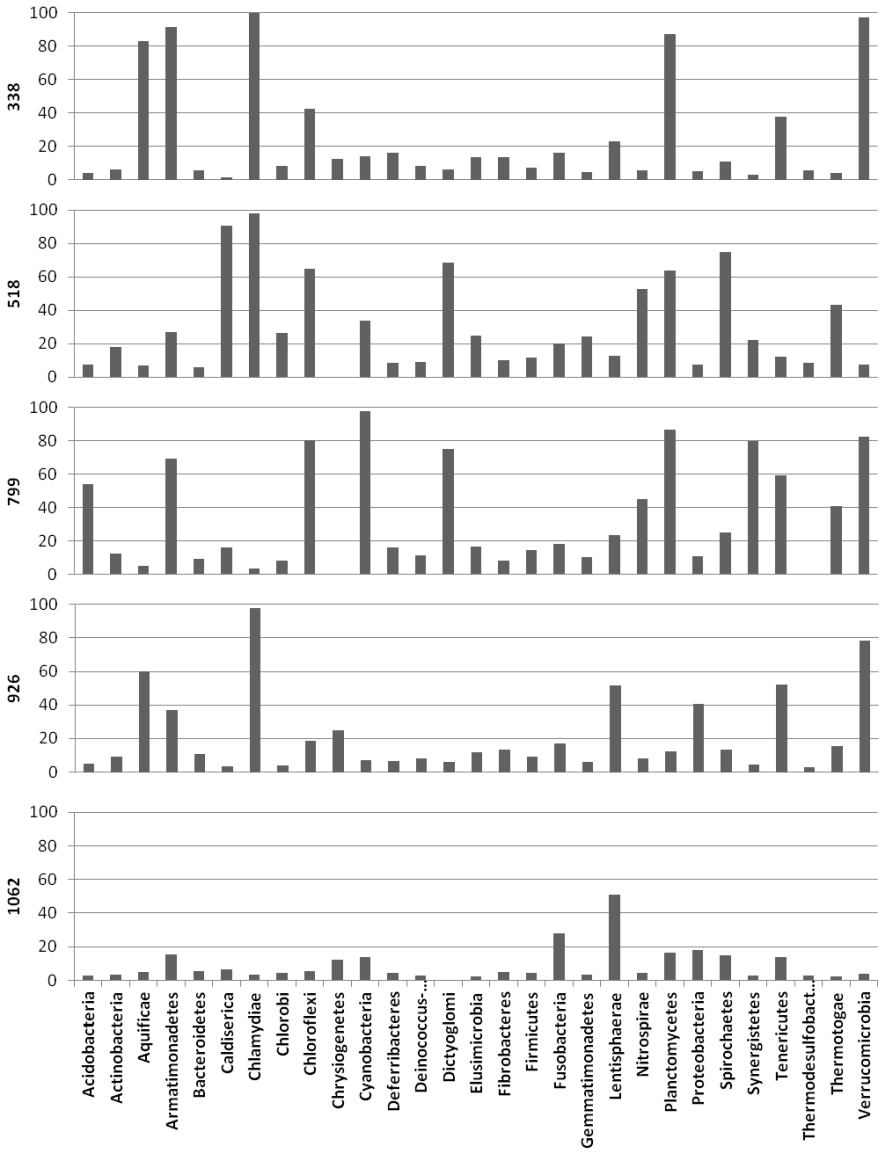


Figure 5.1 Percentage of non-coverage rates in 29 classified bacterial Phyla for the primers analyzed in this study. Non coverage rates were calculated based on the coverage values in the SILVA ssu Ref 113 non redundant database, using SILVA Test Probe with zero mismatches allowed.

Phylogenetic content of short reads

Jeraldo et al. [10] reasoned that the branch length based correlation between trees generated from different tree searches on the same library can be used as a measure for the amount of phylogenetic information contained in a SR. A Pearson Correlation close to one indicates a positive association between branch lengths in trees generated from different tree searches on the same library (i.e. SR(1) and SR(2)), and little deviation from the line of best fit connecting the data points. A high Pearson Correlation will thus be obtained if sequences that are found to be closely together in the SR(1) tree are also found to be closely together in the SR(2) tree. Correlation values close to zero indicate the opposite, i.e. a scattering or deviation from a straight line behavior. The latter case implies that sequences positioned closely together in the SR(1) tree are not necessarily found to be closely together in the SR(2) tree, meaning that the tree generated has high uncertainty with respect to branch lengths. Low correlations thus indicate that the information within the reads is too limited to calculate unequivocal branch lengths for a given sequence library, and as such is insufficient to solve the ML problem. To gain more insight in this matter, we calculated the correlation between two trees generated from the same library for the different libraries investigated. Since full length 16S rRNA gene sequences are the benchmark for constructing phylogenies [44], it was expected that the Pearson Correlation between different tree searches for NFL sequences would be the maximum correlation possible. However, the correlation between two tree searches from one NFL sequence library was 0.93 (Table S5.1) instead of the theoretically expected value of 1.00. This can be explained by the fact that ML trees are calculated using a heuristic method, and therefore there is no guarantee that the tree calculated best represents the sequence data, and thus is the best tree. Representation of sequence data in a phylogenetic tree which is based on heuristics is prone to uncertainties in tree structure, and therefore different tree searches for one and the same sequence library will unavoidably lead to differences in tree structure to some extent. Moreover, the random order in which sequences are added to a maximum parsimony starting tree in RAxML [45] is likely to generate several different starting trees for every new analysis that is started [46], again having implications for the “best tree”. Regardless, as the construction of ML trees from sequence data can only be as good as the phylogenetic information which it is generated from (i.e. the sequence data), we expect that the correlation between trees from different tree searches will be higher as more information is contained within the read. Surprisingly, a higher correlation was observed between two trees that were generated from the same 518f library (i.e. 0.97 (Table S5.1)). However, as explained above, ML is an approximation and there is no guarantee that the NFL tree calculated best represents the sequence data. As such, the possibility exists that the true NFL tree was ‘overlooked’. It is possible, although difficult to tell, that increasing the number of NFL starting trees during the ML calculation process would have resulted in higher correlations between trees obtained

from different tree searches. The search for the best-known ML tree would in that case have started at different points in the vast search space and would have followed different trajectories, thus increasing chances of obtaining ML trees with higher likelihood values. Another possibility is that the initial sampling (two trees on the NFL alignment and two trees on the SR alignment) was too small, and that the higher correlation obtained for 518f reads happened by chance. This considered, we decided to generate five trees for all SR libraries investigated, and three for the NFL library. Table S5.1 shows a correlation of 0.98 between NFL(1) and NFL(3), which shows that our assumption was true and also confirms the upper-limit statement made earlier. Table S5.1 also shows that the high correlation values were maintained with a higher number of tree searches for the 518f library. Still, the upper-limit of 0.98 was not reached; correlation values ranged from 0.93 to 0.97 (coefficient of variation 0.015). This clearly shows that any tree constructed from the 518f reads is very robust with respect to patristic distances. Similarly, high correlations were obtained and maintained for different tree searches from 799r reads (coefficient of variation 0.019) (Table S5.1). These results indicate that any tree constructed from libraries targeting the V4 region of the bacterial 16S rRNA gene (i.e. 799r and 518f) is very stable with respect to branch-lengths. The V6-targeting 1062r read library on the other hand, showed the lowest correlation between trees generated from different tree searches, indicating its low reproducibility and thus phylogenetic content.

Since comparing phylogenetic trees based on correlations between patristic distances is known to have its weaknesses [34], we strengthened our study by additionally applying the recently published vCEED approach [34]. A statistically significant negative correlation was found between results obtained with the vCEED approach (in terms of wRMSD), and those obtained with the Pearson Correlation method for comparisons of trees obtained from different tree searches on the same library ($R= -0.93$, $p<0.0005$). This indicates that the vCEED approach confirmed the results obtained with the Pearson Correlation approach. Similar to the Pearson Correlation approach, the highest degree of fit was found for NFL(1) vs NFL(3). Amongst the SR libraries, the highest degree of fit was observed for the 518f library, followed by the 799r and 1062f libraries. Supporting the observations obtained with the Pearson Correlation approach, the averaged wRMSD and the corresponding coefficients of variation were slightly lower for 799r reads than for 1062f reads (i.e. 0.0103 versus 0.0106, with coefficients of variation being 0.113 and 0.191 respectively) indicating its higher phylogenetic content. The V6 targeting 1062r read library again showed the largest variation among tree searches, which reflects its rather low phylogenetic content.

Conservation of tree topology with different tree searches

To answer the question whether differences in branch length conservation amongst the different SR libraries investigated can be extrapolated to conservation of the tree's branching pattern, differences in topologies between trees generated from different tree searches on each SR library were calculated. Unweighted RF distance calculations showed that the 518f SR trees had the most consistent tree topology, followed by 799r and 1062f reads. Still, the RF distance was around two times higher than the RF distance between trees from different tree searches on the NFL library. The most variant tree topology was calculated for trees generated from the 1062r library, which confirmed the results obtained with patristic distances (Table S5.1).

The difference between RF and WRF values for a given tree comparison provides insight in the nature of differences in tree topology [10]. If the WRF value approximates the RF value, differences mainly occur in high-supported sub-trees, while a WRF value that is much lower than the corresponding RF value indicates that differences mainly occur on less supported sub-trees. Comparing tree topology conservation of the 518f and 799r tree sets with tree topology conservation of the 1062f tree set indicated that the topologies of the former were more conserved than topologies between trees generated from the 1062f library (Table S5.1). However, if penalized for the lower supported clades, trees generated from the 1062f library were more consistent with respect to tree topology conservation. Therefore, differences between the trees generated from different trees searches on the 518f and 799r SR libraries seem to occur on better supported branches than for trees that were generated from the 1062f SR library.

Do SR reflect NFL phylogeny?

With respect to patristic distances

The Pearson Correlation between corresponding patristic distances in trees generated from NFL and SR libraries was used to investigate if a read can be used to infer 16S rRNA gene based phylogeny. The correlation plots (Fig. 5.2) show that with the exception of the 1062r read library, there seemed to be no significant deviation from a straight line behavior, which is reflected by the correlation values given in Table 5.2. This indicates that all reads, with the exception of 1062r, can be used to study 16S rRNA gene based phylogeny. However, in most cases a scattering is observed for large NFL patristic distances, indicating a rather poor association between distant sequences in the SR and NFL trees. Table 5.2 shows that correlations between SR and NFL trees fluctuate with different tree searches. These fluctuations are the combined effect of differences occurring in branch lengths between trees generated from different

tree searches on NFL and SR libraries, which, as mentioned in the previous paragraph, can be related to the phylogenetic content of the reads.

A strong statistically significant negative correlation ($R= -0.93$, $p<0.0005$) indicated that the vCEED approach confirmed the results obtained with the Pearson Correlation method for comparisons between SR and NFL trees. The highest degree of fit was obtained for the 518f and 1062f libraries, closely followed by the 799r library.

Tree comparison	Variable region	PC patristic[a]	wRMSD[b]	Tree Comparison	Variable region	PC patristic[a]	wRMSD[b]
338f (1) vs NFL (1)		0.65	0.0194	799r (1) vs NFL (1)		0.774	0.0151
338f (2) vs NFL (1)		0.69	0.0189	799r (2) vs NFL (1)		0.781	0.0146
338f (1) vs NFL (2)		0.685	0.0182	799r (1) vs NFL (2)		0.779	0.0162
338f (2) vs NFL (2)		0.729	0.0172	799r (2) vs NFL (2)		0.782	0.0158
338f (1) vs NFL (3)		0.655	0.0195	799r (1) vs NFL (3)		0.796	0.015
338f (2) vs NFL (3)	V3	0.714	0.0187	799r (2) vs NFL (3)	V4	0.812	0.0144
338r (1) vs NFL (1)		0.778	0.0163	926f (1) vs NFL (1)		0.724	0.0178
338r (2) vs NFL (1)		0.749	0.0152	926f (2) vs NFL (1)		0.625	0.0191
338r (1) vs NFL (2)		0.731	0.0189	926f (1) vs NFL (2)		0.75	0.0183
338r (2) vs NFL (2)		0.703	0.018	926f (2) vs NFL (2)		0.657	0.0197
338r (1) vs NFL (3)		0.791	0.0164	926f (1) vs NFL (3)		0.744	0.0177
338r (2) vs NFL (3)	V2	0.771	0.0151	926f (2) vs NFL (3)	V6	0.654	0.019
518f (1) vs NFL (1)		0.783	0.0142	926r (1) vs NFL (1)		0.738	0.0166
518f (2) vs NFL (1)		0.799	0.0142	926r (2) vs NFL (1)		0.725	0.0162
518f (1) vs NFL (2)		0.807	0.0143	926r (1) vs NFL (2)		0.718	0.0174
518f (2) vs NFL (2)		0.822	0.0144	926r (2) vs NFL (2)		0.715	0.0174
518f (1) vs NFL (3)		0.805	0.0139	926r (1) vs NFL (3)		0.726	0.0172
518f (2) vs NFL (3)	V4	0.81	0.0142	926r (2) vs NFL (3)	V5	0.751	0.016
518r (1) vs NFL (1)		0.695	0.0177	1062f (1) vs NFL (1)		0.833	0.0138
518r (2) vs NFL (1)		0.698	0.0174	1062f (2) vs NFL (1)		0.825	0.0139
518r (1) vs NFL (2)		0.735	0.0167	1062f (1) vs NFL (2)		0.809	0.0139
518r (2) vs NFL (2)		0.69	0.0177	1062f (2) vs NFL (2)		0.81	0.0141
518r (1) vs NFL (3)		0.704	0.0179	1062f (1) vs NFL (3)		0.827	0.0141
518r (2) vs NFL (3)	V3	0.69	0.0178	1062f (2) vs NFL (3)	V7&8	0.82	0.0141
799f (1) vs NFL (1)		0.774	0.0152	1062r (1) vs NFL (1)		0.603	0.0196
799f (2) vs NFL (1)		0.727	0.0162	1062r (2) vs NFL (1)		0.691	0.0172
799f (1) vs NFL (2)		0.746	0.0178	1062r (1) vs NFL (2)		0.596	0.0213
799f (2) vs NFL (2)		0.697	0.0187	1062r (2) vs NFL (2)		0.711	0.0184
799f (1) vs NFL (3)		0.777	0.0154	1062r (1) vs NFL (3)		0.642	0.0192
799f (2) vs NFL (3)	V5	0.749	0.0161	1062r (2) vs NFL (3)	V6	0.741	0.0167

Table 5.2 Overview of the research parameters that were applied in comparisons of short read and nearly full-length sequence libraries – individual values for each of the tree comparisons.

[a], PC= Pearson Correlation

[b], wRMSD= Weighted Root Mean Square Deviation

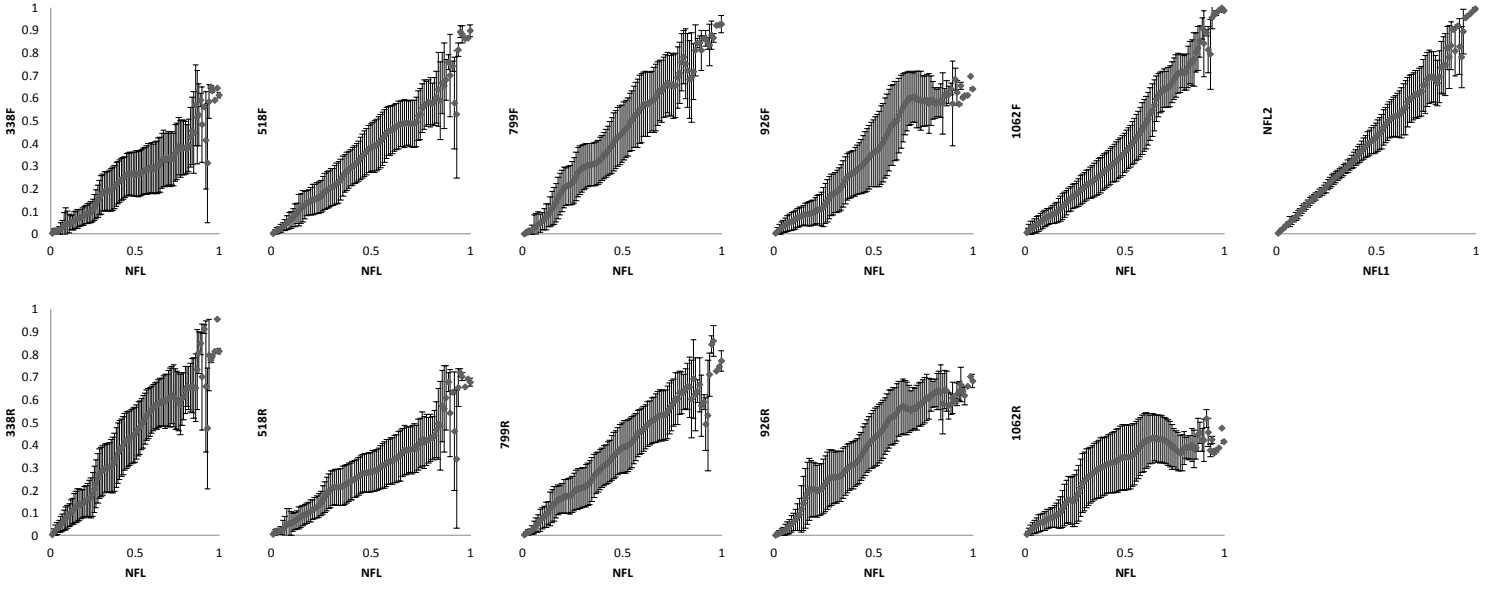


Figure 5.2 The Pearson Correlation between corresponding patristic distances in trees generated from nearly full-length (x-axis) and short read libraries (y-axis) for the different primers investigated. Patristic distances were normalized to a maximum value of one prior to plotting.

With respect to tree topology

To find out whether branch length correlations were conform with consistency of the tree's branching pattern, RF and WRF distances were calculated between NFL and SR trees. The SR libraries that best conserved NFL tree topology were the 518f, 799r and 926r libraries (Table 5.3). The SR libraries that least conserved NFL tree topology were those targeting the V6 region, i.e. 1062r and 926f (Table 5.3). Despite the relatively large RF distances between NFL and 1062r SR trees, the WRF1 and WRF2 distances were relatively small, in the same range of 338f/NFL and 518f/NFL distances. This indicates that a large part of the bipartitions that are unique in the 1062r or NFL tree have a low support value. The 1062f trees, which had the lowest WRF values between trees generated from different tree searches amongst the SR libraries investigated (WRF1, Table S5.1), showed a relatively low conservation of NFL tree topology (RF, Table 5.3). Similarly, the WRF1 and WRF2 distances between 1062f SR trees and NFL trees were high (Table 5.3). These observations show that trees generated from the 1062f library did not conserve NFL topology.

Relation between patristic distances in SR and NFL trees

The Pearson Correlation does not provide information about the extent to which patristic distances in the SR tree approximate corresponding distances in the NFL tree. To address this question we calculated the slope of the best-fitting line forced through the origin of the chart (Table 5.3). A steep slope (slope>>1) indicates that for a given patristic distance in the NFL tree (x-axis), the corresponding distance in the SR tree (y-axis) is higher, which thus means an overestimation of patristic distances in the SR tree. Reads generated from primers 338f, 926f, 1062r, 799f and 518r were found to generally overestimate branch-length distances, while reads generated from primers 926r, 338r, 518f, 1062f and 799r were found to generally underestimate branch-length distances. The 799f and 518r libraries approximated NFL patristic distances best.

Libraries	Variable region	PC [a] pairwise	Slope patristic[b]	Slope pairwise[b]	RF [c]	WRF1[d]	WRF2[e]	RF-WRF1	RF-WRF2
338f vs NFL	V3	0.68	1.46	1.01	1916	636.74	756.32	1279.26	1159.68
338r vs NFL	V2 partially	0.81	0.7	1.07	1916	613.3	743.78	1302.3	1171.82
518f vs NFL	V4	0.83	0.67	1.27	1797	601.73	742.95	1195.47	1054.25
518r vs NFL	V3	0.69	1.07	0.98	1914	637.43	754.86	1276.17	1158.74
799f vs NFL	V5	0.84	1.08	0.75	1938	641.39	755.13	1297.01	1183.27
799r vs NFL	V4 (almost complete)	0.83	0.58	1.32	1833	615.35	735.74	1217.45	1097.06
926f vs NFL	V6	0.72	1.17	1.05	2032	702.96	801.56	1329.44	1230.84
926r vs NFL	V5	0.84	0.82	1.04	1838	606.2	743.73	1232.2	1094.67
1062f vs NFL	V7 & V8 partially	0.78	0.59	0.64	1948	695.49	798.8	1252.51	1149.2
1062r vs NFL	V6	0.72	1.12	1.05	2015	643.48	758.94	1371.32	1255.86

Table 5.3 Overview of the research parameters that were applied in comparisons of short read and nearly full-length sequence libraries.
[a], PC= Pearson Correlation; averaged for different tree comparisons
[b], slope was calculated for SR(1) versus NFL(1)
[c], RF= averaged Robinson Foulds distance between NFL and SR trees
[d], WRF1 = averaged Weighted Robinson Foulds distance between NFL and SR trees based on the sum of the supports of the unique bipartitions
[e], WRF2 = averaged Weighted Robinson Foulds distance between NFL and SR trees based on the sum of the supports of the unique bipartitions plus the diifference of support values amongst the shared bipartitions

Resolving power of SR fragments

In relation to patristic NFL distances

The sizes of the error bars on the averaged SR distances (Fig. 5.2) are an indication for the resolving power of a SR fragment for a given normalized distance in the NFL tree. As mentioned in the methods section, branch lengths in the SR tree were averaged for each 0.01 distance unit interval in the NFL tree and the corresponding standard deviation on branch lengths in the SR tree was calculated. An increasing standard deviation on the averaged SR distances indicates an increasing variety of branch lengths in the SR tree that are associated with an averaged branch length in the NFL tree. In other words, for a particular averaged NFL branch length, a high standard deviation indicates that the phylogenetic information within the reads did not allow to resolve the true branch lengths between all concerning pairs of sequences in the SR tree. In contrast to the Pearson Correlation, the standard deviation provides insight in the variation of patristic distances in the SR tree relative to a given normalized distance in the NFL tree. As such, it provides insight in the resolving power of the read for any normalized patristic distance in the NFL tree. The path of this standard deviation, plotted in function of the patristic distances in the NFL tree, is given for each read library in Fig. 5.3. In general, a scattering is observed at NFL patristic distances larger than 0.8, which is explained by the decreasing amount of patristic distances contributing to each averaged distance interval for larger distances. We should note that for interpretation of the standard deviation curve standard deviations corresponding to distances larger than 0.8 were not taken into account. The y-axis was set at a maximum value of 0.2 in order to gain more detail in the path of the standard deviation curve. Limiting this maximum value caused the loss of some non-informative outlier points at patristic distances larger than 0.8. A general trend is that the standard deviation increases with increasing NFL patristic distance. In some cases (i.e. 518f, 799f, 518r and 799r) the standard deviation reaches a maximum value at a certain NFL branch length, and then fluctuates around this maximum value for increasing patristic distances. This implies that the resolving power generally decreases for distant sequences, and in a number of cases varies around a constant minimum value from a specific NFL patristic distance forward. Libraries generated from the 338f, 518f, 518r, 799r and 1062f primers were found to generally have the lowest standard deviation over the complete range of NFL patristic distances, which means that these libraries have the highest resolving power over all NFL patristic distances. The 926f library peaked to the highest standard deviation amongst all libraries. In the special case of the 1062r library, the resolving power decreased with increasing NFL patristic distance to reach a minimum, but from that value forward increased for even more distant sequences.

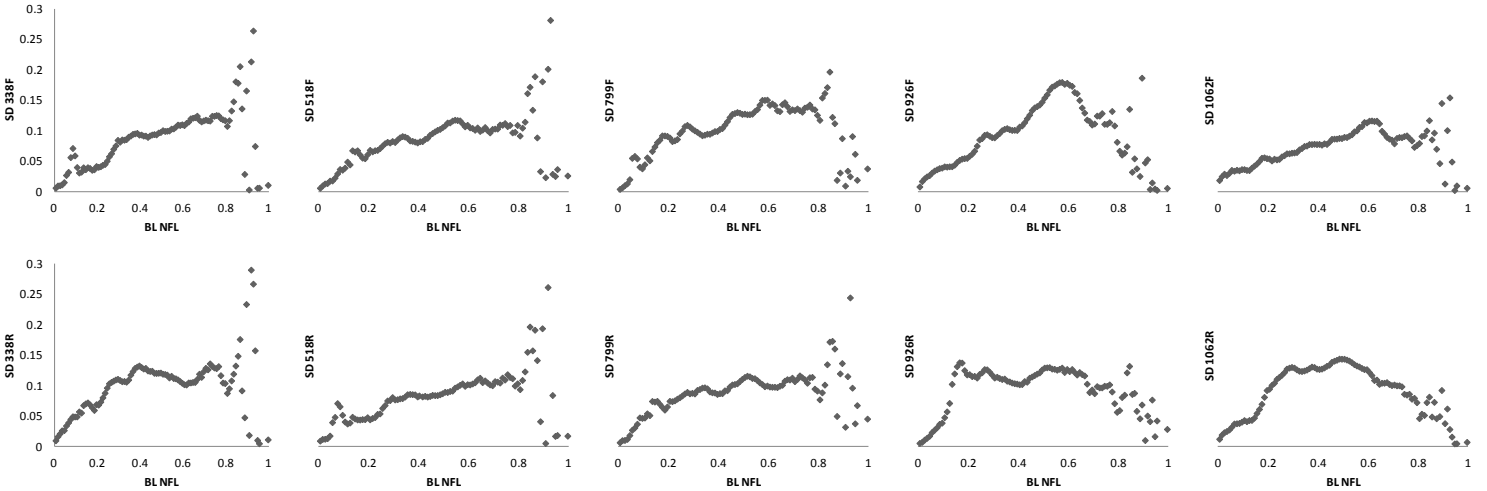


Figure 5.3 The size of the error bars in the short read versus nearly full-length patristic distance correlation plots (Fig. 5.2) (y-axis) for any normalized patristic distance in the NFL tree (x-axis). The error bars represent the resolving power of short reads for any normalized patristic distance in the nearly full-length tree.

In relation to pairwise NFL distances

Fig. 5.4 shows the standard deviation on the averaged pairwise SR distances in function of the pairwise distances in the NFL tree. Similar to the plots for patristic distances, a scattering is observed for normalized pairwise distances larger than 0.6. These points were not taken into account for interpretation. The y-axis was set at a maximum value of 0.2, which caused the loss of some non-informative outlier points. A general trend is that the standard deviation increases with increasing NFL distance. In the case of read 1062r, the standard deviation reaches a maximum value for an NFL distance of approximately 0.4, and then fluctuates around this maximum value for increasing patristic distances. These observations imply that, in general, the resolving power decreases for distant sequences, and in the special case of 1062r varies around a constant minimum value from a specific distance forward. Libraries generated from the 338f, 518f, 518r, 799r and 926r primers were found to generally have the lowest standard deviation over the range of NFL distances up to 0.6, meaning that these libraries have the highest resolving power over all NFL distances in question.

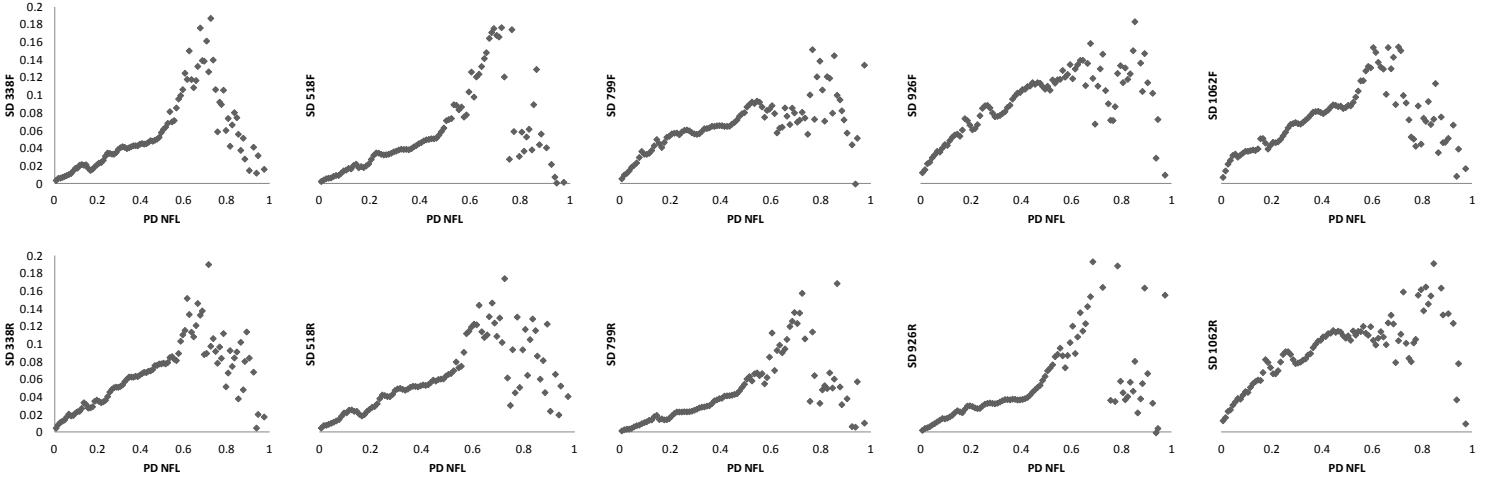


Figure 5.4 The size of the error bars in the short read versus nearly full-length pairwise distance correlation plots (Fig. 5.5) (y-axis) for any normalized pairwise distance in the NFL tree (x-axis). The error bars represent the resolving power of short reads for any normalized pairwise distance in the nearly full-length tree.

OTU richness assessment in SR libraries based on pairwise distances

The Pearson Correlation between pairwise distances in SR libraries and corresponding pairwise distances in their parent NFL library was never close to 1.00. The highest correlations were found for the 338r, 518f, 799f, 799r, 926r and 1062f reads (Fig. 5.5**,** Table 5.3), confirming what was observed for patristic distance correlations between SR and NFL sequences. In each correlation plot (Fig. 5.5) we observe a strong correlation up to normalized pairwise distances of 0.5 to 0.6 on the x-axis. For larger distances there was some degree of scattering, depending on the library. This implies that for sequences with a high degree of similarity within a NFL library, the daughter SR sequences are proportionally similar within the SR library. However, this association is lost for sequences with a low degree of similarity. Since correlations do not provide any information about the extent to which pairwise distances between SR sequences approximate pairwise distances between their parent NFL sequences, we calculated the slope of the line of best fit forced through the origin in the NFL versus SR pairwise distance plots. Youssef et al. [7] found that the slope depends on the proportion of hypervariable, variable and conserved bases in the region of the 16 rRNA gene sequenced. Distances within the 338f and 518r libraries were found to be the best estimators of distances between NFL sequences, with slopes of 1.01 and 0.98 respectively (Table 5.3). Similarly, OTU richness calculated from the 518r and 338f libraries best approximated OTU richness calculated from NFL sequences (Table S5.1). However, no significant relationship was found between OTU richness calculated from the SR libraries, and the slope of the best fitting line forced through the origin (R=0.64, 0.59 and 0.65 for OTU cut-offs of 0.01, 0.02 and 0.03 respectively). This was somehow unexpected, but could have been due to the fact that pairwise distances for OTU assignment were calculated using the Mothur software, while distance correlation plots were based on pairwise distances calculated in RAxML. It was shown previously that distance calculation method and parameters used have a significant effect on OTU richness [4]. Still, regardless of this discrepancy, the data shows a clear effect of the region sequenced on α-diversity in terms of OTU richness (Table S5.1). In each case there was an underestimation of OTU's compared to the NFL sequences. It is clear that these findings argue with the assumption frequently made that distances between short reads are representative for distances between full-length 16S rRNA gene sequences.

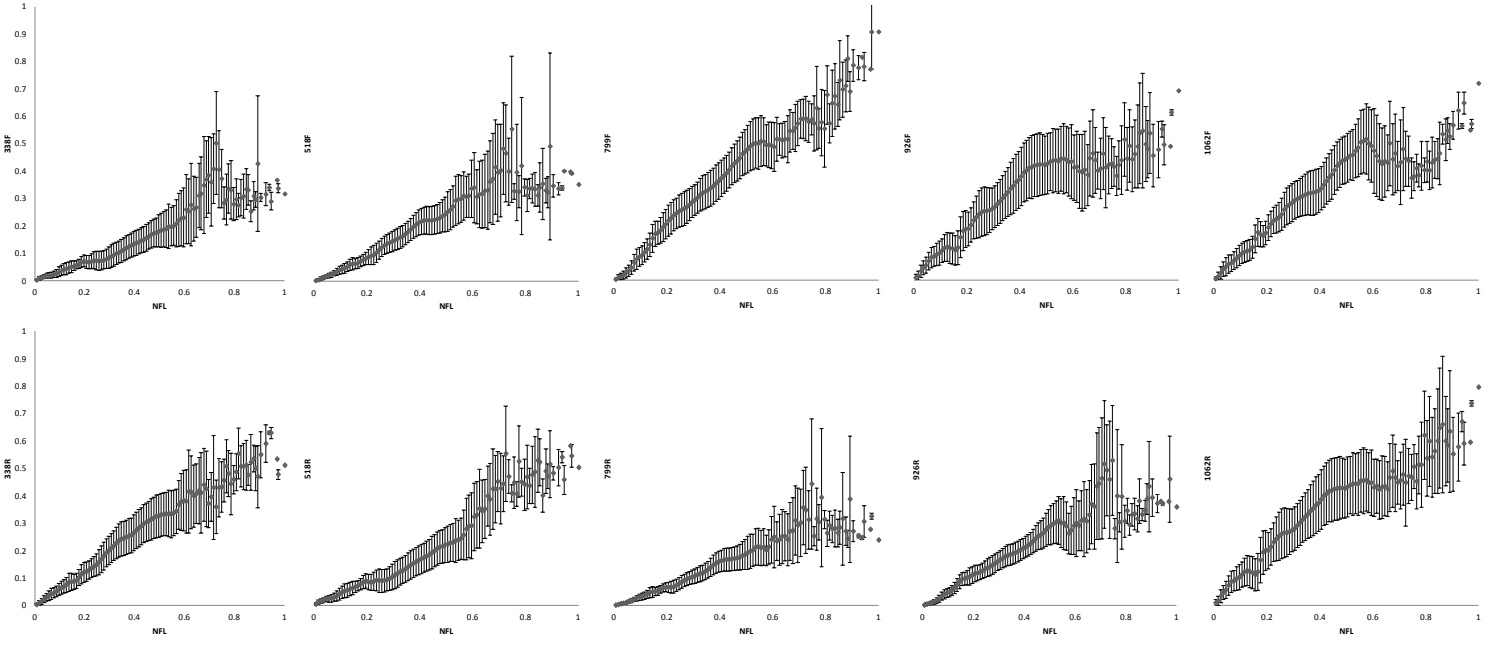


Figure 5.5 The Pearson Correlation between corresponding pairwise distances in nearly full-length (x-axis) and short read libraries (y-axis) for the different primers investigated. Pairwise distances were normalized to a maximum value of one prior to plotting.

Taxonomic assignment of SR sequences

Table 5.4 summarizes results on the taxonomic assignment performance of each SR library. Assignment performance was assessed by comparing identifications for each read within a SR library with identifications obtained for the parent NFL sequences in the NFL library. Taxonomic assignment was performed both at the phylum and genus level. The 518f library was found to generate the highest percentage of correct assignments at the genus level (80.15%), followed by the 338f, 799r and 518r libraries with 76.43%, 76.17% and 76% correct assignments respectively (Table 5.4). These observations confirm results obtained by Liu et al. [5] and Soergel et al. [6]. At the phylum level results were slightly different. The best assignments were obtained with the 518f, 799f, 799r, 926r, 338f and 518r libraries, all of which gave a comparably high number of correct assignments. Although the number of correct assignments obtained with the other SR libraries was lower, the difference was almost negligible. Short read sequences that were identified while the NFL sequence could not, were labeled false positives. The 799f library returned the smallest amount of false positive genus identifications, while the 926f and 1062r libraries returned the highest amount. At the phylum level, the number of false positive assignments was comparable for all libraries. Conversely, a number of SR sequences could not be assigned, while the NFL sequence was in fact assigned. Such SR sequences were labeled false negatives. Both at the genus and phylum level, the 518f library returned the lowest amount of false negatives while the 1062f library returned the highest amount. Based on these results it can be concluded that the 518f library is the best target for assignment of short reads. With the exception of false positives (for which it scored last but one), the 518f library scored best for the different criteria for both genus and phylum level identifications.

	PHYLUM (%)									
Taxonomic assignment	**338f**	**338r**	**518f**	**518r**	**799f**	**799r**	**926f**	**926r**	**1062f**	**1062r**
Correct assignments[a] by read (including unclassified NFL reads)	98.55	97.79	98.98	98.47	98.89	98.81	97.45	98.72	97.11	97.45
Correct assignments[a] by read (excluding unclassified NFL reads)	98.72	98.03	99.14	98.72	99.06	98.97	97.6	98.89	97.26	97.6
False positive reads[b]	0.17	0.26	0.17	0.26	0.17	0.17	0.17	0.17	0.17	0.17
Unclassified reads (total)	1.36	1.79	1.02	1.28	1.19	1.19	2.38	1.36	2.89	2.47
False negative reads[c]	1.02	1.53	0.68	1.02	0.85	0.85	2.04	1.02	2.55	2.13
Unclassified NFL reads	0.34	0.26	0.34	0.26	0.34	0.34	0.34	0.34	0.34	0.34
	GENUS (%)									
Taxonomic assignment	**338f**	**338r**	**518f**	**518r**	**799f**	**799r**	**926f**	**926r**	**1062f**	**1062r**
Correct assignments[a] by read (including unclassified NFL reads)	76.43	72.43	80.17	76	64.09	76.17	71.57	70.38	63.66	72.43
Correct assignments[a] by read (excluding unclassified NFL reads)	75.37	71.02	79.54	74.81	61.57	75.09	70.37	68.61	61.39	71.3
False positive reads[b]	0.94	0.94	1.02	0.85	0.6	0.94	1.19	0.77	0.85	1.19
Unclassified reads (total)	27.91	32.94	24.51	28.17	37.28	28.68	30.3	32.09	40	29.87
False negative reads[c]	20.77	25.79	17.45	20.94	29.79	21.53	23.4	24.77	32.77	22.98
Unclassified NFL reads	7.15	7.15	7.06	7.23	7.49	7.15	6.89	7.32	7.23	6.89

Table 5.4 Taxonomic assignment performance of short read sequence libraries.
[a], percentage of reads that were assigned to the same genus/phylum as in the NFL library
[b], reads that were assigned in the SR library but could not be assigned in the NFL library
[c], percentage of reads that were assigned in the NFL library but could not be assigned in the SR library

5.4 DISCUSSION

The aim of this research was to analyze the suitability of commonly used, published primers targeting dispersed regions of the bacterial 16S rRNA gene for Next Generation Sequencing (NGS). The study targets different aspects that each are involved in the interpretation of NGS data. We started by calculating primer coverage rates for each of the primers analyzed, and continued with the phylogenetic information that is contained within NGS reads. Subsequently, the relation between pairwise distances in NFL and SR sequence libraries was studied to assess the effect on OTU richness. We ended by investigating the taxonomic assignments obtained with each of the SR libraries. In order to do so, we constructed a sequence library composed of 1175 sequences, which served as a representative substitute of the SILVA SSU database. The choice to work with this representative library was motivated by the fact that we did not want to focus on a specific environment, which is inherently biased towards specific taxonomic groups, but instead we aimed at making our results applicable for divergent taxa, and consequently for a variety of environments.

The methodology used allows a thorough evaluation of the scientific outcome that is obtained with sequencing short read fragments generated from primers targeting dispersed regions of the 16S rRNA gene. For the outline of this study, we started by following the reasoning of Jeraldo and colleagues [10] who focused on *de novo* synthesis of phylogenetic trees from short reads to study the implications of information loss which is inherent to sequencing short fragments of the 16S rRNA gene. We extended their well designed approach by checking whether short reads can be used to infer 16S rRNA gene based phylogeny and by assessing whether short reads are reliable estimators of relationships between their parent NFL sequences in terms of patristic distances. Insight in the resolving power of short read fragments for any patristic or pairwise distance between NFL sequences was obtained from standard deviations on averaged short read distances. Next, the relation between pairwise distances between short read fragments and pairwise distances between NFL sequences was studied. This information was used to perceive the effect of sequencing different regions of the 16S rRNA gene on OTU richness and taxonomic assignment accuracy. Additionally the coverage rates of the primers were calculated based on sequences in public 16S rRNA gene databases. We acknowledge the fact that these databases are composed of sequences that were obtained from amplicon sequencing, which makes the results obtained prone to PCR amplification bias. Inclusion of metagenomic data, as performed by Mao and colleagues [11], would have given a superior picture. However, as the emphasis of this study was on

phylogenetic and taxonomic information, we considered this extension of primer coverage rate beyond the scope of this study.

Our results show that the 518f reads that target the V4 region of the bacterial 16S rRNA gene were generally most informative. The correlation value of 0.97 (and the high degree of fit) that was obtained after comparing 518f trees from different tree searches is a very optimistic approximation to the upper limit of 0.98, and indicates the high phylogenetic content of these reads. High correlation values were maintained with an increasing number of tree searches, indicating that the trees generated were very reproducible with respect to patristic distances. Although 518f reads tended to underestimate patristic distances in ML trees, they were found to best reflect 16S rRNA gene based phylogenetic relationships with good resolving power. The 518f reads were found to score best for most of the criteria investigated to assess taxonomic assignment performance. However, nonetheless a high correlation (and degree of fit) was observed between pairwise distances in SR libraries and corresponding pairwise distances in the parent NFL library, reads were not the best estimators of pairwise distances between NFL sequences (cf. slope). This had its effect on OTU richness, for which the 518r and 338f libraries were found to perform better. Furthermore, primer coverage rates showed that the 518 primer is not specific for bacterial 16S rRNA, which implies that contamination with eukaryotic and archaeal 16S rRNA genes may occur.

Since 799r reads also target the V4-region of the 16S rRNA gene, it was not surprising that the primer in question was also found to be a promising instrument for NGS studies. The Pearson Correlation and the degree of fit between patristic distances that were extracted from SR and NFL trees were higher for reads generated with the 799r primer than with the 799f primer. The same was observed for multiple tree searches on the same library. The Pearson Correlation between pairwise distances in the 799f library and the NFL library was similar to the Pearson Correlation between pairwise distances in the 799r library and the NFL library. The high correlation values that were obtained in both cases indicated that both libraries reflect similarities between NFL sequences. Sizes of the error bars in both the patristic and pairwise correlation plots were generally larger in 799f generated reads than in 799r generated reads, indicating a higher resolving power of the 799r reads. The slope of the best fitting line through the origin was 1.08 for the 799f primer, which is a good approximation of NFL patristic distances. The slope calculated for the 799r library, however, was only 0.58, indicating that in general branch lengths were 42% shorter. The 799r reads tended to overestimate differences between sequences, while the 799f reads tended to underestimate differences, with a clear effect on α-diversity. Of both libraries, OTU richness in the 799r library was a better estimator of OTU richness in the NFL library. In terms of taxonomic assignment of SR sequences at the phylum level, performance was comparable for the 799f

and 799r libraries for the different criteria investigated. However, at the genus level the 799r library generally performed better than the 799f library.

Our results illustrate that the 1062f/r primer had the highest coverage rate over the 29 phyla studied. Therefore, this primer is most likely to target the broadest bacterial diversity amongst the primers investigated. However, the 518f library scored best for most of the criteria that allow measuring to which extent the information obtained from short reads is representative for their parent full length sequences. In some cases the use of the 799 primer is recommended in order to avoid the interference caused by co-extracted host organelle DNA. For such cases, the results obtained show that the 799 primer is best used in the reverse direction in order to optimally exploit the information contained within short sequencing reads. However, it was mentioned earlier that in order to exclude the interference of host derived mitochondrial sequences the primer should be used in the forward direction. The consideration between information loss due to the presence of mitochondrial sequences when using the primer in the reverse direction, and information loss due to the less informative region sequenced in the forward direction is a decision that should be driven by the aims of the research.

5.5 ACKNOWLEDGEMENTS

In the first place we would like to thank Dr. Shawn Gomez for adjusting the vCEED script to our needs and Dr. Patricio Jeraldo for providing the script that allowed the extraction of branch-length distances from ML trees. Furthermore we would like to thank Dr. Alexandros Stamatakis for providing answers to our numerous ML related questions. Finally we would like to thank Bart Verheyde for providing scripts that were necessary for data analyses.

5.6 REFLECTING ON THE WORK PERFORMED

Brief summary of work

The bacterial 16S rRNA gene is often used in prokaryotic diversity studies. Current next generation sequencing technologies allow only sequencing parts of the gene, not the full length sequence. However, as different regions of the 16S rRNA gene evolve at different evolutionary rates, the scientific outcome of diversity studies that apply next generation sequencing tools alters with the gene region sequenced. To get insight in the impact of primer choice, and thus the region of the gene sequenced, on the outcome of next generation sequencing efforts, all the unknowns associated with sequencing data, i.e. primer coverage rate, phylogeny, OTU-richness and taxonomic assignment, were implemented in one study for ten well established universal primers (338f/r, 518f/r, 799f/r, 926f/r and 1062f/r) targeting dispersed regions of the bacterial 16S rRNA gene. The manuscript has been accepted for publication in PLOS ONE with minor revisions.

In hindsight

The total primer coverage rate and the coverage rate per phylum were calculated using stringent criteria, i.e. allowing no mismatches. However, there are studies which reported that a restricted amount of mismatches between primer and primer target site may not hamper primer binding, if those mismatches do not occur in the first four bases at the 3' primer end. Therefore, the coverage rates mentioned in this chapter may be an underestimation.

In the conclusions, we highlighted the consideration between using the 799 primer in the forward or in the reverse direction. Use in the forward direction allows separating bacterial from mitochondrial amplicons, while the region sequenced in the reverse direction is more informative. Separation of mitochondrial and bacterial amplicons is performed by gel electrophoresis; hence it is likely that recuperation of bacterial amplicons will result in amplicon loss. Therefore, unless the number of mitochondrial amplicons in the PCR mixture exceeds the number of bacterial amplicons (which could be checked by constructing a small clone library prior to the analysis), the better option is likely to use the 799 primer in the reverse direction.

We did not investigate the coverage rates of primer pairs (e.g. 8f & 518r) used in the amplification of 16S rRNA gene fragments. Since different primers have different coverage rates, their combined coverage rates will be different from the individual primer coverage rates.

Sequences were selected from the SILVA database to construct the representative subset. Subsequently, sequences were trimmed to equal length. Due to the trimming process, binding sites for primers 8f and 27f were cut off. Although it would have been interesting to also check these primers for the different criteria investigated, it was not possible with the chosen subset. Another representative subset would have to be constructed. However, as not all sequences within the database are long enough to contain these primer regions, the question remains whether that subset will still be representative for the SILVA tree of life.

In the Introduction (Part II) it was described that different regions of the 16S rRNA gene show different susceptibilities to chimera formation (§2.3.2), with the V6-V9 region showing a higher chimera rate than the V1-V3 region and the V3-V5 region. This means that PCRs performed with the 338f/r, 518f/r, 799f/r and 926r have lower chances of generating chimeras relative to primers 926f and 1062f/r.

Future research

It was generally assumed that Lateral Gene Transfer (LGT) does not occur in rRNA genes because these radical changes would compromise the structural integrity of the ribosome, resulting in cell death. However, growing evidence has raised doubt about this previous assumption [47,48,49].

The analytical methods that were used in this research allow investigating whether historically LGT was involved in the evolution of the 16S rRNA gene. By analyzing phylogenetic trees built from different segments of the 16S rRNA gene, an answer to this question could be provided. The following approach is suggested:

- Build phylogenetic trees from each 16S rRNA gene segment.
- Extract patristic distances between pairs of taxa and construct Pearson Correlation plots.
- Convert each of the trees to a 3D space using the vCEED approach and make superimposition plots.
- Generate bar plots to easily detect and identify outliers [34]; visual confirmation in the tree remains necessary.

If no LGT occurred during the evolution of any of the segments, the evolution of each of the gene segments should be similar. This would be reflected by a low wRMSD value obtained from the superimposition plots, and a high Pearson Correlation in the Pearson Correlation plots. As it is known that each segment evolves at a different evolutionary rate, the resolution will differ per segment. Hence,

the slope of the best fitting line connecting the data points and forced through the origin in the Pearson Correlation plots will vary with the regions being compared. However, this will not affect the association (and thus the measure to detect LGT) between pairs of sequences in two trees built from different segments of the 16S rRNA gene.

However, as only parts of the trees are sequenced, phylogenetic information within the reads may be a major limitation. As mentioned in this chapter, low phylogenetic information content may create inconsistencies between trees generated from a given sequence library. Inconsistencies between trees may hamper the detection of LGT, since it remains unclear whether low Pearson Correlations or high wRMSD values result from inconsistencies between the trees or from LGT events. Hence, the phylogenetic information content of each of the segments should be checked prior to the analysis.

Remark

The topic of this research deviates from the central focus of this dissertation (which is mainly cultivation based and focuses on bacterial diversity). Initially, an experiment was designed to study the influence of microenvironment on bacterial populations. Unfortunately, due to heavy rainfall at the sampling locations, the samples intended for this research were lost and the experiment had to be cancelled. The study presented in this chapter was set up instead, as it had become clear during the design of the experiment that specific questions could not be answered with information available in literature. One of these questions involved the nature of the information obtained by sequencing with the 799 primer.

5.7 REFERENCES

1. Larsen PE, Gibbons SM, Gilbert JA (2012) Modeling microbial community structure and functional diversity across time and space. FEMS Microbiology Letters 332: 91-98.
2. Gilbert J, Meyer F, Jansson J, Gordon J, Pace N, et al. (2010) The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6th 2010.
3. Lombard N, Prestat E, van Elsas JD, Simonet P (2011) Soil-specific limitations for access and analysis of soil microbial communities by metagenomics. FEMS Microbiology Ecology 78: 31-49.
4. Schloss PD (2010) The Effects of Alignment Quality, Distance Calculation Method, Sequence Filtering, and Region on the Analysis of 16S rRNA Gene-Based Studies. Plos Computational Biology 6.
5. Liu Z, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Research 36.
6. Soergel DAW, Dey N, Knight R, Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. Isme Journal 6: 1440-1444.
7. Youssef N, Sheik CS, Krumholz LR, Najar FZ, Roe BA, et al. (2009) Comparison of Species Richness Estimates Obtained Using Nearly Complete Fragments and Simulated Pyrosequencing-Generated Fragments in 16S rRNA Gene-Based Environmental Surveys. Applied and Environmental Microbiology 75: 5227-5236.
8. Van de Peer Y, Chapelle S, De Wachter R (1996) A Quantitative Map of Nucleotide Substitution Rates in Bacterial rRNA. Nucleic Acids Research 24: 3381-3391.
9. Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. Journal of Microbiological Methods 69: 330-339.
10. Jeraldo P, Chia N, Goldenfeld N (2011) On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. Environmental Microbiology 13: 3000-3009.
11. Mao D-P, Zhou Q, Chen C-Y, Quan Z-X (2012) Coverage evaluation of universal bacterial primers using the metagenomic datasets. BMC Microbiology 12.
12. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. (2012) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research.
13. Wang Y, Qian P-Y (2009) Conservative Fragments in Bacterial 16S rRNA Genes and Primer Design for 16S Ribosomal DNA Amplicons in Metagenomic Studies. Plos One 4.

14. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, et al. (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proceedings of the National Academy of Sciences 82: 6955-6959.

15. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Research 33: D294-D296.

16. Berry D, Ben Mahfoudh K, Wagner M, Loy A (2011) Barcoded Primers Used in Multiplex Amplicon Pyrosequencing Bias Amplification Applied and Environmental Microbiology 78: 612-612.

17. Wu JY, Jiang XT, Jiang YX, Lu SY, Zou F, et al. (2010) Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiology 10: 7.

18. Chelius MK, Triplett EW (2001) The diversity of archaea and bacteria in association with the roots of Zea mays L. Microbial Ecology 41: 252-263.

19. Kumar PS, Brooker MR, Dowd SE, Camerlengo T (2011) Target Region Selection Is a Critical Determinant of Community Fingerprints Generated by 16S Pyrosequencing. Plos One 6: 8.

20. Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, et al. (2008) Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. Plos Genetics 4.

21. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35: 7188-7196.

22. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, et al. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research.

23. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461.

24. Yarza P, Ludwig W, Euzeby J, Amann R, Schleifer K, et al. (2010) Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. Systematic and Applied Microbiology 33: 291-299.

25. Munoz R, Yarza P, Ludwig W, Euzéby J, Amann R, et al. (2011) Release LTPs104 of the All-Species Living Tree. Systematic and Applied Microbiology 34: 169-170.

26. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

27. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32: 1363-1371.

28. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.

29. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 75: 7537-7541.

30. Caporaso G, Kuczynski J, Stombaugh J, Bittinger K, Bushman F, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nature methods 7: 335-336.

31. Yang ZH (1994) Maximum-Likelihood phylogenetic estimation from DNA-sequences with variable rates over sites - Approximate methods. Journal of Molecular Evolution 39: 306-314.

32. Stamatakis A. Phylogenetic models of rate heterogeneity: a high performance computing perspective; 2006 25-29 April 2006. pp. 8 pp.

33. Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24: 2098-2100.

34. Choi K, Gomez SM (2009) Comparison of phylogenetic trees through alignment of embedded evolutionary distances. BMC Bioinformatics 10: 15.

35. Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Mathematical Biosciences 53: 131-147.

36. Robinson DF, Foulds LR (1979) Comparison of weighted labelled trees. In: Horadam AF, Wallis WD, editors. Combinatorial Mathematics VI: Springer Berlin Heidelberg. pp. 119-126.

37. Silvestro D, Michalak I (2012) RaxmlGUI: A graphical front-end for RAxML. Organisms Diversity and Evolution 12: 335-337.

38. Schloss PD, Westcott SL (2011) Assessing and Improving Methods Used in Operational Taxonomic Unit-Based Approaches for 16S rRNA Gene Sequence Analysis. Applied and Environmental Microbiology 77: 3219-3226.

39. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37: D141-D145.

40. Suzuki MT, Giovannoni SJ (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and Environmental Microbiology 62: 625-630.

41. Muyzer G, Dewaal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by Denaturing Gradient Gel Electrophoresis analysis of Polymerase Chain Reaction amplified genes coding for 16S ribosomal RNA. Applied and Environmental Microbiology 59: 695-700.

42. Lane D (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic Acid Techniques in Bacterial Systematics. New York: Wiley. pp. 115-175.

43. Allen AE, Booth MG, Verity PG, Frischer ME (2005) Influence of nitrate availability on the distribution and abundance of heterotrophic bacterial nitrate assimilation genes in the Barents Sea during summer. Aquatic Microbial Ecology 39: 247-255.

44. Ludwig W, Klenk H-P (2001) Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. In: Boone DR, Castenholz RW, Garrity GM, editors. Bergey's Manual of Systematic Bacteriology. 2 ed. New York: Springer. pp. 49-65.

45. Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21: 456-463.

46. Rokas A (2011) Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum Likelihood (RAXML) Program. Curr Protoc Mol Biol Chapter 19: Unit19 11.

47. Miller SR, Augustine S, Le Olson T, Blankenship RE, Selker J, et al. (2005) Discovery of a free-living chlorophyll d-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial small-subunit rRNA gene. Proceedings of the National Academy of Sciences of the United States of America 102: 850-855.

48. Wang Y, Zhang ZS (2000) Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. Microbiology-Uk 146: 2845-2854.

49. Schouls LM, Schot CS, Jacobs JA (2003) Horizontal transfer of segments of the 16S rRNA genes between species of the Streptococcus anginosus group. Journal of Bacteriology 185: 7241-7246.

Author Contributions:
JG and KH designed the experiment. JG wrote the manuscript. ES and JG did the experimental work. AC performed the TaxonGap analyses and wrote the corresponding paragraph in the manuscript; JG performed the remainder of the analyses. AVL, AC, KH and PDV proofread the manuscript.

CHAPTER 6

RPOD GENE SEQUENCE BASED EVALUATION OF CULTURED ANDEAN POTATO-ASSOCIATED PSEUDOMONAS DIVERSITY

Redrafted from: Ghyselinck, J., Coorevits, A., Van Landschoot, A., Samyn E., Heylen, K. & De Vos, P. (2013). rpoD gene sequence based evaluation of cultured Andean potato-associated Pseudomonas diversity. Microbiology, revisions

SUMMARY

Producers and consumers of agricultural products have become increasingly aware of the negative effects associated with the application of agrochemical fertilizers and pesticides. Direct and indirect plant growth-promotion (PGP) by micro-organisms presents a valuable alternative. However, as several plant diseases are difficult to control by currently known biocontrol agents, a continued quest for new agents is required. Considering the well-known direct and indirect PGP properties of *Pseudomonas* members, three cultivation media (Trypticase Soy Agar (TSA), Potato Dextrose Agar (PDA) and *Pseudomonas* Isolation Agar (PIA)) were evaluated for their abilities to grow *Pseudomonas* strains. The rationale was to select for media that allow the retrieval of the highest *Pseudomonas* diversity, as such increasing the chance of isolating PGP candidates. TSA and PDA were found to generate the largest *Pseudomonas* diversity. However, communities obtained with TSA and PDA overlapped, while those obtained with PIA were unique; herewith indicating that the largest diversity is obtained by sampling from either PDA or TSA, and from PIA in parallel. To evaluate OTU-richness (biodiversity) of the isolated *Pseudomonas* members, a thorough investigation of the taxonomic resolution of the 16S rRNA, *rpoD*, *gyrB* and *rpoB* genes was performed. The *rpoD* gene sequences not only contained the highest phylogenetic information and had the highest taxonomic resolution amongst the genes investigated, but also had a gene phylogeny that related well with that of the 16S rRNA gene.

6.1 INTRODUCTION

The genus *Pseudomonas* historically developed as a kind of dumping ground for aerobic, motile Gram-negative rods, and thus lacked a profound classification of its members [1]. Therefore, numerous efforts have been made to reclassify members of the genus. Members, which were originally distributed over the *Alpha-*, *Beta-*, and *Gammaproteobacteria*, were driven back restricted to the *Gammaproteobacteria*. Still, the genus *Pseudomonas* continuously harbors a collection of bacterial strains with very diverse characteristics. Numerous positive traits have been attributed to *Pseudomonas* strains, ranging from denitrification [2] (e.g. in wastewater treatments) to the degradation of toxic components [3,4] and the promotion of plant growth [5,6]. However, the genus *Pseudomonas* also has a negative reputation, as opportunistic human pathogens [7] and plant pathogens [8].

Numerous studies and reviews have demonstrated the role of *Pseudomonas* strains in plant growth-promotion (PGP) applications [5,6,9,10,11,12,13,14]. Both direct and indirect PGP properties have been attributed to the genus. The term 'direct' refers to the positive effect that a bacterial agent has on plant growth itself, for instance by fixation of nitrogen or by production of plant hormones. Indirect PGP, on the other hand, involves the protective effect of PGP agents against pathogenic organisms (also referred to as biocontrol). Both types of PGP can be exerted through a variety of mechanisms. See Lugtenberg et al. [15] for an overview.

Microbial PGP is extremely relevant for agriculture today. The negative impact of agrochemicals on health of both the environment and the consumers, resistance development in pathogenic organisms and high production cost of agrochemicals demand alternative solutions to the crop disease problem. In addition, agrochemical pesticide use, crop rotation and breeding for resistant plant varieties have been found to be insufficient to prevent disease of important crops [6]. Contrary to agrochemicals, the compounds involved in direct and indirect PGP are environmentally-friendly, biodegradable and excreted on or near the plant surface where they are most efficient [15]. While natural suppressive soils seem to be effective only against specific pathogens, introduced biocontrol rhizobacteria are usually effective against a wide range of pathogens. However, the benefit in plant-bacteria interactions depends on environmental factors and can be affected by the nutritional status of the soil, the potential toxic effects of the bacterium, the presence of pathogenic fungi, plant age, induced stress resistance and cross-talk between plant signal transduction pathways [16]. These determining factors may explain why very often a large discrepancy is observed between lab and field performance.

Most biopesticides discovered thus far belong to the genera *Pseudomonas* and *Bacillus* [6]. The versatile nature of *Pseudomonas* strains makes them interesting targets for the development of biopesticides [16]. However, until now, only a few strains of *Pseudomonas fluorescens*, *Pseudomonas putida*, and *Pseudomonas chlororaphis* are commercially available [6]. This is mainly due to the restricted shelf-life of the formulated products. Currently available *Pseudomonas* biocontrol agents tend to lose viability when stored for several weeks. Further exploration for new and stable bacterial PGP agents is thus strongly encouraged.

To further advance the discovery of *Pseudomonas* isolates with PGP properties, three different cultivation media were evaluated for their capacities to grow members of the genus *Pseudomonas*. We assumed that chances to obtain *Pseudomonas* isolates with PGP properties increase with increased cultured *Pseudomonas* diversity. Hence, the medium that showed the largest diversity of *Pseudomonas* isolates was considered the most optimal medium. However, due to the complex taxonomy of the genus *Pseudomonas*, there is no straightforward protocol available for differentiating *Pseudomonas* isolates at the deep taxonomic levels. Hence, it was unclear which taxonomic marker was best suited for the purpose of our study. Therefore, of each of the four biomarkers that were proposed in the identification scheme of Mulet et al. [17] the taxonomic resolution was assessed. Results illustrated that the *rpoD* gene was preferred.

6.2 MATERIALS AND METHODS

Sampling

Potato rhizosphere and root samples were taken from three fields (E1 (Latitude: S 02°37'20.4'', Longitude: W 078°56'04.7''), E2 (Latitude: W 079°09'25.4'', Longitude: S 03°20'15.9'') and E3 (Latitude: W 079°13'32.6'', Longitude: S 03°32'21.8'')) in the Central Andean Highlands of Ecuador. Ten plants were sampled per field. Per plant, 5 g of rhizosphere soil adhering to the potato roots was collected by brushing the roots. Rhizosphere soil samples obtained from different plants were pooled per field. Five ml phosphate buffered saline (PBS) and 10 sterile glass beads (6 mm) were then added to 1 g pooled rhizosphere soil, and the obtained suspension was vortexed for 2 min.

Root fragments of each plant sampled were surface sterilized with 5% sodium hypochlorite for 2 min, then rinsed in autoclaved water and transferred into sterile, sealable plastic bags. Subsequently, roots were cut into small pieces and pooled per field. Fifteen grams of pooled root fragments were then triturated in 50 ml PBS by using a sterile glass rod, and the resulting mixture was incubated for one hour at 28°C with agitation (150 rpm).

Serial dilutions were made (10^{0}-10^{-2}) of the root and rhizosphere soil suspensions and 10^{-1} and 10^{-2} dilutions were plated (100 µl) on ten-fold diluted Trypticase Soy Agar (TSA), Potato Dextrose Agar (PDA) and *Pseudomonas* Isolation Agar (PIA; Difco™, BD). TSA and PDA were supplemented with 0.005% (w/v) cycloheximide to inhibit fungal growth. After 48 h of incubation at 28°C the bulk cultivable fraction was harvested per plate, and collected in eppendorf tubes for DNA extraction.

DNA extraction, PCR and clone-libraries

DNA of the cultivable fraction was extracted per field, per medium and per dilution (e.g. E1 rhizosphere on TSA, dilution 10^{-1}) according to Pitcher et al. [18]. *rpoD* gene amplification PCR was performed in triplicate on each DNA extract according to Mulet et al. [19]. PCR amplicons were purified with the Nucleofast® 96 PCR system (Millipore, Belgium). Triplicate PCR products of the 10^{-1} and 10^{-2} dilutions were then pooled equimolar. Consequently, eighteen samples remained (Table 6.1) in which the *Pseudomonas* diversity was studied using clone libraries that were composed with the pGEM-T Vector System (Promega Benelux, The Netherlands). Sequencing of clones was performed on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, USA). Sequencing products were purified with the BigDye XTerminator® Purification Kit (Applied Biosystems, USA) and sequenced using a 3130xl Genetic Analyzer

(Applied Biosystems, USA). Sequence quality checking and trimming of primer sequences was performed manually in BioNumerics 5.1 (Applied Maths, Belgium). Chimera detection was performed with Chimera Slayer [20].

Rhizosphere samples	Abbreviation	Root samples	Abbreviation
E1 PIA RHIZOSPHERE	E1 RH PIA	E1 PIA ROOT	E1 RO PIA
E1 PDA RHIZOSPHERE	E1 RH PDA	E1 PDA ROOT	E1 RO PDA
E1 TSA RHIZOSPHERE	E1 RH TSA	E1 TSA ROOT	E1 RO TSA
E2 PIA RHIZOSPHERE	E2 RH PIA	E2 PIA ROOT	E2 RO PIA
E2 PDA RHIZOSPHERE	E2 RH PDA	E2 PDA ROOT	E2 RO PDA
E2 TSA RHIZOSPHERE	E2 RH TSA	E2 TSA ROOT	E2 RO TSA
E3 PIA RHIZOSPHERE	E3 RH PIA	E3 PIA ROOT	E3 RO PIA
E3 PDA RHIZOSPHERE	E3 RH PDA	E3 PDA ROOT	E3 RO PDA
E3 TSA RHIZOSPHERE	E3 RH TSA	E3 TSA ROOT	E3 RO TSA

Table 6.1 Overview of the samples analyzed.

Selection of the *rpoD* gene to differentiate *Pseudomonas* isolates

Taxongap analysis (Slabbinck et al., 2008) was performed on type strains representing the different subgroups of the *Pseudomonas fluorescens* group to identify the gene best suited for deep differentiation of *Pseudomonas* isolates. All type strains, and corresponding 16S rRNA, *rpoB*, *gyrB* and *rpoD* genes that were included for the analysis are shown in Table S6.1. The 16S rRNA gene sequences were aligned using the ARB software (Ludwig et al., 2004) with the integrated ARB aligner based on the secondary structures of the 16S rRNA gene. Aligned sequences were exported applying the position variability filter for bacteria (integrated in the software) and re-imported in the Molecular Evolutionary Genetics Analysis (MEGA5) software [21]. Overhangs were trimmed resulting in a final alignment of 1337 positions. A Maximum Likelihood (ML) tree was constructed applying the Jukes-Cantor substitution model, and bootstrap analysis was performed based on 1000 replications.

The three other genes were aligned based on amino acid (AA) sequences using the MEGA5 software. After reconversion into the original nucleotide sequences, overhangs were trimmed resulting in final alignments of 915 positions for *rpoB* gene sequences, 717 positions for *rpoD* gene sequences and 798 positions for *gyrB* gene sequences. ML trees were constructed applying the Jukes-Cantor substitution model (with complete deletion of gaps/missing data), and bootstrap analysis was performed based on 1000 replications. The obtained pairwise similarity matrices were used for Taxongap analysis. Alignments of the four gene sequences were concatenated (3767 positions) using the Seaview v4

software [22]. Species groups to which the species were assigned are also indicated in Table S6.1. Two parameters were evaluated, namely the heterogeneity within species groups, and species group separability.

Phylogeny of the *rpoD* gene

Construction of phylogenetic trees

The TaxonGap analysis suggested the suitability of the *rpoD* biomarker for the evaluation of *Pseudomonas* richness on three different growth media. Hence, further studies were performed on the *rpoD* gene only. An *rpoD* sequence library was constructed which contained all currently available *rpoD* sequences from *Pseudomonas* type strains. Sequences were obtained through query in Straininfo [23] and the PseudoMLSA database [17] (http://www.uib.es/microbiologiaBD/Welcome.php) (Table S6.2). For sequence quality checking, nucleotide sequences were translated to AA sequences using Transeq (http://www.ebi.ac.uk/Tools/emboss/transeq/). The functionality of AA sequences was confirmed with the pBLAST tool of NCBI [24]. Sequence alignment was performed on the AA sequences using the MEGA 5 software [21]. After alignment, AA sequences were reconverted into the original nucleotide sequences. 16S rRNA gene sequences of the same *Pseudomonas* type strains were collected into a library. The 16S rRNA gene sequences were aligned using the ARB software [25] as mentioned above.

Sequences in both libraries were trimmed to obtain maximum overlap between the sequences. ML trees were then constructed from both the 16S rRNA and *rpoD* gene sequence libraries. The software used was RAxML v7.3.5. A ML search was performed under gamma, in combination with rapid bootstrapping under CAT [26]. The substitution model used was GTR. Bootstrapping was performed with 1000 replicates. The command line used for the tree search was the following: raxmlHPC-PHTREADS-SSE3 –T <number of processors> -fa -m GTRGAMMA -N <replicates> -x <seed1> -p <seed2> -s <filename> -n <outputfile>.

Branch length based comparison of phylogenetic trees

To study the phylogenetic relation between two ML trees, the Pearson Correlation (PC) was calculated between patristic distances between corresponding sequence pairs in the two trees. Patristic distances are defined as the length of the shortest path connecting two taxa in a phylogenetic tree. These patristic distances were extracted from the ML trees using a script that was kindly provided by Jeraldo and colleagues [27].

A second method used to study the relation between a pair of phylogenetic trees was based on the vCEED approach that was developed by Choi and Gomez [28]. Distance matrices were generated from the ML trees using the PHYLOCOM software [29]. These matrices were used as inputs for the vCEED script that was written in Matlab. The vCEED script maps taxa to a Euclidean space via metric multidimensional scaling (MDS), thus producing a multidimensional plot in which each point represents one sequence (or taxon) within the phylogenetic tree. This procedure was applied to both trees to be compared. Both embedded point patterns were then superimposed on one another and the degree of fit, which is expressed by the weighted Root Mean Square Deviation (wRMSD), was calculated. A low wRMSD indicated a high degree of fit, and thus a high similarity between trees.

Construction of correlation plots

To plot the correlation between 16S rRNA gene based ML trees and *rpoD* based ML trees graphically, corresponding patristic distances were transferred into a tuple, which formed the coordinate of a point in a plot. Distances were ordered for the 16S rRNA gene tree, and corresponding distances of the *rpoD* tree were rearranged accordingly. Subsequently, a binning step was performed by calculating the averages and standard deviations of corresponding patristic distances in both trees over each patristic distance interval of 0.001 in the 16S rRNA gene tree. Averaged 16S rRNA gene sequence distances and corresponding averaged *rpoD* sequence distances were then plotted in a graph, and the standard deviations on the averaged *rpoD* distances were superimposed as error bars.

Robinson-Foulds distance calculations

The unweighted and weighted Robinson-Foulds (RF and WRF respectively) distances [30] were calculated to gain insight in the topological differences between two phylogenetic trees. RF and WRF distances were calculated by importing the ML trees into RAxML v7.4.2Gui. The RF metric calculates the number of splits that are unique to one of both trees being compared, so it actually describes ancestral differences between trees. The higher the RF value, the lower the amount of shared ancestors. As such, phylogenetic trees are more similar as the RF values decrease. The WRF, however, takes into account the support values of the branches that are unique to one of the trees being compared instead of just counting the number of unique clades. Comparing RF and WRF distance values allows gaining insight in the nature of differences between trees. If the WRF value approximates the RF value for a given tree comparison, this means that differences mainly occur on high supported branches. Conversely, if the WRF value is much smaller than the corresponding RF values, differences between the trees mainly occur on low supported branches.

Evaluation of *Pseudomonas* diversity on different media

Construction of rarefaction curves and taxonomic assignment based on rpoD sequences

The *Pseudomonas* diversity on the different media was evaluated by means of rarefaction curves that were calculated with the Mothur v1.27.0 software [31]. Sequences were assigned to Operational Taxonomic Units (OTUs) using the cluster command and the average neighbor algorithm. The averaged OTU numbers that were retrieved from 1000 iterations were used to construct rarefaction curves. For taxonomic assignment of sequences, the Bayesian classifier that is integrated in the Mothur v1.27.0 software was used. The same *rpoD* gene sequence library that was used for studying *rpoD* phylogeny (but containing one more sequence for which no 16S rRNA gene sequence was available (Table S6.2)) served as a reference database for the taxonomic assignment of environmental *rpoD* sequences. Assignments were performed with the classify.seqs() command on aligned *rpoD* sequences. The bootstrap cutoff for assigning a sequence to a specific taxon was set at 50%.

Fast UniFrac analysis

A total of 1500 *rpoD* sequences that were obtained from all samples and from the different media investigated were merged into one fasta file. Nucleotide sequences were converted to AA sequences and the alignment was performed in MEGA5 [21] using the ClustalW alignment tool. The alignment was checked and adjusted manually. After reconverting the aligned AA sequences back to the original nucleotide sequences, the endpoints of all sequences were trimmed to obtain maximum overlap between the sequences. A ML tree was constructed from the aligned sequences in RAxML v7.3.5 with the same parameters as mentioned above but with 500 bootstraps instead of 1000. The ML tree was then imported in the Fast UniFrac webtool that is available online [32]. Sample clustering was performed, and sample distance matrices were calculated.

Nucleotide sequence accession numbers

The *rpoD* rRNA gene sequence data generated in this study has been deposited in GenBank/EMBL/DDBJ with accession numbers HF931547-HF933112.

6.3 RESULTS

Taxonomic resolution of the *rpoD* gene

The genus *Pseudomonas* is composed of ten phylogenetic groups, each group being a collection of closely related species [17]. In addition, one of these groups, the *Pseudomonas fluorescens* group, consists of nine subgroups [17]. The taxonomic resolution of the 16S rRNA gene fails to differentiate *Pseudomonas* strains at the intrageneric level [33]; i.e. the gene does not allow to distinguish species within a *Pseudomonas* group or subgroup. Hence, the 16S rRNA gene is not suited to adequately measure the diversity of *Pseudomonas* members. However, recently Mulet et al. [17] proposed a Multi Locus Sequence Analysis (MLSA) scheme that was found to be useful in the identification of *Pseudomonas* strains . The scheme is based on sequences of four different genes, namely 16S rRNA, *gyrB*, *rpoD* and *rpoB*. In the present study, the TaxonGap software (Slabbinck et al., 2008) was used to evaluate each of the four biomarker genes for its power to differentiate *Pseudomonas* strains on the one hand, and for its ability to represent the *Pseudomonas* phylogeny on the other hand. Fig. 6.1 shows that species subgroup separability, in other words gene resolution (represented by dark grey bars), was highest for the *rpoD* gene, followed by the *gyrB*, *rpoB* and 16S rRNA genes. Although species subgroup heterogeneity (indicated by the light grey bars) was generally large, thus indicating a high within-subgroup resolution, it mostly did not exceed species subgroup separability for the *rpoD* gene (Fig. 6.1). This finding indicates a high within-subgroup resolution that does not hamper differentiation of *Pseudomonas* subgroups. Additionally, a good correlation was observed between *rpoD* gene based phylogeny and phylogeny based on all four genes of the MLSA-scheme of Mulet et al. (2010). These findings supported the use of the *rpoD* biomarker for measuring the *Pseudomonas* diversity on the growth media. Hence, further studies were performed on the *rpoD* gene only.

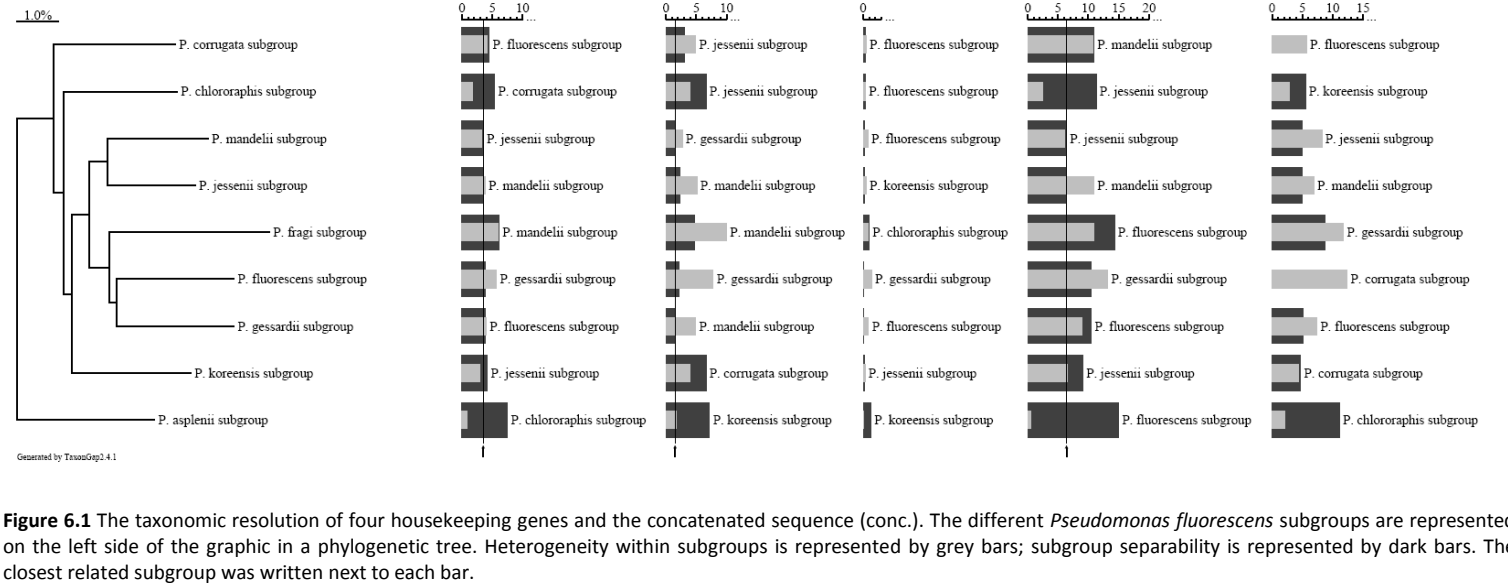


Figure 6.1 The taxonomic resolution of four housekeeping genes and the concatenated sequence (conc.). The different *Pseudomonas fluorescens* subgroups are represented on the left side of the graphic in a phylogenetic tree. Heterogeneity within subgroups is represented by grey bars; subgroup separability is represented by dark bars. The closest related subgroup was written next to each bar.

Phylogenetic content of *Pseudomonas rpoD* sequences

Patristic distance based comparison of phylogenetic trees

The phylogenetic content refers to the amount of phylogenetic information contained within sequences. Sequences with a high phylogenetic content contain sufficient information to build robust phylogenetic trees. In order to study the phylogenetic content of *Pseudomonas rpoD* sequences, we calculated the Pearson Correlation (PC) between patristic distances between corresponding pairs of taxa in two ML trees that were constructed from different tree searches on the same *rpoD* sequence library [27]. A high PC indicates a positive association between patristic distances in both trees being compared, implying that taxa positioned closely together in the first *rpoD* tree are also positioned closely together in the second *rpoD* tree. Conversely, a low PC indicates a very low association between corresponding patristic distances in both trees, which in turn means that pairs of taxa positioned closely together in the first *rpoD* tree are not necessarily positioned closely together in the second *rpoD* tree. A high PC thus indicates that the phylogenetic information within the sequences being studied is sufficient to calculate unequivocal patristic distances between taxa, which in turn points to a high phylogenetic content of the sequences.

Because ML trees are built using a heuristic method, there is no guarantee that the tree calculated is the best representation of the sequence data. As such, a high PC between a pair of *rpoD* trees generated from different tree searches on the same library may have been obtained by chance. To deal with this problem, we calculated four trees from the *rpoD* sequence data, and subsequently calculated the PCs between all possible pairs of trees (Table 6.2). The obtained PCs remained high for all tree comparisons. This indicated that *rpoD* based ML trees were very robust with respect to patristic distances between taxa, illustrating that *Pseudomonas rpoD* sequences contain enough phylogenetic information to construct phylogenetic trees unequivocally. However, the PC method to compare phylogenetic trees is known to have some weaknesses [28], which is why an additionally study was performed by using the vCEED approach [28]. Results are given in Table 6.2 and are expressed in terms of degree of fit (wRMSD) between two trees. A very high correlation was obtained between results generated with the vCEED approach and results obtained with the PC method (R=-0.99).

Tree comparison[a]	PC[b]	wRMSD[c]	Slope patristic[d]	RF[e]	WRF[f]
rpoD(1) vs rpoD(2)	1.0000	0.0000035	NA	0	0
rpoD(1) vs rpoD(3)	0.9994	0.0028	NA	4	1.5
rpoD(1) vs rpoD(4)	0.9994	0.0028	NA	6	2.37
rpoD(2) vs rpoD(3)	0.9994	0.0028	NA	4	1.45
rpoD(2) vs rpoD(4)	0.9994	0.0028	NA	6	2.33
rpoD(3) vs rpoD(4)	1.0000	0.000003	NA	2	0.91
16S(1) vs 16S(2)	0.8588	0.0509	NA	148	13.87
16S(1) vs 16S(3)	0.8924	0.0467	NA	142	12.31
16S(2) vs 16S(3)	0.9820	0.0164	NA	110	7.53
16S(1) vs rpoD(1)	0.7876	0.0506	27.46	212	95.32
16S(1) vs rpoD(2)	0.7876	0.0506	27.46	212	95.68
16S(1) vs rpoD(3)	0.7875	0.0507	27.61	214	96.56
16S(1) vs rpoD(4)	0.7875	0.0507	27.61	214	96.4
16S(2) vs rpoD(1)	0.8122	0.0496	23.02	218	93.33
16S(2) vs rpoD(2)	0.8122	0.0496	23.02	218	93.67
16S(2) vs rpoD(3)	0.8113	0.0498	23.14	220	94.57
16S(2) vs rpoD(4)	0.8113	0.0498	23.14	220	94.36
16S(3) vs rpoD(1)	0.7785	0.0539	25.08	220	96.82
16S(3) vs rpoD(2)	0.7785	0.0539	25.08	220	97.17
16S(3) vs rpoD(3)	0.7777	0.0539	25.21	222	98.02
16S(3) vs rpoD(4)	0.7777	0.0539	25.21	222	97.79

Table 6.2 Overview of research parameters that were used to measure the phylogenetic information within *rpoD* sequences.
[a], The number between brackets refers to the number of the tree that was generated from the sequence library
[b], PC= Pearson Correlation
[c], wRMSD= weighted Root Mean Square Deviation
[d], NA= Not applicable
[e], RF= Unweighted Robinson Foulds
[f], WRF= Weighted Robinson Foulds

The PCs between trees obtained from different tree searches on a given sequence library, can be used as a measure to compare the phylogenetic content of different genes. It was striking that the PC between trees obtained from *Pseudomonas rpoD* sequences was systematically higher than the PC between trees constructed from 16S rRNA gene sequences. As can be observed from the superimposition plots given in Fig. 6.2, the vCEED method similarly showed that the degree of fit was higher for *rpoD* based trees than for 16S rRNA gene based trees.

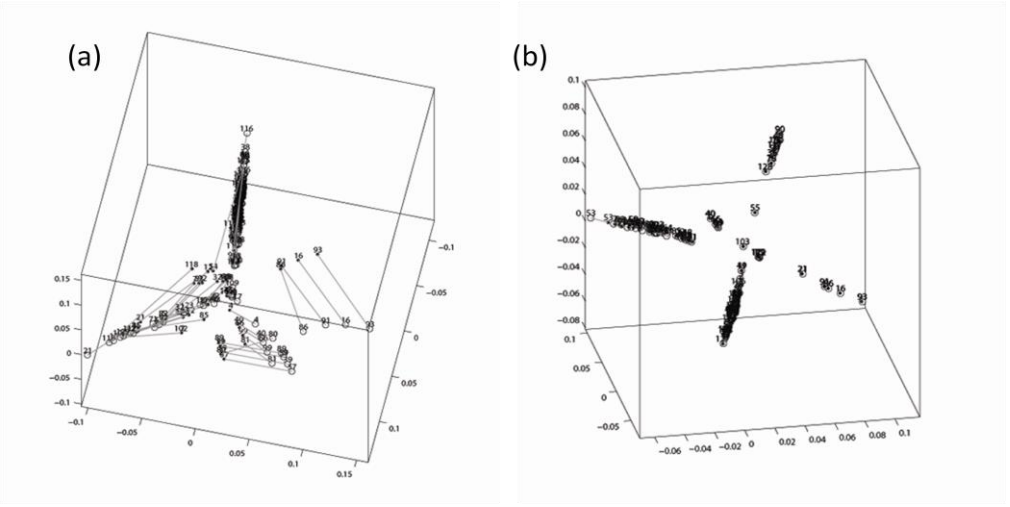


Figure 6.2 Superimposition plots that were created with the vCEED script. (a) Shows the superimposition of two 16S rRNA gene based trees that were generated from two tree searches on the same sequence library. (b) Shows the superimposition of two *rpoD* gene based trees that were generated from two tree searches on the same sequence library. Bars connect corresponding taxa in both trees. Bar length decreases with increasing similarity between trees.

This implies that *rpoD* sequences allow the construction of phylogenetic trees that are more robust than trees built from 16S rRNA gene sequences. Therefore, these findings suggest that there is more phylogenetic information contained within *Pseudomonas rpoD* sequences than within 16S rRNA gene sequences.

Topology based comparison of phylogenetic trees

RF and WRF distances provide information on topological differences between trees, and the nature of those differences. Table 6.2 shows the RF and WRF values for the different tree comparisons. We observed that the RF and WRF distances were significantly smaller between trees generated from different tree searches on *rpoD* sequences than between trees generated from different tree searches on 16S rRNA sequences. Furthermore, differences between RF and WRF values for a given tree comparison were generally larger between trees generated from 16S rRNA gene sequences than between *rpoD* gene based trees. This indicates that differences in tree topology in 16S rRNA gene based trees mainly occur on branches with low bootstrap support values, while differences between *rpoD* gene based trees occur on branches with higher supporting values. In other words, topological differences between 16S rRNA gene based trees are mostly caused by inadequacies of sequences to validate the topology, suggesting that the differences are due to the lower phylogenetic content of 16S rRNA sequences to construct unequivocal tree topologies. As mentioned above, RF distances between trees obtained from different tree searches on the *rpoD* library were smaller, and the branches leading

to differences were better supported. This again shows that the phylogenetic content of *rpoD* sequences is higher compared to that of the 16S rRNA gene sequences. Relative to the averaged RF distances between trees generated from different tree searches on 16S rRNA gene sequences, the averaged RF distances between *rpoD* and 16S rRNA gene trees was only 1.6 times higher. This indicates that the differences between topologies in *rpoD* and 16S rRNA gene based trees did not significantly exceed topological differences between different trees that were generated from a given 16S rRNA gene sequence library.

Is *rpoD* based phylogeny in contradiction to 16S rRNA gene based phylogeny?

The phylogenetic information that is contained within *rpoD* sequences was found to be higher than the phylogenetic information contained within 16S rRNA gene sequences. However, the question still remains whether or not *rpoD* phylogeny is contradictory to 16S rRNA gene based phylogeny. Although microbiologists since long-time have deviated from the assumption that the 16S rRNA gene reflects true phylogenetic relationships between organisms, it is still regarded as the benchmark for reconstructing phylogenetic relationships amongst bacterial genera. Therefore, we considered this question very relevant for this study. To answer this question, the PC was calculated between patristic distances in an *rpoD* ML tree and the corresponding patristic distances in a 16S rRNA gene tree. To avoid that a high correlation would have been obtained by chance, PCs were calculated between all 16S rRNA and *rpoD* gene trees that were obtained from different tree searches on the 16S rRNA and *rpoD* gene libraries respectively. The patristic distance correlations ranged from 0.7777 to 0.8122 (Table 6.2) for the different tree comparisons. To make these numbers visual and to understand their meaning, a correlation plot was constructed for the *rpoD*(1) versus 16S(1) tree comparison (Fig. 6.3). Fig. 6.3 shows a clear positive relation between patristic distances in 16S rRNA trees and corresponding patristic distances in *rpoD* trees, and little deviation from a straight line behavior. This indicates that patristic distances in the *rpoD* gene tree follow corresponding patristic distances in the 16S rRNA gene tree, indicating that *rpoD* gene based phylogeny is, generally spoken, similar to 16S rRNA gene based phylogeny. Also in this case a high correlation was found between wRMSD values and corresponding PC values (R=-0.85), thus giving extra support for the positive association between patristic distances in *rpoD* gene trees and patristic distances in 16S rRNA gene trees.

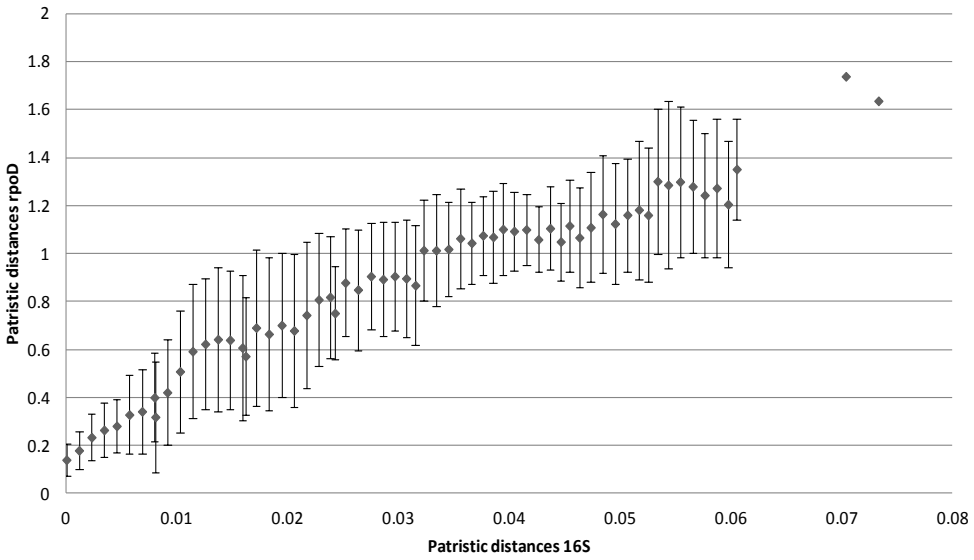


Figure 6.3 The Pearson Correlation plot obtained from corresponding patristic distances in a tree generated from 16S rRNA gene sequences (x-axis) and a tree generated from *rpoD* gene sequences (y-axis). Patristic distances were subjected to a binning step prior to plotting. The binning step was performed by sorting patristic distances for the 16S rRNA gene tree, and subsequently calculating the average patristic distances in the 16S rRNA and *rpoD* trees over each patristic distance interval of 0.001 in the 16S rRNA gene tree. Averaged 16S rRNA distances and corresponding averaged *rpoD* distances were then plotted in this graph, and the standard deviations on the averaged *rpoD* distances were represented by error bars on the chart.

To see how patristic distances in the *rpoD* tree related to patristic distances in the 16S rRNA gene tree, we calculated the slope of the best fitting line connecting the data points and forced through the origin (Table 6.2). On average, patristic distances between *rpoD* sequences were 25.3 times higher than patristic distances between 16S rRNA gene sequences. This indicates a higher evolutionary rate of the *rpoD* gene, which by definition implies that the organisms can be distinguished at a finer taxonomic level based on *rpoD* sequences.

Evaluation of *Pseudomonas* diversity on different media

Construction of rarefaction curves

Pseudomonas diversity on all three growth media was expressed in terms of *rpoD* sequence diversity. In theory, a one-base difference between a pair of sequences may indicate that both sequences originate from different bacterial strains. As we were interested in the *Pseudomonas* diversity at the strain level, an OTU in this work had to be defined as a unique sequence (i.e. similarity cut-off 100%). However, an OTU definition of 99% similarity was used instead to avoid an effect of possible sequencing errors [34]. Since each of the clone libraries differed in size, OTU richness in the different samples was compared by

rarefaction [35,36,37]. Richness estimation based on curve extrapolation methods requires data from relatively well sampled communities [36]. However, this was not the case in this study. Therefore, we decided not to calculate richness estimators based on the obtained rarefaction curves, but use the rarefaction curves directly for data interpretation. To check whether one of the growth media investigated repeatedly came out as the one growing the largest diversity for the different samples analyzed, or conversely, whether the best growth medium varied with the sample being analyzed, we constructed clone libraries for six different samples. As shown by the rarefaction curves in Fig. 6.4, the *Pseudomonas* diversity covered was different for the three media investigated. The outcome also depended on the sample being analyzed. In all but one sample (E2 root), the *Pseudomonas* specific medium (PIA) generated the lowest diversity. In each of the samples either PDA or TSA were found to pick up the largest diversity. Exceptions were the E2 and E3 root samples, where TSA was the least suited medium.

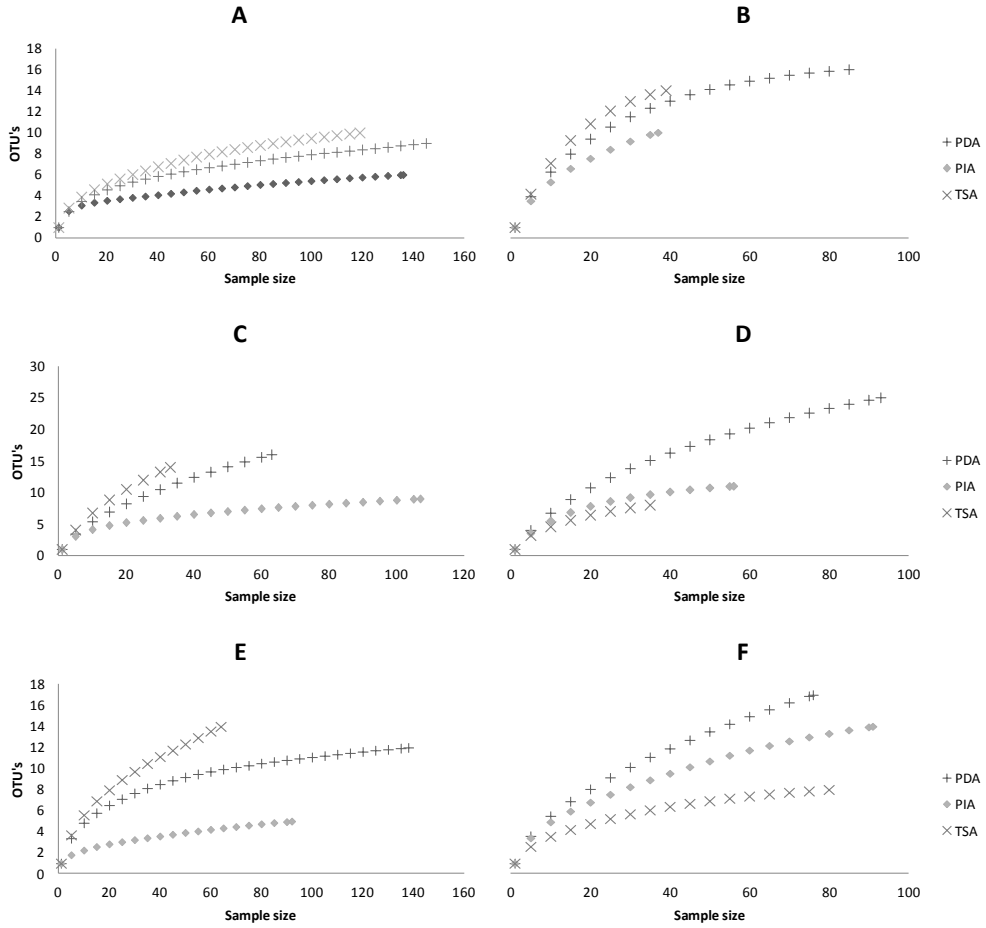


Figure 6.4 Rarefaction curves constructed from *rpoD* sequences that were obtained from three media (TSA, PDA and PIA) in six samples. RH refers to rhizosphere, while RO refers to root communities. (A) E1 Rhizosphere sample, (B) E1 Root sample, (C) E2 Rhizosphere sample, (D) E2 Root sample, (E) E3 Rhizosphere sample, (F) E3 Root sample.

Fast UniFrac analysis

To check whether *Pseudomonas* diversity obtained from the three media overlapped, we used the Fast UniFrac webtool [32]. Fast UniFrac allows the comparison of microbial communities based on phylogenetic information. All the analyses performed were unweighted, i.e. not taking into account sequence abundances. Due to biases that are inherent to working with clone libraries, such as PCR-bias and differences in efficiency of the *E. coli* cells with respect to the uptake of amplicon sequences, the relative abundances obtained were considered not representative for the true abundances in the samples. Therefore, weighted UniFrac analyses were not applicable here. Fig. 6.5 shows that in most cases PIA samples clustered separate from TSA and PDA samples. With the exception of E1 PIA RH and

E3 PIA RO, all PIA samples obtained from both rhizosphere and roots from the three locations clustered together, which indicated that the diversities overlapped. These findings illustrate that nonetheless the diversity obtained with PIA was generally lowest, the medium revealed a different diversity compared to TSA and PDA. Therefore, PIA is an interesting medium to use in parallel with either PDA or TSA. PDA and TSA on the other hand, appeared in the same clusters in most of the cases (i.e. E1 RH, E1 RO, E3 RH and E3 RO), illustrating that the obtained diversities overlapped. The UniFrac distances between the different samples investigated can be consulted in Table 6.3.

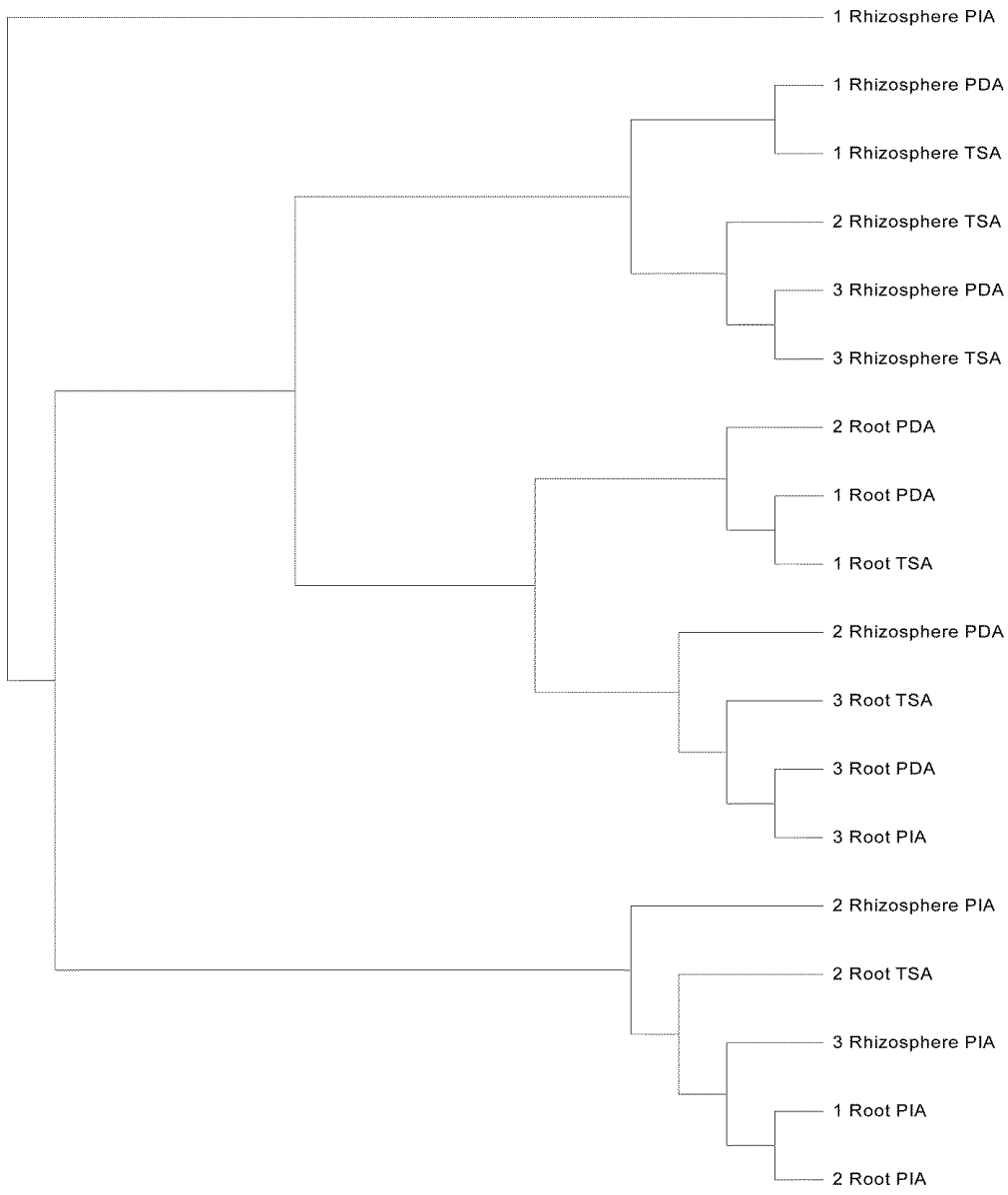


Figure 6.5 Sample clustering as performed by Fast UniFrac.

Sample	E1 RH PDA	E1 RH PIA	E1 RH TSA	E1 RO PDA	E1 RO PIA	E1 RO TSA	E2 RH PDA	E2 RH PIA	E2 RH TSA	E2 RO PDA	E2 RO PIA	E2 RO TSA	E3 RH PDA	E3 RH PIA	E3 RH TSA	E3 RO PDA	E3 RO PIA
E1 RH PDA																	
E1 RH PIA	***0.91***																
E1 RH TSA	***0.50***	***0.92***															
E1 RO PDA	0.78	0.90	0.75														
E1 RO PIA	0.81	0.91	0.80	***0.75***													
E1 RO TSA	0.76	0.89	0.74	***0.33***	***0.69***												
E2 RH PDA	0.63	0.93	0.62	0.57	0.78	0.52											
E2 RH PIA	0.84	0.92	0.84	0.82	0.72	0.82	***0.69***										
E2 RH TSA	0.60	0.93	0.67	0.77	0.71	0.77	***0.67***	***0.68***									
E2 RO PDA	0.85	0.92	0.84	0.59	0.75	0.56	0.70	0.85	0.80								
E2 RO PIA	0.82	0.87	0.84	0.77	0.50	0.73	0.80	0.64	0.69	***0.70***							
E2 RO TSA	0.86	0.82	0.87	0.78	0.76	0.72	0.81	0.80	0.87	***0.82***	***0.60***						
E3 RH PDA	0.71	0.96	0.75	0.77	0.80	0.77	0.67	0.85	0.65	0.76	0.81	0.90					
E3 RH PIA	0.79	0.85	0.81	0.80	0.66	0.79	0.83	0.84	0.82	0.81	0.62	0.74	***0.70***				
E3 RH TSA	0.69	0.91	0.73	0.71	0.77	0.71	0.65	0.83	0.62	0.71	0.74	0.83	***0.25***	***0.68***			
E3 RO PDA	0.75	0.93	0.74	0.65	0.79	0.61	0.52	0.73	0.76	0.67	0.78	0.81	0.68	0.84	0.63		
E3 RO PIA	0.72	0.92	0.70	0.64	0.76	0.63	0.52	0.69	0.74	0.71	0.78	0.81	0.68	0.82	0.63	***0.30***	
E3 RO TSA	0.67	0.91	0.66	0.64	0.78	0.60	0.48	0.74	0.74	0.74	0.80	0.82	0.70	0.79	0.65	***0.37***	***0.34***

Table 6.3 Diversity dissimilarity between the different samples analyzed. Sample names correspond to names given in Table 6.1.

Taxonomic assignment of *rpoD* sequences

To obtain insight in the taxonomic diversity that was picked up from the media, the *rpoD* sequences were assigned using the Mothur v1.27.0 software. The reference dataset used was constructed from *Pseudomonas* type strain *rpoD* sequences. Whether or not *rpoD* sequences allow species identification is still not underpinned, regardless of the high taxonomic resolution and the phylogenetic congruence with 16S rRNA gene based phylogeny that was observed for the *rpoD* gene in this study. Therefore, species names listed in Table 6.4 should be considered tentative, rather than exact identifications.

Species	E1 PDA RH	E1 PIA RH	E1 TSA RH	E1 PDA RO	E1 PIA RO	E1 TSA RO	E2 PDA RH	E2 PIA RH	E2 TSA RH	E2 PDA RO	E2 PIA RO	E2 TSA RO	E3 PDA RH	E3 PIA RH	E3 TSA RH	E3 PDA RO	E3 PIA RO	E3 TSA RO
Pseudomonas azotoformans IAM 1603T					100					27-43	24-44	33-49						
Pseudomonas baetica a390T			90	89-93														
Pseudomonas chlororaphis subsp. *aurantiaca* ATCC 33663T									37									
Pseudomonas chlororaphis subsp. *aureofaciens* LMG 1245T					44					36	44-45							
Pseudomonas corrugata NCPPB 2445T									36-80				50-60		40-57			
Pseudomonas extremaustralis DSM 17835T							29	36										
Pseudomonas frederiksbergensis DSM 13022T	24-27								dec/22				75-80	76-83				
Pseudomonas grimontii CIP 106645T	40-100	89-100	81-100	17-100	20-97			91-94	59			96	78-85	38-97	90-96	48-98	32-99	60-99
Pseudomonas jessenii CIP 105274T	82-100		88-100				92-98		99						75-85			
Pseudomonas koreensis LMG 21318T				nov/91			46			14-82			28-81		59-80	55-99		61-71
Pseudomonas lini CIP 107460T							91-94								37			
Pseudomonas lurida LMG 21995T	33-81	34-70	32-77	21-100	99-100	62-100	40-100	96-100	52-100	38-100	45-100	41-100	91-99		95-99	42-89	85-86	45-60
Pseudomonas marginalis NCPPB 667T				100	100	100	100	100	100	100						100	100	98-100
Pseudomonas migulae CCUG 43165T													16-57		40-86			
Pseudomonas moraviensis DSM 16007T				30			43-56			99-100						44-65	53	
Pseudomonas mucidolens IAM 12406T								100	100		100							
Pseudomonas palleroniana LMG 23076T							100	27-100	100	60			100		79-100	39-100	35-100	50-100
Pseudomonas panacis CIP 108524T					100					99	61					39		
Pseudomonas proteolytica CIP 108464T					100													
Pseudomonas putida ATCC 12633T		100																

Continued on the next page.

Species	E1 PDA RH	E1 PIA RH	E1 TSA RH	E1 PDA RO	E1 PIA RO	E1 TSA RO	E2 PDA RH	E2 PIA RH	E2 TSA RH	E2 PDA RO	E2 PIA RO	E2 TSA RO	E3 PDA RH	E3 PIA RH	E3 TSA RH	E3 PDA RO	E3 PIA RO	E3 TSA RO
Pseudomonas reinekei DSM 18361T										18-78								
Pseudomonas rhizosphaerae LMG 21640T										12								
Pseudomonas rhodesiae LMG 17764T				36-37		35-42				31-34	32	37		35	28			
Pseudomonas saponiphila DSM 9751T													96-100	94-100	99			
Pseudomonas simiae CCUG 50988T				100		100				26-100	23-100	25-100						
Pseudomonas tolaasii NCPPB 2192T							17											
Pseudomonas umsongensis LMG 21317T			96-99						100									
Pseudomonas vancouverensis ATCC 700688T																48		
Pseudomonas veronii LMG 17761T						54-61												

Table 6.4 Bootstrap percentages obtained with *rpoD* sequence assignment. Sample names correspond to names given in Table 6.1.

6.4 DISCUSSION

Plant disease control by microorganisms has received increasing attention the last few decades for several reasons. However, at present a number of plant diseases such as potato late blight disease cannot be controlled efficiently by micro-organisms [38]. As a consequence, producers heavily rely on the application of agrochemicals and a search for new biocontrol agents is required. Many bacterial strains of the genus *Pseudomonas* harbor interesting plant growth-promoting (PGP) properties [5,6,9,12,13,14]. Still, only few PGP *Pseudomonas* strains have been commercialized. This study evaluated two generally used growth media – TSA and PDA – and one *Pseudomonas* specific medium – PIA – for their abilities to grow members of the genus *Pseudomonas*. Since chances of encountering PGP *Pseudomonas* isolates increase with increasing *Pseudomonas* diversity on the media, the growth medium that yielded the largest diversity was considered to be the most interesting medium to conduct PGP studies with.

Of all media investigated, the *Pseudomonas* specific medium (PIA) resulted in the lowest diversity of *Pseudomonas* isolates, and was thus considered the least interesting of the three media tested to conduct isolation campaigns with. TSA on the contrary, outperformed the other two media in all but two samples for which PDA scored best. However, as opposed to TSA, PDA never scored worst. Fast UniFrac [32] analyses showed that in many cases the diversity picked up with PIA differed from the diversities obtained with either TSA or PDA. The latter two were found to show some extent of overlap in most of the samples investigated. Since no relation between choice of growth medium and yield of PGP isolates has been established, our results suggest that the best results may be achieved from cultivations on either PDA or TSA and from PIA in parallel.

Since PGP is a strain specific property, a technique with a highly differentiating power was required to assess *Pseudomonas* diversity. The role of housekeeping genes in resolving the taxonomy of *Pseudomonas* has been established previously. Yamamoto et al. [33] showed, based on combined *gyrB* and *rpoD* sequences, that the genus *Pseudomonas* diverges into two intrageneric clusters IGCI and IGCII, which could be further subdivided into a number of subclusters. Mulet and coworkers [39] found a clear correlation between phylogenetic similarities based on concatenated sequences of the 16S rRNA, *gyrB* and *rpoD* genes on the one hand, and DNA-DNA relatedness values expressed as ΔTm on the other hand for members of the *P. stutzeri* group. Later, the same authors [17] proposed a Multi Locus Sequence Analysis (MLSA) scheme based on concatenated sequences of the 16S rRNA, *rpoB*, *rpoD* and *gyrB* genes, which allowed a thorough identification of *Pseudomonas* isolates at the *Pseudomonas* group or

subgroup level. To select the most interesting gene for the purpose of our study, we used the TaxonGap software [40]. This software allowed calculating and visualizing the heterogeneity of gene sequences within and separability between *Pseudomonas* species subgroups that were part of the larger *Pseudomonas fluorescens* group for each of the four genes. Results showed that the *rpoD* gene had the highest resolution within the *Pseudomonas fluorescens* group. Analyses on gene phylogeny showed similar results, as the slope of the best fitting line forced through the origin and connecting the data points in 16S rRNA gene versus *rpoD* patristic distance correlation plots also indicated a high taxonomic resolution of the *rpoD* gene. Our results confirm previous observations by Parkinson et al. [41], who reported a high resolution of the *rpoD* gene for *Pseudomonas* species belonging to the *Pseudomonas syringae* complex. Similarly, Yamamoto and colleagues [33] found that the phylogenetic distances between *Pseudomonas rpoD* sequences were generally larger than the phylogenetic distances between corresponding *gyrB* sequences, thus indicating its higher resolving power. Although Ghyselinck et al. [42] demonstrated the high taxonomic resolution of MALDI-TOF MS and its potential to perform high-throughput analyses, the technique was not considered for the purpose of this study, since identification with MALDI-TOF MS has not yet been fully optimized. Since MALDI-TOF MS currently lacks a robust reference database, it would have been difficult to focus on *Pseudomonas* only. For the same reasons, other typing techniques that are known to have a high taxonomic resolution were disregarded.

The PC, wRMSD, RF and WRF values that were obtained from comparisons of trees generated from different tree searches on the *rpoD* sequence library indicated a higher phylogenetic content of the *rpoD* compared to the 16S rRNA gene. Furthermore, it was shown that *rpoD* phylogeny was similar to 16S rRNA gene based phylogeny. This observation, however, is in contradiction with results obtained by Yamamoto et al. [43], who showed that the genetic distances in the variable regions of the 16S rRNA gene correlated poorly with corresponding distances between the *rpoD* genes. However, contrary to their interesting approach, this study took into account the complete 16S rRNA gene instead of filtering out its conserved regions to focus on specific variable regions and vice versa. Furthermore, their findings were based on the comparison of pairwise distances, while in our opinion a comparison of phylogenetic distances gives a superior picture as they represent evolutionary history. The latter is necessary to compare gene phylogenies. Ultimately, we were able to base our analyses on a larger set of 129 strains of *Pseudomonas* species, while due to the limited amount of sequences available at that time, Yamamoto and coworkers had to restrict their analyses to 20 sequences [43].

Mulet et al. [19] developed primers that allow to specifically target the *rpoD* gene in a wide range of *Pseudomonas* species. Their primers were designed based on *rpoD* sequences of 35 species representing

the different intrageneric phylogenetic *Pseudomonas* clusters. Subsequent testing of these primers by amplifying the *rpoD* gene of 96 *Pseudomonas* type strains and a well characterized *Pseudomonas* collection of more than 100 strains indicated their universality within *Pseudomonas*. We considered that taxonomic assignment would have given additional insight into primer universality. However, from the 130 *Pseudomonas* type strain *rpoD* sequences that were available, and which our reference database was constructed from, only 29 were mentioned in the assignment report. Considering the low bootstrap values obtained in some identifications (Table 6.4), and because the number of currently known *Pseudomonas* species largely exceeds the 130 *Pseudomonas* species in our reference database, primer universality could not be investigated based on the results obtained. The media used may also have narrowed the number of *Pseudomonas* species. Sequence assignment indicated that a number of *rpoD* sequences showed only limited bootstrap support for assignment to *rpoD* sequences of already known *Pseudomonas* type strains. Considering the unexplored origin of the samples, we reason that this may be attributed to the fact that currently existing *rpoD* databases are underrepresented.

Our results illustrate that *rpoD* gene phylogeny is similar to 16S rRNA gene based phylogeny. In addition, it was found to have the highest taxonomic resolution amongst the four biomarker genes investigated. Still, we cannot exclude the possibility that different strains may share 100% *rpoD* gene sequence similarity, which is why we acknowledge the fact that our measurement of *Pseudomonas* diversity on the agar plates may have been an underestimation. Still, as this underestimation occurred for all three media investigated, we do not believe that this weakness biased the results obtained in this study. Our results show that, either TSA or PDA is recommended when isolation campaigns are performed from one single medium. However, the best choice may depend on the sample being analyzed. More optimal would be to use either TSA or PDA in combination with PIA, considering the different communities obtained from both media. We based this research on the hypothesis that an increased bacterial diversity increases chances of yielding PGP strains.

6.5 ACKNOWLEDGEMENTS

The work was supported by the European Commission's Seventh Framework Program FP7/2007-2013 under grant agreement No. 27552. Kim Heylen is supported by the Flemish Fund for Scientific Research (FWO11/PDO/084).

6.6 REFLECTING ON THE WORK PERFORMED

Brief summary of work

Members of the genus *Pseudomonas* are well-known for their direct and indirect plant growth-promoting (PGP) properties. Therefore, three cultivation media were evaluated for their abilities to grow *Pseudomonas* strains. The rationale was to identify media that allow retrieval of a rich *Pseudomonas* diversity, as such increasing the chance of isolating PGP candidates. To evaluate the biodiversity of the isolated *Pseudomonas* members, a foregoing investigation of the taxonomic resolution of the 16S rRNA, *rpoD*, *gyrB* and *rpoB* genes was performed. The *rpoD* gene sequences were found to contain most phylogenetic information amongst the genes investigated and to have the highest taxonomic resolution. Its gene phylogeny related well with that of the 16S rRNA gene.

In hindsight

This study checked the correlation between 16S rRNA gene based phylogeny and *rpoD* based phylogeny. However, as mentioned in the introduction it is unlikely that the 16S rRNA gene reflects the true phylogeny of any genus (§2.1.3). In fact, the genuine phylogeny of the genus *Pseudomonas* remains thus far unresolved. Mulet et al. (2010) [17] attempted to present an MLSA scheme for *Pseudomonas* strains. However, whether this presents true phylogenetic relationships between members of the genus is doubtful [1]. Ideally, the *rpoD* gene based phylogeny should be measured against a gene combination that was checked against complete genome phylogeny of the genus. However, as this was not available at this time, the comparison could not be made and was based on what is regarded the present benchmark for studying phylogenetic relationships between organisms.

We did not check nor delete singleton sequences. The reason for this is the limited sequencing depth that can be obtained with clone libraries. Unlike deep sequencing, in clone libraries singleton sequences are less likely to present sequencing errors as only a fraction of the amplicons obtained in the PCR mixture are sequenced. Consequently, it is not unlikely that a given amplicon was obtained only once. The situation is different for deep sequencing analyses, in which sequencing depth is much bigger and a more complete picture of the community is obtained. Consequently, singleton sequences are less likely to occur. Still, as we cannot exclude that errors occurred during PCR and sequencing, chimera and sequence quality checking was performed to avoid that erroneous sequences were included in further analyses.

Future research

It would be interesting to compare the diversity of members of the genus *Pseudomonas* as obtained on each of the growth media, with the diversity residing in the original samples. This would allow insight in the fraction of *Pseudomonas* members that can be recovered with the media investigated.

We also did not investigate whether the *Pseudomonas* specific medium was truly *Pseudomonas* specific. Irgasan is a broad spectrum antibiotic and antifungal which is added to the medium, and supposed to be selective for growth of *Pseudomonas* isolates. *Pseudomonas* specificity could have been checked by making a 16S rRNA gene based clone library of the cultivated community on the *Pseudomonas* Isolation Agar.

6.7 REFERENCES

1. Coorevits A (2011) Pseudomonads and bacilli as important spoilage organisms in the dairy industry - a taxonomic study [PhD dissertation]. Ghent: Ghent University. 285 p.
2. Zhang JB, Wu PX, Hao B, Yu ZN (2011) Heterotrophic nitrification and aerobic denitrification by the bacterium Pseudomonas stutzeri YZN-001. Bioresource Technology 102: 9866-9869.
3. Farhadian M, Vachelard C, Duchez D, Larroche C (2008) In situ bioremediation of monoaromatic pollutants in groundwater: A review. Bioresource Technology 99: 5296-5308.
4. Khan MS, Zaidi A, Wani PA, Oves M (2009) Role of plant growth promoting rhizobacteria in the remediation of metal contaminated soils. Environmental Chemistry Letters 7: 1-19.
5. Ghyselinck J, Velivelli SL, Heylen K, O'Herlihy E, Franco J, et al. (2013) Bioprospecting in potato fields in the Central Andean Highlands: Screening of rhizobacteria for plant growth-promoting properties. Systematic and Applied Microbiology.
6. Haas D, Defago G (2005) Biological control of soil-borne pathogens by fluorescent pseudomonads. Nature Reviews Microbiology 3: 307-319.
7. Gershman MD, Kennedy DJ, Noble-Wang J, Kim C, Gullion J, et al. (2008) Multistate Outbreak of Pseudomonas fluorescens Bloodstream Infection after Exposure to Contaminated Heparinized Saline Flush Prepared by a Compounding Pharmacy. Clinical Infectious Diseases 47: 1372-1379.
8. Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, et al. (2012) Top 10 plant pathogenic bacteria in molecular plant pathology. Molecular Plant Pathology 13: 614-629.
9. Daayf F, Adam L, Fernando WGD (2003) Comparative screening of bacteria for biological control of potato late blight (strain US-8), using in-vitro, detached-leaves, and whole-plant testing systems. Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie 25: 276-284.
10. De Curtis F, Lima G, Vitullo D, De Cicco V (2010) Biocontrol of Rhizoctonia solani and Sclerotium rolfsii on tomato by delivering antagonistic bacteria through a drip irrigation system. Crop Protection 29: 663-670.
11. Scherwinski K, Grosch R, Berg G (2008) Effect of bacterial antagonists on lettuce: active biocontrol of Rhizoctonia solani and negligible, short-term effects on nontarget microorganisms. FEMS Microbiology Ecology 64: 106-116.
12. Yan ZN, Reddy MS, Ryu CM, McInroy JA, Wilson M, et al. (2002) Induced systemic protection against tomato late blight elicited by plant growth-promoting rhizobacteria. Phytopathology 92: 1329-1333.

13. Kim H-J, Jeun Y-C (2006) Resistance induction and enhanced tuber production by pre-inoculation with bacterial strains in potato plants against *Phytophthora infestans*. Mycobiology 34: 67-72.
14. Chin-A-Woeng TFC, Bloemberg GV, Lugtenberg BJJ (2003) Mechanisms of biological control of phytopathogenic fungi by *Pseudomonas* spp. In: Stacey G, Keen NT, editors. Plant-microbe interactions. St. Paul: American Phytopathological Society (APS Press). pp. 173-225.
15. Lugtenberg B, Kamilova F (2009) Plant-Growth-Promoting Rhizobacteria. Annual Review of Microbiology 63: 541-556.
16. Preston GM (2004) Plant perceptions of plant growth-promoting Pseudomonas. Philosophical Transactions of the Royal Society B-Biological Sciences 359: 907-918.
17. Mulet M, Lalucat J, Garcia-Valdes E (2010) DNA sequence-based analysis of the Pseudomonas species. Environ Microbiol 12: 1513-1530.
18. Pitcher DG, Saunders NA, Owen RJ (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Letters in Applied Microbiology 8: 151-156.
19. Mulet M, Bennasar A, Lalucat J, Garcia-Valdes E (2009) An rpoD-based PCR procedure for the identification of Pseudomonas species and for their detection in environmental samples. Molecular and Cellular Probes 23: 140-147.
20. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research 21: 494-504.
21. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.
22. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27: 221-224.
23. Dawyndt P, Vancanneyt M, De Meyer H, Swings J (2005) Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources. Ieee Transactions on Knowledge and Data Engineering 17: 1111-1126.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403-410.
25. Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. Nucleic Acids Research 32: 1363-1371.
26. Stamatakis A. Phylogenetic models of rate heterogeneity: a high performance computing perspective; 2006 25-29 April 2006. pp. 8 pp.

27. Jeraldo P, Chia N, Goldenfeld N (2011) On the suitability of short reads of 16S rRNA for phylogeny-based analyses in environmental surveys. Environmental Microbiology 13: 3000-3009.

28. Choi K, Gomez SM (2009) Comparison of phylogenetic trees through alignment of embedded evolutionary distances. BMC Bioinformatics 10: 15.

29. Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24: 2098-2100.

30. Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Mathematical Biosciences 53: 131-147.

31. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 75: 7537-7541.

32. Hamady M, Lozupone C, Knight R (2010) Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. Isme Journal 4: 17-27.

33. Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, et al. (2000) Phylogeny of the genus Pseudomonas: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. Microbiology-Uk 146: 2385-2394.

34. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Applied and Environmental Microbiology 71: 8966-8969.

35. Dunbar J, Takala S, Barns SM, Davis JA, Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Applied and Environmental Microbiology 65: 1662-1669.

36. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM (2001) Counting the uncountable: Statistical approaches to estimating microbial diversity. Applied and Environmental Microbiology 67: 4399-4406.

37. Moyer CL, Tiedje JM, Dobbs FC, Karl DM (1998) Diversity of deep-sea hydrothermal vent Archaea from Loihi seamount, Hawaii. Deep-Sea Research Part Ii-Topical Studies in Oceanography 45: 303-317.

38. Axel C, Zannini E, Coffey A, Guo JH, Waters DM, et al. (2012) Ecofriendly control of potato late blight causative agent and the potential role of lactic acid bacteria: a review. Applied Microbiology and Biotechnology 96: 37-48.

39. Mulet M, Gomila M, Gruffaz C, Meyer J-M, Palleroni NJ, et al. (2008) Phylogenetic analysis and siderotyping as useful tools in the taxonomy of Pseudomonas stutzeri: description of a novel genomovar. International Journal of Systematic and Evolutionary Microbiology 58: 2309-2315.

40. Slabbinck B, Dawyndt P, Martens M, De Vos P, De Baets B (2008) TaxonGap: a visualization tool for intra- and inter-species variation among individual biomarkers. Bioinformatics 24: 866-867.

41. Parkinson N, Bryant R, Bew J, Elphinstone J (2011) Rapid phylogenetic identification of members of the Pseudomonas syringae species complex using the rpoD locus. Plant Pathology 60: 338-344.

42. Ghyselinck J, Van Hoorde K, Hoste B, Heylen K, De Vos P (2011) Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. Journal of Microbiological Methods 86: 327-336.

43. Yamamoto S, Harayama S (1998) Phylogenetic relationships of Pseudomonas putida strains deduced from the nucleotide sequences of gyrB, ropD and 16S rRNA genes. Int J Syst Bacteriol 48: 813-819.

PART IV

Concluding Remarks

I. CONCLUSIONS

This work investigated the plant growth-promotion potential of bacterial isolates obtained from potato fields in the Central Andean Highlands. The large number of isolates, which covered a broad bacterial diversity, was simultaneously used to evaluate MALDI-TOF MS as an alternative dereplication tool for rep-PCR. As many of the plant growth-promoting bacteria were identified as members of the genus *Pseudomonas*, and considering that many previous studies also reported the plant growth-promotion properties of members of the genus, another study was conducted in which the potential of three different media in retrieving a large diversity of *Pseudomonas* members was investigated. In addition, we focused on the effect of primer choice on the outcome of next generation sequencing experiments.

It is well-known that promotion of plant growth and disease suppression by means of agrochemical products has a negative impact on the environment. Hence, there is a need for alternative sustainable approaches. A number of alternative strategies exist, such as direct and indirect plant growth-promotion activity mediated by microorganisms, genetic modification of plants to induce resistance against plant pathogens and RNA interference technology. This PhD study focused on plant growth-promotion mediated by microorganisms. Many of the isolated bacteria with antagonistic properties against *Rhizoctonia solani* and *Phytophthora infestans* were identified as *Pseudomonas* and *Bacillus*, confirming earlier observations. Although many studies report the potential that resides in natural soils, the number of biopesticides on the market is still surprisingly low. This discrepancy can be explained by the existing gap between the industry's needs and the goals of academic research. Shifting academic research from the discovery of new biocontrol agents towards a thorough characterization of already existing agents would likely increase the number of biopesticides on the market. Thorough identification of bacterial agents is a necessity for commercializing microbial biopesticides; therefore, taxonomic labs have an important role to play. Risk assessments of existing biocontrol agents are interesting research topics that can be performed by academic institutions, and which the industry can benefit from. It should be clear that better communication and cooperation between academic institutions and the industry may lead to a significant increase in the number of commercialized biological control agents. This would benefit sustainable agriculture. However, a number of diseases remain difficult to control by means of existing biocontrol agents. For such diseases, the continued search for new biocontrol agents is strongly encouraged.

Considering the fact that thorough identification is one condition required for the commercialization of biopesticides, a new technique was evaluated which significantly reduces time and financial costs in large-scale identification efforts. MALDI-TOF MS represented a valuable alternative for rep-PCR, which is traditionally used for dereplication. Although it requires optimization, MALDI-TOF MS may have the potential to perform high-throughput analyses when it's used in series with a colony picker. As thorough identification is a time consuming process, lowering the number of isolates to be identified will lower personnel costs. Moreover, the consumables needed to perform MALDI-TOF MS analyses are cheaper relative to those needed for rep-PCR. Still, a more detailed study of the taxonomic resolution of the technique is recommended, and may be a topic for future research.

As many of the biocontrol strains were identified as belonging to the genus *Pseudomonas*, three different media were evaluated for their capacities to retrieve members of the genus. The study illustrated that the *Pseudomonas* specific medium did not result in a higher *Pseudomonas* diversity relative to the general media tested. Although a high number of the isolates obtained during the isolation campaign were members of the genera *Pseudomonas* and *Bacillus*, the bacteria cultivated are not necessarily the most important actors in their original environments. Bacteria obtained represent those groups of the community that were able to grow under the cultivation conditions provided. Therefore, design of new media mimicking the natural conditions under which the organisms in the sample reside may lead to the discovery of new biocontrol agents. Hence, the development of new cultivation media to search for biocontrol bacteria may be an interesting field of research for the future.

It was shown that the degree to which culture independent diversity studies reflect true bacterial diversity is significantly determined by the extent to which biases and errors occur in the processes ranging from DNA extraction to sequencing. Culture independent studies are interesting because they allow the screening of samples for the presence of specific bacterial groups or properties. Positive samples may then be subjected to creative isolation campaigns to retrieve the bacterial strains searched for. Because this PhD work was performed in a taxonomic lab, we were interested in studying the effect of primer choice on the outcome of next generation sequencing efforts. It was observed that the gene region of the 16S rRNA gene sequenced had an important impact on the results obtained. This requires that researchers involved are fully aware of the effect of their decisions on the end result of the experiment. The work also shows that it is necessary to step aside from the general assumption that partial 16S rRNA gene fragments are representative for full length 16S rRNA gene sequences. This assumption often leads to the generalization of the 97% rule, which is applicable only for full length

sequences. However, considering the errors and biases that occur during the process, the 97% OTU cutoff rule can serve as an error buffer instead.

Regardless of the important role that once was attributed to members of the rare biosphere, it is important to critically approach rare biosphere representing sequences. Previous work shows that only a fraction of these sequences represent genuine members of the rare biosphere, while a significant proportion is simply the result of errors occurring during PCR and sequencing. It is a prerequisite that any researcher involved in culture independent studies is aware of the weaknesses inherent to the technique.

II. LOOKING BACK

The outline and structure of the work performed within the frame of this PhD dissertation may seem somewhat untraditional. The reason for this deviation lies in a number of decisions that were made at the beginning of the project, and which were driven by a lack of background on the research topic. In hindsight, some decisions made at the beginning of the project were not the best ones at that time, and more in-depth preparation of an experiment was required. It is beyond doubt that this is a skill which evolves in the 4-year period that turns a master into a PhD. Hence, with the current knowledge, the trajectory followed would have been different.

During the process, it became apparent that specific answers to questions that arose during the setup of an experiment were not available in the literature at that time. For instance, a pyrosequencing experiment was set up that required the amplification of bacterial 16S rRNA gene sequences from plant material. The 799 primer seemed an interesting instrument for this application. However, no information was available on the effect of the primer on the end result of the experiment. In general, I noticed that in order to set up specific experiments and not make the same mistakes as the ones made at the beginning of the project, more information was needed.

As a result, with my current knowledge I would have followed a different trajectory. In my opinion, it is interesting to start with an exploratory cultivation independent diversity study of the samples, as this allows the screening for the presence of taxonomic groups of organisms, or specific traits. Amplicon sequencing or metagenomics could be performed to check whether the sample contains taxonomic groups that were not discovered or cultured before, and to design probes based on these sequences to screen a large number of samples. Alternatively, one may be interested in members representing the rare biosphere. In this particular case, a sequence reality check would simultaneously generate the full length sequence and allow the design of specific probes (§2.3.2). Following screening of the samples, efforts could be directed to the isolation of members of the groups of interest, by applying the methods mentioned in the introduction (§2.3.4). Cultivation could either be achieved by trial-and-error, or by a directed approach, applying genome sequences to deduce the organism's metabolic requirements. Whether or not the organisms targeted represent yet cultured or uncultured organisms, newly developed (or existing) media could be evaluated for their abilities to retrieve a large diversity of the organism of interest, similar to the methodology that was used in the *Pseudomonas* evaluation study. This could guide the medium selection for isolation campaigns on a larger scale. The large-scale isolation

campaign could be performed by a colony picker. Subsequent dereplication of the isolates with MALDI-TOF MS, would significantly shorten the time required for identification of the isolates. After identifying and verifying whether the isolated members truly represent the organisms targeted, a screening campaign for plant growth-promotion properties could be set up. Bacteria testing positive for desired traits in the lab, the greenhouse and in the field should then be characterized in-depth to assess information on identity, plant and animal pathogenicity, organism life-cycle, as well as the other criteria mentioned in the section 'Reflecting on the work performed' of chapter 4.

PART V

Supplementary Material

CHAPTER 4

Table S4.1 Overview of the gene accession numbers and strain numbers of the *Pseudomonas* type strains of which the gene sequences were used in the MLSA study to identify the PGP *Pseudomonas* strains.

Species	*gyrB* accession number	Strain number	*rpoB* accession number	Strain number	*rpoD* accession number	Strain number
Pseudomonas abietaniphila	FN554166	ATCC 700689T	AJ717416	CIP 106708T	FN554447	ATCC 700689T
Pseudomonas aeruginosa	AB039386	IFO 12689T	AJ717442	LMG 1242T	AB039607	IFO 12689T
Pseudomonas agarici	AB039457	NCPPB 2289T	AJ717477	LMG 2112T	AB039563	NCPPB 2289T
Pseudomonas alcaliphila	FN554167	LMG 23134T	AJ717463	CIP 108031T	FN554448	LMG 23134T
Pseudomonas amygdali	AB039462	NCPPB 2607T	AJ717462	LMG 2123T	AB039509	NCPPB 2607T
Pseudomonas anguilliseptica	FN554168	LMG 21629T	FN554726	LMG 21629T	FN554449	LMG 21629T
Pseudomonas antarctica	FN554169	LMG 22709T	FN554727	LMG 22709T	FN554450	LMG 22709T
Pseudomonas argentinensis	FN554170	LMG 22563T	FN554728	LMG 22563T	FN554451	LMG 22563T
Pseudomonas asplenii	AB039455	NCPPB 1947T	AJ717432	LMG 2137T	AB039593	NCPPB 1947T
Pseudomonas azotifigens	FN554174	DSM 17556T	FN554729	DSM 17556T	FN554455	DSM 17556T
Pseudomonas azotoformans	AB039411	IAM 1603T	AJ717458	CIP 106744T	AB039547	IAM 1603T
Pseudomonas balearica	AB039394	DSM 6083T	AJ717480	CIP 105297T	AB039605	DSM 6083T
Pseudomonas borbori	FN554175	LMG 23199T	FN554730	LMG 23199T	FN554456	LMG 23199T
Pseudomonas brassicacearum	AM084675	CFBP 11706T	AJ717436	CIP 107059T	AM084334	CFBP 11706T
Pseudomonas brenneri	FN554176	DSM 15294T	AJ717482	CIP 106646T	FN554457	DSM 15294T
Pseudomonas cannabina	FN554177	LMG 5096T	AJ717453	CIP 106140T	FN554458	LMG 5096T
Pseudomonas caricapapayae	AB039454	NCPPB 1873T	AJ717437	LMG 2152T	AB039507	NCPPB 1873T
Pseudomonas cedrina	FN554178	DSM 17516T	AJ717424	CIP 105541T	FN554459	DSM 17516T
Pseudomonas chlororaphis aurantiaca	FN554171	ATCC 33663T	AJ717421	CIP 109718T	FN554452	ATCC 33663T
Pseudomonas chlororaphis aureofaciens	FN554172	LMG 1245T	AJ717426	LMG 1245T	FN554453	LMG 1245T
Pseudomonas cichorii	AB039434	NCPPB 943T	AJ717418	LMG 2162T	AB039526	NCPPB 943T
Pseudomonas congelans	FN554179	LMG 21466T	FN554731	LMG 21466T	FN554460	LMG 21466T
Pseudomonas corrugata	AB039460	NCPPB 2445T	AJ717487	LMG 2172T	AB039566	NCPPB 2445T
Pseudomonas costantinii	FN554180	LMG 22119T	FN554732	LMG 22119T	FN554461	LMG 22119T
Pseudomonas cremoricolorata	FN554181	DSM 17059T	AJ717476	CIP 107616T	FN554462	DSM 17059T
Pseudomonas extremorientalis	FN554182	LMG 19695T	FN554733	LMG 19695T	FN554464	LMG 19695T
Pseudomonas flavescens	FN554183	LMG 18387T	AJ717468	CIP 104204T	FN554465	LMG 18387T
Pseudomonas fluorescens	AB178888	IAM12022T	AJ717451	CIP 69.13T	AB039545	IAM 12022T
Pseudomonas fragi	FN554184	ATCC 4973T	AJ717444	LMG 2191T	FN554466	ATCC 4973T
Pseudomonas frederiksbergensis	AM084676	DSM 13022T	AJ717465	CIP 106887T	AM084335	DSM 13022T
Pseudomonas fulva	AB039395	IAM 1529T	AJ717419	CIP 106765T	AB039586	IAM 1529T
Pseudomonas fuscovaginae	FN554185	LMG 2158T	AJ717433	LMG 2158T	FN554467	LMG 2158T
Pseudomonas gessardii	FN554186	CIP 105469T	AJ717438	CIP 105469T	FN554468	CIP 105469T
Pseudomonas graminis	FN554187	LMG 21611T	AJ717429	CIP 105897T	FN554469	LMG 21611T

Continued on the next page.

Species	*gyrB* accession number	Strain number	*rpoB* accession number	Strain number	*rpoD* accession number	Strain number
Pseudomonas grimontii	FN554188	CIP 106645T	AJ717439	CIP 106645T	FN554470	CIP 106645T
Pseudomonas guineae	FN554189	LMG 24016T	FN554734	LMG 24016T	FN554471	LMG 24016T
Pseudomonas indica	FN554190	LMG 23066T	AJ717481	CIP 107714T	FN554472	LMG 23066T
Pseudomonas jessenii	FN554191	CIP 105274T	AJ717447	CIP 105274T	FN554473	CIP 105274T
Pseudomonas jinjuensis	FN554192	LMG 21316T	FN554735	LMG 21316T	FN554474	LMG 21316T
Pseudomonas knackmussii	FN554193	LMG 23759T	FN554736	LMG 23759T	FN554475	LMG 23759T
Pseudomonas koreensis	FN554194	LMG 21318T	FN554737	LMG 21318T	FN554476	LMG 21318T
Pseudomonas libanensis	FN554195	CIP 105460T	AJ717454	CIP 105460T	FN554477	CIP 105460T
Pseudomonas lini	FN554196	CIP 107460T	AJ717466	CIP 107460T	FN554478	CIP 107460T
Pseudomonas lundensis	FN554197	LMG 13517T	AJ717428	CIP 103272T	FN554479	LMG 13517T
Pseudomonas lutea	FN554198	LMG 21974T	FN554738	LMG 21974T	FN554480	LMG 21974T
Pseudomonas luteola	FN554199	LMG 21607T	AJ717452	CIP 102995T	FN554481	LMG 21607T
Pseudomonas mandelii	FN554200	LMG 2210T	AJ717435	CIP 105273T	FN554482	LMG 2210T
Pseudomonas marginalis	AB039448	NCPPB 667T	AJ717425	LMG 2210T	AB039575	NCPPB 667T
Pseudomonas marincola	FN554201	JCM 14761T	FN554739	JCM 14761T	FN554483	JCM 14761T
Pseudomonas mediterranea	AM084678	CFBP 5447T	AJ717449	CIP 107708T	AM084337	CFBP 5447T
Pseudomonas mendocina	AJ633103	ATCC 25411T	AJ717440	LMG 1223T	AJ633567	ATCC 25411T
Pseudomonas meridiana	FN554203	CIP 108465T	FN554740	CIP 108465T	FN554485	CIP 108465T
Pseudomonas migulae	FN554204	CCUG 43165T	AJ717446	CIP 105470T	FN554486	CCUG 43165T
Pseudomonas mohnii	AM293561	IpA-2T	FN554741	CCUG 53115T	FN554487	CCUG 53115T
Pseudomonas monteilii	FN554205	DSM 14164T	AJ717455	CIP 104883T	FN554488	DSM 14164T
Pseudomonas moorei	AM293560	RW10T	FN554742	CCUG 53114T	FN554489	CCUG 53114T
Pseudomonas moraviensis	FN554206	DSM 16007T	FN554743	DSM 16007T	FN554490	DSM 16007T
Pseudomonas mosselii	FN554207	ATCC BAA-99T	FN554744	ATCC BAA-99T	FN554491	ATCC BAA-99T
Pseudomonas mucidolens	AB039409	IAM 12406T	AJ717427	LMG 2223T	AB039546	IAM 12406T
Pseudomonas nitroreducens	FN554208	ATCC 33634T	AJ717448	CIP 106747T	FN554492	ATCC 33634T
Pseudomonas oleovorans	AB039396	IFO 13583T	AJ717461	LMG 2229T	AB039601	IFO 13583T
Pseudomonas orientalis	FN554209	DSM 17489T	AJ717434	CIP 105540T	FN554493	DSM 17489T
Pseudomonas oryzihabitans	FN554210	LMG 7040T	AJ717470	CIP 102996T	FN554494	LMG 7040T
Pseudomonas otitidis	FN554211	DSM 17224T	FN554745	DSM 17224T	FN554495	DSM 17224T
Pseudomonas pachastrellae	FN554212	CCUG 46540T	FN554746	CCUG 46540T	FN554496	CCUG 46540T
Pseudomonas palleroniana	FN554213	LMG 23076T	FN554747	LMG 23076T	FN554497	LMG 23076T
Pseudomonas panacis	FN554214	CIP 108524T	FN554748	CIP 108524T	FN554498	CIP 108524T
Pseudomonas panipatensis	FN554215	CCM 7469T	FN554749	CCM 7469T	FN554499	CCM 7469T
Pseudomonas parafulva	FJ418638	BCRC 17511T	AJ717471	CIP 107617T	FN554500	DSM 117004T
Pseudomonas peli	FN554217	LMG 23201T	FN554750	LMG 23201T	FN554501	LMG 23201T

Continued on the next page.

Species	*gyrB* accession number	Strain number	*rpoB* accession number	Strain number	*rpoD* accession number	Strain number
Pseudomonas pertucinogena	DQ350613	JCM 11950T	AJ717441	LMG 1874T	EF596883	JCM 11590T
Pseudomonas plecoglossicida	FN554218	CIP 106493T	AJ717456	CIP 106493T	FN554503	CIP 106493T
Pseudomonas poae	FN554219	LMG 21465T	FN554751	LMG 21465T	FN554504	LMG 21465T
Pseudomonas proteolytica	FN554220	CIP 108464T	FN554752	CIP 108464T	FN554505	CIP 108464T
Pseudomonas pseudoalcaligenes	AB039397	IFO 14167T	AJ717430	LMG 1225T	AB039602	IFO 14167T
Pseudomonas psychrophila	FN554221	DSM 17535T	AJ717464	CIP 107901T	FN554506	DSM 17535T
Pseudomonas psychrotolerans	FN554222	LMG 21977T	FN554753	LMG 21977T	FN554507	LMG 21977T
Pseudomonas putida	FJ418635	BCRC 10459T	AJ717474	LMG 2257T	AB039581	ATCC 12633T
Pseudomonas reinekei	AM293559	MT1T	FN554754	CCUG 53116 T	FN678362	DSM 18361T
Pseudomonas resinovorans	FN554223	LMG 2774T	AJ717479	LMG 2774T	FN554509	LMG 2774T
Pseudomonas rhizosphaerae	FN554224	LMG 21640T	FN554755	LMG 21640T	FN554510	LMG 21640T
Pseudomonas rhodesiae	FN554225	LMG 17764T	AJ717431	CIP 104664T	FN554511	LMG 17764T
Pseudomonas salomonii	FN554226	LMG 22120T	FN554756	LMG 22120T	FN554512	LMG 22120T
Pseudomonas savastanoi	AB039469	NCPPB 639T	AJ717422	CIP 103721T	AB039514	NCPPB 639T
Pseudomonas simiae	FN554227	CCUG 50988T	FN554757	CCUG 50988T	FN554513	CCUG 50988T
Pseudomonas straminea	AB039410	IAM 1598T	FN554758	LMG 21615T	AB039600	IAM 1598T
Pseudomonas synxantha	AB039415	IFO 3913T	AJ717420	LMG 2335T	AB039550	IFO 3913T
Pseudomonas syringae	AB039428	PDDCC 3023T	FN554759	ATCC 19310T	AB039516	PDDCC 3023T
Pseudomonas taetrolens	AB039412	IAM 1653T	AJ717423	LMG 2336T	AB039523	IAM 1653T
Pseudomonas thermotolerans	FN554228	CIP 107795T	FN554760	CIP 107795T	FN554514	CIP 107795T
Pseudomonas thivervalensis	AM084679	CFBP 11261T	AM084680	CFBP 11261T	AM084338	CFBP 11261T
Pseudomonas tolaasii	AB039423	NCPPB 2192T	AJ717467	LMG 2342T	AB039561	NCPPB 2192T
Pseudomonas tremae	FN554229	LMG 22121T	FN554761	LMG 22121T	FN554463	LMG 22121T
Pseudomonas trivialis	FN554230	LMG 21464T	FN554762	LMG 21464T	FN554515	LMG 21464T
Pseudomonas umsongensis	FN554231	LMG 21317T	FN554763	LMG 21317T	FN554516	LMG 21317T
Pseudomonas vancouverensis	FN554232	ATCC 700688T	AJ717473	CIP 106707T	FN554517	ATCC 700688T
Pseudomonas veronii	FN554233	LMG 17761T	AJ717445	CIP 104663T	FN554518	LMG 17761T
Pseudomonas viridiflava	AB039427	PDDCC 2848T	FN554764	ATCC 13223T	AB039520	PDDCC 2848T
Pseudomonas xanthomarina	AM905836	CCUG 46543T	FN554765	CCUG 46543T	AM905872	CCUG 46543T

Table S4.2 Overview of the origin and both direct and indirect plant growth promotion properties of the 58 bacterial isolates with antagonistic activity.

			Antagonism[b]		Fungal cell wall degrading enzymes[c]					Plant growth promotion				
Field[a]	Genus	Isolate	*R. solani*	*P. infestans*	Chitinase	Protease	Cellulase	Glucanase	Siderophore production[c]	IAA[d]	ACC[e]	PO_4^{3-} sol[f]	HCN[c]	NH_3[c]
B1	*Pseudomonas*	R-41947	34.54	81.49	-	+	-	-	+	91.21	40	2	+	+
		R-41955	33.34	29.62	-	+	-	-	+	73.36	130	1	-	-
		R-41973	26.9	49.8	-	+	-	-	+	60.5	100	2	-	-
	Bacillus	R-41857	30.12	81.49	-	+	-	-	+	N	60	1	-	-
		R-41858	36.95	77.04	-	+	-	-	+	N	N	1	-	-
		R-41859	37.75	69.62	-	+	-	-	+	N	20	1	-	-
		R-41958	34.94	65.18	-	+	-	-	+	N	160	0	-	+
	Flavobacterium	R-41965	30.93	44.98	-	+	-	-	-	N	70	0	-	-
B2	*Pseudomonas*	R-41739	34.94	30.59	-	-	-	-	+	156.93	20	3	-	-
		R-41757	38.95	43.84	+	+	-	-	+	N	20	5	-	+
		R-41761	47.98	30.59	-	+	-	-	+	N	130	2	-	+
		R-41998	48.59	61.05	-	+	-	-	+	N	220	2	-	-
		R-42010	34.54	81.53	-	+	-	-	+	N	N	1	-	-
		R-42020	28.11	80.72	-	+	-	-	+	N	30	2	+	+
		R-42027	45.78	55.42	-	+	-	-	+	N	200	3	-	-
	Bacillus	R-41753	30.12	66.67	-	+	-	-	+	N	N	1	-	-
		R-41849	35.34	35.56	+	+	-	-	+	N	60	1	-	+
		R-41850	28.92	88.16	-	+	-	-	+	N	140	1	-	+
		R-41855	36.95	75.56	-	+	-	-	+	N	20	1	-	-
B3	*Pseudomonas*	R-41777	42.17	75.81	-	+	-	-	+	65.5	80	3	-	-
		R-41805	33.34	100	-	+	-	-	-	N	40	2	+	+

Continued on the next page.

Field[a]	Genus	Isolate	Antagonism[b]		Fungal cell wall degrading enzymes[c]				Siderophore production[c]	Plant growth promotion			HCN[c]	NH_3[c]
			R. solani	*P. infestans*	Chitinase	Protease	Cellulase	Glucanase		IAA[d]	ACC[e]	PO_4^{3-}sol[f]		
B3	*Pseudomonas*	R-42058	47.39	69.48	-	+	-	-	+	N	230	3	-	+
		R-42071	26.9	78.31	-	+	-	-	+	N	N	3	+	+
	Bacillus	R-41787	41.77	58	-	+	-	-	+	N	N	0	-	-
		R-41798	44.58	59.82	-	+	-	-	+	N	90	0	-	+
		R-41806	30.12	42.93	-	+	-	-	+	N	N	0	-	-
B4	*Pseudomonas*	R-42085	32.53	84.73	-	+	-	-	+	76.21	N	2	-	-
		R-42086	53.01	66.67	-	-	-	-	+	232.64	80	5	-	+
		R-42090	31.72	85.14	-	+	-	-	+	74.07	N	2	+	+
		R-42091	26.51	79.92	-	+	-	-	+	N	N	2	-	-
		R-42098	30.12	84.34	-	+	-	-	+	91.21	40	3	+	+
		R-42137	36.95	83.94	-	+	-	-	+	N	N	2	-	-
		R-43978	31.33	37.75	-	-	-	-	+	88.36	140	2	-	-
	Bacillus	R-41815	44.58	88.89	-	+	-	-	+	N	60	0	-	+
		R-42116	35.34	67.47	+	+	-	-	+	N	40	1	-	-
		R-42124	38.16	55.02	-	+	-	-	+	N	N	0	-	-
	Pedobacter	R-41842	24.9	40	-	-	-	-	-	N	N	0	-	-
	Enterobacter	R-42089	31.72	53.41	-	-	-	-	+	N	N	2	-	-
		R-42141	53.41	50.2	-	-	-	-	+	N	50	2	-	+
	Curtobacterium	R-42100	30.52	66.66	-	+	-	+	+	136.21	N	1	-	-
		R-42111	35.34	57.43	-	+	-	+	+	117.64	N	1	-	-
P1	*Pseudomonas*	R-42286	32.93	0	-	-	-	-	+	N	N	2	-	-

Continued on the next page.

Field[a]	Genus	Isolate	Antagonism[b]		Fungal cell wall degrading enzymes[c]				Siderophore production[c]	Plant growth promotion			HCN[c]	NH_3[c]
			R. solani	*P. infestans*	Chitinase	Protease	Cellulase	Glucanase		IAA[d]	ACC[e]	PO_4^{3-}sol[f]		
P1	*Pseudomonas*	R-42287	34.94	0	-	-	-	-	+	N	N	2	-	-
	Bacillus	R-42276	36.54	62.65	-	+	-	-	+	N	N	3	-	-
		R-42277	38.95	75.9	-	+	-	-	+	N	N	2	-	-
		R-42278	32.53	49.8	-	+	-	-	-	N	N	2	-	-
		R-42289	27.71	39.36	-	+	-	-	+	N	N	2	-	-
P2	*Bacillus*	R-42292	43.37	73.1	+	+	-	-	+	N	N	2	-	-
	Paenibacillus	R-42302	37.75	57.83	-	+	+	+	+	N	N	1	-	-
P3	*Pseudomonas*	R-42357	44.98	63.86	-	+	-	-	+	N	230	4	-	+
		R-42358	47.8	48.19	-	+	-	-	+	N	N	3	-	-
	Bacillus	R-42363	42.98	83.53	-	+	-	-	+	N	N	1	-	-
P4	*Pseudomonas*	R-43582	41.36	43.77	-	+	-	-	+	N	310	3	-	+
		R-43628	51	40.96	-	+	-	-	+	N	230	2	-	+
		R-43631	45.39	31.33	-	+	-	-	+	N	150	3	-	+
		R-43638	52.2	28.55	-	+	-	-	+	N	190	2	-	+
	Bacillus	R-43629	30.93	72.69	+	+	-	-	+	N	20	0	-	-
		R-43639	45.78	66.66	+	+	-	-	+	N	N	0	-	-

[a], B1-4= Bolivia, field 1-4; P1-4= Peru field 1-4

[b], Percentage antagonism, calculated from

$$\frac{\text{(Total growth of the control (=diameter of petri dish) - measured growth with bacteria)}}{\text{Total growth}} \times 100\%$$

[c], + Represents production on plate assay

[d], IAA: Indole-3-acetic acid production ($mg.ml^{-1}$); N: No activity detected

[e], ACC: 1-Aminocyclopropane-1-carboxylate deaminase activity (nmol (α-ketobutyrate).$mg^{-1}.h^{-1}$); N: No activity detected

[f], PO_4^{3-}sol: Phosphate solubilization; 0= 0 mm, 1= 1-5 mm, 2= 6-10 mm, 3= 11-15 mm, 4= 16-20 mm, 5= >20 mm clearing zone on plate

CHAPTER 5

Table S5.1 Overview of research parameters that were used to measure the phylogenetic information contained within short read sequences and the OTU richness calculated from each library.

Libraries[a]	Variable region	PC patristic[b]	wRMSD[c]	RF[d]	WRF1[e]	WRF2[f]	RF-WRF1	RF-WRF2	OTU 0.01 cutoff[g]	OTU 0.02 cutoff[g]	OTU 0.03 cutoff[g]
NFL(1) vs NFL(2)	V1-V9	0.928	0.0098	585.3	121.07	155.98	462.93	428.02	-	-	-
NFL(1) vs NFL(3)		0.979	0.0041								
NFL(2) vs NFL(3)		0.943	0.0091								
338f(1) vs 338f(2)	V3	0.799	0.0135	1260.6	92.97	118.42	1167.63	1142.18	0.86	0.87	0.89
338f (1) vs 338f (3)		0.697	0.0182								
338f (1) vs 338f (4)		0.767	0.0156								
338f (1) vs 338f (5)		0.911	0.0098								
338f (2) vs 338f (3)		0.858	0.0143								
338f (2) vs 338f (4)		0.821	0.0137								
338f (2) vs 338f (5)		0.819	0.0139								
338f (3) vs 338f (4)		0.685	0.0178								
338f (3) vs 338f (5)		0.789	0.0157								
338f (4) vs 338f (5)		0.802	0.0154								
338r(1) vs 338r (2)	V2	0.846	0.0141	1359	85.17	110.33	1273.83	1248.67	0.82	0.84	0.84
338r (1) vs 338r (3)		0.851	0.0140								
338r (1) vs 338r (4)		0.735	0.0191								
338r (1) vs 338r (5)		0.828	0.0138								
338r (2) vs 338r (3)		0.914	0.0100								
338r (2) vs 338r (4)		0.642	0.0193								
338r (2) vs 338r (5)		0.828	0.0123								
338r (3) vs 338r (4)		0.699	0.0194								
338r (3) vs 338r (5)		0.826	0.0129								
338r (4) vs 338r (5)		0.729	0.0175								

Continued on the next page.

Libraries[a]	Variable region	PC patristic[b]	wRMSD[c]	RF[d]	WRF1[e]	WRF2[f]	RF-WRF1	RF-WRF2	OTU 0.01 cutoff[g]	OTU 0.02 cutoff[g]	OTU 0.03 cutoff[g]
518f(1) vs 518f (2)	V4	0.970	0.0062	1033.8	92.76	122.92	941.04	910.88	0.79	0.79	0.81
518f (1) vs 518f (3)		0.969	0.0059								
518f (1) vs 518F(4)		0.949	0.0077								
518f (1) vs 518f (5)		0.956	0.0070								
518f (2) vs 518f (3)		0.952	0.0076								
518f (2) vs 518f (4)		0.931	0.0086								
518f (2) vs 518f (5)		0.934	0.0084								
518f (3) vs 518f (4)		0.937	0.0082								
518f (3) vs 518f (5)		0.942	0.0075								
518f (4) vs 518f (5)		0.956	0.0074								
518r(1) vs 518r (2)	V3	0.905	0.0112	1245.6	91.78	118.82	1153.82	1126.78	0.86	0.85	0.87
518r (1) vs 518r (3)		0.660	0.0206								
518r (1) vs 518r (4)		0.957	0.0069								
518r (1) vs 518r (5)		0.871	0.0117								
518r (2) vs 518r (3)		0.660	0.0201								
518r (2) vs 518r (4)		0.892	0.0115								
518r (2) vs 518r (5)		0.813	0.0151								
518r (3) vs 518r (4)		0.653	0.0211								
518r (3) vs 518r (5)		0.661	0.0213								
518r (4) vs 518r (5)		0.839	0.0130								

Continued on the next page.

Libraries[a]	Variable region	PC patristic[b]	wRMSD[c]	RF[d]	WRF1[e]	WRF2[f]	RF-WRF1	RF-WRF2	OTU 0.01 cutoff[g]	OTU 0.02 cutoff[g]	OTU 0.03 cutoff[g]
799f(1) vs 799f (2)	V5	0.888	0.0106	1300.2	85.86	112.82	1214.34	1187.38	0.67	0.61	0.59
799f (1) vs 799f (3)		0.821	0.0130								
799f (1) vs 799f (4)		0.941	0.0095								
799f (1) vs 799f (5)		0.941	0.0092								
799f (2) vs 799f (3)		0.914	0.0096								
799f (2) vs 799f (4)		0.822	0.0126								
799f (2) vs 799f (5)		0.817	0.0130								
799f (3) vs 799f (4)		0.741	0.0155								
799f (3) vs 799f (5)		0.740	0.0159								
799f (4) vs 799f (5)		0.929	0.0084								
799r(1) vs 799r (2)	V4	0.92	0.0098	1143.6	99.29	128.54	1044.31	1014.06	0.81	0.77	0.79
799r (1) vs 799r (3)		0.91	0.0116								
799r (1) vs 799r (4)		0.89	0.0118								
799r (1) vs 799r (5)		0.95	0.0108								
799r (2) vs 799r (3)		0.95	0.0083								
799r (2) vs 799r (4)		0.93	0.0088								
799r (2) vs 799r (5)		0.92	0.0109								
799r (3) vs 799r (4)		0.93	0.0096								
799r (3) vs 799r (5)		0.92	0.0106								
799r (4) vs 799r (5)		0.91	0.0111								

Continued on the next page.

Libraries[a]	Variable region	PC patristic[b]	wRMSD[c]	RF[d]	WRF1[e]	WRF2[f]	RF-WRF1	RF-WRF2	OTU 0.01 cutoff[g]	OTU 0.02 cutoff[g]	OTU 0.03 cutoff[g]
926f(1) vs 926f (2)	V6	0.871	0.0129	1423	103.08	127.38	1319.92	1295.62	0.81	0.77	0.79
926f (1) vs 926f (3)		0.841	0.0145								
926f (1) vs 926f (4)		0.858	0.0161								
926f (1) vs 926f (5)		0.930	0.0118								
926f (2) vs 926f (3)		0.836	0.0132								
926f (2) vs 926f (4)		0.847	0.0157								
926f (2) vs 926f (5)		0.851	0.0132								
926f (3) vs 926f (4)		0.820	0.0173								
926f (3) vs 926f (5)		0.863	0.0136								
926f (4) vs 926f (5)		0.849	0.0165								
926r(1) vs 926r (2)	V5	0.857	0.0136	1228.4	86.39	113.04	1142.01	1115.36	0.73	0.69	0.7
926r (1) vs 926r (3)		0.870	0.0160								
926r (1) vs 926r (4)		0.819	0.0143								
926r (1) vs 926r (5)		0.783	0.0151								
926r (2) vs 926r (3)		0.884	0.0140								
926r (2) vs 926r (4)		0.924	0.0082								
926r (2) vs 926r (5)		0.812	0.0143								
926r (3) vs 926r (4)		0.769	0.0162								
926r (3) vs 926r (5)		0.844	0.0155								
926r (4) vs 926r (5)		0.729	0.0150								

Continued on the next page.

Libraries[a]	Variable region	PC patristic[b]	wRMSD[c]	RF[d]	WRF1[e]	WRF2[f]	RF-WRF1	RF-WRF2	OTU 0.01 cutoff[g]	OTU 0.02 cutoff[g]	OTU 0.03 cutoff[g]
1062f(1) vs 1062f (2)	V7&8	0.95	0.0078	1212.6	75.51	102.48	1137.09	1110.12	0.68	0.64	0.6
1062f (1) vs 1062f (3)		0.88	0.0105								
1062f (1) vs 1062f (4)		0.90	0.0105								
1062f (1) vs 1062f (5)		0.93	0.0082								
1062f (2) vs 1062f (3)		0.87	0.0107								
1062f (2) vs 1062f (4)		0.91	0.0103								
1062f (2) vs 1062f (5)		0.90	0.0097								
1062f (3) vs 1062f (4)		0.78	0.0147								
1062f (3) vs 1062f (5)		0.89	0.0106								
1062f (4) vs 1062f (5)		0.84	0.0129								
1062r(1) vs 1062r (2)	V6	0.742	0.0164	1432.8	107.86	130.78	1324.94	1302.02	0.79	0.82	0.84
1062r (1) vs 1062r (3)		0.708	0.0179								
1062r (1) vs 1062r (4)		0.776	0.0152								
1062r (1) vs 1062r (5)		0.832	0.0163								
1062r (2) vs 1062r (3)		0.792	0.0155								
1062r (2) vs 1062r (4)		0.817	0.0145								
1062r (2) vs 1062r (5)		0.770	0.0170								
1062r (3) vs 1062r (4)		0.698	0.0173								
1062r (3) vs 1062r (5)		0.830	0.0139								
1062r (4) vs 1062r (5)		0.722	0.0172								

[a], NFL= Nearly Full-Length
[b], PC= Pearson Correlation
[c], wRMSD= Weighted Root Mean Square Deviation
[d], RF = average Robinson Foulds distance between 5 best ML trees
[e],WRF1 = average Weighted Robinson Foulds distances between 5 best ML trees based on the sum of the supports of the unique bipartitions
[f], WRF2 = average Weighted Robinson Foulds distance between 5 best ML trees based on the sum of the supports of the unique bipartitions plus the difference of support values amongst the shared bipartitions
[g], The ratios of the number OTU's obtained with short read sequence libraries to the number of OTU's obtained with the nearly full-length library

CHAPTER 6

Table S6.1 List of *Pseudomonas* strains used for TaxonGap analysis.

Species name	Accession number				Subgroup
	16S rRNA	*rpoB*	*rpoD*	*gyrB*	
Pseudomonas asplenii^T	AB021397	AJ717432	AB039593	AB039455	*P. asplenii* subgroup
Pseudomonas fuscovaginae^T	FJ483519	AJ717433	FN554467	FN554185	*P. asplenii* subgroup
Pseudomonas chlororaphis subsp. *aurantiaca*^{T}	DQ682655	AJ717421	FN554452	FN554171	*P. chlororaphis* subgroup
Pseudomonas chlororaphis subsp. *aureofaciens*^{T}	AY509898	FJ652689	FN554453	FN554172	*P. chlororaphis* subgroup
Pseudomonas chlororaphis subsp. *chlororaphis*^{T}	Z76673	FJ652691	AB039549	FJ652718	*P. chlororaphis* subgroup
Pseudomonas brassicacearum^T	AF100321	AJ717436	AM084334	AM084675	*P. corrugata* subgroup
Pseudomonas corrugata^T	D84012	AJ717487	AB039566	AB039460	*P. corrugata* subgroup
Pseudomonas kilonensis^T	AJ292426	AJ717472	---	---	*P. corrugata* subgroup
Pseudomonas thivervalensis^T	AF100323	AM084680	AM084338	AM084679	*P. corrugata* subgroup
Pseudomonas antarctica^T	AJ537601	FN554727	FN554450	FN554169	*P. fluorescens* subgroup
Pseudomonas azotoformans^T	D84009	AJ717458	AB039547	AB039411	*P. fluorescens* subgroup
Pseudomonas cedrina subsp. *cedrina*^{T}	AF064461	AJ717424	FN554459	FN554178	*P. fluorescens* subgroup
Pseudomonas cedrina subsp. *fulgida*^{T}	AJ492830	HE586401	HE586449	---	*P. fluorescens* subgroup
Pseudomonas costantinii^T	AF374472	FN554732	FN554461	FN554180	*P. fluorescens* subgroup
Pseudomonas extremorientalis^T	AF405328	FN554733	FN554464	FN554182	*P. fluorescens* subgroup
Pseudomonas fluorescens^T	D84013	AJ717451	AB039545	D86016	*P. fluorescens* subgroup
Pseudomonas grimontii^T	AF268029	AJ717439	FN554470	FN554188	*P. fluorescens* subgroup
Pseudomonas libanensis^T	AF057645	AJ717454	FN554477	FN554195	*P. fluorescens* subgroup
Pseudomonas lurida^T	AJ581999	HE586402	HE586451	---	*P. fluorescens* subgroup
Pseudomonas marginalis^T	Z76663	AJ717425	AB039575	AB039448	*P. fluorescens* subgroup
Pseudomonas orientalis^T	AF064457	AJ717434	FN554493	FN554209	*P. fluorescens* subgroup
Pseudomonas palleroniana^T	AY091527	FN554747	FN554497	FN554213	*P. fluorescens* subgroup
Pseudomonas panacis^T	AY787208	FN554748	FN554498	FN554214	*P. fluorescens* subgroup
Pseudomonas poae^T	AJ492829	FN554751	FN554504	FN554219	*P. fluorescens* subgroup
Pseudomonas rhodesiae^T	AF064459	AJ717431	FN554511	FN554225	*P. fluorescens* subgroup
Pseudomonas salomonii^T	AY091528	FN554756	FN554512	FN554226	*P. fluorescens* subgroup
Pseudomonas simiae^T	AJ936933	FN554757	FN554513	FN554227	*P. fluorescens* subgroup
Pseudomonas synxantha^T	D84025	AJ717420	AB039550	AB039415	*P. fluorescens* subgroup
Pseudomonas tolaasii^T	AF255336	AJ717467	FN645158	FN645137	*P. fluorescens* subgroup
Pseudomonas trivialis^T	AJ492831	FN554762	FN554515	FN554230	*P. fluorescens* subgroup
Pseudomonas veronii^T	AF064460	AJ717445	FN554518	FN554233	*P. fluorescens* subgroup
Pseudomonas fragi^T	AF094733	AJ717444	FN554466	FN554184	*P. fragi* subgroup
Pseudomonas lundensis^T	AB021395	AJ717428	FN554479	FN554197	*P. fragi* subgroup
Pseudomonas psychrophila^T	AB041885	AJ717464	FN554506	FN554221	*P. fragi* subgroup
Pseudomonas taetrolens^T	D84027	AJ717423	AB039523	AB039412	*P. fragi* subgroup

Continued on the next page.

Species name	Accession number				Subgroup
	16S rRNA	*rpoB*	*rpoD*	*gyrB*	
Pseudomonas brenneri^T	AF268968	AJ717482	FN554457	FN554176	*P. gessardii* subgroup
Pseudomonas gessardii^T	AF074384	AJ717438	FN554468	FN554186	*P. gessardii* subgroup
Pseudomonas meridiana^T	AJ537602	FN554740	HE586433	HE586495	*P. gessardii* subgroup
Pseudomonas mucidolens^T	D84017	AJ717427	AB039546	AB039409	*P. gessardii* subgroup
Pseudomonas proteolytica^T	AJ537603	FN554752	FN554505	FN554220	*P. gessardii* subgroup
Pseudomonas jessenii^T	AF068259	AJ717447	FN554473	FN554191	*P. jessenii* subgroup
Pseudomonas mohnii^T	AM293567	FN554741	FN554487	AM293561	*P. jessenii* subgroup
Pseudomonas moorei^T	AM293566	FN554742	FN554489	AM29560	*P. jessenii* subgroup
Pseudomonas reinekei^T	AM293565	FN554754	FN554508	AM293559	*P. jessenii* subgroup
Pseudomonas umsongensis^T	AF468450	FN554763	FN554516	FN554231	*P. jessenii* subgroup
Pseudomonas vancouverensis^T	AJ011507	AJ717473	FN554517	FN554232	*P. jessenii* subgroup
Pseudomonas koreensis^T	AF468452	FN554737	FN554476	FN554194	*P. koreensis* subgroup
Pseudomonas moraviensis^T	AY970952	FN554743	FN554490	FN554206	*P. koreensis* subgroup
Pseudomonas frederiksbergensis^T	FR750403	AJ717465	AM084335	AM084676	*P. mandelii* subgroup
Pseudomonas lini^T	AY035996	AJ717466	FN554478	FN554196	*P. mandelii* subgroup
Pseudomonas mandelii^T	AF058286	AJ717435	FN554482	FN554200	*P. mandelii* subgroup
Pseudomonas migulae^T	AF074383	AJ717446	FN554486	FN554204	*P. mandelii* subgroup

Table S6.2 *rpoD* and 16S rRNA gene accession numbers of *Pseudomonas* strains used for taxonomic assignments and phylogenetic tree comparisons.

***Pseudomonas* species**	**Strain number**	**Accession number *rpoD***	**Accession number 16S**
Pseudomonas abietaniphila	ATCC 700689T	FN554447	AJ011504
Pseudomonas aeruginosa	IFO 12689T	AB039607	HE978271
Pseudomonas agarici	NCPPB 2289T	AB039563	AJ308298
Pseudomonas alcaligenes	IFO 14159T	AB039606	HM190231
Pseudomonas alcaliphila	LMG 23134T	FN554448	AB030583
Pseudomonas amygdali	CFBP 3205T	JN185893	Z76654
Pseudomonas anguilliseptica	LMG 21629T	FN554449	AB021376
Pseudomonas antarctica	LMG 22709T	FN554450	AJ537601
Pseudomonas argentinensis	MG 22563T	FN554451	AY691188
Pseudomonas arsenicoxydans	CECT 7543T	HE800488	FN645213
Pseudomonas asplenii	NCPPB 1947T	AB039593	AB021397
Pseudomonas avellanae	CIP 105176T	FN554454	---
Pseudomonas azotifigens	DSM 17556T	FN554455	AB189452
Pseudomonas azotoformans	IAM 1603T	AB039547	D84009
Pseudomonas azotoformans	LMG 21611T	FN554469	D84009
Pseudomonas baetica	a390T	FN678357	FM201274
Pseudomonas balearica	DSM 6083T	AB039605	U26418
Pseudomonas bauzanensis	DSM 22558T	HE800489	GQ161991
Pseudomonas benzenivorans	DSM 8628T	HE800490	FM208263
Pseudomonas borbori	LMG 23199T	FN554456	AM114527
Pseudomonas brassicacearum	CFBP 11706T	AM084334	AF100321
Pseudomonas brenneri	DSM 15294T	FN554457	AF268968
Pseudomonas caeni	CECT 7778T	HE800491	EU620679
Pseudomonas cannabina	LMG 5096T	FN554458	AJ492827
Pseudomonas caricapapayae	NCPPB 1873T	AB039507	D84010
Pseudomonas cedrina	DSM 17516T	FN554459	AF064461
Pseudomonas cedrina subsp. *fulgida*	LMG 21467T	HE586449	AJ492830
Pseudomonas chlororaphis subsp. *aurantiaca*	ATCC 33663T	FN554452	DQ682655
Pseudomonas chlororaphis subsp. *aureofaciens*	LMG 1245T	FN554453	FJ652608
Pseudomonas chlororaphis subsp. *chlororaphis*	IFO 3904T	AB039549	Z76673
Pseudomonas cichorii	NCPPB 943T	AB039526	JX913784
Pseudomonas citronellolis	NCIMB 12783T	AB039604	AB021396
Pseudomonas composti	CECT 7516T	FR716577	FN429930
Pseudomonas congelans	LMG 21466T	FN554460	AJ492828
Pseudomonas corrugata	NCPPB 2445T	AB039566	D84012

Continued on the next page.

***Pseudomonas* species**	**Strain number**	**Accession number *rpoD***	**Accession number 16S**
Pseudomonas costantinii	LMG 22119T	FN554461	AF374472
Pseudomonas cremoricolorata	DSM 17059T	FN554462	AB060137
Pseudomonas cuatrocienegasensis	LMG 24676T	FR716578	EU791281
Pseudomonas deceptionensis	M1T	GU936596	GU936597
Pseudomonas delhiensis	RLD1T	HE800493	DQ339153
Pseudomonas duriflava	KCTC 22129T	HE800494	EU046271
Pseudomonas extremaustralis	DSM 17835T	JN589935	AJ583501
Pseudomonas extremorientalis	LMG 19695T	FN554464	AF405328
Pseudomonas ficuserectae	JCM 2400T	AB039501	AB021378
Pseudomonas flavescens	LMG 18387T	FN554465	U01916
Pseudomonas fluorescens	IAM 12022T	AB039545	D84013
Pseudomonas fragi	ATCC 4973T	FN554466	AF094733
Pseudomonas frederiksbergensis	DSM 13022T	AM084335	AJ249382
Pseudomonas fulva	IAM 1529T	AB039586	AB046996
Pseudomonas fuscovaginae	LMG 2158T	FN554467	FJ483519
Pseudomonas gessardii	CIP 105469T	FN554468	AF074384
Pseudomonas grimontii	CIP 106645T	FN554470	AF268029
Pseudomonas guineae	LMG 24016T	FN554471	AM491810
Pseudomonas indica	LMG 23066T	FN554472	AF302795
Pseudomonas japonica	JCM 21532T	HE577795	AB126621
Pseudomonas jessenii	CIP 105274T	FN554473	AF068259
Pseudomonas jinjuensis	LMG 21316T	FN554474	AF468448
Pseudomonas kilonensis	52020T	AM084336	AJ292426
Pseudomonas knackmussii	LMG 23759T	FN554475	AF039489
Pseudomonas koreensis	LMG 21318T	FN554476	AF468452
Pseudomonas libanensis	CIP 105460T	FN554477	AF057645
Pseudomonas lini	CIP 107460T	FN554478	AY035996
Pseudomonas lundensis	LMG 13517T	FN554479	AB021395
Pseudomonas lurida	LMG 21995T	HE586451	AJ581999
Pseudomonas lutea	LMG 21974T	FN554480	AY364537
Pseudomonas mandelii	LMG 21607T	FN554481	AF058286
Pseudomonas marginalis	NCPPB 667T	AB039575	Z76663
Pseudomonas marginalis	LMG 2210T	FN554482	Z76663
Pseudomonas marincola	JCM 14761T	FN554483	AB301071
Pseudomonas mediterranea	CFBP 5447T	AM084337	AF386080
Pseudomonas meliae	CCUG 51503T	FN554484	AB021382

Continued on the next page.

***Pseudomonas* species**	**Strain number**	**Accession number *rpoD***	**Accession number 16S**
Pseudomonas mendocina	ATCC 25411T	AJ633567	D84016
Pseudomonas meridiana	CIP 108465T	FN554485	AJ537602
Pseudomonas migulae	CCUG 43165T	FN554486	AF074383
Pseudomonas mohnii	CCUG 53115T	FN554487	AM293567
Pseudomonas monteiliin	DSM 14164T	FN554488	AF064458
Pseudomonas moorei	DSM 12647T	FN678363	AM293566
Pseudomonas moraviensis	DSM 16007T	FN554490	AY970952
Pseudomonas mosselii	ATCC BAA99T	FN554491	AF072688
Pseudomonas mucidolens	IAM 12406T	AB039546	D84017
Pseudomonas nitroreducens	ATCC 33634T	FN554492	AM088473
Pseudomonas oleovorans	IFO 13583T	AB039601	D84018
Pseudomonas oleovorans subsp. *lubricantis*	RS1T	EF667505	DQ842018
Pseudomonas orientalis	DSM 17489T	FN554493	AF064457
Pseudomonas oryzihabitans	LMG 7040T	FN554494	D84004
Pseudomonas otitidis	DSM 17224T	FN554495	AY953147
Pseudomonas pachastrellae	CCUG 46540T	FN554496	AB125366
Pseudomonas palleroniana	LMG 23076T	FN554497	AY091527
Pseudomonas panacis	CIP 108524T	FN554498	AY787208
Pseudomonas panipatensis	CCM 7469T	FN554499	EF424401
Pseudomonas parafulva	DSM 117004T	FN554500	AB046999
Pseudomonas pelagia	CECT 7689T	FN908495	EU888911
Pseudomonas peli	LMG 23201T	FN554501	AM114534
Pseudomonas pertucinogena	JCM 11590T	EF596883	AB021380
Pseudomonas plecoglossicida	CIP 106493T	FN554503	AB009457
Pseudomonas poae	LMG 21465T	FN554504	AJ492829
Pseudomonas pohangensis	DSM 17875T	HE800498	DQ339144
Pseudomonas proteolytica	CIP 108464T	FN554505	AJ537603
Pseudomonas psychrophila	DSM 17535T	FN554506	AB041885
Pseudomonas psychrotolerans	LMG 21977T	FN554507	AJ575816
Pseudomonas putida	ATCC 12633T	AB039581	D84020
Pseudomonas reinekei	DSM 18361T	FN678362	AM293565
Pseudomonas resinovorans	LMG 2774T	FN554509	Z76668
Pseudomonas rhizosphaerae	LMG 21640T	FN554510	AY152673
Pseudomonas rhodesiae	LMG 17764T	FN554511	AF064459
Pseudomonas salomonii	LMG 22120T	FN554512	AY091528
Pseudomonas saponiphila	DSM 9751T	HE800499	FM208264
Pseudomonas savastanoi	NCPPB 639T	AB039514	AB021402

Continued on the next page.

***Pseudomonas* species**	**Strain number**	**Accession number *rpoD***	**Accession number 16S**
Pseudomonas segetis	IMSNU 14101T	HE800500	AY770691
Pseudomonas seleniipraecipitans	LMG 25475T	HE800501	FJ422810
Pseudomonas simiae	CCUG 50988T	FN554513	AJ936933
Pseudomonas straminea	IAM 1598T	AB039600	D84023
Pseudomonas stutzeri	CCUG 11256T	AJ631316	AF094748
Pseudomonas synxantha	DSM 18928T	JN589943	D84025
Pseudomonas syringae pv. *syringae strain*	CECT 4429T	JX867790	HM190217
Pseudomonas taeanensis	KCTC 22612T	HE800502	FJ424813
Pseudomonas taetrolens	IAM 1653T	AB039523	D84027
Pseudomonas taiwanensis	DSM 21245T	HE577796	EU103629
Pseudomonas thermotolerans	CIP 107795T	FN554514	AJ311980
Pseudomonas thivervalensis	CFBP 11261T	AM084338	AF100323
Pseudomonas tolaasii	NCPPB 2192T	AB039561	AF255336
Pseudomonas tremae	LMG 22121T	FN554463	AJ492826
Pseudomonas trivialis	LMG 21464T	FN554515	AJ492831
Pseudomonas tuomuerensis	JCM 14085T	AB571152	DQ868767
Pseudomonas umsongensis	LMG 21317T	FN554516	AF468450
Pseudomonas vancouverensis	ATCC 700688T	FN554517	AJ011507
Pseudomonas veronii	LMG 17761T	FN554518	AF064460
Pseudomonas viridiflava	PDDCC 2848T	AB039520	AY180972
Pseudomonas vranovensis	DSM 16006T	HE577793	AY970951
Pseudomonas xanthomarina	CCUG 46543T	AM905872	AB176954

CURRICULUM VITAE

PERSONALIA

Jonas Ghyselinck

Huisgaverstraat 31

9750 Ouwegem

Born December 1, 1985, Ghent, Belgium

EDUCATIONAL BACKGROUND

2009 - present PhD studies at the Laboratory of Microbiology, Faculty of Sciences, Department of Biochemistry and Microbiology; Ghent University. PhD dissertation "Understanding the composition and role of the prokaryotic diversity in the potato rhizosphere for crop improvement in the Andes"
Promotor: Prof. Dr. Paul De Vos

2008 - 2009 Theoretical part of the Specific Teacher's Training (SLO) at the Faculty of Psychology and Educational Sciences; Ghent University.

2007 - 2008 Master of Science in "Applied Microbial Systematics" (graduated magna cum laude). Faculty of Sciences, Department of Biochemistry and Microbiology; Ghent University. Dissertation: "Antimicrobial resistance in human and broiler chicken *Escherichia coli* isolates"

2003 - 2007 Master Industrial Engineering Biochemistry, Environmental Biotechnology (graduated cum laude). Faculty of Bioscience Engineering; University College Ghent - Ghent University. Dissertation: "Optimization of a pilot-scale ultrafiltration apparatus and feasibility study on the re-use of vegetable rinsing water obtained from activated sludge wastewater treatments".

1997 - 2003 Koninklijk Atheneum Erasmus Deinze (Science/Mathematics)

WORK EXPERIENCE

February 2009 – March 2009	Full-time teacher (assistant) at Master Industrial Engineering Biochemistry, Environmental Biotechnology. Faculty of Bioscience Engineering, University College Ghent - Ghent University.

SCIENTIFIC OUTPUT

A1 Publications

Ghyselinck, J., Van Hoorde, K., Hoste, B., Heylen, K. & De Vos, P. (2011). Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. J Microbiol Meth **86**(3), 327-336

Ghyselinck, J., Velivelli, S. L. S., Heylen, K., O'Herlihy, E., Franco, J., Rojas, M., De Vos, P. & Prestwich, B. D. (2013). Bioprospecting in potato fields in the Central Andean Highlands: Screening of rhizobacteria for plant growth-promoting properties. Syst Appl Microbiol **36**(2), 116-127

Ghyselinck, J., Pfeiffer, S., Heylen, K., Sessitsch, A. & De Vos, P. (2013). The effect of primer choice on the outcome of Next Generation Sequencing studies. PLOS ONE, minor revisions

Ghyselinck, J., Coorevits, A., Van Landschoot, A., Heylen, K. & De Vos, P. (2013). *rpoD* gene sequence based evaluation of cultured Andean potato-associated *Pseudomonas* diversity. Microbiology, revisions

C1 Publications

Velivelli, S., O'Herlihy, E., Janczura, B., Prestwich, B. D., **Ghyselinck, J.** & De Vos, P. (2012). Efficacy of Rhizobacteria on plant growth-promotion and disease suppression in vitro. Acta Horticulturae **961**, 525-532

Oral presentations

Ghyselinck, J. (2010). Cultivation dependent diversity assessment of bacterial populations. Symposium "Utilization of beneficial microorganisms in agricultural systems: exploring microbial diversity for novel application". February 16, 2010. Lima, Peru.

Ghyselinck, J. (2012). Cultivation dependent diversity assessment of bacterial populations. Lectures on plant growth-promotion by micro-organisms. February 24, 2012. Cochabamba, Bolivia.

Poster presentations and abstracts

Ghyselinck, J.,De Bock, J., De Vos, P. & Heylen, K. (2010). Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. Molecular dialogue in host-parasite interaction. November 18, 2010. Brussels, Belgium.

Ghyselinck, J., O'Herlihy, E., Heylen, K., Velivelli, S. L. S., De Vos, P. & Prestwich, B. D. (2011). Screening for antagonistic and plant growth promoting properties in bacteria isolated from the Central Andean Highlands. Spring symposium NVvM/NVMM. April 18-20, 2011. Arnhem, Netherlands.

Ghyselinck, J., O' Herlihy, E., Heylen, K., Velivelli, S. L. S., Prestwich, B. D. & De Vos, P. (2011). Screening for antagonistic and plant growth-promoting properties in bacteria isolated from the Central Andean Highlands. Ecology of Soil Microorganisms - Microbes as Important Drivers of Soil Processes. April 27 – May 1, 2011. Prague, Czech Republic.

Velivelli, S. L. S., O' Herlihy, E., Bogumila, J., Prestwich, B. D., **Ghyselinck, J.** & De Vos, P. (2011). Efficacy of Rhizobacteria on Plant Growth Promotion and Disease Suppression In vitro. VII International Symposium on In Vitro Culture and Horticultural Breeding. September 18-22, 2011. Ghent, Belgium.

Ghyselinck J., O' Herlihy, E., Heylen K., Velivelli, S. L. S., Prestwich, B. D. & De Vos, P. (2011). Screening for antagonistic and plant growth-promoting properties in bacteria isolated from the Central Andean Highlands. Life, Death and Survival of Micro-organisms. November 16th, 2011. Brussels, Belgium.

Ghyselinck, J., O'Herlihy, E., Prestwich, B. D., Velivelli, S. L. S., Van Hoorde, K., Hoste, B., Heylen, K. & De Vos, P. (2012). Bacterial Diversity and Plant Growth-Promotion Potential in Andean Potato Rhizosphere Soils. The Power of the Small. August 19-24, 2012. Copenhagen, Denmark.

Ghyselinck, J., Velivelli, S. L. S., Heylen, K., O'Herlihy, E., De Vos, P. & Prestwich, B. D. (2012). Bioprospecting in Potato Fields in the Central Andean Highlands: The Selection of Biocontrol Agents with Plant Growth-Promoting Properties. Posttranscriptional regulation and epigenetics in microorganisms. November 30, 2012. Brussels, Belgium.