Structure and assembly cues of Arabidopsis root-inhabiting bacterial communities and comparative genomics of selected Rhizobium members

Inaugural-Dissertation

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwisssenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Matthias Rott
aus Troisdorf

Köln, Oktober 2012

Die vorliegende Arbeit wurde am Max-Planck-Institut für Pflanzenzüchtungsforschung in Köln in der Abteilung für Pflanze-Mikroben Interaktionen (Direktor: Prof. Dr. P. Schulze-Lefert) angefertig.

Berichterstatter:	Prof. Dr. Paul Schulze-Lefert
	Prof. Dr. Martin Hülskamp
	Prof. Dr. Allan Downie
Prüfungsvorsitzender:	Prof. Dr. Ulf Ingo Flügge
Tag der Disputation:	03.12.2012

Yo, Adrian – I did it
(*R. Balboa*)

Publications

Buch, F., **M. Rott**, S. Rottloff, C. Paetz, I. Hilke, M. Raessler, and A. Mithoefer. "Carnivorous Nepenthes Plants Generate Sterile Pitcher Fluids Unsuitable for Microbial Growth: New Findings Strengthen an Old Hypothesis." *Annals of Botany* (2013).

Bulgarelli, D.*, **M. Rott***, K. Schlaeppi*, E. Ver Loren van Themaat*, N. Ahmadinejad, F. Assenza, P. Rauf, B. Huettel, R. Reinhardt, E. Schmelzer, J. Peplies, F. O. Gloeckner, R. Amann, T. Eickhorst, and P. Schulze-Lefert. "Revealing Structure and Assembly Cues for Arabidopsis Root-Inhabiting Bacterial Microbiota." *Nature* 488, (2012): 91-5.

Lebeis*, S. L., **M. Rott***, J. L. Dangl, and P. Schulze-Lefert. "Culturing a Plant Microbiome Community at the Cross-Rhodes." *New Phytol* 196 (2012): 341-344.

Spanu, Pietro D., J. C. Abbott, J. Amselem, T. A. Burgis, D. M. Soanes, K. Stueber, E. V. L. v. Themaat, J. K. M. Brown, S. A. Butcher, S. J. Gurr, M.-H. Lebrun, C. J. Ridout, P. Schulze-Lefert, N. J. Talbot, N. Ahmadinejad, C. Ametz, G. R. Barton, M. Benjdia, P. Bidzinski, L. V. Bindschedler, M. Both, M. T. Brewer, L. Cadle-Davidson, M. M. Cadle-Davidson, J. Collemare, R. Cramer, O. Frenkel, D. Godfrey, J. Harriman, C. Hoede, B. C. King, S. Klages, J. Kleemann, D. Knoll, P. S. Koti, J. Kreplak, F. J. Lopez-Ruiz, X. Lu, T. Maekawa, S. Mahanil, C. Micali, M. G. Milgroom, G. Montana, S. Noir, R. J. O'Connell, S. Oberhaensli, F. Parlange, C. Pedersen, H. Quesneville, R. Reinhardt, **M. Rott**, S. Sacristan, S. M. Schmidt, M. Schoen, P. Skamnioti, H. Sommer, A. Stephens, H. Takahara, H. Thordal-Christensen, M. Vigouroux, R. Wessling, T. Wicker, and R. Panstruga. "Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism." *Science* 330, (2010): 1543-1546.

* These authors contributed equally to this work

Table of Contents

Chapter II

Comparative genomics of selected *Rhizobium* members of the Arabidopsis root-associated bacterial microbiota

Abstract

The plant root defines the interface between a multicellular eukaryote and soil, one of the richest microbial ecosystems on earth. Soil bacteria are able to colonize the root surface and even multiply inside roots as benign endophytes. Some of these bacteria modulate plant growth and development, with implications ranging from enhanced nutrition to resistance against pathogens.

In this study a high-resolution methodology based on pyrosequencing of the bacterial 16S rRNA marker gene was adopted to characterize and compare soil and root-inhabiting bacterial communities. Our results show that roots of *Arabidopsis thaliana*, grown in different natural soils under controlled environmental conditions, are preferentially colonized by Proteobacteria, Bacteroidetes and Actinobacteria, and each bacterial phylum is represented by a dominating class or family of bacteria. Soil type defines the composition of root-inhabiting bacterial communities whereas the host genotype modulates their profiles only to a limited extent. Furthermore, bacterial communities associated to wooden sticks, representing a metabolically inactive lignocellulosic matrix for bacterial colonization, were analyzed to deconvolute actively recruited from opportunistic root microbiota members. This comparison showed that plant cell wall features appear to provide a sufficient cue for the assembly of app. 40% of the Arabidopsis root microbiota. This root- and wood-shared sub-community was mainly composed by Betaproteobacteria. In contrast, specifically recruited members of the root-inhabiting bacteria, mostly Actinobacteria, depended on cues from metabolically active host cells, defining a root-specific sub-community.

This culture-independent survey of the Arabidopsis root-associated bacteria was utilized to guide a targeted cultivation-based approach resulting in the isolation of members of both sub-communities of the Arabidopsis root-inhabiting bacterial microbiota. Several *Rhizobium spp*., which are either members of the shared or specific sub-communities, were isolated from Arabidopsis roots in pure culture and tested for plant growth promotion capabilities. Upon inoculation in a gnotobiotic system, *Rhizobium* spp. isolated from the specific sub-community increased shoot fresh weight of Arabidopsis, while a *Rhizobium* strain from the shared sub-community was not able to promote plant growth. Comparative whole-genome analysis of independent exemplars of the isolated *Rhizobium* spp. revealed differential gene enrichments among members of both sub-communities. Particularly, the amplification and divergence of transcription factor genes might represent a signature of differential habitat adaptation.

This culture-based approach backed by a broad scale culture-independent survey sets the stage to advance from descriptive characterization of bacterial communities to testing the functional significance of plant-microbiota interactions.

Zusammenfassung

Die Pflanzenwurzel bildet die Kontaktfläche eines Eukaryoten und Boden, einem der artenreichsten Habitate auf der Erde. Bodenbakterien können sowohl die Wurzeloberfläche, als auch das Wurzelinnere besiedeln. Einige dieser Bakterien sind in der Lage, das Wachstum und die Entwicklung der Pflanze zu beeinflussen, mit Auswirkungen auf die Nährstoffaufnahme bis hin zur Schädlingsresistenz.

In der vorliegenden Arbeit wurde eine kultivierungsunabhängige auf Pyrosequenzierung beruhende Methode angepasst, um bakterielle Gemeinschaften des Bodens und der Wurzel zu charakterisieren und zu vergleichen. Die Untersuchungen zeigten, dass *Arabidopsis thaliana* Wurzeln, die unter kontrollierten Umweltbedingungen in verschiedenen Böden angezogen wurden, überwiegend von Proteobakterien, Bakteroidetes und Actinobakterien besiedelt werden. Jedes dieser Phyla wird dabei durch eine Klasse oder Familie von Bakterien dominiert. Die Zusammensetzung der Wurzel-besiedelnden Bakteriengemeinschaften wird hauptsächlich durch den Bodentyp bestimmt, während der Genotyp der Pflanze nur einen geringen Einfluss ausübt. Weiter haben wir die bakteriellen Gemeinschaften in der lebenden Wurzel mit denjenigen von metabolisch inaktivem Holz verglichen. Dieser Vergleich zeigte, dass Zellwandbestandteile wie Lignocellulose als Besiedelungssignal für 40% der wurzelbesidelnden Bakterien ausreichend sind. Diese Teilgemeinschaft, die Pflanzenwurzeln und Holz gemeinsam haben, besteht hauptsächlich aus Betaproteobakterien. Der Vergleich mit Holz zeigte auch eine hauptsächlich aus Actinobakterien bestehende wurzelspezifische Teilgemeinschaft auf, die im Gegenzug Besiedelungssignale von metabolisch aktiven Wurzeln benötigt.

Diese kultur-unabhängige Untersuchung wurde für einen zielgerichteten Kultivierungsansatz zum Studieren der wurzel-assoziierten Bakterien verwendet. Dabei wurden mehrere *Rhizobium* spp. Bakterien aus den spezifischen und auch gemeinsamen Teilgemeinschaften isoliert. Diese Isolate wurden auf die Fähigkeit hin Pflanzenwachstum zu verstärken getestet. Inokulation von Arabiopsis mit *Rhizobium* spp. der spezifischen Teilgemeinschaft führte zu einem erhöhten Frischgewicht des Sprosses, während ein unspezifisches *Rhizobium* sp. nicht in der Lage war, das Pflanzenwachstum zu verstärken. Die vergleichende Genomanalyse unabhängig isolierter Exemplare der *Rhizobium* spp. zeigte eine spezifische Anreicherungen von Gengruppen in den jeweiligen Teilgemeinschaften. Diese Gengruppen beinhalten vor allem Transkriptionsfaktor-codierende Gene, was auf eine Habitatanpassung durch Gen-Duplikationen und Gen–Diversifizierungen von Transkriptionsfaktoren hindeutet.

Dieser Kultur-abhängge Ansatz, auf der Basis des Kultur-unabhängigen Zensus der bakteriellen Gemeinschaft, ist die Grundlage für den Fortschritt von einer rein deskriptiven Charakterisierung, hin zu einer funktionellen Analyse der Interaktionen zwischen Pflanzen und ihrer Mikrobiota.

Chapter I
Structure and assembly cues of Arabidopsis root-inhabiting bacterial communities

1 Introduction

Eukaryotic hosts live in close association to bacteria and fungi. Microbial cells associated to a humans are estimated to outnumber the host cells by an order of magnitude (Savage, 1977) and their genomes comprise up to 200 times more genes than the genome of the host (Xu *et al.*, 2003). Some of these genes encode for metabolic capabilities that were never evolved by the host. Human beings conveyed delicate tasks to their associated microbiota, as for example the breakdown of food in the digestive tract, especially in carbohydrate metabolism (Hooper *et al.*, 2002; Backhed *et al.*, 2005). This renders eukaryotes "superorganisms" and their phenotypes are most likely a result of the combined interactions of eukaryotic host genes and genetic elements of prokaryotic colonizers that have evolved during millions of years of co-existence (Backhed *et al.*, 2005). Deeper insights in such host-microbiota exchanges at a community level became recently possible with to the advent of next generation sequencing in combination with phylogenetic markers such as the 16S rRNA gene (Edwards *et al.*, 2006). This technical revolution sets the stage for new discoveries in host-microbiota interactions beyond the well-characterized binary host-pathogen or host-symbiont interactions.

1.1 Methodologies to study bacterial communities

1.1.1 DNA based methodologies

Studying bacterial communities was limited to culture-based techniques until Woese and Fox introduced gene encoding for the small subunit ribosomal (SSU) RNA as marker for bacterial phylogeny (Woese & Fox, 1977). This gene is usually referred to as 16S rRNA gene, as the SSU has a sedimentation rate of 16 Svedberg units (S) when centrifuged. The ribosomal RNA of the SSU forms alternating secondary structures of hairpins and loops that result from complementary sequence- and variable sequence-stretches, respectively. The complementary gene segments are highly conserved among the bacterial kingdom whereas the so-called hyper-variable regions are extremely diverse between bacterial taxa and can be used for phylogeny analyses. PCR primers can be designed to hybridize to conserved regions and to generate amplicon pools that span the hyper-variable regions of a broad taxonomic diversity of bacterial species. In this way the sequence information of the amplicon population can then be used to determine the composition

of the analyzed bacterial communities. Before the rise of next generation sequencing bacterial community structures were analyzed using low-resolution fingerprinting techniques. (i) By denaturing gradient-gel electrophoresis (DGGE) which is based on amplicon pools that can be separated by electrophoresis according to differential melting behavior as a function of their GC content (Muyzer *et al.*, 1993). (ii) By terminal Restriction Fragment Length Polymorphism (T-RFLP) which is based on the position of the restriction site closest to a labeled end of the amplicon pool. After a restriction digest the size of each of the individual resulting terminal fragments is detected using a DNA sequencer (Liu *et al.*, 1997). (iii) By automated ribosomal RNA intergenic spacer analysis (ARISA) which is a related methodology that does not rely on the 16S rRNA gene itself, but on PCR amplification of the variable region next to it, termed intergenic spacer region. Generated amplicons differ in length and these length polymorphisms can be detected by a capillary sequencer (Fisher & Triplett, 1999). The resulting signals, bands in the electrophoresis or peaks detected by the sequencer, are defined as operational taxonomic units (OTUs) and changes in bacterial community composition are interpreted as a function of presence/absence and relative abundance these OTUs. A limitation of these methodologies is that they do not provide direct sequence information of the amplified 16S rRNA genes and hence no taxonomical information. Indirect taxonomy information can be obtained via clone libraries from isolated bands of OTUs, which is not feasible at larger scales. Automated systems can assign low-resolution taxonomic ranks to some DGGE and T-RFLP signals but not down to the genus or species level. (Spiegelman *et al.*, 2005). Changes in bacterial community composition as a response to a given influence can therefore be detected, but only to a limited extend linked with taxonomical information.

In contrast, next generation sequencing technologies applied to an amplicon population provide the sequence information of each individual DNA molecule of the amplicon pool which can be used to obtain a phylogeny assignment from taxonomic databases (Edwards *et al.*, 2006). Available databases are GreenGenes (DeSantis *et al.*, 2006), SILVA (from Latin *silva*, forest) (Pruesse *et al.*, 2007) and ribosomal database project (RDP) (Cole *et al.*, 2009). Each database provides a constantly growing collection of bacterial and archeal 16S rRNA sequences. The quality of obtained taxonomy information, however, reflects the database representation of the analyzed habitat. For example, bacterial sequences derived from environmental samples like soil are notoriously difficult to classify, as these habitats are highly diverse and underrepresented in the databases (Gans *et al.*, 2005). As a consequence, the resulting phylogeny does often not contain the taxonomic depth down to the genus level. To overcome database representation limitations and to maximize the number of sequences analyzed, reads are typically clustered at a sequence identity of ≥97% to define OTUs. Communities are often compared based on the relative

abundance (RA) of each OTU, which is calculated as the proportion of reads clustered to a given OTU of all reads in a sample. Secondly, the number of sequences per sample is corrected for differences in sequencing depth between samples by rarefication, i.e. the same number of reads is randomly sub-sampled in each sample.

1.1.2 Fluorescence in situ hybridization

Fluorescence *in situ* hybridization (FISH) allows the visualization of specific taxa of bacteria for studying their community structure, spatial distribution and responses to changes in the environment. Bacteria are non-invasively visualized in their natural environment, e.g. on plant roots (Blazejak *et al.*, 2005a) or for quantification purposes equally distributed on filter papers (Eickhorst & Tippkotter, 2008). The detection is based on the hybridization of a specific oligonucleotide probe fused to a fluorescent dye to the RNA of a specific bacterial marker (Amann *et al.*, 1990). Typically, targeted markers are, similar as for PCR-based techniques, the bacterial 16S rRNA or 23S rRNA sequence, which are the RNA components of the small and large ribosomal subunit, respectively. Because instead of genomic DNA RNA molecules are targeted, signals detected by FISH are a proxy for the metabolically active proportion of the community members. For the design of FISH probes the same rules apply as for the design of 16S rRNA gene targeting PCR primers. Probes targeting conserved regions detect a broad diversity of bacterial species, as for example the probe EUB338, which detects a wide range of Eubacteria (Amann *et al.*, 1990). In contrast, probes targeting a taxon-specific sequence in the hyper-variable region of the 16S rRNA molecule permit the detection of bacteria corresponding to the respective specific phylum, order or class. For example, Betaproteobacteria are visualized with the class-specific probe Bet42a (Manz *et al.*, 1992). Despite modern software packages like ARB (from Latin *arbor*, tree) (Ludwig *et al.*, 2004) a perfect probe design is difficult and probes suffer from false negative and false positive signals (Siyambalapitiya & Blackall, 2005). Another limitation of this technique is the low fluorescent signal intensity from single cells, particularly in environmental samples with high auto-fluorescence. To overcome this limitation, the direct labeling of the probes with a fluorescent dye can be replaced by fusing the probe with a horseradish peroxidase. After hybridization with the RNA molecules this probe-coupled enzyme catalyzes the binding of hundreds of fluorescent tyramides which results in a significant amplification of the signal intensity (Pernthaler *et al.*, 2002). This variant of the FISH technique is referred to as *catalyzed reporter deposition fluorescence in situ hybridization* (CARD-FISH). FISH has been successfully applied to environmental samples like seawater (Alonso & Pernthaler, 2006), soil (Eickhorst & Tippkotter, 2008) and roots (Buddrus-Schiemann *et al.*, 2010).

1.2 Soil bacterial communities

Natural soils harbor up to 10^3 to 10^7 bacterial species and 10^9 bacterial cells per gram of soil (Gans *et al.*, 2005). The composition of bacterial populations in soils are influenced by nutrient content (Fierer *et al.*, 2007), soil texture (Sessitsch *et al.*, 2001) and by pH, which is one of the main structure determinants of bacterial assemblages in soil (Fierer & Jackson, 2006; Rousk *et al.*, 2010). Typical bacterial phyla found to dominate bacterial communities in soil are Proteobacteria, Acidobacteria, Actinobacteria, Firmicutes and Bacteroidetes (Janssen, 2006; Nacke *et al.*, 2011). These bacteria and their genetic potential represent a reservoir of functional traits for beneficial plant-microbe interactions. A known phenomenon is the occurrence of soils that suppress by a so far unknown mechanism plant diseases after several years of culturing the same crop species in the same field. This disease suppressiveness could be linked to microbial activities already in 1973, since these soils lose their activity upon pasteurization (Shipton *et al.*, 1973) and a recent study could correlate activity to specific bacterial phyla (Mendes *et al.*, 2011).

1.3 Rhizosphere bacterial communities

Roots of plants are constantly exposed to diverse bacterial communities in soil including potentially pathogenic and beneficial bacteria. In 1904, Lorenz Hiltner introduced the concept of the rhizosphere effect, which refers to a microcosm of soil around the root that is under direct influence of the plant and was recognized as a hotspot of bacterial abundance and activity (Hiltner, 1904). Today, most studies define the rhizosphere pragmatically as soil particles remaining firmly attached to the root during a sampling procedure.

The composition of rhizosphere bacterial communities is highly dissimilar from communities in the surrounding soil, as e.g. shown for strawberry and oilseed rape (Costa *et al.*, 2006), trees (DeAngelis *et al.*, 2009; Uroz *et al.*, 2010; Gottel *et al.*, 2011) or crop plants (Bouffaud *et al.*, 2012; Knief *et al.*, 2012). The structure of rhizosphere bacterial communities is furthermore dependent on the plant species (Marschner, 2004; Bezemer *et al.*, 2006; Costa *et al.*, 2006; Haichar *et al.*, 2008). Cultivar-/genotype specific accumulation of bacteria has been reported for multiple plant species including potato (Weinert *et al.*, 2011), where it seems to be related to tuber starch content (Inceoglu *et al.*, 2011). Rhizosphere communities harbor the full continuum of interactors with the plant: pathogens and symbionts at the extremes and mutualistic and commensalistic bacteria filling the gap in between (Raaijmakers *et al.*, 2009). For the interaction with the host, symbiotic bacteria like *Rhizobium* and pathogens like *Pseudomonas syringae* suppress their host's innate immune system by so-called effectors, proteins that are secreted into the host cells and inhibit a host immune response (Bartsev *et al.*, 2004; Hubber *et al.*, 2004; Jones & Dangl, 2006).

The interaction of plants with bacteria in the range between these specialists is however less well understood.

Many bacteria in the rhizosphere exhibit plant growth promoting effects and are therefore termed plant growth promoting rhizobacteria (PGPR). These growth promoting effects can be due to increase of plant available nutrients by bacterial activities (Vassilev *et al.*, 2006), interference with the plant's phytohormone levels (Glick *et al.*, 2007a; Spaepen & Vanderleyden, 2011), increasing drought stress resistance (Dimkpa *et al.*, 2009) or biocontrol of soil-borne pathogens. Biocontrol takes place via two main mechanisms, which are competition for space and nutrients like iron (Kloepper *et al.*, 1980) and antagonism by the production of antimicrobial compounds (Weller, 2007). Likewise, rhizosphere bacteria such as fluorescent *Pseudomonas* spp. improve plant health by the induction of a systemic resistance (ISR) by priming the plants' innate immune system throughout the whole plant (Weller *et al.*, 2002; Berendsen *et al.*, 2012). In this way plants also become more resistant to foliar pathogens and are protected by associated beneficial bacteria without direct contact of those with the pathogen (Pieterse *et al.*, 2003). In agricultural systems, soil-borne pathogens present a major threat to yield and are more recalcitrant to control measures compared to above-ground pathogens, since they are less accessible for pesticide treatments (Brühl, 1987). Therefore, it is desirable to find new opportunities of pest control, potentially by exploiting the plants' ability to influence the bacterial communities in the rhizosphere. Understanding the underlying mechanisms of how rhizosphere microbial communities are affected by the plant is the first step towards this direction.

Most of soil-borne bacteria are heterotroph and depend on exogenous supply of carbon as energy source as well as other nutrients for growth and development. The rhizosphere effect, i.e. the enrichment of bacterial abundance in the vicinity of the root, negatively correlated with soil nutrient content and the highest enrichment of bacteria occurred in low nutrient soils close to parts of wheat roots where the concentration of root exudates was the highest (Semenov *et al.*, 1999). This indicates that plant roots supplement the soil through rhizodeposition providing what is lacking in soil. Rhizodeposits comprise actively secreted compounds, like flavonoids and glucosinolates (Steinkellner *et al.*, 2007), and passively leaking root exudates like sugars, organic acids (Kamilova *et al.*, 2006), peptides or respiratory CO_2 (Shibistova *et al.*, 2002). Additionally root tips produce mucilage and slough off cells to reduce friction while penetrating soil (Vicre *et al.*, 2005; Li *et al.*, 2011). Depending on plant age and species, rhizodeposition is estimated to sum up to 10-40% of the total carbon fixed by photosynthesis (Lynch & Whipps, 1990; Jones *et al.*, 2009). This considerable carbon investment tempts to speculate about a fitness advantage linked with rhizodeposition, which could be chemical communication of the host plants with the associated microbes. CO_2 fixed by the plant from the atmosphere and released into soil by rhizodeposition is

metabolized by bacteria within a few hours as shown by stable isotope probing (Ostle *et al.*, 2003; Staddon, 2004). Some compounds have been characterized to mediate beneficial binary host-microbe interactions, for example root secreted flavonoids that attract symbionts like rhizobia (Downie, 2010). Likewise, the colonization of the rhizoplane, the root surface, by the beneficial rhizobacterium *Bacillus subtilis* FB17 is promoted by specific compounds of the root exudates. Arabidopsis reacts to foliar infection by the pathogen *Pseudomonas syringae* DC3000 with the increased secretion of L-malic acid in the root exudates. This compound is responsible for the attraction of *Bacillus subtilis* FB17 from the soil in a dose dependent manner (Rudrappa *et al.*, 2008). Chemotaxis describes the ability to sense and react to chemical compounds (e.g. emitted by the plant) by subsequent motility towards or away from it and is a main bacterial trait for successful host-microbe interactions (de Weert *et al.*, 2002). Root exudates contain also compounds with repellent activities towards specific bacteria. Bais and colleagues identified rosmarinic acid in the root exudates of sweet basil exerting antimicrobial activities against a range of soil-borne bacteria including the opportunistic pathogen *Pseudomonas aeruginosa* (Bais *et al.*, 2002). The type of bacterial response to root exudates can also be specific to different host cultivars: *Pseudomonas aeruginosa* PAO1 expresses differential subsets of chemotaxis genes in reaction to exudates of the sugar beet variety *Celt* in comparison to exudates of the variety *Roberta* (Mark *et al.*, 2005). In conclusion, rhizodeposition is sensed by bacteria and clearly influences bacteria in the rhizosphere, but if this happens in a targeted way depending on the plants' needs has still to be shown.

1.4 Root-inhabiting bacterial communities

Specific bacteria of the rhizosphere can colonize the rhizoplane, i.e. the surface of the root, and also the inner root tissue. The existence of such has been postulated already in 1887 by M. L. V. Gallipe (Galippe, 1887). Bacteria that enter the inner tissue and do not harm the host are referred to as endophytes (Hallmann *et al.*, 1997). Operationally endophytes are defined through the isolation from plant tissue after surface sterilization, typically with bleach and/or ethanol. Since the efficiency of these treatments is difficult to measure, the endophytic lifestyle of an isolate should be confirmed by transferring Koch's postulates from pathogens to endophytes (Koch, 1893) and using for example fluorescently tagged derivatives to detect bacterial cells within the host tissue in re-colonization experiments. If such approaches fail one should consider referring to a root-associated (Reinhold-Hurek & Hurek, 2011) or root-inhabiting isolate. In contrast to mutualistic Rhizobia, endophytes do not reside intracellularly or encapsulated by a host

membrane (Reinhold-Hurek & Hurek, 2011). Endophytes have also been shown to exhibit plant growth-promoting effects such as for example by improving nutrient availability, conferring drought stress resistance, production of the growth hormone auxin (Taghavi *et al.*, 2009) or lowering the plants' ethylene levels (Hardoim *et al.*, 2008).

Endophytes are thought to be specialized members of the rhizosphere bacterial community that first colonizes the rhizoplane and from there the endosphere, i.e. the inner plant tissue. These bacteria are hypothesized to enter the root via root hairs or through fissures in the exodermis, the outermost cell layer of the root. Root surface fissures can be caused by injuries or at the outgrowth of secondary roots (Reinhold-Hurek & Hurek, 1998).

Once inside the root tissue endophytes form a very intimate relationship with their host and therefore, this interaction has to be tightly controlled. Endophyte community composition is determined by plant developmental stage (van Overbeek & van Elsas, 2008), specific tissue, host species, host genotype (Graner *et al.*, 2003; Tan *et al.*, 2003; van Overbeek & van Elsas, 2008; Hardoim *et al.*, 2011) and the surrounding soil type (Conn & Franco, 2004; Hardoim *et al.*, 2011). The occurrence of endophytes in the presence of the plant innate immune system appears paradox, since the immune system is designed to trigger defense responses upon detection of conserved microbe associated molecular patterns (MAMPs) (Boller & Felix, 2009). These MAMPs, including the peptide sequence 'flg22', an epitope of bacterial flagella, are conserved among endophytic bacteria like *Azoarcus sp*. (Buschart *et al.*, 2012). The mutant strain *fliC/fliB* of *Salmonella enterica* serovar Typhimurium lacks flagella and colonizes *Medicago* to higher cell numbers, indicating that the perception of MAMPs presents a mechanism to limit endophyte proliferation (Iniguez *et al.*, 2005). In contrast, non-motile mutant strains of *Salmonella enterica* serovar Newport that possess inactive or lack flagella were reported to show decreased colonization efficiency and spreading in Arabidopsis (Cooley *et al.*, 2003). Therefore, spreading might be important to move inside the host tissue for efficient colonization which requires flagella. This in turn presents a conflicting situation with the need to avoid recognition by the host immune receptors. The plant innate immune modulating phytohormones jasmonic acid (JA) and ethylene were reported to contribute to control of host-endophyte interactions. For example, the exogenous application of JA decreased the colonization efficiency of endophytic *Azoarcus* sp. strain BH72 in rice (Miche *et al.*, 2006). In *Nicotiana attenuata* mutants impaired in ethylene synthesis or perception harbor a less diverse bacterial culturable bacterial community than the wild type (Long *et al.*, 2010). The observation that endophytes, similar as pathogens, have a defined host range and the ability of some endophytes to enhance their host's resistance against pathogens by priming the host immune system (Jetiyanon & Kloepper, 2002) suggests that the control of endophytes is integrated and/or supervised by the plant's immune system. The

discrimination of friends from foes in this context is crucial for plants and remains an intriguing open question in current immune related science.

1.5 Microbial communities associated with *Arabidopsis thaliana*

Arabidopsis thaliana (hereafter referred to as Arabidopsis), a *Brassicaceae* species, is a cosmopolite and widely used model system due to its short generation cycle, low space and nutrient requirements and comparatively simple diploid genome (Meinke *et al.*, 1998). Additionally, protocols for transformation are available for genetic research and an extensive collection of mutants easy to maintain due to Arabidopsis' capability of self-fertilization. Importantly, the genome has been sequenced as first plant genome and is one of the best-annotated genomes in plant science (TheArabidopsisGenomeInitiative, 2000). In the course of the "1001 genomes project" genome sequencing of different ecotypes adapted to habitats all over the world is in progress for providing insights into the natural variation of Arabidopsis (Weigel & Mott, 2009; Cao *et al.*, 2011). Arabidopsis, as all other *Brassicaceae* species, lacks interactions with mycorrhizal fungi and does not form nodules to benefit from nitrogen fixing rhizobia, which makes it an ideal simplified system to study host-bacteria interactions at the community level.

In comparison to other plant species like potato or cereal crops few studies have been conducted to analyze microbial communities inhabiting Arabidopsis roots or its rhizosphere. A reason might be that the Arabidopsis root system is relatively small and delicate to sample from natural soils. Reproducible protocols for analyzing the Arabidopsis associated microbiota as developed in this study and other laboratories (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012) set the stage to infer first principles of plant microbe interactions at the community level and to understand e.g. how the control of bacterial communities is integrated into the host's innate immune system.

The Arabidopsis rhizosphere bacterial community was reported to be different from the microbiota in the surrounding soil based on T-RFLP fingerprinting (Hein *et al.*, 2008; Micallef *et al.*, 2009). As a cautionary note, the rhizosphere sampling protocol of these studies included root tissue and therefore potentially also endophyte bacteria. Hein and co-workers define the rhizosphere as "root system with remaining adhering soil" and Micallef *et al*. sample by scratching the rhizosphere off the roots with scalpel blades. Using this protocol Micallef *et al*. identified genotype-dependent community compositions among eight Arabidopsis accessions. Indicator T-RFLP signals, however, that would be digitally specific for only one accession were not reported.

The hypothesis that mutations altering the plant's immune system might lead to differential compositions in the plant-associated microbiota has also been tested in Arabidopsis. By culture-independent DGGE fingerprinting and a culture-dependent approach Doornbos *et al*., 2011,

tested Arabidopsis lines impaired in the production or perception of the phytohormones jasmonic acid (JA), ethylene (ET) and salicylic acid (SA), all involved in plant defense (Robert-Seilaniantz *et al.*, 2011). Dependent on soil type mutants in these defense pathways seemed to harbor a lower bacterial cell density in their rhizospheres. The induction of the plant immune system by a single foliar application of methyl jasmonate (MeJA) or salicylic acid (SA) did however not alter numbers of culturable cells in the rhizosphere samples of wild-type Arabidopsis (Doornbos *et al.*, 2011). These findings do not point to a marked influence of activated defenses by these phytohormones on rhizosphere bacterial communities. Soil type independent reduction in culturable bacterial cell numbers was only observed the *C*onstitutive expressor of *PR* genes 1 (*CPR1*) mutant that exhibits a hyper-accumulation of a specific type of immune receptors, intracellular nucleotide-binding site leucine-rich repeat (NB-LRR) -type receptors, and constitutive SA defense responses (Bowling *et al.*, 1994). Since NB-LRR immune responses can act largely independent of phytohormone signaling (Tsuda *et al.*, 2009) the reduced bacterial titers observed in *cpr1* might result from constitutive SA-independent NB-LRR auto-activity (Doornbos *et al.*, 2011). The *cpr1* mutant was also tested by fingerprinting in an independent study in which peaks in T-RFLP profiles that differentiated the mutant from wild-type as T-RFLP profiles also discriminated wild-type from the mutant *npr1* (Hein *et al.*, 2008). NPR1 is essential for regulating SA–dependent gene expression during systemic acquired resistance (Dong, 2004). The colonization by endophytes induces transcriptome changes in Arabidopsis and interestingly, some of the differentially regulated genes are related to the host's immune system (Wang *et al.*, 2005; Miche *et al.*, 2006). Of note, in these studies Arabidopsis were inoculated with bacterial strains that were initially isolated from different host species. Although these isolates have been tested to retain colonization competence and plant growth promotion activity on Arabidopsis, such results should be interpreted with care, especially since endophytes have been also hypothesized to present "disarmed pathogens" (Krause *et al.*, 2006). Thus, a deeper understanding of the influence of the host's innate immune system of root-associated bacteria is still lacking.

The influence of glucosinolate compounds on the Arabidopsis-associated microbiota was also studied (Bressan *et al.*, 2009). This is a class of phytochemicals showing activity against microbes and insects (Halkier & Gershenzon, 2006) which is produced by Arabidopsis among a diverse array of other secondary metabolites. Glucosinolate production has been almost exclusively reported from plants of the order Capparales that includes the *Brassicaceae* (Halkier & Gershenzon, 2006). Arabidopsis was transformed with the sorghum cytochrome P450 gene CYP79A1 that leads to the production of a tyrosine-derived glucosinolate normally not present in wild-type Arabidopsis plants of the ecotype Columbia (Bak *et al.*, 1999). This ectopic accumulation of p-hydroxybenzylglucosinolate correlated with changes in the rhizosphere community as determined

by DGGE in combination with stable isotope probing, indicating that plants influence their associated microbiota through secondary metabolites such as e.g. glucosinolates (Bressan *et al.*, 2009). A shortcoming of this work is that the authors did not test an independent transgenic line to exclude that overexpression of this compound leads to pleiotropic effects that then influenced the microbial communities.

ATP-binding cassette (ABC) proteins transport a wide range of compounds across membranes (Rea, 2007) and are possibly involved in the secretion of root exudates (Badri *et al.*, 2012). Badri and colleagues screened a panel of seven Arabidopsis ABC transporter mutants for differential rhizosphere community structures by ARISA fingerprinting. The ABC transporter mutant *abcg30* induced changes in the bacterial community determined by low-pass pyrosequencing 16S ribotyping after two generations of plant growth in the same natural soil. Root exudates of the mutant *abcg30* contained elevated levels of organic acids and lower amounts of sugars compared to wild-type, possibly correlating the shift in community composition with the alterations in the root exudate profile (Badri *et al.*, 2009).

Most of the aforementioned studies lack full-factorial replicates in independent soil batches from the same soil type. These studies provide snapshots of Arabidopsis-microbiota interactions, but only robust and reproducible phenotypes above experiment-to-experiment variation will permit to identify underlying first principles.

1.6 Aim of the study

Published studies on soil and host-associated microbiota often rely on ecological indices to detect differences between microbial communities and do not validate their findings in full factorial replicates. The lack of identification of bacterial species or taxonomic units which are indicative for a given treatment, genotype or soil type makes it difficult to infer general principles underlying these host-microbe interactions. This is also partly due to the use of low-resolution methodologies like DGGE or T-RFLP fingerprinting and the use of sampling protocols that do not clearly compartmentalize and define the sampled bacterial sub-communities. The aim of this first part of my PhD project was to develop a robust methodology to reproducibly characterize the bacterial microbiota associated to Arabidopsis roots consistent over full factorial replicates. The high-resolution pyrosequencing of 16S rRNA gene amplicons was applied to root-inhabiting and rhizosphere communities of the model plant *Arabidopsis thaliana* in comparison with the soil microbiota. The developed methodology was used to determine the effect of different soil types and host genotypes on the rhizosphere- and root-inhabiting bacterial microbiota. Inclusion of

metabolically inactive wooden sticks that mimic cell wall features of plant roots in the experiments allowed to infer first hints on assembly cues of root-associated bacterial communities.

2 Results

Parts of the presented results were published together with data generated by other members of Paul Schulze-Lefert's group at the MPIPZ (Bulgarelli *et al.*, 2012). For the conclusive interpretation of my data it is necessary to present all data published in the aforementioned publication.

My contribution to the data presented in this thesis:

Generation of 16S ribotyping amplicon libraries and analysis of pyrosequencing reads together with Davide Bulgarelli and Klaus Schläppi for Figures 1, 2, 4-11.

CARD-FISH analysis of root, soil, and rhizosphere samples, Figures 14-16

Processing of pyrosequencing data and computational analysis was coordinated by Emiel Ver Loren van Themaat.

2.1 Culture-independent characterization of the Arabidopsis root-associated bacterial microbiota via 16S ribotyping

Since only a minor fraction of the soil-inhabiting bacteria can be isolated with current cultivation techniques (Amann *et al.*, 1995; Keller & Zengler, 2004), the bacterial root-associated microbiota was characterized by pyrosequencing-based 16S ribotyping.

To gain insights into the bacterial community associated with Arabidopsis roots in natural soils, *Arabidopsis thaliana* ecotypes Landsberg (Ler) and Shakdara (Sha) were grown in two natural soils of contrasting geochemistry that originate from the locations Cologne (clay and silt-rich) and Golm (sand-rich) (see Material and Methods (M&M) section 4.1.1). Plants were grown at a defined planting density of nine plants per pot in the greenhouse under controlled conditions (see M&M 4.1.2). At early flowering stage we sampled the root system and prepared the following four compartments for characterization of their bacterial communities: (1) the soil compartment, representing a pool of bacteria present in the soil, uninfluenced by any plant. The soil compartment was sampled from a plant-free pot that was exposed to the same conditions as planted pots; (2) the rhizosphere compartment, which represents soil that tightly adhered to the roots and is thought to be influenced by the plant through rhizodeposits. We collected the rhizosphere compartment by sequentially washing the root with detergent containing phosphate buffer and collecting the soil after centrifugation; (3) the root compartment, which consists of root tissue depleted of soil particles and epiphytic bacteria by sequential washing and sonication treatments. Therefore, this compartment is enriched for root-inhabiting bacteria that are likely to interact closely with the plant. Additionally, wooden sticks made of the softwood birch (Betula) or the hardwood beech (Fagus) were incubated in soil in parallel with the Arabidopsis plants. We reasoned that part of the root tissue consists of dead cells that are in the process of being replaced by new cells due to secondary growth and therefore merely represent a lignocellulosic

matrix for bacterial colonization (Dolan *et al.*, 1993). This cell material can be mimicked by wood sticks because their main component is lignocellulose. Comparisons of wood- and root-associated bacterial communities might lead to the identification of bacteria that depend on metabolically active roots to interact with. The wooden sticks were sampled following the same methodology applied for the root compartment including washing steps and sonication, resulting in the fourth compartment: (4) 'wood'.

A total of 111 samples derived from all of the aforementioned compartments were profiled by sequencing which resulted in a total of 4.3 million sequence reads. We referred to reads that do not contain errors in the barcode and the primer sequences and with a defined length of 315 bp as high-quality sequences. For high-quality sequences also ambiguous bases were removed to obtain reliable taxonomic assignments. Additionally, high-quality sequences have a maximum of 10% nucleotides with a Phred score ≤ 27. Finally, we based our subsequent analysis on 1.6 million of high-quality sequences (Table 1). For analysis of the detected bacterial communities their relative abundance (RA), i.e. the proportion of sequences assigned to a given phylum, order, etc. of all detected sequences, was calculated.

Table 1: Overview of the sequencing effort. Number of generated raw and high-quality sequence reads of the profiled samples per compartment and soil. HQ = high quality

Compartment	Soil	# of samples	#of raw reads	# of HQ reads
root	Cologne	18	921,273	420,138
rhizosphere	Cologne	15	380,253	117,124
soil	Cologne	13	437,995	164,078
wood stick	Cologne	12	345,812	156,020
root	Golm	14	669,805	283,567
rhizosphere	Golm	12	411,934	125,873
soil	Golm	12	736,887	227,620
wood stick	Golm	15	416,330	131,236
total		**111**	**4,320,289**	**1,625,656**

2.2 Taxonomical composition of bacterial communities of soil, rhizosphere, root and wood compartments

The taxonomic composition of bacterial communities was determined for the high quality sequences by classification using the SILVA database (Pruesse *et al.*, 2007) and the RDP database (Wang *et al.*, 2007). Assignments using SILVA resulted in increased taxonomic depth, especially for soil and rhizosphere samples. For example, 10 % or 98 % of soil sequences were classified down to

the taxonomical level of the class using RDP or SILVA, respectively (Figure 1). For this reason we used the SILVA classification pipeline for further analysis.

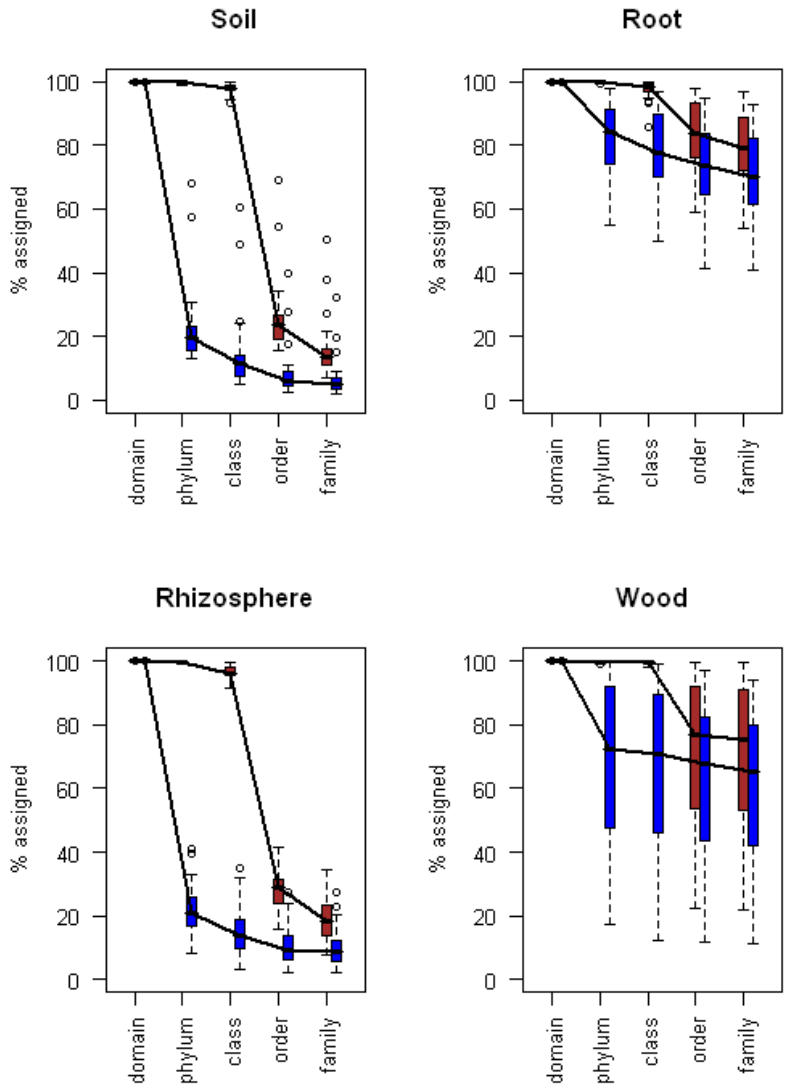


Figure 1: Percentage of reads assigned to taxonomical bins for each compartment using the SILVA (red) or RDP (blue) database. High quality sequences were assigned to the indicated taxonomic levels using the RDP and SILVA databases. The boxplots display the percentage of classified reads of all samples at different taxonomical ranks for the indicated compartments.

The taxonomic classification via the SILVA database identified a total of 43 bacterial and two archeal phyla and divisions of which eight bacterial phyla showed an average relative abundance of >5 ‰ over the total dataset (Supplementary table 1). The classification revealed an anomalous abundance of reads assigned to Chloroflexi in all samples, with an average relative abundance of app. 80 % for all soil and rhizosphere samples and app. 35 % for all root samples (Figure 2).

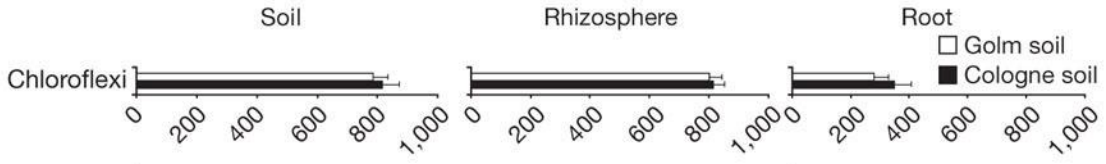


Figure 2: Average relative abundance (‰ ± s.e.m.) of the phylum Chloroflexi in the indicated compartments and soil types

It is likely that such an overrepresentation of Chloroflexi, which has never been described before (Janssen, 2006; Nacke *et al.*, 2011; Fierer *et al.*, 2012), is due to a bias introduced by the PCR primers 799F2 – 1193R utilized for amplification of the 16S rRNA gene in this study. We had developed the forward PCR primer 799F2 based on the previously described PCR primer 799F (Chelius & Triplett, 2001) to avoid massive co-amplification of plant organellar 18S rDNA that was observed using the parental PCR primer. We experimentally tested a possible PCR-bias by comparative PCR primer analysis (see Appendix) and PCR-independent CARD-FISH analysis on soil samples, which will be described later (see section 2.4.3). The comparative PCR primer analysis shows that the introduced base pair substitution does not distort the results of the overall community composition except for the phylum of Chloroflexi (see Appendix). Therefore, the PCR primer pair 799F2-1193R was used throughout the study and reads assigned to the phylum Chloroflexi were removed *in silico* before further analysis.

Soil and rhizosphere bacterial communities were dominated by Acidobacteria, Planctomycetes, Actinobacteria and Proteobacteria. In both tested soils Proteobacteria were significantly enriched in root-inhabiting bacterial communities compared to soil- and rhizosphere bacterial communities. Bacteroidetes were significantly enriched in Golm soil-grown plants and Actinobacteria in Cologne soil-grown plants (Figure 3b). Within the root-inhabiting Proteobacteria Betaproteobacteria families were overrepresented compared to Alphaproteobacteria and Gammaproteobacteria (Figure 3b). Likewise, a single family dominated each of the other root-inhabiting phyla: Flavobacteriaceae and Streptomycetaceae dominated in the phyla Bacteroidetes and Actinobacteria, respectively (Figure 3b). The Streptomycetaceae, Rhizobiaceae and the families belonging to the Betaproteobacteria are significantly enriched in the root compared to soil and rhizosphere compartments in both soils tested. Consistent with the whole phylum of Bacteroidetes the family of Flavobacteriaceae was significantly enriched only in Golm soil compared to soil and rhizosphere communities (Figure 3b). In conclusion, root-inhabiting bacterial communities are highly dissimilar from soil and rhizosphere bacterial communities and are influenced by the surrounding soil type.

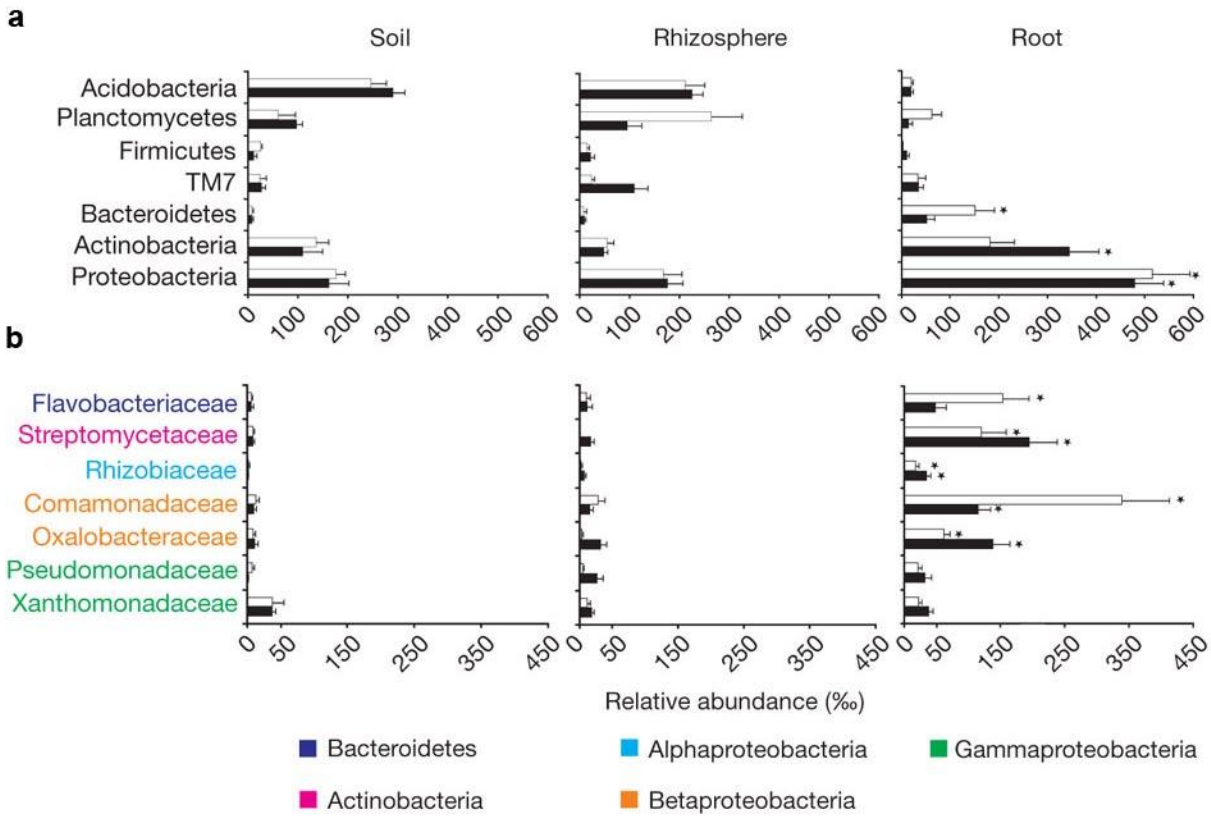


Figure 3: (a) Average relative abundance (‰ ± s.e.m.) of the dominant phyla in root compartments (>5 ‰) in the indicated compartments and soil types. **(b)** Average relative abundance (‰ ± s.e.m.) of families belonging to the three dominant phyla in the root compartment. Average relative abundances are calculated after removal of reads assigned to Chloroflexi. Asterisks indicate significant enrichment (FDR<0.05) in the root compartment compared to soil and rhizosphere compartments.

2.3 OTU analysis

As mentioned above soil and rhizosphere samples could not be unambiguously classified at lower taxonomic levels using two different databases (Figure 1). This is indicative of an insufficient database representation of the biodiversity of soil-borne bacteria (Gans *et al.*, 2005). To increase resolution and to maximize the number of reads analyzed, all high quality sequences of all compartments were clustered at ≥97% sequence identity using the web-based tool PyroTagger (Kunin & Hugenholtz, 2010) and each cluster was defined as an operational taxonomic unit (OTU). OTUs at ≥97% sequence identity correspond at best to the taxonomic rank of the genus (Schloss & Handelsman, 2005). The most abundant sequence within each OTU was defined as the OTU representative sequence and classified using the SILVA database. Bacterial diversity measured as OTU richness was estimated in the three compartments by rarefaction analysis and revealed the greatest number of OTUs in soil (app. 2,000 OTUs detected in total), followed by rhizosphere and root compartments that both show a reduced richness (each app. 1,000 OTUs detected in total) (Figure 4).

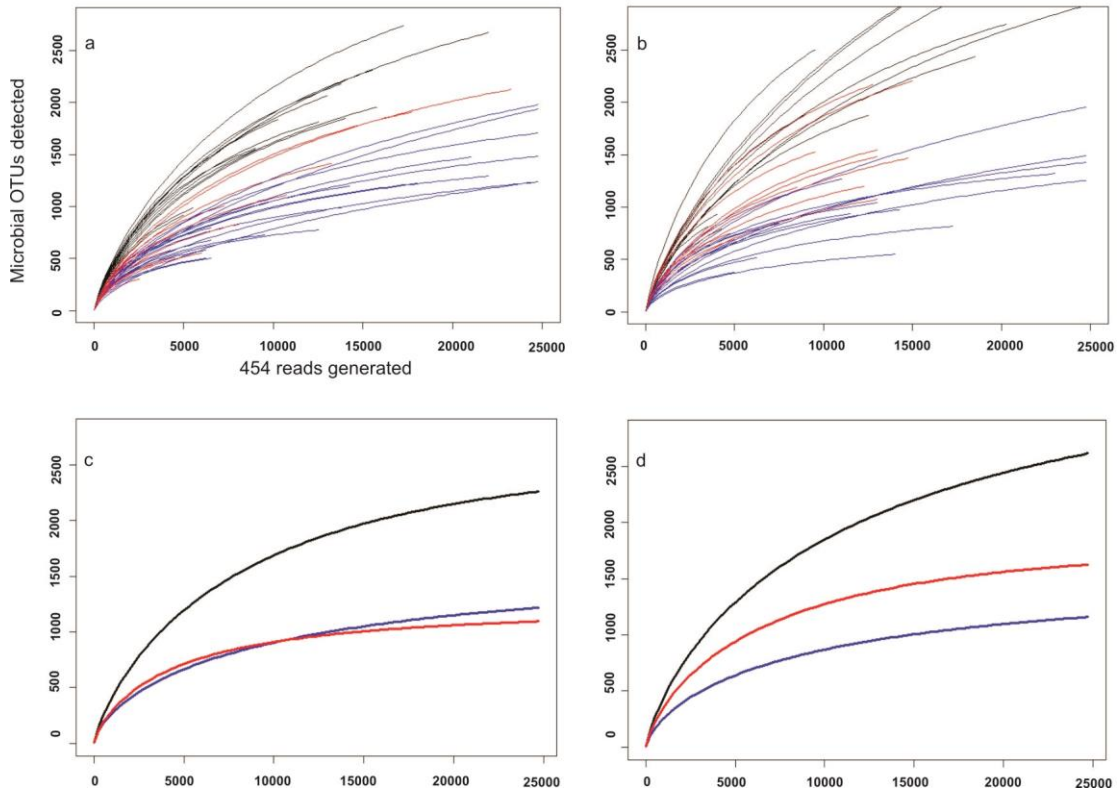


Figure 4: OTU richness in the compartments soil (black), rhizosphere (red) and root (blue). (**a**) OTU richness of each sample in Cologne soil, (**b**) OTU richness of each sample in Golm soil. (**c** and **d**) Rarefaction curves for estimated averages of samples per compartment in Cologne (**c**) and Golm (**d**) soil.

The overrepresentation of Chloroflexi was also apparent after clustering reads to OTUs and consequently OTUs assigned to the phylum Chloroflexi were depleted for subsequent analysis. The removal of OTUs belonging to Chloroflexi affected the number of sequences per sample, especially for soil and rhizosphere samples. After depletion 14 samples had less than 1,000 sequence reads. Sequence reads of three pairs of these samples sharing the same experimental variables were pooled to reach at least 1000 sequence reads in a pooled sample. Consequently, the final dataset comprises 100 samples with 22 of the 'soil' compartment (number of replicates: Cologne nC=12, Golm nG=10), 19 'rhizosphere' (nC=9, nG=10), 32 'root' (nC=18, nG=14) and 27 'wood' (nC=12, nG=15). To correct for the different sequencing depths among the samples the raw OTU counts were rarefied to 1,000 counts per sample.

2.3.1 Technical reproducibility of OTUs quantification by 16S ribotyping

We experimentally determined the lower limit for a reproducible quantification of OTUs by profiling three technical replicates of a soil sample. These replicates contained different barcodes and were PCR amplified and sequenced in parallel. We chose a soil sample, because soil communities are more complex and harbor a greater diversity of bacteria than root communities (Figure 4). For each of the replicates approximately 15,000 sequence reads were generated,

clustered to OTUs with PyroTagger and used for pair-wise comparisons (Figure 5). Abundant but not the low-abundant OTUs could be robustly quantified (Figure 5a) and Spearman rank correlation analysis between technical replicates revealed that only OTUs with a minimal relative abundance of 5 ‰ were reliably measured (Figure 5b). Therefore, only OTUs with a relative abundance ≥ 5 ‰ were considered for in depth analysis of factors influencing root-inhabiting bacterial communities. That does not mean however, that the presence of OTUs with a RA of < 5 ‰ is a sequencing artifact, but only that reproducible quantification of those is not possible.

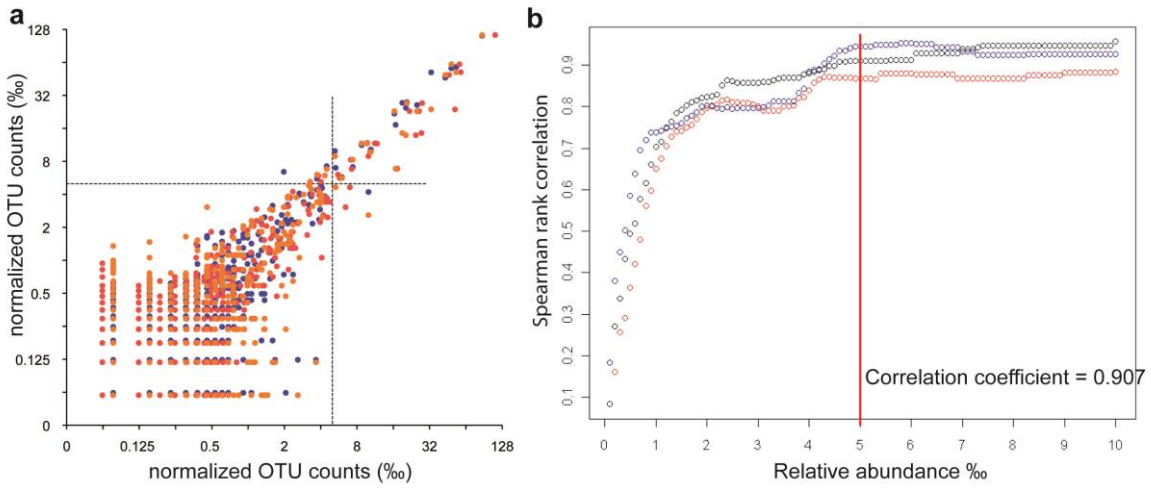


Figure 5: The RA threshold of 5 ‰ for OTUs was determined using technical replicates of a soil sample. **(a)** The abundance of every OTU is plotted for the three pair-wise combinations (barcode (bc)#1 vs. bc#2 red dots, bc#1 vs. bc#3 blue dots, and bc#2 vs. bc#3 orange dots). The dashed line indicates the 5 ‰ threshold. **(b)** Non-parametric Spearman rank correlation of data shown in (a).

2.3.2 Analysis of global community compositions

To compare the composition of the identified bacterial communities between the studied compartments a hierarchical cluster dendrogram based on Bray-Curtis distance was generated (Figure 6). This takes into account the presence/absence and the RA of all community members in the respective samples. The root compartment renders the root-inhabiting microbiota highly dissimilar from the communities retrieved from rhizosphere and soil compartments, as reflected by two distinct clusters that are formed between root samples and on the other hand soil and rhizosphere samples (Figure 6). Closer examination showed a soil type-dependent effect on both soil and rhizosphere communities. The seasonal variation in within soil and rhizosphere samples per soil type is smaller than the difference between the two tested soil types (Figure 6).

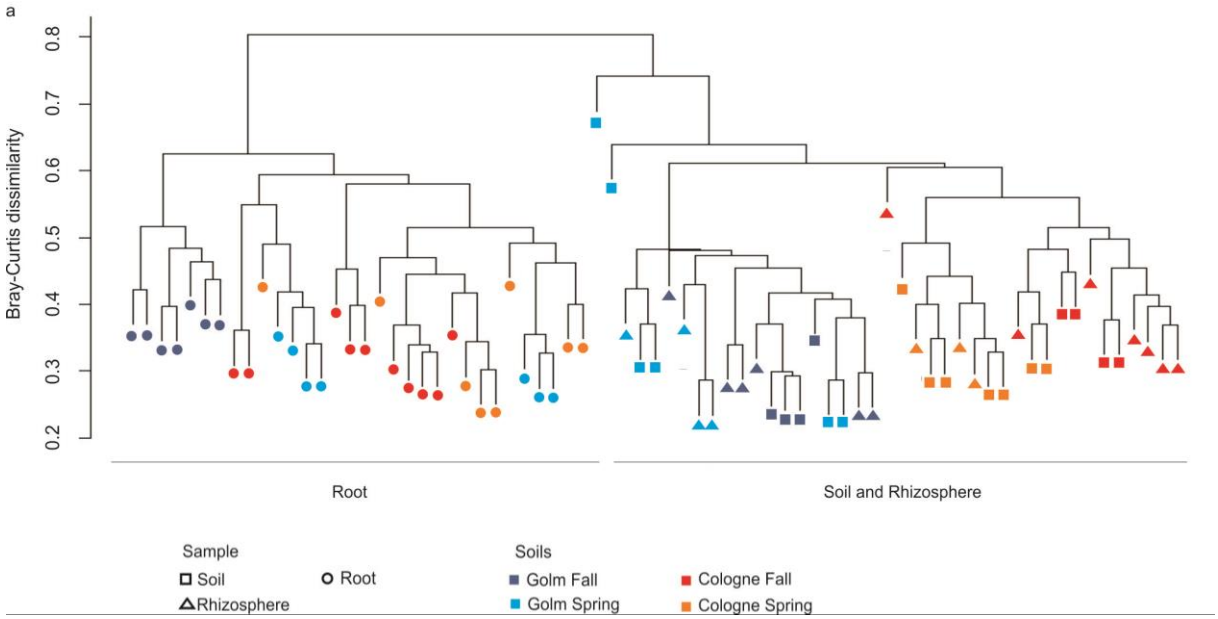


Figure 6: OTU dissimilarity of the bacterial communities identified in soil, rhizosphere and root compartments. The Cluster dendrogram was generated using a Bray-Curtis dissimilarity matrix applied to the RA of dominant OTUs (RA ≥5‰, log2 transformed RA data). OTU counts were rarefied to 1,000 counts per sample.

A differentiation between unplanted soil and the rhizosphere microbiota across seasonal soil batches was not evident from the cluster dendrogram, since e.g. rhizosphere samples of Cologne fall and Cologne spring soil grown plants did not cluster together. However, single OTUs in the rhizosphere distinct from the respective unplanted soil compartment are detectable when looking at samples obtained from within the same soil batch (Figure 7). One OTU differentiates soil from rhizosphere in spring Cologne soil and 22 OTUs differentiate soil from rhizosphere in fall Cologne soil, while two of them are shared between the seasonal samplings. For Golm soils, 17 and 11 OTUs differentiate soil from rhizosphere in spring and fall soils, respectively, while none of these are shared between the seasonal samplings (Figure 7).

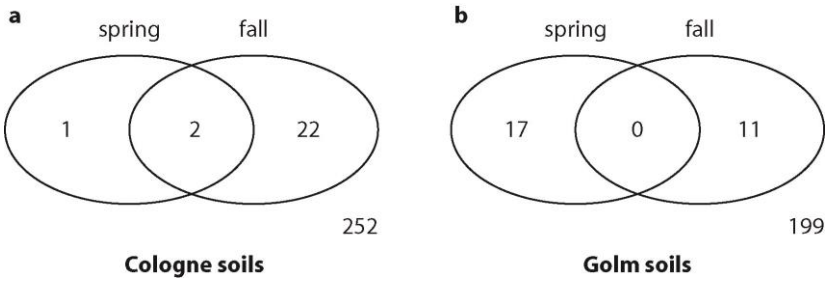


Figure 7: Rhizosphere effects identified in seasonal samplings of Cologne and Golm soils. Venn diagrams display the number of OTUs (RA >5 ‰) that significantly (FDR <0.05) differentiate soil and rhizosphere compartments in the indicated seasonal soil samplings. Numbers at the bottom right of the plots indicate the number of OTUs included in the analysis that do not fulfill the statistical criteria. **(a)** Cologne soil and **(b)** Golm soil.

This differentiation is reflected by a significant rhizosphere effect according to PERMANOVA analysis of the Bray-Curtis distance matrix (Table 2), suggesting that a rhizosphere effect is obscured by soil batch-to-batch variation.

Table 2: F- statistic obtained from permutational multivariate analysis of variance of the Bray-Curtis distance dissimilarity matrix. Asterisks indicate the significance of the calculated F value (***<0.001, **<0.01, *<0.05)

Cologne soil			
	soil	**root**	**rhizosphere**
root	26.013***		
rhizosphere		13.653***	
soil			2.852**

Golm soil			
	soil	**root**	**rhizosphere**
root	24.868***		
rhizosphere		19.821***	
soil			2.440***

2.3.3 Defining community members enriched in the root compartment

The Bray Curtis distance analysis showed a significant difference between root compartments and soil- and rhizosphere compartments at the level of whole bacterial communities. To identify individual OTUs responsible for the observed community differentiation, a linear model analysis was used to determine indicator OTUs for each tested compartment. In Cologne soil 56 OTUs and in Golm soil 44 OTUs significantly differentiate root-inhabiting bacterial communities from soil- and rhizosphere communities (Figure 8). Soil and rhizosphere compartments share a large proportion of OTUs, namely 252 and 199 OTUs in Cologne and Golm soil, respectively (Figure 8 and Figure 7).

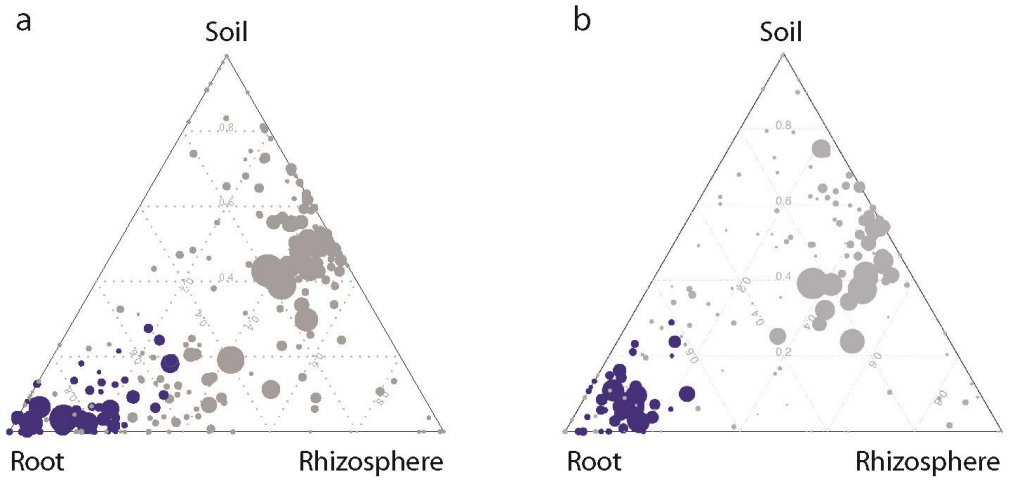


Figure 8: Arabidopsis assembles a distinctive root-inhabiting bacterial microbiota. Compartment specificity and RA of OTUs (≥ 5 ‰) determined in samples from **(a)** Cologne soil and **(b)** Golm soil. In these ternary plots each circle represents one OTU. The size of each circle represents its RA (weighted average). The position of each circle is determined by the contribution of the indicated compartments to the total RA. The dotted grid and numbers inside the plot indicate 20% increments of contribution from each compartment. Dark blue circles mark OTUs significantly enriched in the root compartment (FDR<0.05).

2.3.4 Defining community members specifically enriched by a living plant

The root-inhabiting bacterial community is significantly different from the bacterial communities in the surrounding soil. To obtain insights in potential plant-derived assembly cues for the assembly of this root-inhabiting microbiota, surface sterilized wooden splinters were incubated in both tested soils. These wooden splinters represent metabolically inactive lignocellulosic matrices for bacterial colonization, therefore mimicking part of the chemical properties of root tissue, without the typical characteristics of a living plant root, like e.g. root exudates. The associated bacterial communities of wooden sticks derived of two genotypes, the softwood birch and the hardwood beech, were determined.

The comparison of the root with the wood compartment revealed that 26 of the 56 root-enriched OTUs in Cologne soil and 22 of 44 root-enriched OTUs in Golm soil were equally or even more abundant in the wood compartment (Figure 9). This bacterial sub-community, that was found in roots and wooden sticks was termed ‘lOTUs’, for lignocellulosic matrix-associated OTUs (Figure 9, light blue color coded). The remaining 30 and 22 OTUs in Cologne and Golm soil, respectively, were significantly enriched in the root in comparison to soil-, rhizosphere as well as wood compartment. Members of this specifically root-enriched sub-community were termed ‘rOTUs’ (Figure 9, dark blue color coded). A third sub-community, designated ‘wOTUs’ for wood OTUs comprised 36 and 38 OTUs in Cologne and Golm soil experiments, respectively, that were specifically enriched in the wood compartment and are not enriched in the root compartment (Figure 9, orange color coded).

The three identified sub-communities of rOTUs, lOTUs and wOTUs are defined by distinct taxonomic structures. In Cologne soil Proteobacteria represent the vast majority of wOTUs and lOTUs sub-community members (~95% and 100%, respectively) while Actinobacteria were largely underrepresented. In contrast, Actinobacteria dominated the rOTUs community (Figure 9). In Golm soil results were qualitatively similar, except of a large proportion of Bacteroidetes in both rOTUs and in wOTUs sub-communities. This reflects that Bacteroidetes were generally found to have a higher abundance in the root-inhabiting bacterial communities of Golm soil grown plants (compare Figure 3). In both soils Actinobacteria are specifically enriched in roots, while for Golm soil grown plants also Bacteroidetes are characteristic for the rOTUs sub-community.

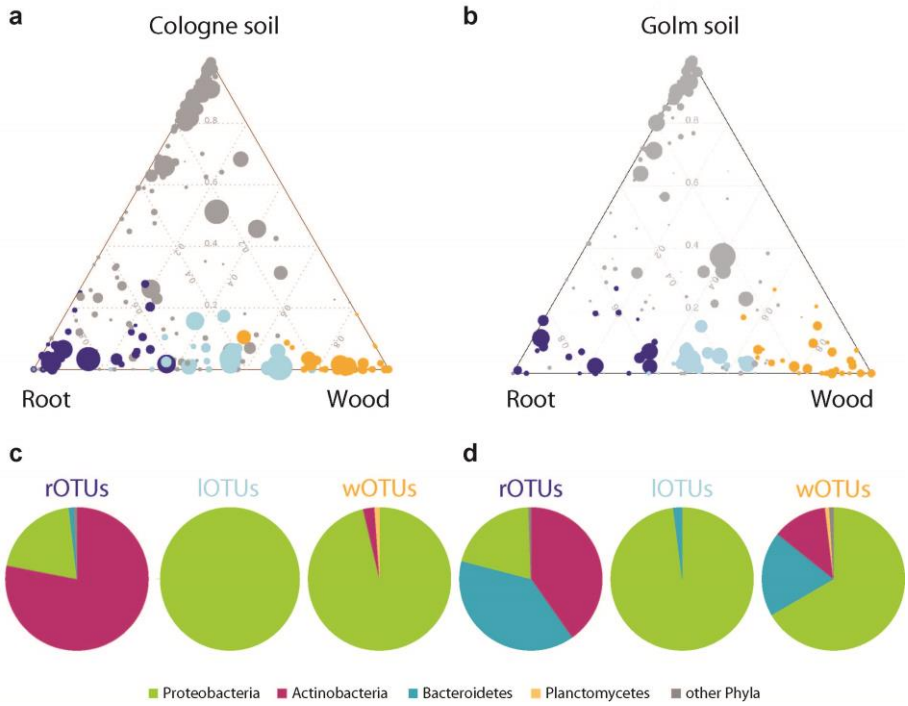


Figure 9: Ternary plot similar to Figure 8 including the wood compartment in **(a)** Cologne soil and **(b)** Golm soil. Specifically root-enriched OTUs (dark blue, rOTUs), wood-enriched community members (orange, wOTUs) and OTUs shared by root and wood compartments (light blue, lOTUs) (FDR<0.05). Taxonomic composition of rOTUs, lOTUs, and wOTUs subcommunities are given in the pie charts for **(c)** Cologne soil and **(d)** Golm soil. The size of each segment in the chart is proportional to the cumulative relative abundance of OTUs assigned to the indicated taxa.

2.3.5 Influence of soil type on root-inhabiting bacterial communities

To examine a potential role of soil type on the root-inhabiting bacterial assemblage, we compared the bacterial profiles obtained from plants grown in Cologne and Golm soils. Within the rOTUs communities, 9 OTUs were enriched in roots of both tested soil types. 21 and 13 OTUs were significantly enriched only in roots derived from Cologne or Golm soils in comparison to the

corresponding unplanted soil, respectively. Within the lOTUs community, 14 OTUs were enriched in both tested soil types. 12 and 8 OTUs were significantly enriched only in the lOTUs sub-community derived from Cologne or Golm soils, respectively (Figure 10).

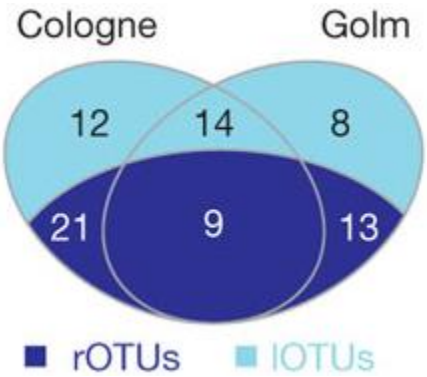


Figure 10: Numbers of lOTUs (light blue) and rOTUs (dark blue) in the indicated soils (FDR<0.05).

For 15 of all root-inhabiting OTUs (lOTUs plus rOTUs) in both soils a significant differential enrichment between the soil types could be observed, supporting the conclusion that the soil type strongly defines the structure of root-inhabiting bacterial communities (see Appendix, Supplementary Figure 1).

2.3.6 Influence of the plant genotype on the root-inhabiting bacterial communities

To examine a potential host-genotype effect on the composition of root-inhabiting bacterial communities two Arabidopsis ecotypes, Landsberg (Ler) and Shakdara (Sha), have been investigated. A single OTU, classified as *Actinocorallia sp.* and belonging to the Actinobacteria phylum, was specifically root-enriched in Cologne and Golm soil grown plants (Figure 11a) and showed a differential accumulation between the two Arabidopsis ecotypes tested (Figure 11b). This differential accumulation of an approximately 10-fold difference in RA was robust over independent experiments using Cologne soil collected during fall or spring. A similar differential accumulation trend was found between these ecotypes grown in Golm soil although it was statistically not significant.

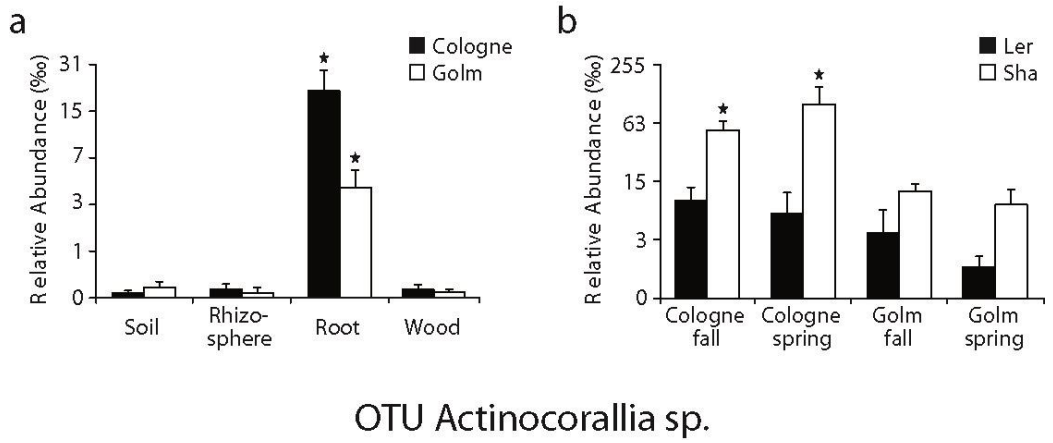


Figure 11: (a) RA of OTU *Actinocorallia sp.* in the indicated compartments (mean ± s.e.m). Asterisks indicate significant enrichment in the root compartment over all other indicated compartments (FDR<0.05). **(b)** Differential RA of OTU *Actinocorallia sp.* in the root compartment of the indicated Arabidopsis ecotypes (mean ± s.e.m.). Asterisks indicate significant differences (FDR<0.05).

End-point PCR analysis using PCR primers designed on the basis of the representative sequence of the OTU *Actinocorallia sp.* that are specific for Actinomycetales independently validated their presence in soil-grown roots of Sha and Ler (Figure 12b). PCR amplicons were also detectable in unplanted soil samples but not in surface-sterilized and crushed seeds of either accession (Figure 12b), suggesting root-inhabiting Actinocorallia are recruited from soil and not passed on via the seeds from one generation to the next. Bacterial DNA could also not be detected in axenically grown plants and surface sterilized seeds using the primer combination 799F2-1193 utilized for 16S ribotyping (Figure 12a).

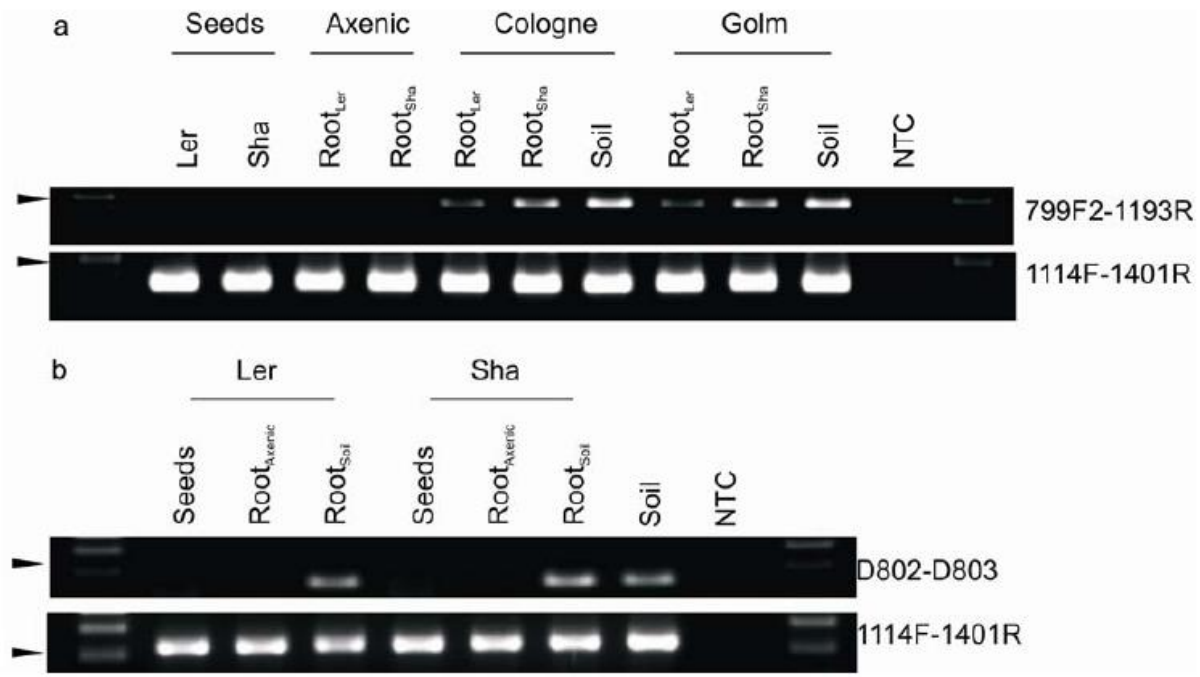


Figure 12: (a) An end-point PCR was performed using with the indicated DNA templates and the primers 799F2 and 1193R (upper lane). No amplification products were obtained using DNA extracted from roots of axenically grown plants or surface-sterilized crushed seeds. The indicated primer combination 1114F and 1401R (Heuer *et al.*, 1997) (lower lane) amplifies 16S rRNA gene sequences from A. thaliana plastids as loading control. Arrowheads indicate the position of a 400 bp fragment size as estimated by a DNA size marker. **(b)** An end-point PCR was performed with the indicated DNA templates and primers. Primers D802 and D803 specifically amplify Actinomycetales 16S rRNA genes. Arrowheads indicate the position of a 200 bp fragment. NTC, non-template negative control.

2.4 Investigation of rhizoplane and soil bacterial communities by imaging techniques

The Arabidopsis rhizoplane, i.e. the root surface and therefore the interface between host tissue and rhizosphere soil, was largely eliminated by sonication during sample preparation for the pyrosequencing-based 16S rRNA gene survey (compare Figure 17) in order to enrich for root-inhabiting bacteria. To illustrate the presence of bacteria present on the rhizoplane we applied scanning electron microscopy (SEM) and secondly, to gain insights into their taxonomy and spatial distribution, we employed *catalyzed reporter deposition - fluorescence in situ hybridization* (CARD-FISH).

2.4.1 Detection of bacteria attached to the Arabidopsis rhizoplane by scanning electron microscopy

Scanning electron microscopy (SEM) analysis of the rhizoplane revealed morphologically diverse bacteria-like structures. (Figure 13 a to e). These structures were often embedded in a matrix which could be an extracellular matrix that is often produced by bacteria colonizing various surfaces (Flemming & Wingender, 2010).

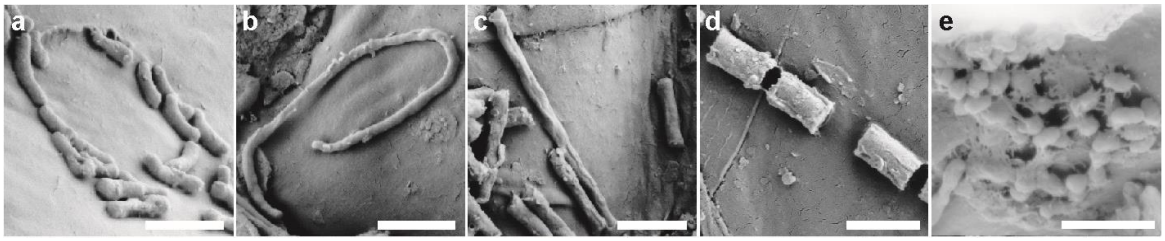


Figure 13: (a to **e)** Scanning electron micrographs of bacteria-like structures on the rhizoplane. Scale bars=1 µm

2.4.2 Investigation of bacteria attached to the Arabidopsis rhizoplane with *catalyzed reporter deposition – fluorescence in situ hybridization*

To characterize the bacteria that were found on the rhizoplane with SEM, *catalyzed reporter deposition – fluorescence in situ hybridization* (CARD-FISH) was applied.

A high density of bacteria was visualized by using the probe EUB338 that detects the majority of known Eubacteria (Figure 14a). No particulate CARD-FISH signals were found using the reverse complement probe of EUB338 as a negative control (Figure 14b). All three taxa that were found to dominate the root-inhabiting community according to pyrosequencing-based 16S ribotyping were also detected on the rhizoplane by CARD-FISH (Betaproteobacteria, Bacteroidetes and Actinobacteria; Figure 14 c to e). Each CARD-FISH probe detected a distinct colonization as for example bacteria detected with a Betaproteobacteria specific probe tended to form clusters or aggregates, while signals detected using a Bacteroidetes specific probe were evenly distributed over the whole root surface. Hybridizations using an Actinobacteria specific probe lead to detection of filamentous signals on the rhizoplane. Particulate CARD-FISH signals were also not found upon hybridization of roots of axenically grown plants using the EUB338 probe indicating that CARD-FISH indeed visualizes bacteria (Figure 14f).

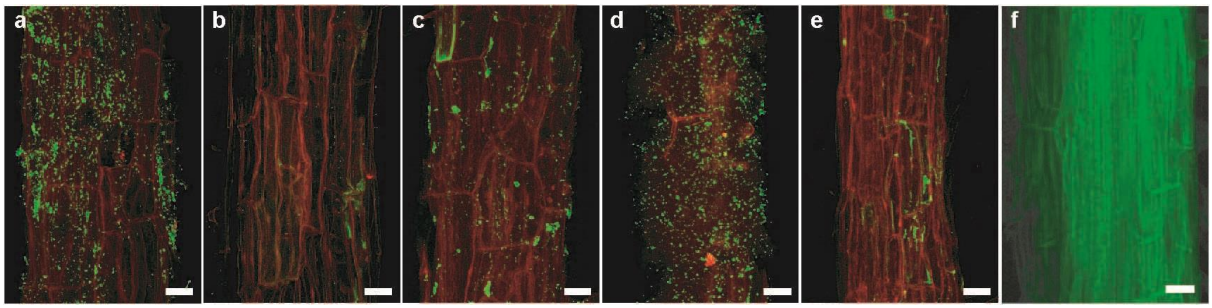


Figure 14: (a to **e)** CARD-FISH detection of bacteria (green, AlexaFluor488) on the root surface (red, root autofluorescence) of soil grown plants by confocal laser scanning microscopy. **(a)** most Eubacteria detected with probe EUB338. **(b)** negative control with reverse complementary probe of EUB338 (NONEUB). **(c)** Betaproteobacteria detected with probe BET42a. **(d)** Bacteroidetes detected with probe CF319a. **(e)** Actinobacteria detected with probe HGC69a. **(f)** Hybridization of axenically grown root with probe EUB338. Gain increased to capture background fluorescent signals resulting from tissue autofluorescence and/or AlexaFluor-tyramide conjugate binding to plant cell walls. Scale bars= 20 µm.

In order to show that signals detected by CARD-FISH were indeed bacteria and not randomly occurring artifacts, double-labelings with the EUB probe and the Betaproteobacteria specific probe BET42a were performed. The CARD-FISH protocol was conducted twice with the same root specimen successively using different fluorescent dyes. As expected more signals could be detected with the EUB probe than with the BET42a probe. As seen also in single labelings signals detected with BET42a tend to cluster. Nearly all BET42a signals co-localized with the EUB signals. No particulate signals could be detected if the same root specimen was labeled twice using the NONEUB control probe subsequently with both fluorophores indicating that the double labeling procedure does not lead to false positive signals (Figure 15).

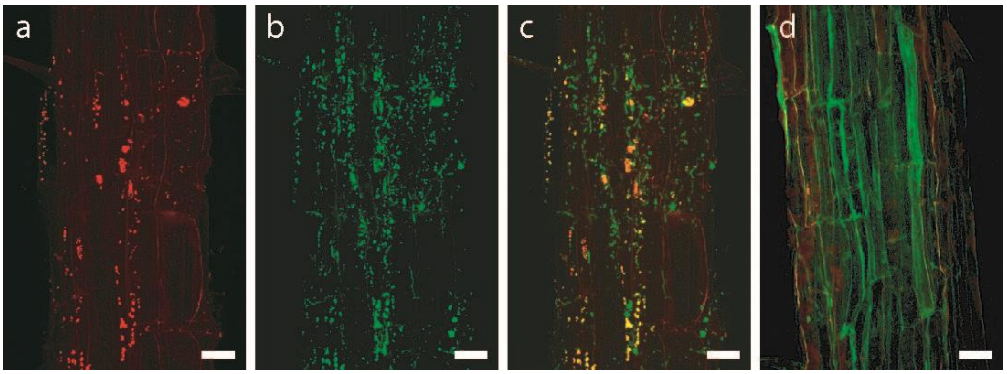


Figure 15: Simultaneous CARD-FISH detection of Betaproteobacteria bacteria (red, AlexaFluor546) and most Eubacteria (green, AlexaFluor488) on the root surface by confocal laser scanning microscopy. **(a)** Betaproteobacteria detected with probe BET42a. **(b)** Most Eubacteria detected with probe EUB338. **(c)** Merged (a and b), yellow signals result from co-localization of the BET42a and EUB signals. **(d)** Negative control with reverse complementary probe of EUB338 (NONEUB) used twice in a double-labeling with both fluorophores. Scale bars= 20 µm.

2.4.3 Investigation of bacteria in unplanted soil and rhizosphere samples with CARD-FISH

Soil and rhizosphere samples were also investigated by CARD-FISH to obtain a PCR and pyrosequencing independent survey of their inhabiting bacterial communities. The same phylum-specific CARD-FISH probes used for the rhizoplane were also utilized for soil and rhizosphere samples (see section 2.4.2 and Table 6). Soil and rhizosphere samples were spotted in equal amounts on individual polycarbonate filters. For hybridization experiments with the respective probes the filters were cut into equal-sized segments to allow quantitative comparisons between the different bacterial taxa. The number of detected signals for the respective taxa was then represented as percent of the signals detected with the EUB probe, representing a proxy of all bacteria present in the sample.

Using the taxa specific probes fewer signals in soil and rhizosphere samples were detected compared to the EUB338 probe (Figure 16). Betaproteobacteria and Actinobacteria were calculated to make up about 15% of the total bacterial community of both soil and rhizosphere samples. For unplanted soil these values are in the same range as determined by 16S ribotyping, while for rhizosphere samples 16S ribotyping showed a lower abundance of Actinobacteria (app. 5%, compare Figure 3). Quantification of Bacteroidetes showed a different detected RA between CARD-FISH and 16S ribotyping data. CARD-FISH estimates an abundance of 9% and 18% for unplanted soil and rhizosphere samples, respectively, while a relative abundance of < 5% was determined by 16S ribotyping for this phylum in both compartments. Using a probe that detects Chloroflexi an abundance of 5% and 7% was detected in soil and rhizosphere samples, respectively, which is in contrast to a relative abundance of app. 80%, as determined by 16S ribotyping. Therefore, CARD-FISH analysis in accordance with previously published studies (Janssen, 2006; Nacke *et al.*, 2011; Fierer *et al.*, 2012) supported the justification to remove Chloroflexi reads from the pyrosequencing dataset.

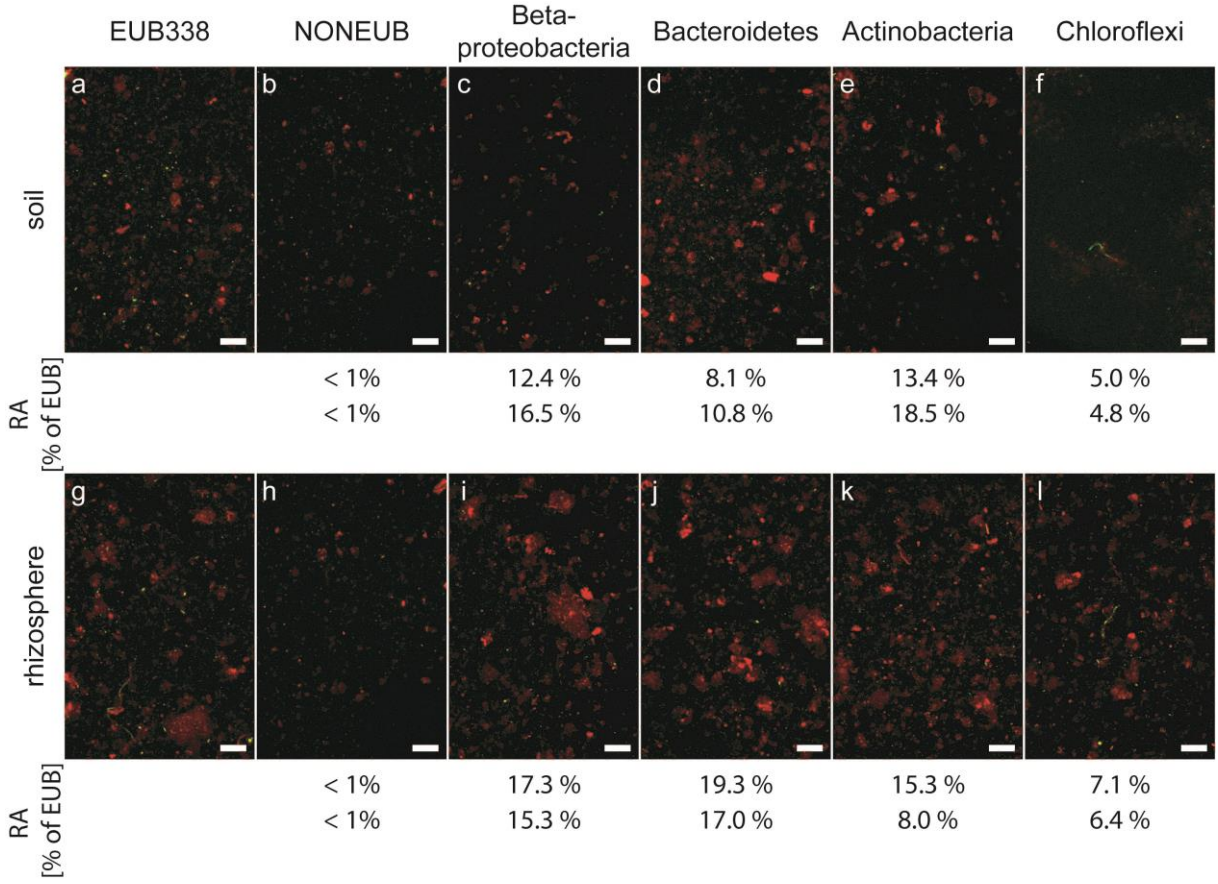


Figure 16: CARD-FISH of soil and rhizosphere samples. Pictures show soil and rhizosphere samples (red, soil autofluorescence) spotted in equal amounts on individual polycarbonate filters visualized by confocal laser scanning microscopy. Targeted taxa of the hybridization experiments are indicated on the top. The top panel (**a** to **f**) depicts microscopic images of the Cologne soil compartment, the lower panel (**g** to **l**) of the Cologne rhizosphere compartment. Numbers beneath the pictures depict relative abundances of bacterial taxa obtained in two replicates. These were

calculated as the relative proportion of signals obtained after hybridizations with an indicated taxon-specific and the EUB probe on filter segments derived from the same sample. Cell counts were conducted by Hannes Schmidt at the University of Bremen.

3 Discussion

Eukaryotes and microbes have co-existed and co-evolved for millions of years (Thrall *et al.*, 2001). Through their roots, land plants are exposed to soil, one of the richest microbial ecosystems on earth (Tringe *et al.*, 2005). Members of the soil biota can colonize the rhizosphere and even multiply inside roots as endophytes that modulate plant growth and development (Hardoim *et al.*, 2008). 16S ribotyping by pyrosequencing amplicon libraries generated by PCR amplification of the bacterial phylogenetic marker 16S rRNA gene allows the comparison of such bacterial communities associated to roots, rhizosphere and soil at high resolution. At the same time 16S ribotyping permits first conclusions about the biological significance of the findings, since, in contrast to low-resolution fingerprinting techniques like DGGE or T-RFLP profiles, it allows to obtain taxonomic information of the studied bacterial communities. In this study, 16S ribotyping was established to characterize bacterial communities associated to roots of the model plant Arabidopsis and its rhizosphere in comparison to the microbiota of unplanted soil in a reproducible way across full factorial replicates. Two Arabidopsis ecotypes were grown in two natural microbe-rich soils of contrasting geochemistry to study to which extend root-inhabiting bacterial communities are influenced by the host genotype and by environmental component soil type. Furthermore, bacterial communities associated to wooden sticks, representing a metabolically inactive lignocellulosic matrix for bacterial colonization, were analyzed to deconvolute the complexity of the bacterial microbiota specifically associated to Arabidopsis roots.

3.1 Defining a robust 16S ribotyping procedure for characterization of rhizosphere and root microbiota

Profiling microbial communities via 16S ribotyping involves DNA extraction, PCR amplification of the bacterial 16S rRNA gene and sequencing of amplicon libraries via pyrosequencing. All of these steps bear risks of introducing biases to the final result. It has been shown that the choice of the method of DNA extraction can influence the determined composition of the analyzed bacterial communities (Morgan *et al.*, 2010) which is why the same DNA extraction kit was used to extract DNA from all samples of the different compartments. Secondly, the choice of the PCR primer pair is another critical step, as an analysis of different hyper-variable regions can lead to over- or underestimation of bacterial species richness and evenness, hence resulting in biased ribotype profiles (Wang & Qian, 2009; Engelbrektson *et al.*, 2010). The choice of PCR primers is also constrained by the sequencing technique: in this study we employed pyrosequencing that allows the sequencing of amplicons with a maximum read length of 400 bp. In the present study the hypervariable regions V5 and V6 have been analyzed, which are capable to recapitulate the

composition of bacterial communities in environmental and host-associated samples (Huse *et al.*, 2008). Additionally, PCR primers can only be designed based on sequences known up to date, potentially omitting a huge number of bacterial species that have not been described so far. This is also reflected by the low coverage of sequences obtained from environmental samples like soil or sea water in established databases like RDP or SILVA (Gans *et al.*, 2005). Therefore, a true "universal PCR primer pair" targeting all bacterial 16S rRNA gene sequences in a given ecosystem, does not exist, and data or comparisons of different studies have to be analyzed in the light of the used PCR primer pair. The presence of host DNA in the template DNA adds another layer of complexity, as co-amplification of host DNA might alter the determined composition of the bacterial communities (Huys *et al.*, 2008). A single nucleotide exchange was introduced into a previously published PCR primer 799F (Chelius & Triplett, 2001) which resulted in exclusion of amplification of the host organelle 18S rRNA gene. However, this modification yielded a disproportionate abundance of reads assigned to the phylum Chloroflexi of app. 80% in the whole dataset (Figure 2). Chloroflexi have been reported to represent members of the soil microbiota but typically at a cumulative relative abundance below 5% (Janssen, 2006; Fierer *et al.*, 2012). CARD-FISH analysis of soil and rhizosphere samples revealed that the observed disproportion of Chloroflexi reads in this study is likely to be a consequence of the nucleotide exchange. Nevertheless, analysis of a subset of soil and root samplers with an additional PCR primer set revealed that the rank distribution and relative abundance of bacterial taxa are robust against the observed PCR primer bias, with the exception of reads assigned to the phylum Chloroflexi (see appendix). Therefore, sequence reads directly classified or reads clustered to an OTU which was classified as Chloroflexi were removed *in silico* from the analysis.

The utilized sequencing technology of pyrosequencing introduces sequence errors, especially in case of longer homopolymers, i.e. repetitive stretches of a single base, or by wrong base-calling resulting in artificially inflated bacterial diversity in analyzed samples (Kunin *et al.*, 2010). Such biases can be reduced by applying strict quality trimming measures.

Nevertheless analyzing individual sequence reads is hampered by an underrepresentation of bacterial 16S rRNA gene sequences detected in environmental samples in the aforementioned databases RDP and SILVA. To increase the number of analyzed reads and the resolution of the analysis sequence reads can be clustered to OTUs which at the same time absorbs undetected sequencing errors. Aim of this study was to infer the role of soil type and host genotype in shaping bacterial communities associated with plant roots, which premises a robust quantification of all community members. We determined the lower limit of reproducibility based on correlation analysis of technical replicates. The PCR based pyrosequencing approach only allows a robust quantification of OTUs with RA of ≥ 5‰ and therefore, only OTUs above this threshold were

analyzed. In summary, 16S ribotyping respecting the aforementioned quality standards combined with OTU clustering and a minimum abundance threshold of 5‰ is a robust methodology to investigate Arabidopsis associated bacterial communities.

3.2 The taxonomical composition of the bacterial communities of soil, rhizosphere and roots

To get an overview of the taxonomical structure of soil, rhizosphere and root-inhabiting bacterial communities all sequencing reads were classified using the SILVA database (Pruesse *et al.*, 2007). Both Cologne and Golm soil bacterial communities were dominated by Acidobacteria, Planctomycetes, Actinobacteria and Proteobacteria (Figure 3), phyla that can be typically found in forest- grassland- and agricultural soils (Fierer *et al.*, 2012). In terms of RA of these phyla are not significantly different between the tested soils. Both soils are abandoned agricultural soils utilized for crop cultivation in the past and have minor differences in the content of macronutrients (Table 4). More importantly, the pH values of both soils are around pH 7 and therefore neutral (Table 3). The pH has been shown to be the major edaphic factor shaping soil bacterial communities (Rousk *et al.*, 2010). The main difference between the tested soils is their physical texture: Cologne soil contains very small clay and loam particles, while Golm soil is characterized by larger sand particles (Table 3). In accordance with our observation, soil texture played a minor role for the profile of bacterial communities and is overridden by the influence of pH (Nacke *et al.*, 2011). The detection of globally similar compositions of the bacterial communities in the two tested soils is therefore not an unexpected finding although the collection sites of the two soils were app. 500 km apart.

On the basis of the overall taxonomical structure the rhizosphere bacterial community resembles the community of unplanted soil (Figure 3). Major shifts in community composition, as observed by DGGE in other plant species like *Brassica napus* (Costa *et al.*, 2006) could not be observed. A reason might be that the soil sphere that is influenced might be very small due to the delicate nature of the Arabidopsis root system. The sampled rhizosphere compartment might therefore contain a substantial amount of uninfluenced soil. However, also in an investigation of oak trees by pyrosequencing-based 16S ribotyping differences between rhizosphere and surrounding soil bacterial communities could only be detected at higher resolution but not at the phylum level (Uroz *et al.*, 2010).

Arabidopsis root-inhabiting bacterial communities are dominated by Proteobacteria, Actinobacteria and Bacteroidetes. All of these phyla have been shown previously to be very abundant in endophytic bacterial communities, e.g. of potato (Manter *et al.*, 2010), while

Proteobacteria have been found before to dominate poplar as well as rice endophyte communities (Sessitsch *et al.*, 2001; Gottel *et al.*, 2011). Proteobacteria are significantly enriched in root-inhabiting bacterial communities against soil and rhizosphere communities in both soils, while Actinobacteria are significantly enriched in bacterial communities of Cologne- and Bacteroidetes in Golm soil grown plants, respectively. This shows the influence of the soil type on root-inhabiting bacteria, which becomes more apparent in the OTU-based analysis and will be discussed later in detail. A single family of bacteria dominated each of the aforementioned phyla (Figure 3), while other detected bacterial families in the same phylum are less abundant. This indicates that whole bacterial families might be under influence of Arabidopsis roots and form building blocks of root-inhabiting bacterial communities.

The rhizoplane, i.e. the root surface, was largely eliminated by the sonication treatment during sample preparation for 16S ribotyping to enrich for Arabidopsis endophytes. Nevertheless, rhizoplane colonization often precedes colonization of the inner root tissue (Rosenblueth & Martinez-Romero, 2006). Thus, the detection of bacteria on the rhizoplane links root-inhabiting bacteria to the reservoir of soil bacteria (Compant *et al.*, 2010). All phyla that dominate the root-inhabiting bacterial communities by 16S ribotyping were also detected abundantly by CARD-FISH on the rhizoplane of Arabidopsis grown in Cologne soil. No particulate hybridization signals could be detected using the reverse complement probe of EUB338 on soil-grown roots, as well as using EUB338 in hybridizations using axenically grown roots as negative controls. Each probe targeting a specific taxon showed a different spatial pattern of signals. In double labelings signals of probes targeting a specific taxon co-localize with signals detected by the EUB338 probe that targets a wide range of bacterial phyla. Together, these findings show that obtained hybridizations signals indeed detect bacteria and not unspecific signals, like soil particles. Moreover, the different probes used show indeed different subclasses of bacteria. CARD-FISH analysis supports the idea that the phyla members, which dominate the root-inhabiting community also abundantly colonize the rhizoplane. This suggests that the root-inhabiting bacteria originate from the surrounding soil and colonize the root compartment via the rhizoplane.

Chloroflexi were detected by CARD-FISH in Cologne soil and rhizosphere samples with an app. RA of 5% in contrast to 80% as detected by 16S ribotyping. Such a strong discrepancy points towards a bias in detection of this phylum introduced by the utilized PCR primers 799F2-1193R, as previously discussed. CARD-FISH hybridization signals for Actinobacteria and Proteobacteria in unplanted soil largely reflect and thus confirm the results obtained by 16S ribotyping. Both methodologies detect RAs of app. 15% for Actinobacteria as well as for Proteobacteria. Bacteroidetes were more abundantly detected by CARD-FISH than by 16S ribotyping. Slight differences between RA obtained by 16S ribotyping and hybridization with CARD-FISH probes,

however, can be explained by the characteristics inherent to the respective methodology. Both techniques make use of short DNA molecules targeting the 16S sequence. FISH probes target the rRNA molecules and therefore represent a proxy of the metabolic state of the visualized bacteria. In contrast, PCR primers amplify the bacterial 16S rRNA gene and therefore DNA, a molecule that is much more stable than RNA and could be amplified from dead bacteria as well as from bacteria in resting states, such as spores. All nucleotides, irrespective of their targeted molecule, have been designed to detect a specific range of bacteria. Since such a design cannot be done perfectly, both methodologies will introduce a bias to the observed composition of microbial communities. PCR primers might not target all bacterial phyla evenly, while FISH probes might not be able to target a whole phylum without also producing false positive signals (Siyambalapitiya & Blackall, 2005).

3.3 Defining the assembly cues for the root-inhabiting microbiota

To maximize the number of analyzed reads and the resolution of the analysis all sequencing reads were clustered to OTUs at a sequence identity of ≥97%. The composition of OTUs in each of the tested compartments was then compared to find OTUs that are differentially accumulated in one of the compartments. This also allows the retrieval of OTUs indicative for a given condition, like soil- or genotype. As already shown by classification using the SILVA database, and consistent with studies using other plant species, this comparison revealed a striking impact of the host-associated bacterial microbiota (Berg & Smalla, 2009; Hardoim *et al.*, 2011; Inceoglu *et al.*, 2011). Roots are preferentially colonized by OTUs classified as Actinobacteria, Proteobacteria and, especially Golm soil-grown roots, by OTUs classified as Bacteroidetes. Wooden sticks are preferentially colonized by Proteobacteria and only to a minor extent by Bacteroidetes and Actinobacteria (Figure 9). Similar results were obtained by another laboratory by investigating the root-inhabiting bacterial communities of eight Arabidopsis accessions grown in two natural soils. Lundberg and colleagues also used pyrosequencing-based 16S ribotyping, a similar sampling protocol and a similar computational pipeline. Despite of the utilization of a different primer pair to characterize root-inhabiting bacterial communities the root-inhabiting microbiota in their setting is dominated by OTUs classified as Actinobacteria or Proteobacteria, while the differences between unplanted soil and rhizosphere were less pronounced (Lundberg *et al.*, 2012). Interestingly, members of the Actinobacteria and Bacteroidetes were reported to be associated with disease suppressive soils as they were more abundantly detected in suppressive than in conducive soils (Mendes *et al.*, 2011).

Comparisons between bacterial communities inhabiting roots and wooden sticks allow the identification of sub-communities based on the assembly cues they react to. A first subgroup, termed lOTUs, colonizes both living roots and dead wood indicating that these community members are attracted by cell wall features like lignocellulose. lOTUs members have been classified as Betaproteobacteria and Bacteroidetes. Interestingly, members of these two phyla are characterized as copiotrophic soil bacteria, i.e. they compete successfully only when organic resources are abundant (Fierer *et al.*, 2007). Thus, colonization of Arabidopsis roots by the lOTUs sub-community may reflect their ability to proliferate in the presence of polysaccharide polymers and might contribute to the decomposition of organic matter after plant death. The results predict that the lOTUs sub-community is not Arabidopsis-specific, but may consist of saprophytic bacteria that would naturally be found on any plant root or plant debris in the tested soils. Within the lOTUs community *Burkholderia* are well represented. Members of this genus are often saprophytic and at the same time can have plant growth promotion capacity by secreting a vast array of secondary metabolites into the soil (Vial *et al.*, 2007). Consistent with our prediction of this sub-community being less specialized to a particular plant species *Burkolderia* Bu72, an endophyte initially isolated from lupine (Barac *et al.*, 2004), is also able to exhibit plant growth promoting effects in poplar.

A second sub-community, termed rOTUs, comprises bacteria that are significantly enriched in the bacterial community associated to living plant roots which suggests that cues from metabolically active host cells are needed for their root colonization and persistence. rOTUs sub-community members were mostly classified as Actinobacteria. Actinobacteria may provide probiotic functions for the plant as, for example, Actinobacteria are known to produce a vast diversity of antimicrobial compounds (Basilio *et al.*, 2003) and endophytic Actinobacteria were shown to protect chickpea from the pathogens *Phytophthora*, *Pythium* and *Botrytis* (Misk & Franco, 2011). Additionally members of the Actinobacteria are capable of producing the plant growth hormone auxin (Merzaeva & Shirokikh, 2010). The production of auxin is a widely accepted mode of plant growth promotion by rhizobacteria (Spaepen & Vanderleyden, 2011).

A third sub-community, ‘wOTUs’, mainly comprised by Proteobacteria, was only significantly enriched in the wood compartment, but not in the root compartment. Members of the wOTUs community might exploit lignocellulose, but lack the competence to colonize living Arabidopsis roots, potentially because they might be repelled by root exudates compounds or outcompeted by other bacteria in the root habitat.

Together these findings support the idea that the assembly and maintenance of the root bacterial microbiota in Arabidopsis is a host process comprising attractant and repellent activities. The main process how plants influence bacteria in their surrounding soil is rhizodeposition for which

both attractant and repellent compounds have been described. Carbon fixed by plants from the atmosphere and deposited in soil through roots has been shown by stable isotope probing with labeled $^{13}CO_2$ to be utilized by bacteria in the rhizosphere and therefore influences the microbial community in the vicinity of the root (Prosser *et al.*, 2006; Li *et al.*, 2011). Rhizodeposits can account for 40% of the total carbon assimilated from the atmosphere by plants depending on plant species, age and environmental factors (Lynch & Whipps, 1990). Displaying such tremendous costs it is likely that a fitness benefit for the plant is linked to rhizodeposition, potentially recruitment of beneficial soil-borne bacteria. These rhizodeposits comprise actively secreted compounds like strigolactones and flavonoids (Steinkellner *et al.*, 2007), passively leaking chemicals like sugars, amino acids and organic acids, lysed or sloughed off root cells (Vicre *et al.*, 2005; Li *et al.*, 2011) and gases like respiratory CO_2 (Shibistova *et al.*, 2002). Rhizodeposits provide a source of nutrients and function as chemoattractants for bacteria, making them move towards the exudates. Chemotaxis towards root exudates is one of the major traits for successful root colonization as for example shown in tomato - *Pseudomonas* interactions (de Weert *et al.*, 2002). On the other hand root exudates also contain repellent activities towards specific bacteria. Bais and colleagues identified rosmarinic acid as a component of sweet basil root exudates upon challenge with *Pythium ultimum* exhibiting antimicrobial activity against an array of soil-borne bacteria including the opportunistic pathogen *Pseudomonas aeruginosa* (Bais *et al.*, 2002). Similarly, naphtoquinones secreted by *Lithospermum* were found to have biological activity against soil-borne bacteria (Brigham *et al.*, 1999). The same compounds can even have contrasting effects on different bacterial phyla. Eilers and colleagues added common compounds of root exudates, namely glucose, glycine and citric acid, to microcosms containing natural soils and monitored changes in the microbial communities via 16S ribotyping. While the relative abundance of Acidobacteria was reduced, members of the Proteobacteria, Actinobacteria and Bacteroidetes were enriched upon all three treatments. Interestingly, the final outcome of the enrichment was strongly soil type dependent (Eilers *et al.*, 2010). Since root-associated bacteria and endophytes colonize the root tissue via the rhizosphere and the rhizoplane (Compant *et al.*, 2010), rhizodeposition is very likely also an important trait for shaping the root inhabiting bacterial community defined in this study and a possible explanation for the observed differences between root-, wood- and unplanted soil-inhabiting bacterial communities.

The rhizosphere microbial communities were found to be significantly different from the communities retrieved from unplanted soil by PERMANOVA analysis (Table 2) and to have a reduced OTU richness. A reduced richness in bacterial diversity is commonly found for rhizosphere communities and it is hypothesized that only a subset of soil-inhabiting bacteria are rhizocompetent, i.e. able to colonize the rhizosphere (Costa *et al.*, 2006). We did not identify

OTUs that differentiate these two compartments in all seasonal batches of both soils tested (Figure 7). While for Cologne soil two rhizosphere-OTUs are retrieved both from samples of plants grown in spring and in fall soils, no rhizosphere-OTUs could be detected in Golm soil that fulfill these criteria. Hence, the rhizosphere effect is subtle, which is consistent with the findings of Lundberg et al., 2012. Investigation of the Arabidopsis rhizosphere conducted by Badri et al. 2009, detected a rhizosphere effect only after two generations of plant growth in the same soil, indicating an increasing influence on the soil with time. The presented investigation does not exclude that a pronounced rhizosphere effect could be observed after extended incubation time or in soil types to which they are adapted to. On the other hand, partitioning of photosynthetically fixed carbon to the rhizosphere decreases with plant age (Jones *et al.*, 2009), which is why younger roots might exert a more pronounced effect on rhizosphere bacterial communities.

3.4 The soil type influences the composition of root-inhabiting bacterial communities

To examine a potential role of the soil type on the composition of root-inhabiting bacterial communities profiles from roots grown in Cologne and Golm soil were compared. Of all root-inhabiting bacteria (lOTUs plus rOTUs) app. 70% are significantly enriched in only one of the two tested soils (Figure 10). Additionally, app. 17% of these OTUs are significantly different between bacterial communities of Cologne and Golm soil grown roots (Figure 10). This identifies soil type as the most important environmental variable influencing the quantitative and qualitative composition of the Arabidopsis root-inhabiting microbiota. With 33 and 21 soil-specific regulated root-enriched OTUs for Cologne and Golm soil, respectively, the soil type influences root-inhabiting bacterial communities stronger than the host genotype which influences a single OTU, as discussed later. This is consistent with other recently published studies (Inceoglu *et al.*, 2011; Weinert *et al.*, 2011; Lundberg *et al.*, 2012). Soil type-dependent root colonization processes observed in this study support the conclusion that a subset of, if not all, root-inhabiting bacteria originated from the microbiota reservoir present in soil and might reflect different bacterial start inocula of the tested natural soils. In addition we were not able to detect bacterial DNA in surface sterilized seeds (Figure 12) which strengthens the claim that root-inhabiting bacteria are soil-borne and not transmitted via the seeds, which is another way how endophytes can be transferred from one plant generation to the next (Johnston-Monje & Raizada, 2011). Whether soil-type dependent root-inhabiting bacterial communities are a plant's reaction towards different abundances of (micro-)nutrients or an action to attract antagonists to a soil-specific pathogen is an intriguing question that needs further experimentation.

3.5 The host genotype influences root-inhabiting bacterial communities to a minor extent

We tested the influence of the host genotype on the root-inhabiting bacterial community by comparing communities of two ecotypes of Arabidopsis. Ecotypes are genetically distinct geographic varieties within the same species that are adapted to different habitats. One of the tested ecotypes, Ler, was originally collected in Poland, the other, Sha, in central Asia. The adaption to different habitats leads to small scale differentiations in the genome. Correlating phenotypic traits to single nucleotide polymorphisms between ecotype genomes is a common strategy to prove the genetic basis of an observed trait and in the end to identify the responsible gene(s).

Comparison of the community profiles of the two accessions identified a single OTU, classified as *Actinocorallia sp.*, that showed a significant differential accumulation. This colonization process is under quantitative control of the host genotype. Its ultimate characteristic is still determined by the soil type and/or the specific composition of bacterial communities of the respective soil type as shown by the differential accumulation in the two experimental soils. We therefore concluded that host genotype takes minor control on the composition of root-inhabiting communities compared to soil type (see above).

This is in agreement with similar results obtained by Lundberg and colleagues. Testing eight Arabidopsis accessions only twelve OTUs could be identified that differentially accumulate between these ecotypes. These results are reminiscent of the mouse gut microbiota for which the host genotype quantitatively, but not qualitatively, contributes to its structure (Benson *et al.*, 2010). Ecotype/cultivar specific root-inhabiting bacterial communities have also been also shown in other species like potato (Manter *et al.*, 2010). Genotype dependent differences in the root-inhabiting bacterial communities between ecotypes or cultivars might be explained by compounds of the aforementioned root exudates. Micallef and colleagues (2009) hypothesized that differences in the microbial communities of the rhizosphere of eight Arabidopsis accessions detected by T-RFLP and ARISA might be due to differences in their root exudate profile (Micallef *et al.*, 2009). Similarly, a T-DNA insertion knock out of the ABC transporter ABCG30 in Arabidopsis resulted in an altered root exudate profile and induced changes in the rhizosphere bacterial communities after two generations of plant growth in the same soil (Badri *et al.*, 2009). The influence of genes in signaling pathways involved in the plant innate immune system on endophytes has also been investigated. Iniguez et al, 2005, show that the Arabidopsis mutant *npr1* impaired in salicylic acid signaling can increase the colonization efficiency of the inner root tissue by *Klebsiella pneumoniae* 342 *in vitro* (Iniguez *et al.*, 2005). Long et al, 2010, tested the

influence of another signaling hormone in the plant innate immune system, ethylene, on the composition of bacterial endophyte communities of plants grown in natural soil by a culturing-based approach. Arabidopsis mutants impaired in ethylene signaling accumulated different bacterial strains than wild-type plants and one bacterial strain isolated from a mutant impaired in ethylene perception was only able to re-colonize this particular mutant but not the wild-type in a gnotobiotic system (Long *et al.*, 2010).

Hence, one plausible hypothesis is that the host genotype influences root-associated bacteria at two levels: bacteria from the soil reservoir are attracted to or repelled from the rhizosphere via root exudates or rhizodeposition and in a next step only a subset is allowed to enter the inner root tissue. This is supported by the reduced OTU richness in the root-inhabiting bacterial communities in comparison to soil (Figure 4) (Gottel *et al.*, 2011). A potential gatekeeper might be the plant innate immune system. Part of this question is the unresolved paradox of how the plant innate immune system discriminates friends from foes. The innate immune system triggers immune responses upon the detection of conserved microbe associated molecular patterns which are also present in beneficial bacteria. In depth characterization of root-inhabiting bacterial communities of Arabidopsis mutants impaired in central parts of their innate immune system as well as a survey of mutants that alter root exudation would help to elucidate underlying principles of shaping host-associated bacterial communities.

The next level of analysis, not based on the 16S rRNA sequence, would comprise functional genes of the microbiota in a given microhabitat like the plant root. Instead of 16S ribotyping, sequencing the full metagenome would result in the identification of enriched functional genes instead of enriched OTUs or bacterial species. This will lead to functional community insights, since several bacterial species might share the same functions useful for the plant.

4 Material and Methods

4.1 Material

4.1.1 Collection and storage of experimental soils

Experimental soils were collected at the Max Planck Institute for Plant Breeding Research (50.958 N/ 6.856 E, Cologne, Germany) or at the Max Planck Institute of Molecular Plant Physiology (52.416 N/ 12.968 E, Potsdam – Golm, Germany). To evaluate seasonal variation of each experimental soil two harvests were performed, one in fall and another in spring. In Cologne the fall soil was harvested in September 2009 and the spring soil in March 2010, referred to as 'Cologne fall' and 'Cologne spring' soils hereafter. The Golm soils were collected in March ('Golm spring') and September 2010 ('Golm fall'), respectively. At both sites the soils are from experimental agricultural sites that were not exposed to agrochemicals for several years. The topsoil (10-15 cm) was removed and the layer between -15 and -30 cm was harvested (shovel excavator), followed by an intermediate storage of one week at ambient temperature for drying. The experimental soils were homogenized and subsequently stored in 400 l containers at 4°C in the dark until further use. Soils were sieved with at 5 mm followed by 2 mm mesh and filled into 7 x 7 x 9 cm plastic pots for experiments. Physical and chemical properties for each soil type and harvest were determined at the "Labor für Boden- und Umweltanalytik" (Eric Schweizer AG, Thun, Switzerland). The soil analyses are listed in Table 3 and Table 4.

Table 3: Soil parameters of the experimental soils Cologne and Golm.

Soil	C. org. (%)[1]	Clay (%)	Silt (%)	Sand (%)	pH	Classification[2]
Cologne fall	1.4	13.4	37.3	49.3	7	loam
Cologne spring	1.5	12.4	37.7	49.9	6.9	loam
Golm fall	1	4.2	4.2	91.6	7.1	sand
Golm spring	0.8	3.7	6	90.3	7.3	sand

[1] organic carbon

[2] soil texture classification according FAO

Table 4: Macronutrient content of experimental soils Cologne and Golm.

Soil	Nitrate[1]	Phosphorus[1]	Potassium[1]	Calcium[1]	Magnesium[1]
Cologne fall	14.9	8.6	27.8	52.3	8.2
Cologne spring	21.6	7.7	31	62.1	9.6
Golm fall	13.7	12.9	27.8	84.8	5.4
Golm spring	21.8	11.8	31	108	7.2

[1] mg/kg, determined with H_2O extraction

4.1.2 Plant material

Arabidopsis thaliana (here referred to as Arabidopsis) seeds from the accessions Landsberg *erecta* (L*er*), Shakdara (Sha) and Columbia-0 (Col-0) were kindly provided by Dr. Maarten Koornneef, Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne (Germany).

4.1.3 Oligonucleotides

Oligonucleotides used for PCR amplification are listed in Table 5 and were purchased from Sigma (Hamburg, Germany) or Invitrogen (Karlsruhe, Germany). Oligonucleotide probes fused to a horsereddish peroxidase used for CARD-FISH are listed in Table 6 and were purchased from Biomers (Ulm, Germany).

Table 5: Oligonucleotides used for PCR amplifications

name	sequence (5' - 3')	reference
799F	AAC MGG ATT AGA TAC CCK G	Chelius & Triplett, 2001
799F2	AAC MGG ATT AGA TAC CCG G	Bulgarelli et al, 2012
D802	TCC ACG GTT TCT GCG TCG CA	Bulgarelli et al, 2012
D803	GCC TTG GTA AGG TTC TTC GCG TTG C	Bulgarelli et al, 2012
1114F	YAA CGA RCG MRA CCC	Bulgarelli et al, 2012
1193R	ACG TCA TCC CCA CCT TCC	Buchholz-Cleven et al, 1997
1401R	CGG TGT GTA CAA GGC CC	Heuer et al, 1997

Table 6: Oligonucleotide probes used for CARD-FISH

probe name	targeted taxa	sequence (5' - 3')	FA[1]	reference
EUB338	most Eubacteria	GCT GCC TCC CGT AGG AGT	35	(Amann *et al.*, 1990)
NONEUB	negative control	ACT CCT ACG GGA GGC AGC	35	(Wallner *et al.*, 1993)
BET42a	Betaproteobacteria	GCC TTC CCA CTT CGT TT	35	(Manz *et al.*, 1992)
BET42a competitor	-	GCC TTC CCA CAT CGT TT	35	(Manz *et al.*, 1992)
HGC69a	Actinobacteria	TAT AGT TAC CAC CGC CGT	25	(Roller *et al.*, 1994)
Actinobacteria comp.	-	TAT AGT TAC GGC CGC CGT	25	(Roller *et al.*, 1994)
CFX1223	Chloroflexi	CCA TTG TAG CGT GTG TGT MG	35	(Bjornsson *et al.*, 2002)
CF319a	Bacteroidetes	TGG TCC GTG TCT CAG TAC	35	(Manz *et al.*, 1996)

[1] formamide concentration (%)

4.1.4 Enzymes

DNA free DFS Tag polymerase was purchased from Bioron (Ludwigshafen, Germany).

4.1.5 Chemicals and antibiotics

Laboratory grade chemicals and reagents were purchased from Sigma-Aldrich (Deisenhofen, Germany), Roche (Mannheim, Germany), Roth (Karlsruhe, Germany), Merck (Darmstadt, Germany), Invitrogen (Karlsruhe, Germany), Serva (Heidelberg, Germany) unless otherwise stated.

4.1.6 Buffers and solutions

Buffers and solutions used in this study are described within each method. If not otherwise stated, buffers were prepared in ddH_2O and aqueous solutions were sterilized by autoclaving at 121 °C for 20 min.

4.2 Methods

4.2.1 Culture-independent characterization of Arabidopsis root-associated bacteria

4.2.1.1 Agarose gel electrophoresis of DNA

DNA fragments were separated by agarose gel electrophoresis in gels consisting of 1 - 3 % (w/v) agarose (Bio-Budget Technologies, Krefeld, Germany) supplied with ethidium bromide solution (1:40000) in TAE buffer (400 mM Tris; 10 mM EDTA; 200 mM acetic acid; pH 8.5).

4.2.1.2 Seed surface sterilization

Seeds were surface-sterilized by a washing series in ethanol (2 x 70% and 96%). The ethanol was removed and seeds were dried in a sterile hood.

4.2.1.3 Plant growth conditions

The seeds were sown on natural soil at a density of nine plants per pot. Weeds germinating from seeds naturally present in the experimental soils were manually removed upon identification. Pots were randomly distributed in trays and placed in a greenhouse with the following conditions: 16 hours light (day) and 8 hours dark (night), 22°C during the day and 18°C during the night at a relative humidity of 70%. Plants were watered periodically every other day to maintain for each

soil type a maximum water retention capacity between 50-70%. Plants were maintained in the greenhouse until early flowering stage.

4.2.1.4 Soil incubation of wooden sticks

Natural, untreated birch toothpicks (*Betula*, Fackelmann, Hersbruck, Germany) and untreated beech wood sticks (*Fagus*, Hermann Metz KG, Quickborn, Germany) were utilized as wood compartment. Wood specimens were sterilized using the sterilization protocol described above and inserted at a minimal depth of 3 cm into the natural soils. Soil-filled pots with wood samples together with pots containing plant samples as well as unplanted pots were all subjected to the same watering regime and climatic conditions in the greenhouse.

Arabidopsis plants were grown in liquid culture as previously reported (Reintanz *et al.*, 2001) and used to generate samples of axenically grown plants.

4.2.1.5 Sampling of 'soil', 'rhizosphere', 'root' and 'wood' compartments (Figure 17)

The soil compartment was sampled from the center of an unplanted pot after removing the top 0.5-1 cm of soil, frozen in liquid nitrogen and stored at -80°C until DNA extraction.

Arabidopsis plants were manually harvested from the pots and large soil aggregates were removed by shaking the roots. Nine plants from one pot, represent one 'sample' used for pyrosequencing. To standardize the sampling and because in natural soils (not amended by sand) fragile root tips can be lost during preparation, 3 cm root segments were dissected just below the rosette with a sterile scalpel to maximize reproducibility. Root sections of all 9 plants derived from a single pot were pooled into a 15 ml Falcon tube containing 2.5 ml sterile Silwet L-77 amended PBS buffer (PBS-S; 130 mM NaCl, 7 mM Na_2HPO_4, 3 mM NaH_2PO_4, pH 7.0, 0.02% Silwet L-77) and washed on a shaking platform for 20 min at 180 rpm. Roots were transferred to a new 15 ml Falcon tube containing 2.5 ml PBS-S. The washing buffer was subjected to centrifugation (1,500 x *g*, 20 min) and the resulting pellet was defined as rhizosphere compartment, frozen in liquid nitrogen and stored at -80°C. After a second washing step (20 min, 180 rpm), roots were transferred to another 15 ml Falcon tube (2.5 ml PBS-S) and were sonicated for 10 cycles consisting of 30 second pulses at 160 W and breaks of 30 seconds (Bioruptor Next Gen UCD-300, diagenode, Liège, Belgium). After removing from the PBS-S buffer the washed and sonicated roots were frozen and defined as the root compartment.

Samples of the wood compartment were prepared similar as the root samples by collecting a 3 cm toothpick segment that was inserted in and exposed to soil and collected at the same time as the Arabidopsis roots. Because of the larger diameter of the beech sticks, 4 sticks were used for a

Fagus sample compared to 5 sticks for a *Betula* sample. The wood samples were washed twice and sonicated exactly as the root samples. Prior to homogenization all frozen wood specimen were smashed into small fragments using pre-cooled mortar and pestle.

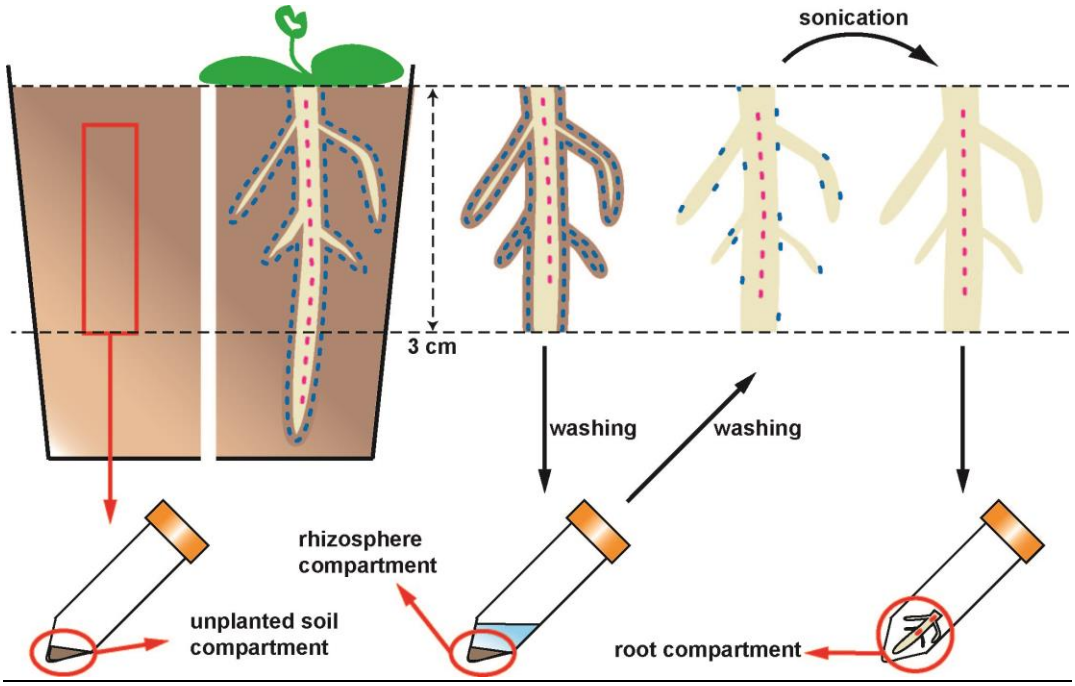


Figure 17: Fractionation protocol for sampling of root, rhizosphere and soil compartments.

4.2.1.6 DNA extraction of 'soil', 'rhizosphere', 'root' and 'wood' compartments

Samples were homogenized utilizing a Precellys®24 tissue lyzer (Bertin Technologies, Montigny-le-Bretonneux, France) at 6,200 rotations per second for 30 seconds per cycle. Frozen samples of the root compartment were pre-homogenized prior to DNA extraction for two cycles with freezing in liquid nitrogen between repetitions, whereas the wood samples were pre-pulverized with mortar and pestle in liquid nitrogen. DNA was extracted from all samples using the FastDNA ® SPIN for soil kit (MP Biomedicals, Solon, USA) according to the manufacturers' instructions except that the homogenization step (step #4; PBS- and MT buffers) was performed using the tissue lyzer with the conditions mentioned above.

4.2.1.7 Generation of 16S rRNA gene amplicon libraries

Bacterial DNA pyrosequencing was based on ~400 bp amplicons generated by the PCR primers 799F2 (see below for details) and 1193R (see Table 5)(Chelius & Triplett, 2001) spanning the hypervariable regions V5-V6-V7 of the 16S rRNA gene. A dedicated touchdown PCR cycling

protocol (Table 7) was employed. For multiplexed pyrosequencing this primer was extended at the 5´ end with a sample-specific, error-tolerant 6-mer barcode (N´s) followed by a sequence containing a *Sfi*I site required for the ligation of 454 adapters (see below; GATGGCCATTACGGCCNNNNNN-799F2). The sequence of the 1193R primer was extended at the 5´ by a sequence containing a *Sfi*I site for ligation of 454 adapters (GGTGGCCGAGGCGGCC). PCRs were performed for each sample in a volume of 25 µl with 30 ng of template DNA. PCR components include 1.25 U DFS-Taq DNA polymerase (Bioron, Ludwigshafen, Germany), 1x incomplete reaction buffer, 0.3% BSA, 2 mmol of $MgCl_2$, 200 µmol of dNTPs and 400 nmol of each primer. PCR reactions were mixed in a laminar flow hood and amplified following a touch-down program (Table 7). Amplicons from multiple replicated reactions of the same sample were pooled and purified from 1.5% agarose gels using the Qiaquick gel extraction kit (Qiagen, Hilden, Germany). Purified DNA was quantified with NanoDrop (ThermoScientific, Wilmington, USA) and equimolar amounts of barcoded amplicons of up to 27 samples were pooled for one library.

Table 7: Touchdown PCR program used for amplification of the bacterial 16S rRNA gene

step	temperature [°C]	time [seconds]	cycles
1	94	600	
2	94	60	
3	58	60	
4	72	15	5
5	94	60	
6	57	60	
7	72	30	5
8	94	60	
9	56	60	
10	72	45	5
11	94	60	
12	55	60	
13	72	60	25
14	72	600	
15	15	pause	

4.2.1.8 Pyrosequencing of 16S rRNA gene amplicon libraries

Pyrosequencing was conducted by the Max Planck Genome Centre Cologne. Amplicons containing asymmetrical *Sfi*I restriction enzyme sites at the 5' and 3' ends were prepared for the ligation of 454-specific adaptors by digestion for 1 h at 50°C with *Sfi* I (NEB, Frankfurt, Germany), followed by purification employing the MinElute PCR purification kit (Qiagen, Hilden, Germany). The 454

adaptors A and B with compatible *Sfi*I sites at 3' ends were ligated to the amplicons overnight at 16°C with 1 U T4 DNA ligase (Roche, Mannheim, Germany) followed by heat inactivation (10 min at 65°C). 454 compatible amplicons were purified from unligated adapters after size fractionation on 2% agarose gels with the Qiaquick gel extraction kit (Qiagen, Hilden, Germany). Amplicon libraries were immobilized on beads and clonally amplified using the GS FLX Titanium LV emPCR Kit (Lib-L). The libraries were then sequenced using the GS FLX Titanium Sequencing Kit XLR70 and GS FLX Titanium PicoTiterPlate Kit. All kits used were purchased from Roche and used according to the manufacturers' protocol.

4.2.1.9 Bioinformational analysis

4.2.1.9.1 Data analysis pipeline I based on taxonomy

To perform a hierarchical taxonomic rank assignment for each sequence, a pipeline was implemented to pre-process raw sequences analogous to the quality and length trimming of PyroTagger (Kunin & Hugenholtz, 2010). The sequences were trimmed to a length of 315 bp. Only sequences that matched perfectly to barcode and primer sequences were included in the analysis. Additionally, all sequences containing ambiguous bases were removed to obtain reliable taxonomic rank assignments. A quality filter allowed 10% of nucleotides with a Phred score ≤ 27 per sequence. Sequences that passed the pre-processing (referred to as high-quality sequences) were taxonomically assigned using (1) SILVA (Pruesse *et al.*, 2007) and (2) RDP (Wang *et al.*, 2007). Taxonomical assignment using the SILVA database was conducted by Jörg Peplies (Ribocon GmbH, Bremen, and Germany) and using the RDP database by Nahal Ahmadinejad (MPIPZ, Cologne, Germany).

(1) Preprocessed sequence reads from 454 sequencing were further processed (quality control and alignment) by the bioinformatics pipeline of the SILVA project. Briefly, reads with more than 2% of ambiguities or 2% of homopolymers, respectively, were removed. Remaining reads were aligned against the SSU rRNA seed of the SILVA database release 108 (http://www.arb-silva.de/documentation/background/release-108) whereupon non-aligned reads have not been considered further for downstream analysis. Using this strategy, putative non SSU rRNA gene reads and other artifacts within the data set could be extracted. Subsequently, remaining reads were (1) dereplicated, (2) clustered, and (3) classified. Dereplication (identification of identical reads ignoring overhangs) was done with cd-hit-est of the cd-hit package 3.1.2 (http://www.bioinformatics.org/cd-hit) using an identity criterion of 1.00 and a word size of 8. Remaining sequences were clustered again with cd-hit-est using an identity criterion of 0.98 (same word size). A random read of each cluster was used as a reference for taxonomic classification done by a local BLAST search against the SILVA SSURef 106 NR dataset

(http://www.arb-silva.de/projects/ssu-ref-nr/) using blast-2.2.22+ (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with standard settings. The full SILVA taxonomic path of the best blast hit has been assigned to the reads in case the value for (% sequence identity + % alignment coverage)/2 were at least 93.0. In the final step, the taxonomic path of each cluster reference read was mapped to the additional reads within the corresponding cluster plus the corresponding replicates, identified in the previous analysis steps (1) and (2), to finally obtain (semi-)quantitative information (number of individual reads representing a taxonomic path) on the composition of the original PCR amplicon pool.

(2) High-quality sequences were taxonomically assigned using the command-line version of the classifier algorithm (version 2.2, provided by the Ribosomal Data Project (RDP, Release 10 Update 27/August 2011)). The taxonomy assignment comprised the six ranks 'domain', 'phylum', 'class', 'order', 'family' and 'genus'. The classifier algorithm reports a confidence value to the assignment at each rank. Sequences were considered as classified above a confidence value threshold > 0.5.

The classified sequences per taxa and per sample generated with both databases were counted at each rank and normalized with the number of classified sequences at a given rank. In the compartments soil, rhizosphere, and root, each separated by soil type, the mean and standard error of the normalized counts were computed and the per mille values of the abundant phyla and families were plotted for SILVA. Taxa counts at the family level were subjected to the same linear model analysis described below (see section 4.2.1.9.6) to identify bacterial taxa differentially accumulating between the tested compartments.

4.2.1.9.2 Data analysis pipeline II based on OTUs (PyroTagger)

16S rRNA gene amplicon sequences were processed using PyroTagger (Kunin & Hugenholtz, 2010). The default quality parameters for Titanium pyrosequencing were used and all DNA sequences were trimmed at 315 bp according to the read length quality distribution plot generated by the software. Consequently, the hypervariable regions V5 and V6 of the 16S rRNA gene were analyzed. The software performs a linkage cluster at 97% sequence identity and these clusters defined bacterial operational taxonomic units (OTUs). Importantly, all samples were co-clustered in the same PyroTagger run. The counts of individual OTUs from the PyroTagger output file "cluster_classification" trimmed for potential chimeras was used for subsequent analysis. To normalize the data for each sample OTU counts were divided by the total counts of all OTUs within that sample and multiplied by 1000 resulting in relative abundance (RA) expressed as per mille. A $\log_2$ transformation ($\log_2$(RA + 1)) of the data was performed for statistical comparisons. Based on technical reproducibility of OTU detection and abundance (compare Figure 5), an

inclusion threshold of >5 ‰ RA was defined such that OTUs have to be above this threshold in at least one sample. Normalization, transformation and thresholding were performed using a custom R script. The most abundant sequence of each OTU was taxonomically assigned using the SILVA database as an OTU representative sequence.

4.2.1.9.3 Rarefaction curves

Rarefaction curves for each sample were estimated individually as well as compartment-specific rarefaction curves using means at each sampling size for all rarefaction curves of samples belonging to that compartment. The function "calculateRarefaction" of the R package "ShotgunFunctionalizeR" (Kristiansson *et al.*, 2009) was used to calculate the rarefaction curves.

4.2.1.9.4 Bray-Curtis dendrogram

The raw OTU counts were rarefied to 1,000 counts per sample employing the function "rrarefy" of the R package "Vegan". Log_2 transformed RA values were used to calculate a Bray-Curtis distance dissimilarity matrix using the function "vegdist" of the R package "Vegan". The dissimilarity matrix was used to generate corresponding cluster dendrograms using the function "hclust" of the R package "Vegan" specifying the 'average' clustering mode.

4.2.1.9.5 Significance of Bray-Curtis dissimilarity matrix

Permutational multivariate analysis of variance (PERMANOVA) was used to identify whether significant differences exist between bacterial communities of two or more compartments. Pair wise comparison among the soil, rhizosphere and root compartments in each of the soils tested was performed using the same Bray-Curtis distance dissimilarity matrix (rarefied) generated for the cluster dendrogram and the function "adonis" of the R package "Vegan". The statistical significance of the F test was assessed with 10,000 permutations.

4.2.1.9.6 Statistical analysis

The statistical analysis of the generated pyrosequencing data was, if not stated otherwise, mainly conducted by Emiel Ver Loren van Themaat or based on methods developed or suggested by him.
To identify differentially abundant OTUs in bacterial communities between compartments linear statistics on RA values (log_2, > 5 ‰ threshold, non-rarefied) was employed using the R package "limma". The dataset was divided by Cologne and Golm soil samples and separately analyzed and filtered. For each OTU a linear model was fitted with the compartment as explanatory variable.

Differentially abundant OTUs between two compartments were calculated using moderated *t*-tests (Smyth, 2004). The moderated *t*-test uses a Bayesian model to shrink or expand the standard error of each OTU towards a common value, thereby borrowing information of the variance of other OTUs. This approach is especially suited when the number of measurements per sample is large but sample sizes are small. The resulting *P*-values were adjusted for multiple hypotheses testing using the Benjamini-Hochberg correction method. To detect OTUs more abundant in one compartment compared to two or more compartments the intersection of corresponding pair-wise comparisons was taken.

To find OTUs differentially enriched in roots between the tested soil types interaction terms were used. The interaction between soil type and compartment was modeled using the whole data set (Cologne plus Golm samples). A linear model was fit for each of the eight combinations between soil type (Cologne, Golm) and compartment (root, rhizosphere, soil, wood). The significance of the interaction term was then calculated for pair-wise comparisons between two compartments and two soil types. For example, soil type-dependent differential OTU enrichments (Figure 10) were found by comparing the fold changes between the two soil types for root *versus* rhizosphere as well as root *versus* soil.

For comparative purposes, a non-parametric *t*-test as implemented in Metastats (White *et al.*, 2009) was applied to our data to detect differentially abundant OTUs between two sample groups. This test estimates the null distribution of *t* by permuting sample labels and thereby removing the normality assumption. The input data are relative abundances and 5,000 permutations were calculated.

4.2.1.9.7 Ternary plots

The concept of applying ternary plots for illustration of our data was developed by Dr. Klaus Schläppi, MPIPZ. Ternary plots were prepared using the mean of the $\log_2$ transformed RA values per compartment. The function “ternaryplot” from the R package “vcd” was modified to plot the point size of each OTU proportional to its weighted sum instead of the total RA-sum. The weights were determined for each OTU by the relative contribution of each compartment to the geometric mean of the RA, thereby adjusting for the location of the point in the ternary plot.

The R scripts used for computational analyses are available via http://www.mpipz.mpg.de/162701/R_scripts

4.3 Investigation of rhizoplane- and soil-inhabiting bacteria via imaging techniques

4.3.1 Scanning electron microscopy (SEM)

SEM micrographs were generated at the Central Microscopy Unit of the MPIPZ. Specimens for SEM were prepared from plants grown and harvested as described above, except that single roots were subjected to the washing procedure. Non-sonicated and sonicated roots were fixed in 4 % glutaraldehyde in PBS buffer (140 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.0) overnight at 4°C. The samples were briefly washed in PBS and gradually dehydrated using an ethanol series (20, 30, 50, 70, 96 % ethanol). Subsequently, specimens were kept overnight in water-free ethanol at 4°C. Critical point drying was performed using a Bal-Tec CPD 030 critical point dryer (Leica, Wetzlar, Germany). The pressure was decreased slowly to prevent tissue damage. Root samples were mounted on stubs and sputter coated with palladium before imaging with a Zeiss Supra 40VP SEM (Carl Zeiss NTS, Oberkochen, Germany). Wooden samples were directly mounted on SEM stubs and sputter-coated with palladium followed by SEM imaging in the probe chamber.

4.3.2 Catalyzed reporter deposition Fluorescence *in situ* hybridization (CARD-FISH)

4.3.2.1 Sample collection and root fixation

Plants were grown and harvested as described above. Roots from two plants were washed in 1.5 ml PBS-S and the sonication step was omitted. Washed roots were transferred into 1.5 ml PBS buffer (140 mM NaCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) amended with 4 % formaldehyde (Roti-Histofix, Roth, Karlsruhe, Germany) and fixed for 2.5 h at 4°C. Roots were then washed twice in PBS buffer for 15 min at RT and stored in fresh PBS buffer at 4°C. Fixed roots were utilized for CARD-FISH within two days. All steps were carried out in 2 ml Eppendorf reaction tubes. Roots were carefully transferred from tube to tube to avoid tissue damage resulting in high background staining.

4.3.2.2 Hybridization and signal amplification

CARD-FISH was carried out as described before (Pernthaler *et al.*, 2002; Eickhorst & Tippkotter, 2008) with minor modifications. For permeabilization of bacterial cell walls, fixed roots were placed in sterile petri dishes, overlaid with lysozyme solution (Roth, Karlsruhe, Germany; 10 mg/ml in 0.05 M EDTA pH 8.0, 0.1 M Tris-HCl pH 8.0) and incubated for 60 min at 37°C. After rinsing with sterile water, roots were transferred to a new petri dish, overlaid with achromopeptidase solution (Sigma, Steinheim, Germany; 60 units/ml in 10 mM NaCl, 10 mM Tris-

HCl, pH 8) and incubated for 30 min at 37°C. Endogenous peroxidases were inactivated by treating roots with Methanol containing 0.15% H_2O_2 for 30 min at RT. After washing in EtOH (96% v/v) roots were dried at RT and stored in 1.5 ml Eppendorf tubes at 4°C until hybridization.

All hybridization probes (Supplementary Table 5) were conjugated to a horse-radish-peroxidase (HRP) at the 5' end (Biomers, Ulm, Germany). The detection of Betaproteobacteria and Actinobacteria were performed together with the recommended competitor probes (no HRP) to increase specificity (Table 6). Roots were equilibrated in 1.5 ml PBS buffer for 15 min at RT prior to hybridization. Hybridization was carried out in 500 µl hybridization buffer (0.9 M NaCl, 20 mM Tris–HCl pH 8.0, 10% (w/v) dextran sulphate, 2% (w/v) blocking reagent (Roche, Mannheim, Germany), 0.1% (w/v) SDS, 25-35% (v/v) formamide (Supplementary Table 5)) containing 75 ng probe. Hybridization was carried out in 0.5 ml Eppendorf tubes for 2.5 h at 35°C in an incubator rotating at 12 rpm. After this hybridization step probes were transferred into 15 ml Falcon tubes and washed in 14 ml of pre-warmed washing buffer (80 mM NaCl, 5 mM EDTA pH 8.0, 20 mM Tris–HCl pH 8.0 and 0.01% (w/v) SDS) for 15 min at 37°C and 12 rpm. A second washing step was performed at RT in 14 ml 0.05% (v/v) Triton X-100 amended PBS.

The following steps were performed in the dark for signal amplification. Reactions were conducted in 500 µl amplification buffer (1x PBS pH 7.4, 0.0015% H_2O_2, 0.1% (w/v) blocking reagent) containing 1.5 µg of tyramide conjugated to AlexaFluor 488 or AlexaFluor 546 (Invitrogen, Darmstadt, Germany) for 30 min at 37°C in 0.5 ml Eppendorf tubes. Tyramide-AlexaFluor conjugates were prepared as previously described (Pernthaler & Pernthaler, 2005). Roots were subsequently transferred to 15 ml Falcon tubes and washed in 14 ml Triton-amended PBS buffer for 15 min at RT and stored in PBS at 4°C until visualization by confocal laser scanning microscopy (CLSM).

For double labeling experiments according to (Blazejak *et al.*, 2005b), the complete CARD FISH protocol was conducted twice with the same root specimen using successively hybridization probes conjugated to different fluorescent dyes. Prior to hybridization with the second probe the root samples were incubated in 1.5 ml of 0.01 M HCl for 15 min at RT, followed by washing of the specimen for 10 min in PBS buffer to inactivate the HRP of the first hybridization probe.

4.3.2.3 Visualization

Visualization was performed using a CLSM 510 Meta microscope (Carl Zeiss, Oberkochen, Germany). AlexaFluor 488 was excited with a 488 nm laser line and detected with a BP 500-530 nm filter. For visualization of the root surface or the soil background on filters samples were excited with the 543 nm laser line and autofluorescence emission was detected using the BP 565-615 nm filter. Vertical image stacks of up to 25 µm (1 µm steps) were recorded and projected

onto a single layer. In double labeling experiments the two channels were recorded sequentially to avoid cross talk between two fluorophores.

4.3.2.4 CARD-FISH on unplanted soil and rhizosphere on filters

CARD-FISH on unplanted soil and rhizosphere was conducted analogous to root hybridizations. Soil and rhizosphere for cell counts on polycarbonate filters was sampled in the same manner as for the 16S rRNA gene pyrosequencing survey. To reach a sufficient amount of rhizosphere 3 pots were pooled into each CARD-FISH rhizosphere sample. Samples were then fixed in 4% formaldehyde for 2.5 h, spotted on 0.2 µm GTTP Isopore Membrane Filters (Millipore, Schwalbach, Germany) and pretreated for CARD-FISH as described previously (Eickhorst & Tippkotter, 2008). Filters were cut into equal-sized segments and each segment was hybridized with a different taxon-specific probe (Table 6) to examine the relative abundance of the tested bacterial taxa in soil and rhizosphere compartments. After hybridization filters were embedded in Vectashield mounting medium H1200 containing DAPI stain (Vector Laboratories, Burlingame, USA). Hybridization signals overlapping with the DAPI signals were counted in at least 10 fields of view (FOVs) of each hybridized filter segment. Relative abundances of bacterial taxa were then calculated as the relative proportion of the signals in the hybridizations using the specific probes in comparison to the hybridization signals using the EUB probe on a filter segment derived from the same sample. Cell counts were conducted by Hannes Schmidt, Institute of soil science, University of Bremen.

Chapter II

Comparative genomics of selected *Rhizobium* members of the Arabidopsis root-associated bacterial microbiota

1 Introduction

Many bacteria found in natural soil are able to exhibit beneficial effects on plant growth. These microbes, designated plant growth promoting rhizobacteria (PGPR), exert their effect in the rhizosphere, after colonizing the root surface or as endophytes from within the root tissue and influence their host through a number of different mechanisms. Effects can be either direct, e.g. by increasing availability of nutrients, or indirect, e.g. by protecting the plant of pathogenic microbes (Lugtenberg & Kamilova, 2009). PGPR bacteria have great potential as an additive or even as surrogates of chemical fertilizers for improving crop yield in sustainable agriculture and first bacterial strains are already applied in the field (Banerjee *et al.*, 2006). However, some bacteria that are potent PGPR in gnotobiotic systems fail to exert their effects in the field. This scenario underlines the need to understand the functions and actions of PGPRs in the context of the endogenous plant bacterial microbiota. To gain deeper insights into this field PGPR have to be isolated in pure culture to test their metabolic potential in laboratory and field conditions. Additionally, their full genomes have to be obtained to identify genes involved in beneficial host-microbe interactions.

1.1 Rhizosphere competence

In order to interact with their host PGPR have first to colonize the rhizosphere to eventually proceed to the rhizoplane or the inner root tissue (Compant *et al.*, 2010). As already discussed the rhizosphere is a hotspot of microbial abundance and activity due to its high nutrient content provided by plant rhizodeposition. At the same time bacterial species richness is reduced in comparison to the surrounding soil (see chapter I, section 2.3.2), showing that only a subset of the soil-borne bacteria is able to occupy this plant-derived habitat. The ability to outcompete other bacteria in colonization of the rhizosphere to access its nutrients has been termed “rhizosphere competence”. Important traits thereof are motility, chemotaxis, attachment to the root and stress resistance. “Clusters of Orthologous Groups” (COGs) related to chemotaxis are overrepresented and less diverse in rhizosphere bacterial communities in comparison to communities of unplanted soil supporting both that chemotaxis is an important trait of rhizosphere competence and that only a subset of all soil bacteria can colonize the rhizosphere (Buchan *et al.*, 2010; Barret *et al.*,

2011). Likewise, COGs of flagella biosynthesis are overrepresented in rhizosphere bacteria (Barret *et al.*, 2011). Well studied PGPR like *Pseudomonas fluorescens*, *Pseudomonas putida* and *Azospirillum brasilense* are severely affected or incapable of root colonization if their flagella are mutated (De Weger *et al.*, 1987; Broek *et al.*, 1998) and disruption of the transduction from sensing signals to movement towards the signal by mutation of the *cheA* gene in *P. fluorescens* leads to a strongly reduced competitive root colonization ability (de Weert *et al.*, 2002). The rhizosphere is an oxidative environment as roots secrete organic acids to solubilize nutrients from the soil (Dakora & Phillips, 2002) and PGPR that enter the root tissue might encounter reactive oxygen species (ROS) which are one of the defense mechanisms of the plant's innate immune system (Torres *et al.*, 2002). Therefore, counteracting oxidative stress contributes to rhizosphere competence and consequently mutations in superoxide dismutases in *P. putida* have been shown to reduce its root colonization efficiency (Kim *et al.*, 2000).

1.2 Mechanisms of plant growth promotion

1.2.1 Direct mechanisms

Making nutrients available is one mode of action of PGPR. A well-studied example is the fixation of atmospheric nitrogen by symbiotic Rhizobia, which will be discussed later (see section 1.3). Also other bacterial species are diazotroph, i.e. able to fix atmospheric nitrogen. *Azoarcus* sp. BH72, an endophyte of Kallar grass (Hurek *et al.*, 2002), *Klebsiella pneumoniae* in wheat (Iniguez *et al.*, 2004) and so far unidentified N_2-fixing diazotrophs in field-grown sugar cane (Urquiaga *et al.*, 2012) provide significant amounts of nitrogen to their host plant. Free living N_2 fixing bacteria, however, are unlikely to contribute significant quantities of nitrogen to plants considering the high metabolic costs of nitrogen fixation (Shanmugam *et al.*, 1978) and the strong competition for plant derived nutrients in the rhizosphere. Thus, nitrogen fixation in free living diazotroph can be seen as mere survival strategy.

Another major essential macronutrient next to nitrogen is phosphate. It exists as organic and inorganic forms in soil, but only 5% of the total phosphate is available to plants. Converting insoluble to soluble, bioavailable phosphates is an important mechanism of PGPR to improve plant yields (Rodriguez *et al.*, 2006). Plants take up phosphate primarily as inorganic HPO_4^{2-} and $H_2PO_4^-$. Some PGPR can convert insoluble forms of phosphate into these plant accessible forms by modulating the pH of the soil by producing organic acids, like gluconic acid by Pseudomonas and a number of Burkholderia strains (Rodriguez & Fraga, 1999; Delvasto *et al.*, 2008). Also Rhizobia can solubilize phosphate by organic acids like *Rhizobium leguminosarum* and *Sinorhizobium meliloti* (Halder *et al.*, 1990; Halder & Chakrabartty, 1993). Another way to make soluble phosphate

available from insoluble organic phosphates is by enzymatic activity of phosphatases and phytases (Rodriguez *et al.*, 2006).

Other well-studied mechanisms of direct plant growth promotion by bacteria which are independent of provision of nutrients involve the modulation or interference with the plant's phytohormone levels. The gaseous phytohormone ethylene is involved in the plant developmental cycle, in the response to abiotic stress and in plant-microbe interactions (Wang *et al.*, 2002). Ethylene inhibits root elongation and transport of the root growth inducing phytohormone auxin as part of stress responses of the plant. Bacteria can interfere with the plant's ethylene levels by breaking down the ethylene precursor 1-aminocyclopropane-1-carboxylate (ACC) or by inhibiting enzymes of the ethylene biosynthesis pathway, namely the ACC synthase or β–cystathionase. The presence of ACC deaminases has been shown amongst others in the PGPR *Pseudomonas*, *Variovorax*, *Burkholderia*, and *Rhizobium* (Glick *et al.*, 2007b). By breaking down ACC bacteria form an ACC sink and reduce ethylene levels in the root releasing the inhibitory effects of ethylene that was synthesized as a response to a given biotic or abiotic stress ameliorating the initial stress response. This in turn leads to increased root and shoot length and increased biomass. Until now ACC deaminases could only be detected within the bacterial cytoplasm, therefore it is likely that the distance of the bacteria to the root influences the growth promotion effect, with true endophytes having the strongest growth promotion capacity (Glick *et al.*, 2007b).

A second well-studied example of a phytohormone that is influenced by PGPR is auxin. Auxin is a class of phytohormones that is the main regulator of plant growth, for shoot elongation as well as for root proliferation (Teale *et al.*, 2006). Auxin production has been shown e.g. for *Azospirillum brasilense*, *Azospirillum lipoferum*, *Enterobacter cloacae*, *Pseudomonas putida* and *Pantoea agglomerans* (Spaepen & Vanderleyden, 2011). Bacteria can produce auxin mainly via three different pathways, starting from tryptophan and resulting in indole-3-acetic acid (IAA), the major naturally occurring auxin *in planta*. IAA is then released from the bacteria and can e.g. alter root morphology, as shown for *A. brasilense*. This PGPR colonizes the root surface and induces root growth and an increase in total root surface which is hypothesized to lead to a better nutrient and water acquisition by the plant and in the end to an increased yield (Spaepen & Vanderleyden, 2011). IAA production by PGPR is influenced by environmental factors that might be plant derived. In *A. brasilense* the expression of the gene encoding for an indole-3-pyruvate decarboxylase (IPCD), a key enzyme that catalyzes the last step of IAA biosynthesis, is increased in acidic pH (Vande Broek *et al.*, 2005). Rhizosphere soil is typically acidic because of organic acids secreted by roots and proton extrusion through root cell membranes (Dakora & Phillips, 2002). Expression of *ipcd* in *A. brasilense* as well as in other bacteria like *P. putida* is also increased by exogenously

applied IAA precursor tryptophan (Patten & Glick, 2002). Tryptophan in turn, dependent on the plant species, can be secreted by the root into the rhizosphere (Kravchenko *et al.*, 2004).

1.2.2 Indirect mechanisms of plant growth promotion

Several described PGPR do not directly provide nutrients to a plant or form close plant-microbe interaction, but promote plant growth by actively antagonizing plant pathogens or by competing with them for nutrients and niches.

Antagonism means the production of antimicrobial compounds by PGPR that suppress the pathogen and therefore protect the plant from disease. Antimicrobial compounds that have been identified to be responsible for plant growth promotion include hydrogen cyanide (HCN), produced by the eight out of nine rhizosphere *Pseudomonas spp*. isolated from different crop plants (Ahmad *et al.*, 2008). Disruption of HCN biosynthesis leads to a reduced plant protection (Haas & Keel, 2003) and transfer of the HCN biosynthesis structural genes into the non-cyanogenic Pseudomonas strain P3 also transfers biocontrol properties against the fungus *Thielaviopsis basicola* in a gnotobiotic system (Voisard *et al.*, 1989). Another antimicrobial compound produced PGPR Pseudomonas is 2,4-diacetylphloroglucinol (DAPG) which suppresses take-all disease in wheat, caused by *Gaeumannomyces graminis var. tritici* and can be detected in the rhizosphere (Bonsall *et al.*, 1997; Kwak *et al.*, 2012). Whether antimicrobial compounds indeed kill plant pathogens *in situ* is at least under debate, since concentrations of the compounds outside of the producing microcolonies are very low and probably not sufficient to terminate the pathogen (Haas & Keel, 2003). Nevertheless, sublethal concentrations might still have physiological effects on the target organism, for instance blocking cell-cell signaling of the human pathogen *P. aeruginosa* in chronic lung infections when treated with antibiotics (Tateda *et al.*, 2001). Part of bacterial cell-cell signaling is the so-called quorum sensing (QS). Bacteria secrete and sense signal molecules and if a certain threshold (a quorum) is reached gene expression is altered. In this way biological processes that only promise success if a certain cell density reached are synchronized, like infection of the host tissue. As an example, in the genus of plant pathogens Xanthomonas virulence factors like degradative enzymes and adhesins are only expressed if a sufficient cell density is achieved (Buttner & Bonas, 2010). One of the main secreted signaling molecules in QS is *N*-acylhomoserine lactone (AHL). A vast array of bacteria is able to interfere with QS signals, also known as quorum quenching, e.g. by degrading AHL with lactonase enzymes (Uroz *et al.*, 2009) and might disturb QS controlled infection processes (Lugtenberg & Kamilova, 2009). However, a direct link of quorum quenching to plant growth promotion has not been shown.

PGPR bacteria can also suppress diseases by competing for nutrients and niches with pathogens. In soil, iron availability is low as it is predominately found as insoluble ferric ions leading to a fierce competition (Loper & Henkels, 1997). Bacteria secrete iron chelating compounds, siderophores, to make iron bioavailable. The siderophore-Fe^{3+} complex is subsequently bound to receptors at the bacterial cell wall, the iron is taken up and released inside the bacterial cell as bioavailable Fe^{2+} (Hider & Kong, 2010). Bacterial siderophores can inhibit the growth of fungal pathogens in the rhizosphere in acidic soils, were the concentration of bio-accessible iron is low, as for example shown for *P. putida* WCS358. In co-inoculation with *Fusarium oxysporum* Fo47 it is able to suppress the density of a pathogenic *F. oxysporum* at the root, while a *P. putida* WCS358 mutant impaired in production of the siderophore pseudobactin does not exhibit a significant effect on the pathogen (Duijff *et al.*, 1999).

1.2.3 Induced systemic resistance

Another mechanism how PGPR counteract pathogens is priming the plant's innate immune system. This induced systemic resistance (ISR) is a state in which the above ground part of the plant is able to trigger an accelerated immune reaction in response to pathogen attack, if the same plant has been colonized before by an ISR inducing PGPR. The priming PGPR itself however does not cause a defense response. A well-studied example for ISR inducing bacteria is *Pseudomonas fluorescens* WCS417. Although MAMPs are present in WCS417 no defense responses are activated upon root colonization with WCS417 alone and inoculation with WCS417 leads to suppression of flg22-mediated defense responses. WCS417 inhibits local immune responses in Arabidopsis by secreting low molecular weight molecules into the plant's apoplast (Millet *et al.*, 2010). At the same time WCS417 initiates a systemical signaling cascade that confers resistance against a broad spectrum of pathogens like *Pseudomonas syringae* pv. *tomato*, *Botrytis cinerea* and *Ralstonia solanacearum* (Ran *et al.*, 2005; Van der Ent *et al.*, 2008). Amongst others ISR inducing capabilities have been also shown for Bacilli, Rhizobia and Serratia (Reitz *et al.*, 2000; Ryu *et al.*, 2004; Domenech *et al.*, 2007). In Arabidopsis *P. fluorescens* WCS417 induction of ISR is mediated by the root specific transcription factor MYB72 and dependent of the phytohormones jasmonic acid and ethylene (Van der Ent *et al.*, 2008), but the systemic signal that leads to a primed innate immune system in the aerial parts of the plant has not been identified.

1.3 Rhizobia

Rhizobia are often defined as a diverse group of soil bacteria capable of fixing atmospheric nitrogen (N_2), in symbiosis with legumes. By this definition, which is solemnly based on the

metabolic capacity of bacteria rather than phylogenetic affiliation to specific taxa, the group of rhizobia contains (as of April 2012) 98 species in 13 genera. Most of the rhizobia described so far fall into the genera *Rhizobium*, *Mesorhizobium*, *Ensifer* (synonymous *Sinorhizobium*) and *Bradyrhizobium* in the class of α-Proteobacteria. Additional bacterial species capable of fixing N_2 in symbiosis have been found in the class of Burkholderiales belonging to the β-Proteobacteria. The listed genera contain many more bacterial species that are not able to fix nitrogen, but still could be called rhizobia based on a phylogenetic definition. In this thesis a Rhizobium is defined as a bacterium classified as member of the order of Rhizobiales, irrespective of its nitrogen fixing capabilities.

1.3.1 Rhizobia as mutualistic symbionts of legumes

In mutualistic rhizobium-legume symbiosis roots develop so called nodules, specialized organs encapsulating the bacterial symbiont in intracellular compartments. Within these nodules rhizobia fix atmospheric nitrogen (N_2) and provide it as plant-available ammonium (NH_4^+) to their host. In turn bacteria are supplied with carbon in form of dicarboxylates like malate, succinate and fumarate, amino acids and micronutrients like Fe^{2+}, Mg^{2+} and SO_4^{2-}.

Rhizobia are soil-borne bacteria and are attracted to roots by root-secreted flavonoids (Hassan & Mathesius, 2012). Rhizobia infect legumes by attaching to root hairs by biofilm formation, via cracks in the epidermis or via interstitial infections between epidermal cells (Gonzalez *et al.*, 1996; Gage, 2004). Upon perception of flavonoids rhizobia start to secrete nodulation (Nod) factors, signaling compounds that are received by the plant, most likely in its epidermis (Goedhart *et al.*, 2000). Nod factors have a chitin backbone attached to a terminal sugar and can carry modifications that differ between rhizobial species determining the host range of a given *Rhizobium* sp. (Denarie *et al.*, 1996) and are encoded by *nod* genes typically located on a symbiotic plasmid. Perception of Nod factors by the host legume in turn leads to the start of morphogenesis of the nodule compartment within the cortical cell layer of the root and at the same time to structural changes in the root hair via two independent signaling pathways (Oldroyd & Downie, 2008). The root hair entraps the Rhizobium by curling and initiates the formation of an infection thread (Esseling *et al.*, 2003; Gage, 2004). This infection thread is an invagination of the root hair tip that leads to an intracellular tunnel through the root hair and grows towards the cortical cell layer of the root (Gage, 2004). During progression of the infection thread the bacteria multiply within the growing infection thread. Rhizobia populations in a nodule are mostly clonal, since either already at the entrapment or during growth of the infection thread a single bacterial cell is selected to proceed with the infection (Gage, 2002; Gage, 2004). In the meantime after

perception of Nod factors close to the site of bacterial infection the nodule primodium is initiated in the cortical cell layer of the root (Oldroyd & Downie, 2008). Two types of nodules can be distinguished, determinate nodules in e.g. Phaseolus, Glycine, and Lotus that lose meristematic activity shortly after initiation and indeterminate nodules in e.g. Medicago, Pisum, and Trifolium that maintain an active apical meristem that continuously produces new cells for growth. The host responding cells start mitosis and to enlarge. This enlargement of the cells provokes the characteristic morphological changes of the root anatomy, designated nodules. Bacteria from the infection thread are released into these cells, are encapsulated with a plant derived membrane and start to differentiate into so-called bacteroids (Oldroyd *et al.*, 2011). Once rhizobia have differentiated into bacteroids they shift their metabolism towards nitrogen fixation and even housekeeping genes, like biosynthesis of amino acids, are switched of (Karunakaran *et al.*, 2009). From then on bacteroids are dependent on nutrient supply from the roots. Bacteroids in indeterminate nodules of *Medicago truncatula* have an altered cell membrane that is more permeable than their free-living counterparts (Mergaert *et al.*, 2006), cannot de-differentiate into the free-living state again and cannot reproduce since it is not possible to re-isolate them from nodules, while bacteroids from determinate nodules of *Lotus japonicus* or *Phaseolus vulgaris* are impermeable to propidium iodide, and can be regrown from nodules.

Fixing nitrogen occurs via the enzyme complex nitrogenase which is made of two separate protein components, a dinitrogenase reductase and a dinitrogenase. The dinitrogenase reductase hydrolyses MgATP to MgADP and transfers the resulting electrons via an Fe-S cofactor to the dinitrogenase which then catalyzes the reduction of N_2 (Seefeldt *et al.*, 2009). The genes encoding for the components of the nitrogenase are as the Nod factors encoding genes located on the symbiotic plasmid. The *nifH* gene encodes the nitrogenase reductase and is a prerequisite for nitrogen fixation and therefore its presence/absence is typically used to predict a strain's putative capability of fixing nitrogen. Fixing nitrogen via nitrogenase is a dioxygen sensitive reaction, since the dioxygen oxidizes the Fe-S cofactor of the enzyme irreversibly inhibiting the nitrogenase. Nodules are therefore filled with leghemoglobin, a high affinity oxygen carrier that keeps concentrations of oxygen in the nodule low but provides enough oxygen for respiration (Appleby, 1984). Due to the color of the leghemoglobin functional nitrogen fixing nodules appear pinkish.

1.3.2 Rhizobia in interactions with non-leguminous plants

Additionally to the well-studied symbiosis with legumes rhizobia are also able to colonize non leguminous plants as monocots like rice (Hardoim *et al.*, 2011) and maize (Rosenblueth & Martinez-Romero, 2004), but also dicots like rapeseed and Arabidopsis (O'Callaghan *et al.*, 2000;

Santaella *et al.*, 2008). Rhizobia have been shown to thrive as endophytes within rice and even spread from the root tissue to the aerial part of the plant (Chi *et al.*, 2005). Colonization does not require *nod* genes, but is stimulated by flavonoids in specific cases like *Azorhizobium caulinodans* ORS571. Addition of specific flavonoids to the growth medium during inoculation leads to an enhanced colonization of Arabidopsis and *Brassica napus*, also in mutated strains that are impaired in biosynthesis of Nod factors as well as strains carrying a mutation in the *nodD* gene, the main flavonoid sensor and Nod factor biosynthesis regulator in rhizobium-legume interactions (Gough *et al.*, 1997; O'Callaghan *et al.*, 2000). *Bradyrhizobium* strain ORS278 is a natural rice endophyte and nitrogen fixation could be detected within the rice tissue by acetylene reduction although rice is not able to form nodules. Re-colonization in greenhouse experiments leads to increased total plant mass (Chaintreuil *et al.*, 2000).

Plant growth promotion in association with non-leguminous plant can also occur independently of nitrogen fixation (Biswas *et al.*, 2000). Potential modes of plant growth promotion not related to biological nitrogen fixation resemble the mechanisms found in PGPR of other bacterial clades. Growth of non-leguminous plants is proposed to be directly influenced by the synthesis of phytohormones, interference with plant ethylene synthesis and improved uptake of nutrients like phosphate. Bacteria can lower the plants' ethylene level by breaking down the ethylene precursor ACC via the enzyme ACC deaminase. This enzyme has been detected in a range of well-studied rhizobia including *R. leguminosarum* bv. *viciae*, *R. hedysari*, *R. japonicum*, *B. elkani*, *R. gallicum*, *R. radiobacter*, *M. loti* and *S. meliloti* (Sullivan *et al.*, 2002; Ma, W *et al.*, 2003; Okazaki *et al.*, 2004; Madhaiyan *et al.*, 2006; Duan *et al.*, 2009). Direct production of phytohormones like IAA, the most important representative of the class of auxins, could be detected amongst others in *B. japonicum*, *M. loti*, *R. leguminosarum* and *S. meliloti* (Sekine *et al.*, 1988; Yanni *et al.*, 2001; Chi *et al.*, 2005; Chandra *et al.*, 2007). However, it still has to be shown whether the presence of the ACC deaminase and/or production of auxin are indeed the cause for observed plant growth promoting effects in field and greenhouse experiments. Antoun and colleagues tested a collection of 266 rhizobia and bradyrhizobia for potentially plant growth promoting traits not related to nitrogen fixation. The majority of the tested strains were next to producing the phytohormone IAA and siderophores also able to solubilize phosphate (Antoun *et al.*, 1998). Plant growth promotion by phosphate solubilization has been shown for *M. mediterraneum.* Barley grown in soil supplemented with insoluble phosphate accumulated more phosphate and showed an increased dry matter upon inoculation with *M. mediterraneum* in growth chamber experiments (Peix *et al.*, 2001).

Indirectly, rhizobia are able to decrease or inhibit deleterious effects of pathogenic microorganisms by synthesizing antibiotics and/or fungicidal compounds, by competing for

nutrients, e.g. by siderophore or by induction of ISR. The *Rhizobium etli* strain G12 is able to induce a systemic response in potato that significantly reduced infection by the nematode pathogen *Globodera pallida*. Split root system experiments could show that Lipid polysaccharides (LPS) extracted from *R. etli* are sufficient to prime the plants' immune system (Reitz *et al.*, 2000). Likewise, inoculation of rice with *Rhizobium leguminosarum* bv. *phaseoli* RRE6 leads to a protection of the plant from infection with *Rhizoctonia solani* probably by an enhanced production of phenolic compounds (Mishra *et al.*, 2006). Thus, the order of Rhizobiales contains many species that are not able to fix nitrogen due to the lack of *nod* and *nif* genes that still harbor tremendous potential for application as biological fertilizers in the field via the described mechanisms.

1.4 Pan-genomics of bacteria

The advent of next generation sequencing techniques decreased prices DNA sequencing and made large scale whole genome sequencing feasible. While up to 2005 public databases contained 239 full bacterial genomes (Medini *et al.*, 2005), nowadays within the framework of the human microbiome project alone sequencing of 3000 bacterial species is planned[1] (Peterson *et al.*, 2009). Therefore, not only type strains for specific model systems are sequenced, but sequencing projects can cover the full bacterial diversity of a given habitat or multiple exemplars or strains of the same species.

In 2005 Tettelin and colleagues introduced the concept of bacterial pan-genomes. The pan-genome is the combined global gene repertoire within a population at a specified taxonomic level, usually within a species, though this can be extended to higher levels, such as genus. The authors compared the gap-closed of three and the draft genomes of five strains of the human pathogen *Streptococcus agalactiae* to determine to which extend different strains within one species share their genomes and on the other hand how many genes are unique for the individual strains. They estimated that the core-genome, i.e. the genes present in every sequenced strain, comprised 1,806 genes plus a dispensable genome of on average 439 genes per strain, that are absent in one or more of the other strains. Sequencing of each additional strain adds genes to the overall *S. agalactiae* gene pool that had not been observed before. Remarkably, the extrapolated curve based on the eight sequenced strains reaches a nonzero asymptotic value of 33 new genes per added genome. The pan-genome of *S. agalactiae* is therefore "open", because a very large number of genomes would be necessary to determine the entire gene repertoire (Tettelin *et al.*, 2005). The information of the genome comparison led to the development of a new broad

[1] http://www.hmpdacc.org/reference_genomes/reference_genomes.php

spectrum vaccine. A mixture of surface-exposed proteins encoded not within the core- but within the variable, strain specific gene set was highly effective against multiple pathogenic strains of *S. agalactiae* in a mouse model (Maione *et al.*, 2005). Since then pangenomes of multiple human-associated bacteria like *Escherichia coli* (Rasko *et al.*, 2008) as well as plant-associated bacteria like *Pseudomonas fluorescens* have been analyzed (Loper *et al.*, 2012), including the pangenome analysis of 26 Sinorhizobium and Bradyrhizobium strains (Tian *et al.*, 2012), all revealing open types of pangenomes reflecting the adaptation to specific habitats or niches of the individual strains. The analyzed Sinorhizobium and Bradyrhizobium strains share the same host, soybean. However, no genes could be identified that are common within this group but distinct from Rhizobia colonizing different hosts. On the other hand lineage specific genes within the Sinorhizobium and Bradyrhizobium clade could be shown to be extensively recruited, pointing towards the evolution of independent colonization strategies (Tian *et al.*, 2012).

1.5 Aim of the study

Culture-independent characterization of the Arabidopsis root-associated bacterial communities by 16S ribotyping led to the identification of bacterial subgroups that either depend on assembly cues of living Arabidopsis roots or depend on cell wall components for colonization. Members of the bacterial rOTUs sub-community depend on assembly cues of living Arabidopsis roots while members of the bacterial lOTUs sub-community colonized living roots as well as wooden sticks and therefore seem to represent saprophytes that are likely to colonize a wide range of plant derived material in soil.

To get closer insights into mechanisms underlying these differential colonization behavior community members of both subgroups were to be cultivated in pure culture. Isolated bacteria can then be investigated in re-colonization experiments. These experiments can show if bacteria that have been identified before as members of the different bacterial sub-communities are able to re-colonize Arabidopsis roots in a gnotobiotic system and if so if the colonization has any effect on plant growth and morphology. Furthermore bacteria-induced effects on Arabidopsis can be compared between rOTUs and lOTUs sub-community members.

The identification of bacteria and their allocation to the different sub-communities was based on the 16S rRNA sequence which does not provide any information about metabolic capabilities of the bacteria. Sequencing full genomes of isolates can reveal their genetic potential and might allow first insight in how these bacteria interact with their plant host and how they have adapted to different habitats. Full genome information also provides information about the genetic variability inherent to bacteria that all have been according to their 16S rRNA gene sequence assigned to the same OTU.

2 Results

2.1 Isolation of Arabidopsis root-inhabiting bacteria

To isolate Arabidopsis root-associated bacteria roots of Arabidopsis grown in natural Cologne and Golm soil were harvested, washed and sonicated in consistency to the sampling procedure used for culture-independent 16S ribotyping. The root material enriched for endophytes was macerated in buffer which was then plated in appropriate dilutions on different growth media. Colony forming units (CFUs) were typed by sequencing their 16S rRNA genes. The obtained sequences were classified via the RDP database to derive their taxonomy information.

In total, 1682 CFUs were closer examined. 77.72% of all CFUs for which 16S sequences could be obtained were classified as Proteobacteria, 10.43% as Actinobacteria, 9.46 % as Firmicutes and 1.64 % as Bacteroidetes (Figure 18a). The order of Pseudomonadales and Rhizobiales dominated within the phylum Proteobacteria, each making up app. 26% of all isolated CFUs. Isolated CFUs classified as Actinobacteria were almost exclusively Actinomycetales (10.36%), and CFUs of the phylum of Firmicutes belonged almost entirely to the order of Bacillales (9.24%). With 1.49% the order of Flavobacteriales dominated within the phylum of Bacteroidetes (Figure 19b).

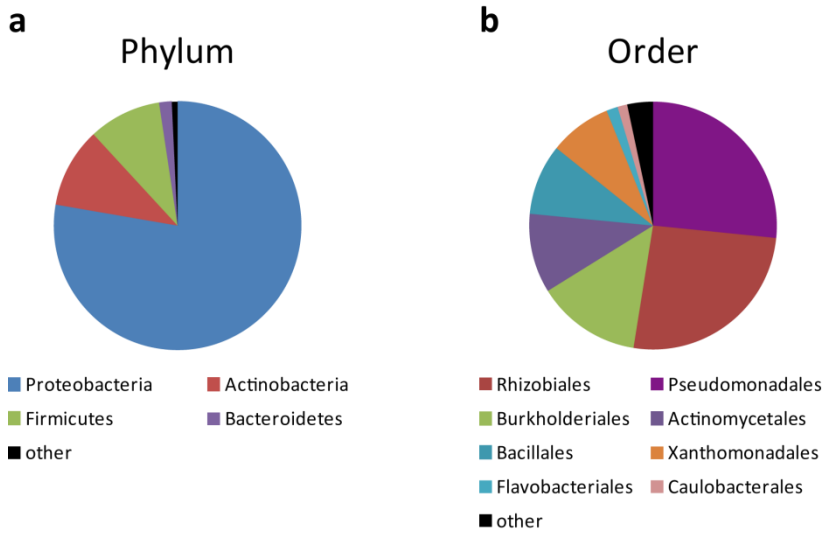


Figure 18: Overview of the **(a)** Phylum assignments and **(b)** Order assignment of all characterized isolates

In total 13 different growth media were used for the isolation of root-associated bacteria. The taxonomic composition of the obtained CFUs depended on the medium used, although the phylum of Proteobacteria dominated the bacterial populations obtained from all media except YEDC, 869, Flour1 and Flour2. 60% of the colonies growing in Flour1 and Flour2 growth medium however could not be identified by sequencing the 16S rRNA gene, either because the utilized

primer pair was not able to amplify the gene of this specific species or because obtained sequences could not be classified using the RDP database. The growth media 1/10 strength 869 and 1/10 strength LB seemed to be especially suited for isolation of Firmicutes, a phylum that was less well obtained from the other media utilized. The proportion of cultivated Actinobacteria was relatively low, although four of the used growth media, YPDA, Flour1, Flour2 and YEDC, were described to enrich Actinobacteria (Coombs & Franco, 2003)(Figure 19).

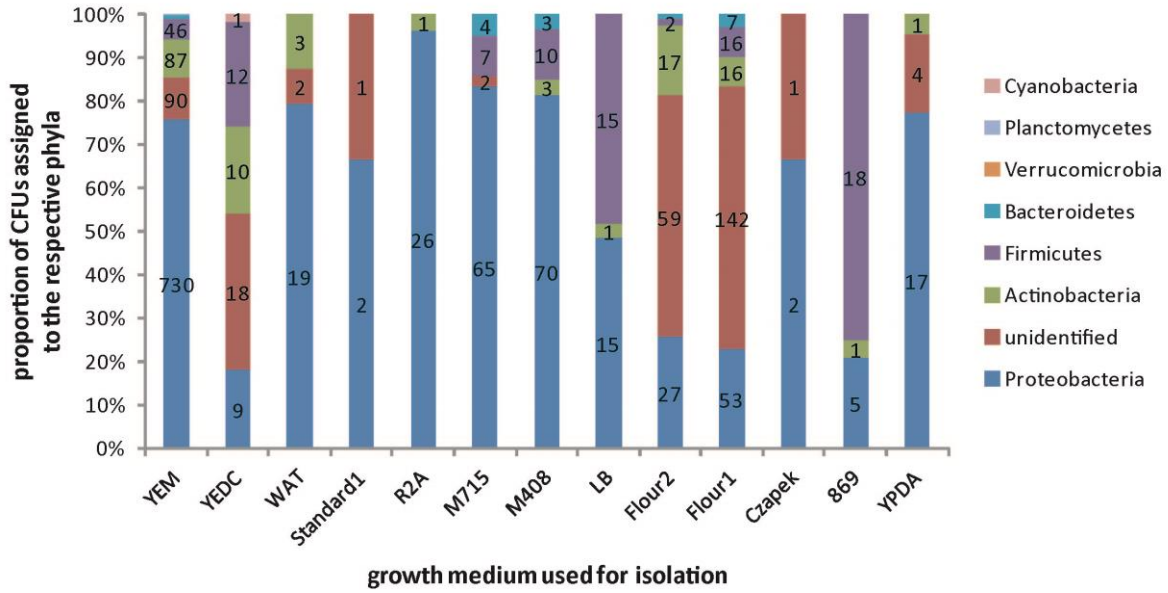


Figure 19: Overview of the taxonomy of characterized CFUs isolated using the indicated growth media. Numbers within each bar represent the number of CFUs assigned to the respective phylum.

To identify whether an isolated CFU matches an OTU that has been identified during the 16S ribotyping survey of root-associated and soil-inhabiting bacterial communities their 16S rRNA gene sequence was blasted against a database containing all OTU representative 16S rRNA gene sequences. A CFU that shares 97% of its 16S rRNA gene sequence over 200 bp with an OTU representative sequence was defined as a member of this OTU.

Of the in total 1682 CFUs 1118 could be matched with an OTU representative sequence in the 16S ribotyping dataset. Of these 237 CFUs were matched to an OTU that was identified to be a member of the rOTUs sub-community, another 395 CFUs were based on their 16S rRNA sequence members of the lOTUs sub-community. 486 of the CFUs matched to OTUs in the dataset that do not belong to any of these sub-communities. The most abundantly isolated OTUs were OTU_{27}, which was classified as a *Pseudomonas sp.*, followed by two members of the Rhizobiales family, OTU_{219} and OTU_{13}. OTU_{27} and OTU_{13} are members of the lOTUs sub-community, while OTU_{219} is a member of the rOTUs sub-community. Prominent isolates that are not members of the aforementioned sub-communities are OTU_{2544}, OTU_{21697}, and OTU_{266}, classified as *Bacillus* sp.,

Variovorax sp. and *Stenotrophomonas* sp., respectively. 564 isolates could not be assigned to an OTU of the ribotyping survey, either because their 16S rRNA gene sequence could not be obtained or their 16S sequence matched no representative in the database (Figure 20).

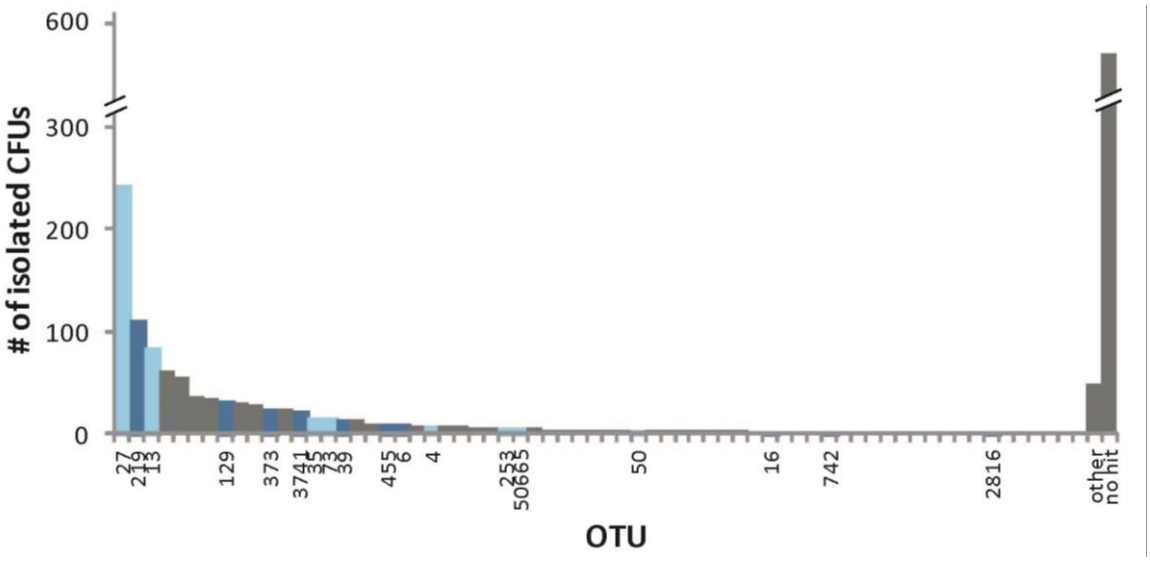


Figure 20: Overview of the total number of CFUs isolated of all isolated OTUs. Members of the lOTUs sub-community are depicted in light blue, members of the rOTUs sub-community in dark blue. OTUs that do not belong to one of those subcommunities are depicted in grey. Numbers at the x-axis indicate the OTUs belonging to rOTUs or lOTUs subcommunities

Most of the CFUs that were matched to members of the rOTUs sub-community were classified as Proteobacteria. Amongst these, Rhizobiales were the most represented order (OTU_{219}, OTU_{129} and OTU_{742}), followed by Xanthomonadales (OTU_{373} and OTU_{455}) and a single CFU that was classified as a Burkholderiales (OTU_{39}). Actinobacteria form the second most abundant group in the isolated fraction of rOTUs members. Within these Streptomycetaceae dominate (OTU_{3741}, OTU_{6} and OTU_{2816}) followed by two single OTUs classified as *Terrabacter* and *Microbacterium*, respectively. A single rOTUs Bacteroidetes could be isolated (OTU_{16}), belonging to the Flavobacteria.

The 16S ribotyping survey indicates that the lOTUs sub-community is largely dominated by members of the phyla Proteobacteria and Bacteroidetes (see chapter I, section 2.3.2). Consistently, most of the isolated lOTUs members are classified as Proteobacteria. 252 of the Proteobacteria CFUs were classified as Pseudomonadales (OTU_{27} and OTU_{4}). 93 CFUs within the lOTUs fraction were classified as Rhizobiales (OTU_{13}, OTU_{50} and OTU_{136}). Seven different Burkholderiales OTUs were isolated, but at a comparatively low number of individual CFUs (OTU_{35}, OTU_{73}, OTU_{253}, OTU_{50665}, OTU_{2292}, OTU_{325} and OTU_{983}). The only CFU classified as Bacteroidetes was a Flavobacterium (OTU_{38844}) (Figure 21). However, it has to be noted that these results are based

on the 16S rRNA gene sequence of the CFU after initial isolation. These CFUs have not been purified and may contain a mixture of different bacterial species.

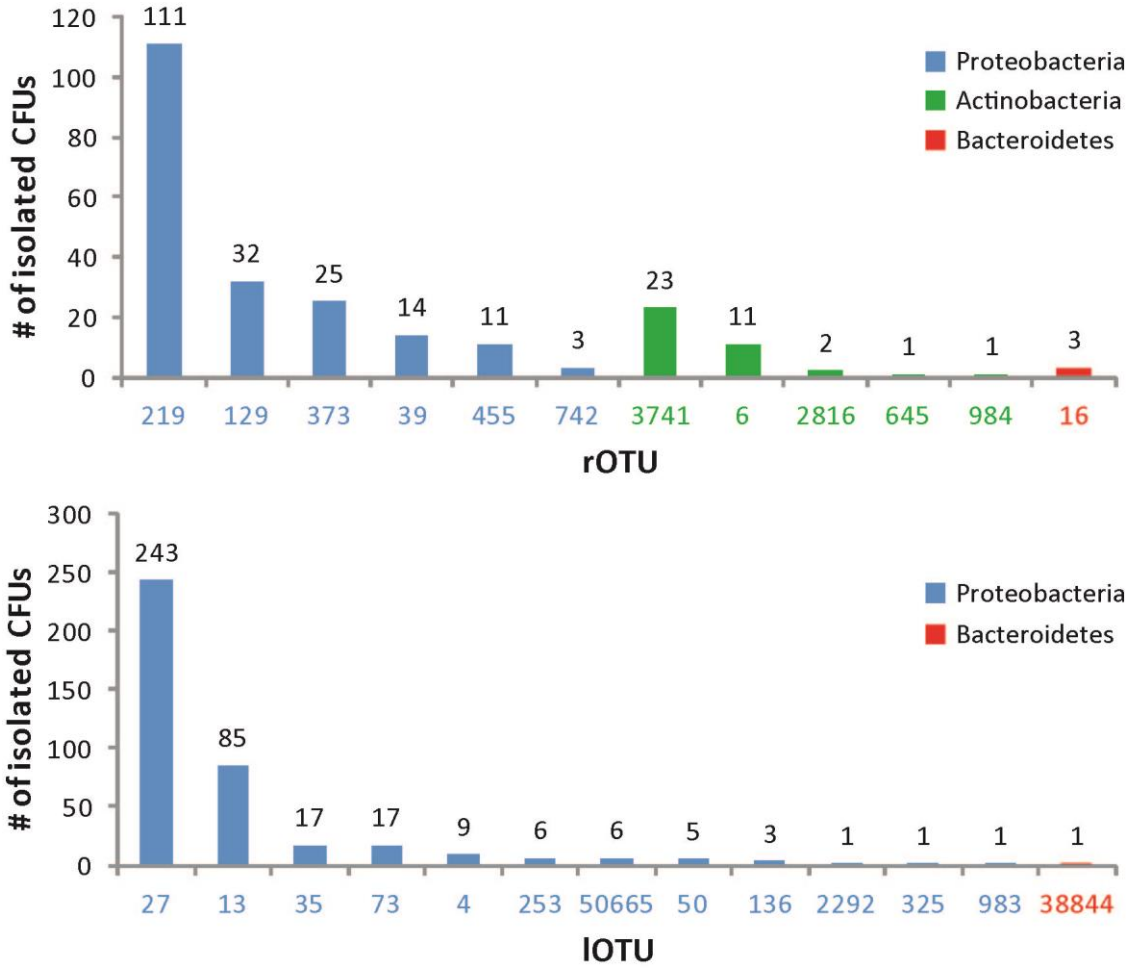


Figure 21: Overview of the total number of CFUs isolated of OTUs belonging to **(a)** the rOTUs sub-community or **(b)** the lOTUs sub-community. Numbers above the bars indicate the total number of CFUs assigned to the respective OTU. Colors of the bars and OTU identifiers indicate the taxonomy of the respective OTU.

2.2 Host interactions of Arabidopsis-associated Rhizobia

In the culture-independent as well as in the culture-dependent approach to assess Arabidopsis root-associated bacteria Rhizobia were abundantly detected. Rhizobia are a well-studied group of bacteria because several species have been shown to be able to fix atmospheric nitrogen and do so after engaging in symbiotic relationships with legumes (Oldroyd *et al.*, 2011). Rhizobia have also been shown to exert plant growth promoting effects independently of nitrogen fixation in association with non-leguminous plants that are not able to form nodules (Bhattacharjee *et al.*, 2008). Arabidopsis as well does not form nodules and is therefore not able to benefit from the nitrogen fixing capabilities of Rhizobia. The reason why Rhizobia are able to colonize Arabidopsis roots and to which extend this is a beneficial interaction is therefore interesting to study.

The isolated Rhizobium strains $Rhizobium_{129}$ and $Rhizobium_{219}$ are members of the rOTUs sub-community in the culture-independent 16S ribotyping survey of the Arabidopsis associated microbiota. They are significantly enriched in the root-inhabiting community in comparison to rhizosphere, unplanted soil and wood associated microbial communities. This strongly suggests that they interact specifically with living Arabidopsis roots. The isolated strain $Rhizobium_{13}$ is a member of the lOTUs sub-community, it is enriched in Arabidopsis root- and wood-associated bacterial communities in comparison to unplanted soil and rhizosphere communities, but no significant difference could be detected between its relative abundance in root and wood communities. It might be therefore able to exploit lignocellulose and to colonize roots and wood in soil. It does not react specifically to assembly cues of living plants but is also not repelled from roots. Comparison of the rOTUs and lOTUs Rhizobia can lead to insights into general principles of host-bacteria interactions.

2.2.1 Re-colonization of Arabidopsis with representative exemplars of Arabidopsis root-associated Rhizobia

To test whether the isolated Rhizobium strains can re-colonize Arabidopsis one representative each of $Rhizobium_{129}$, $Rhizobium_{219}$ $Rhizobium_{13}$ and *Rhizobium leguminosarum* strain 8401 (Downie *et al.*, 1983), hereafter referred to as *R. leguminosarum*, was transformed with a plasmid encoding for the red fluorescent protein (RFP). RFP expression in bacteria allows their visualization in respect to their localization on or in the root tissue. In a gnotobiotic system using square petri dishes RFP expressing bacteria were soaked into filter paper that was placed between the FM plant growth medium and the plant. All tested strains were able to re-colonize Arabidopsis roots in the tested conditions (Figure 22).

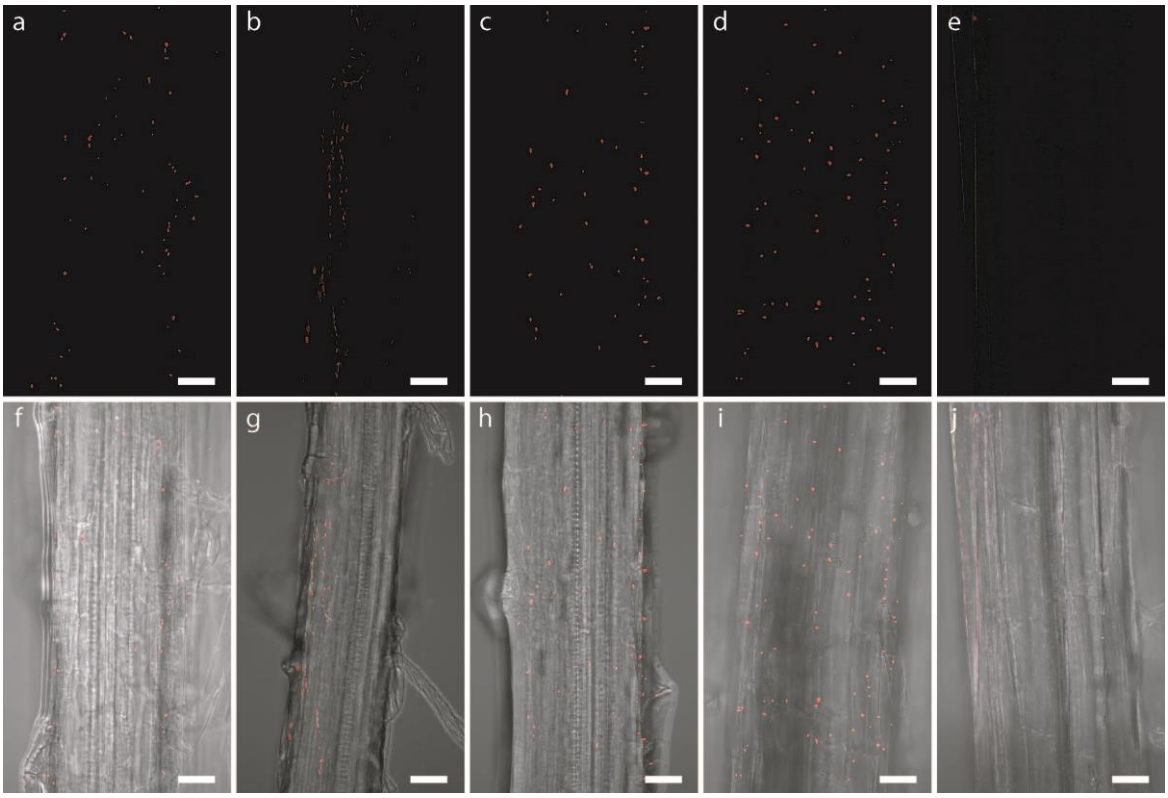


Figure 22: Re-colonization of Arabidopsis roots by *Rhizobium* spp. in a gnotobiotic system. **(a)** Rhizobium$_{13}$ **(b)** Rhizobium$_{129}$ **(c)** Rhizobium$_{219}$ **(d)** *R. leguminosarum* **(e)** $MgCl_2$ control. Size bar 20 µm

To check whether the tested Rhizobium strains can colonize the inner root tissue optical sections of the full diameter of an Arabidopsis root colonized with Rhizobium were made by confocal microscopy. Under the tested conditions RFP expressing Rhizobium strains were only detected on the Arabidopsis root surface but not in the interior of the root tissue (Figure 23).

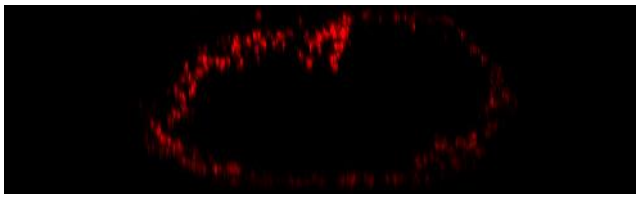


Figure 23: Rhizobium$_{129}$ colonizes the rhizoplane and does not enter the inner root tissue. The picture shows the view into the root along the axis of the root. RFP signals could only be detected on the root surface.

2.2.2 Plant growth promotion capacity of Arabidopsis root-associated Rhizobia in Arabidopsis

To test the plant growth promotion potential of the isolated root colonizing Rhizobia, Arabidopsis were inoculated with representative exemplars of the rOTUs members Rhizobium$_{129}$, Rhizobium$_{219}$ and the lOTUs member Rhizobium$_{13}$, as well as with *R. leguminosarum* for

comparison. The bacteria were embedded at different cell densities in fahraeus growth medium (FM) in square petri dishes. Arabidopsis seeds were then germinated on these FM plates and the shoot fresh weight was monitored after 21 days after sowing.

The two tested members of the rOTUs sub-community, Rhizobium$_{129}$ and Rhizobium$_{219}$, as well as *Rhizobium leguminosarum* were able to induce an increase in shoot fresh weight relative to plant inoculated with $MgCl_2$. Inoculation with Rhizobium$_{129}$ was led to an three fold enhancement in shoot fresh weight relative to $MgCl_2$ treated plants both at an inoculation density of 10^2 and 10^5 cells/ml FM medium. Inoculation with Rhizobium$_{219}$ and *R. leguminosarum* led to a 2.5 fold increase of shoot fresh weight, but only applying the higher inoculation density of 10^5 cells/ml FM medium. No shoot fresh weight increase could be detected after inoculation with the member of the lOTUs sub-community Rhizobium$_{13}$, independent of the applied inoculation density. Embedded *E. coli* cells and heat-treated Rhizobium$_{129}$ cells did also not affect plant growth (Figure 24 and Figure 25).

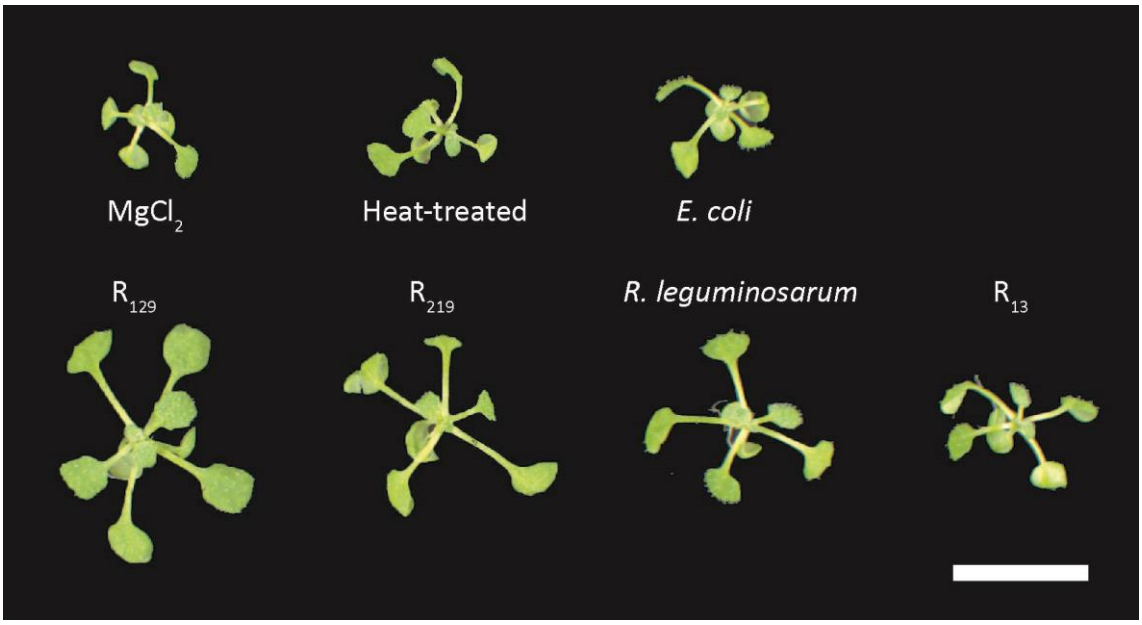


Figure 24: Plant growth promotion in response to inoculation with the indicated Rhizobium strains. Inoculation density 10^5 cells/ml FM medium. Size bar 1 cm.

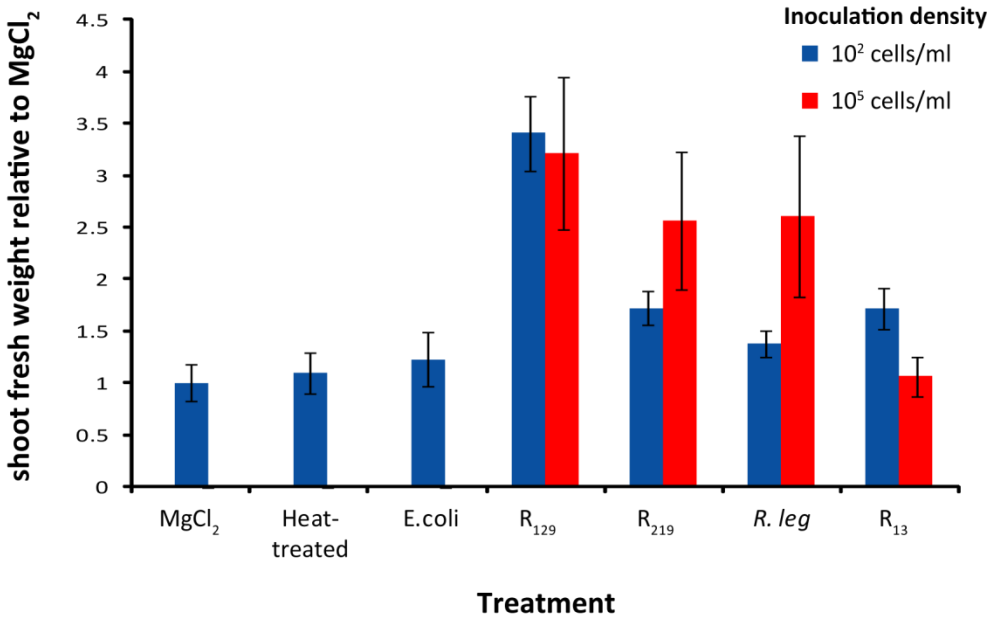


Figure 25: Plant growth promotion in response to inoculation with the indicated Rhizobium strains. Shoot fresh weight relative to $MgCl_2$ treated control

2.3 Genome analysis of Arabidopsis-associated Rhizobia

The plant growth promotion experiments show differences in the outcome of Arabidopsis-Rhizobium interaction, dependent on the Rhizobium strain. Both Rhizobium strains that were identified as members of the rOTUs sub-community were able to promote plant growth and led to an increased shoot fresh weight, while the Rhizobium strain that was not root specific but also abundantly colonized wooden sticks was not able to enhance plant growth. Exploitation of the genomes of members of the rOTUs and lOTUs sub-communities might lead to the identification of key functions in plant-microbe interactions, like root colonization and the establishment of the interaction with the host plant.

2.3.1 The genome of Rhizobium$_{129}$

Since Rhizobium$_{129}$ showed the strongest capacity of plant growth promotion its genome will be discussed in detail as a representative of the Arabidopsis associated rhizobia.

2.3.1.1 Genome structure and general features

The genome of Rhizobium$_{129}$ was sequenced by Illumina 100 bp single reads which were supported by 454 FLX+ reads, resulting in 95 contigs with an average 413 fold coverage and a N50 of 320,297.73 bp. The total contig length was 6,721,213 bp, which is in the range of other sequenced Rhizobia, like *Rhizobium etli* CFN42 (6.53 Mbp (Gonzalez *et al.*, 2006)) and *Sinorhizobium meliloti* (6.7 Mbp (Galibert *et al.*, 2001)) but smaller than the genome of *Rhizobium leguminosarum* (7.75 Mbp (Young *et al.*, 2006)). Since gaps between the contigs were not closed the number of potential plasmids cannot be determined based on the sequence, however five *repA* homologs could be identified in the genome, pointing to the presence of one main chromosome and four plasmids. Coding sequences were predicted using the RAST server service (Aziz *et al.*, 2008) which predicts a total of 6469 coding sequences and 49 RNAs. These coding regions cover 86% of the whole-genome sequence. All predicted coding sequences were additionally annotated using the Blast2Go software package (Conesa *et al.*, 2005). For 5900 the 6469 predicted proteins a blast-hit could be obtained of which 693 sequences are hypothetical proteins with no assigned functions. For 569 predicted proteins no homologs could be identified in the Blast database.

2.3.1.2 Genes potentially involved in Rhizobium$_{129}$-host interactions

Rhizobium$_{129}$ was detected in the unplanted soil, rhizosphere and in the root-associated microbial community. It is therefore likely that it is soil-borne, moves towards the plant and subsequently colonizes its host in high abundance.

For root colonization chemotaxis and motility is indispensable. There are two chemotaxis related operons in the genome of Rhizobium$_{129}$ (*cheDBRWAY*), of which one is in juxtaposition of one flagella biosynthesis operon (*flgABCGHI fliEFGIMNP*). A second detected operon is located on another contig and contains the missing components of flagella biosynthesis (*flgDEJKL fliKQR).* Additionally there are at least nine *dppA* homologs, peptide transporters playing a role for chemotaxis towards peptides (Abouhamad *et al.*, 1991).

Iron acquisition is essential for competition in the rhizosphere, as well as in endophytic growth and its active transport is particularly important (Neilands, 1995). The Rhizobium$_{129}$ genome encodes 13 putative genes involved in siderophore synthesis. Iron-siderophore complexes can then be taken up, e.g. via ATP binding cassette (ABC) transporters. In total 551 predicted proteins were found to be involved in transport of which 228 proteins were assigned to ABC transport. Amongst the putative transporters a variety of sugar transporters could be detected.

The plant's defense strategy against bacterial pathogens includes the formation of reactive oxygen species (ROS), nitric oxide and phytoalexins. The Rhizobium$_{129}$ genome encodes two superoxide dismutases: *sodA*, a manganese superoxide dismutase and *sodC*, a copper/zinc superoxide dismutase. It also encodes three catalases, *KatG*, catalase c and another not further characterized catalase which are both homologues of catalases present in Agrobacterium. Additionally two members of the alkylhydroperoxidase like family could be identified, two chloride peroxidases, a non-heme haloperoxidase, a thioredoxin reductase and a non-heme chloroperoxidase. A putative organic peroxide resistance protein, *ohr*, located next to its organic peroxide sensor/regulator (*ohrR*) was also identified. Rhizobium$_{129}$ seems to be able to detoxify free radical nitric oxide due to an operon involved in nitric oxide reduction. Twenty putative glutathione S-transferase genes (GST) involved in detoxification of a wide variety of both endogenous and exogenous electrophilic compounds could be detected based on homology searches which is in the range of other well-studied PGPR such as *Variovorax paradoxus* (Han *et al.*, 2011). Eight genes were found coding for putative RND family transporters. RND transporters were shown to be required for the export of apple tree phytoalexins by the plant pathogen *Erwinia amylovora* in successful infection of the host plant (Burse *et al.*, 2004).

Rhizobium$_{129}$ colonizes the Arabidopsis root surface and it is known that also other Rhizobia form biofilms on the rhizoplane prior to colonization of the inner root tissue. The establishment of biofilm formation was shown to correlate with the colonization efficiency (Fujishige *et al.*, 2006).

The Rhizobium$_{129}$ genome contains an operon coding for genes involved in the assembly of flp pili, a subfamily of type IV pili (*pilin, tadV, rcpC, rcpA, hyp1, tadZ, hyp4, tadABCD* and *cpaD*). Flp pili are important for biofilm formation, adhesion to surfaces and mutations in the tadA2 gene leads to impaired infection capacities of *R. solanacearum* in potato (Kachlany *et al.*, 2000; Wairuri *et al.*, 2012). The genome of Rhizobium$_{129}$ also encodes a putative filamentous hemagglutinin. These proteins are important in adhesion and virulence in both plant pathogens like *Xanthomonas axonopodis* (Gottig *et al.*, 2009) and animal pathogens like *Moraxella catarrhalis* (Balder *et al.*, 2007).

The genome does not encode any *nod* genes and only one of the *nif* genes: *nifU*, of which its described function is to serve as a scaffold for the [Fe-S] cluster formation during nitrogenase maturation (Johnson *et al.*, 2005).

2.3.2 The core- and pan-genome of Rhizobium$_{129}$

During the culture-independent characterization of the Arabidopsis root-associated bacterial communities via 16S ribotyping 16S rRNA gene sequences were clustered at a sequence identity of 97% and these clusters were defined as OTUs, which is estimated to correspond to a bacterial genus (Schloss & Handelsman, 2005). Nevertheless, an OTU is a mathematical construct based on a single gene that does not reveal metabolic capacities of a given bacterium. Metabolic capacities can only be speculated about indirectly by the retrieved taxonomy, if an OTU was classified as belonging to a taxon that is known for specific functional genes.

Within the kingdom of bacteria genome plasticity is enormous. Bacteria can acquire and exchange genes via horizontal gene transfer (HGT), i.e. transfer of genetic material that does not involve vertical gene transfer during replication, meaning the passing of genes from the parental generation to the offspring. In HGT in bacteria genetic information can be transferred via plasmids and phages which integrate into the bacterial genomes, even between distant bacterial species (Juhas *et al.*, 2009).

To gain insight into genetic variation within a group of bacteria that all would be clustered into the same OTU based on their 16S rRNA gene sequence, the genome of six isolated exemplars of Rhizobium$_{129}$ were sequenced (Table 8). The chosen exemplars were all isolated from individual roots and represent therefore independent colonization events and are unlikely to be representatives of the same clonal population. Within the 290 bp window of the 16S rRNA gene sequence analyzed by 16S ribotyping four single nucleotide polymorphisms (SNPs) differentiating the six exemplars can be detected, over the full length sequence of 1473 bp ten SNPs are present.

Table 8: Overview of the sequencing effort of the isolated Rhizobium$_{129}$ exemplars. * marks the first isolated exemplar that was used in re-colonization experiments

OTU	exemplar	# of contigs	N50	average coverage	sequencing technology
129	**A**	323	72,207.07	164.4	Illumina 100 bp single reads
	B	306	90,156.94	385	Illumina 100 bp single reads
	C	243	95,171.38	308.1	Illumina 100 bp single reads
	D	285	91,391.95	311.4	Illumina 100 bp single reads
	E*	95	320,297.73	413.6	Illumina 100 bp single reads and 454 FLX+
	F	220	104,106.58	369	Illumina 100 bp single reads

The total assembly size as well as the number of predicted genes in the draft genomes of the six Rhizobium$_{129}$ exemplars did not significantly differ. The number of *repA* homologs however ranged from three in exemplar B to six in exemplar D (Table 9).

Table 9: Overview of general genome features of the isolated exemplars of Rhizobium$_{129}$ exemplars. * marks the first isolated exemplar of the respective strain that were used in re-colonization experiments

OTU	exemplar	total contig length	GC content	# of predicted genes	# of repA homologs
129	**A**	6,432,461	60.07	6396	5
	B	6,205,570	60.79	6115	3
	C	6,438,330	60.07	6393	5
	D	6,680,015	60.11	6418	6
	E*	6,721,213	60.1	6469	5
	F	6,445,196	60.12	6186	4

To estimate the genetic diversity inherent to the six exemplars of Rhizobium$_{129}$ their genomes were aligned in a pairwise manner using the MUMmer software package (Delcher *et al.*, 2002). Whole genome alignment revealed that Rhizobium$_{129}$ exemplars can be divided into two subgroups consisting of exemplars A, B, C and on the other hand D, E and F. Within the subgroup$_{DEF}$ 89 – 99 % of all nucleotides can be aligned in pairwise comparisons, while app. 33% of all nucleotides of the genomes of subgroup$_{DEF}$ can be aligned with genomes of members of subgroup$_{ABC}$. Within subgroup$_{ABC}$ genomes of exemplars A and C show a high degree of alignment of 99%. Exemplar B has a more diverse genome, app. 60% of the nucleotides can be aligned to the genomes of exemplars A and C (Table 10).

Table 10: Pairwise alignments of the genomes of exemplars of Rhizobium$_{129}$. The number in a cell gives the fraction [%] of the total assembly size of a test genome that could be aligned to the respective reference genome. The test genome is given in the first column, the respective reference genome in the first row

exemplar	A	B	C	D	E	F
A	-	60.98	99.6	32.93	32.66	33.13
B	62.67	-	62.34	33.08	35.0	35.46
C	99.56	61.03	-	32.99	32.77	33.11
D	31.9	35.36	32.03	-	99.66	85.92
E	31.8	33.09	31.93	99.71	-	85.9
F	33.36	34.68	33.38	89.81	89.86	-

Considering only the segments of the genomes that could be aligned to each other in pairwise comparisons, the percentage average nucleotide identity (ANI) is of course much higher, but still the two subgroups can be distinguished. Within subgroup$_{DEF}$ the average sequence identity is 96-99%, in comparison to app. 86.5% in the alignment with members of subgroup$_{ABC}$. Within subgroup$_{ABC}$ exemplars A and C the sequence identity is 99.9% while exemplar B falls off with 89.1 and 89.2 % of sequence identity with exemplar A and C, respectively (Table 11).

Table 11: Pairwise alignments of the genomes of exemplars of Rhizobium$_{129}$ within the aligned fraction of the total genome. The number in a cell gives the average nucleotide identity in percent in pairwise comparisons

exemplar	A	B	C	D	E	F
A	-	89.1	99.95	86.3	86.4	86.4
B	-	-	89.2	86.4	86.6	86.8
C	-	-	-	86.7	86.7	87.3
D	-	-	-	-	99.9	95.9
E	-	-	-	-	-	95.9
F	-	-	-	-	-	-

The aligned fraction of the total genomes might mostly comprise coding sequences. To check specifically predicted coding sequences all predicted proteins were compared by BLASTing them against a reference genome. The genome of Rhiobium$_{129}$ exemplar E was chosen as a reference since its assembly quality was the highest. Comparisons using the “compare sequence based” function within the RAST server support the distinction of the six Rhizobium$_{129}$ exemplars into two subgroups (Figure 26).

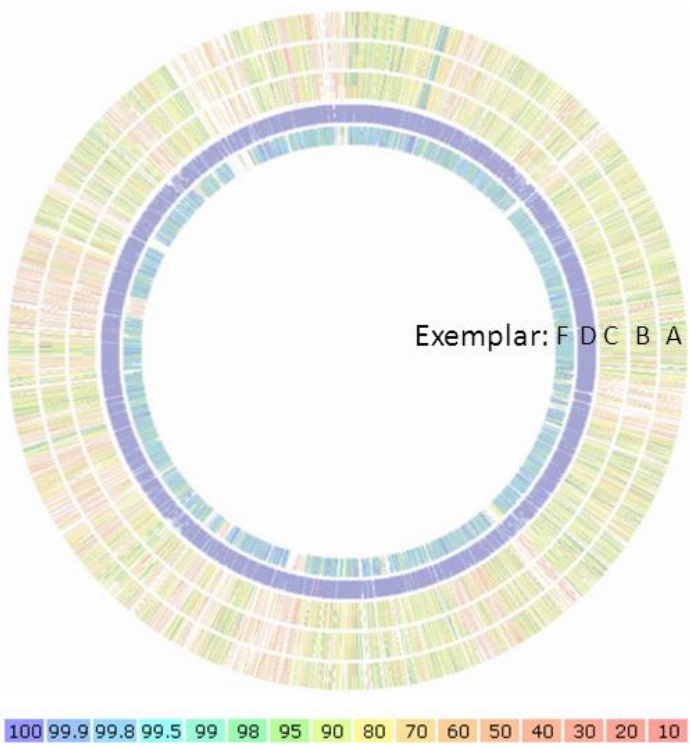


Figure 26: Representation of the amino acid sequence homology of predicted coding sequences in BLAST comparisons of all exemplars against the predicted coding sequences of exemplar E as reference. Color code gives the degree of homology in percent

The size of the core-genome, i.e. the number of genes conserved in all six exemplars and the size of the pan-genome, i.e. the estimated total size of the gene repertoire of Rhizobium$_{129}$ was estimated using a binomial mixture model (Snipen *et al.*, 2009). An all-against-all BLASTing is performed for all predicted coding sequences, and only alignments with at least 50% identity along at least 50% of both sequences are considered. Two sequences are defined to belong to the same gene family if both their reciprocal alignments fulfill this 50-50-cutoff rule. Gene families are then defined to belong to the core- or dispensable genome depending on of if members of these gene families can be detected in all analyzed genomes or only in a subset.

The number of genes belonging to core-genome families in Rhizobium$_{129}$ was 3405±29 (average ± standard deviation), depending on the number of gen family members within the individual Rhizobium$_{129}$ exemplars. 693±484 genes could be identified that belong to gene families that are unique for a given Rhizobium$_{129}$ exemplar and cannot be detected in any other exemplar. Exemplar B shows an outstanding number of 1630 unique genes, which reflects the comparatively low fraction of aligned nucleotides in the alignment with the genomes of other exemplars (Table 12). The total number of all detected gene families in the six Rhizobium$_{129}$ exemplars is app. 11,000. Based on the number of core- and dispensable genes the pan-genome of Rhizobium$_{129}$ is estimated to comprise 30,000 genes families in total.

Because of the separation of the six Rhizobium$_{129}$ exemplars into two subgroups, the number of genes was calculated that belong to gene families unique for one of the subgroups. Exemplars of subgroup$_{ABC}$ have on average 679±11 group specific genes, while exemplars of subgroup$_{DEF}$ have 1337±15 group specific genes (Table 12), although the overall genome size and the overall number of predicted coding sequences does not significantly differ between the isolated exemplars (Table 9).

Table 12: Number of genes in the genome of the individual exemplars of Rhizobium$_{129}$ classified as core genes, exemplar-unique genes or genes assigned as subgroup specific

	A	**B**	**C**	**D**	**E**	**F**
Core genes	3391	3381	3408	3423	3451	3375
Unique genes	574	1630	525	367	315	745
Subgroup genes	683	666	868	1338	1351	1322

The subgroup specific genes of exemplars A and E have been analyzed in detail as representatives of their respective subgroups. The genome of exemplar A encodes 683 subgroup-specific genes of which 64 are predicted be secreted out of the cell by SignalP 4.0 server analysis (Petersen *et al.*, 2011). For 42 of the 683 genes no BLAST hit could be obtained and 117 are annotated as conserved hypothetical proteins for which no function could be assigned. For 316 predicted coding sequences could be mapped to biological processes by the Blast2Go software using gene ontology (GO) terms. Exemplar E has 1351 subgroup specific genes, of which 180 have no homologs in the BLAST database and 229 are assigned conserved hypothetical proteins. 436 predicted coding sequences could be mapped to biological processes. 100 of the subgroup genes of exemplar E are predicted to be secreted out of the cell by SignalP. Analysis of the mapped predicted coding sequences within exemplars A and E reveals a similar GO term composition for subgroup-specific gene-sets (Figure 27).

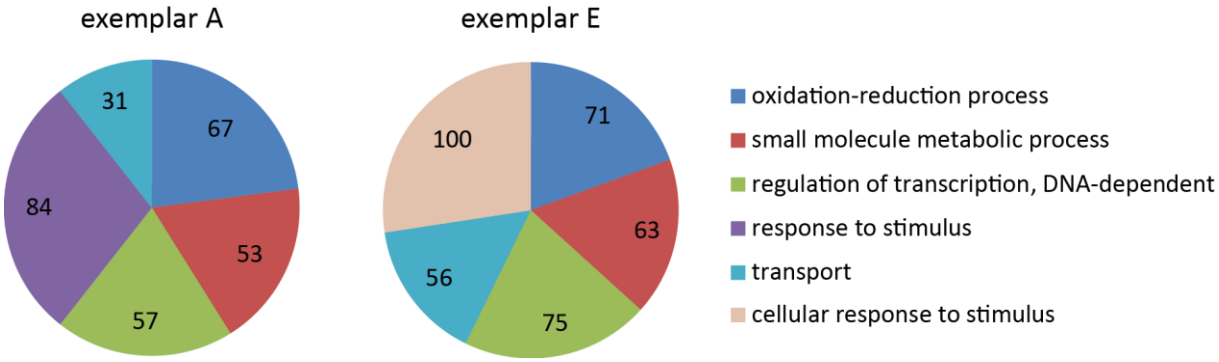


Figure 27: GO term assignments of abundant (≥4% of all mapped genes) subgroup specific predicted genes within the representative exemplars A and E. Numbers within the segments depict the number of predicted genes assigned to the respective GO term

In both subgroup-specific gene sets 'transport', 'oxidation reduction process', 'small molecule process', 'regulation of transcription' and '(cellular) response to stimulus' are the most abundantly determined biological processes, each showing similar proportions in the two gene-sets (Figure 27).

To check whether specific gene functions are enriched in the subgroup-specific gene sets a GO term enrichment analysis was conducted of the respective subgroup-specific genes against the whole genome of Rhizobium$_{129}$ exemplars A and E, respectively. In exemplar A's subgroup-specific gene set 'sequence specific DNA binding proteins' are significantly enriched (Fisher's exact test, FDR ≤0.05). Similarly, in exemplar E's gene set predicted proteins assigned to the GO term categories 'DNA-dependent transcription', 'RNA biosynthesis' and 'regulation of cellular processes' are enriched. Therefore, in both exemplars proteins predicted to be involved in regulation of transcription are overrepresented in the subgroup specific gene set in comparison to the rest of the predicted coding sequences.

2.3.3 Comparison of genomes of the rOTUs Rhizobia and lOTUs Rhizobia

The isolation of root-specific Rhizobia that rely on assembly cues of living root cells and on the other hand Rhizobia that do not interact specifically with living root tissue allows insights into the underlying mechanisms of this differential colonization behavior. Differences in direct Rhizobium-Arabidopsis interactions have been already shown by the differential plant growth promotion capabilities of the tested strains (see section 2.2.2). To elucidate the reason for the observed effects the genomes of four independently isolated lOTUs Rhizobium$_{13}$ exemplars as well as one genome of rOTUs Rhizobium$_{219}$ exemplar was sequenced via Illumina sequencing (Table 13).

Table 13: Overview over sequencing effort of the isolated exemplars of Arabidopsis associated *Rhizobium sp.* strains and exemplars. * marks the first isolated exemplar of the respective strain that was used in re-colonization experiments

OTU	exemplar	# of contigs	N50	average coverage	sequencing technology
219	**A***	311	218,987.22	317.2	Illumina 100 bp paired end reads
13	**A**	336	129,720.83	222.1	Illumina 100 bp single reads
	B	150	170,038.17	259.1	Illumina 100 bp single reads
	C*	122	203,823.38	311.4	Illumina 100 bp single reads
	D	192	121,330.96	162.8	Illumina 100 bp single reads

The total assembly size of the genome of Rhizobium$_{219}$, classified as member of the rOTUs sub-community, is 7.85 Mb with 7586 predicted coding sequences which is app. 1 Mb and 1000 predicted genes larger than the genome of Rhizobium$_{129}$. The four exemplars of Rhizobium$_{13}$ can be divided into two subgroups based on total assembly size, predicted genes and homologs of the *repA* gene. Exemplars A and B have a total assembly size of app. 7.5 Mb, 7,500 predicted genes and based on the number of *repA* homologs three plasmids. Exemplars C and D on the other hand have a smaller genome of 6.2 – 6.5 Mb and 6,200 - 6,500 predicted genes encoded on a chromosome and two plasmids as estimated by the number of *repA* homologs (Table 14).

Table 14: Overview of general genome features of the isolated exemplars of Arabidopsis associated *Rhizobium sp.* strains and exemplars. * marks the first isolated exemplar of the respective strain that were used in re-colonization experiments

OTU	exemplar	total contig length	GC content	# of predicted genes	# of *repA* homologs
219	**A***	7,854,397	60.88	7586	6
13	**A**	7,490,787	60.69	7384	4
	B	7,774,088	59.98	7761	4
	C*	6,521,343	60.16	6512	3
	D	6,176,320	60.35	6183	3

To identify genes potentially involved in the differentiation of the isolated Rhizobium strains into rOTUs and lOTUs sub-communities and the associated root-colonization behavior a genome comparison between all rOTUs exemplars (Rhizobium$_{129}$, exemplars A-F and Rhizobium$_{219}$) and all lOTUs exemplars (Rhizobium$_{13}$, exemplars A-D) was conducted.

The genomes isolated rOTUs exemplars encode 612±37 genes that were classified to belong to a rOTUs enriched gene family, while the lOTUs genomes encode 1363±154 lOTUs enriched gene family members (Table 15).

Table 15: Number group enriched genes in the isolated *Rhizobium* exemplars. SGEG = subgroup specific genes

OTU	**R_{129}**						**R_{219}**	**R_{13}**			
Exemplar	A	B	C	D	E	F	A	A	B	C	D
SGEG	623	538	622	628	640	619	362	1160	1130	1505	1456

The 640 rOTUs enriched genes of Rhizobium$_{129}$ exemplar E include 152 conserved hypothetical genes and 55 genes for which no BLAST hit could be obtained. 202 predicted proteins can be mapped to a biological function via GO term assignments. Based on prediction of the presence and location of signal peptide cleavage sites in the amino acid sequences of the 640 rOTUs enriched genes using the SignalP 4.0 server 63 genes are predicted to be secreted from the bacterial cell. The genome of lOTUs Rhizobium exemplar C encodes 1505 lOTUs enriched genes of which 121 have no homolog in the BLAST database and 303 are conserved hypothetical proteins. 590 predicted proteins can be mapped to a biological function via GO term assignments. In total 134 predicted proteins are predicted to be secreted.

As already observed for subgroups within the Rhizobium$_{129}$ pan-genome also rOTUs and lOTUs enriched gene family members are mostly assigned to the GO terms ‘oxidation-reduction process’, ‘transcription, DNA dependent’ and ‘transport’. The lOTUs enriched gene subset additionally comprises genes assigned as ‘ATP catabolic process’ and ‘response to ‘stimulus’, while the rOTUs enriched gene subset comprises additionally genes assigned to the GO terms ‘small molecule metabolic process’ and ‘regulation of cellular process’ (Figure 28).

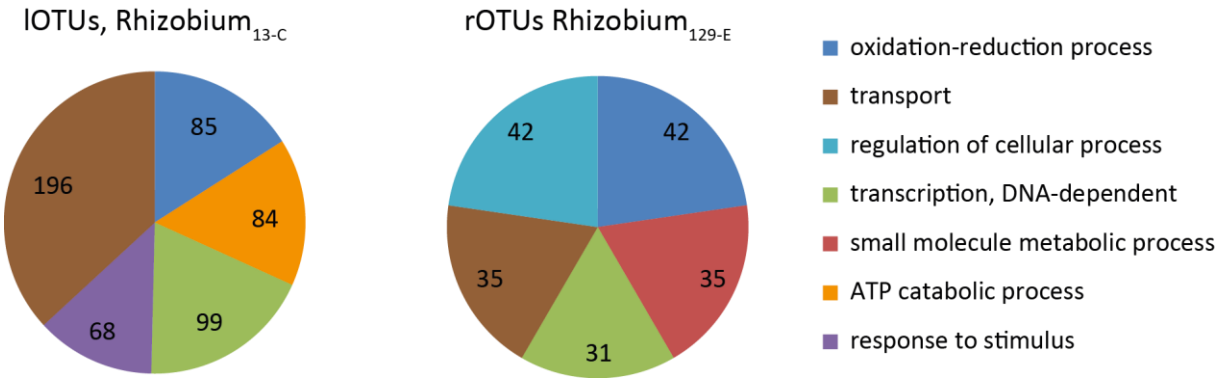


Figure 28: GO term assignments of abundant (≥4% of all mapped genes) subgroup specific predicted genes within the in the lOTUs representative Rhizobium$_{13}$ exemplar C and rOTUs representative Rhizobium$_{129}$ exemplar E. Numbers within the segments depict the number of predicted genes assigned to the respective GO term

GO term enrichment analysis of the 640 rOTUs specific gene family members of $Rhizobium_{129}$ exemplar E within all predicted coding sequences revealed no enrichment for a specific GO term.
GO term enrichment analysis of the 1505 lOTUs specific gene family members of $Rhizobium_{13}$ exemplar C within against its full genome reveals a significant enrichment of genes with the assigned GO terms ‘sequence-specific DNA binding transcription factor activity’, ‘ATP-binding cassette (ABC) transporter complex’, ‘ATP catabolic process’ and ‘peptide transport’.

3 Discussion

3.1 Isolation of Arabidopsis root-inhabiting bacteria

In order to gain deeper insight into the bacterial communities associated with Arabidopsis roots a cultivation-based approach was started. Root material was harvested consistently to the sampling procedure of the culture-independent 16S ribotyping approach, macerated in sterile buffer and plated on a range of bacterial growth media. Obtained CFUs were characterized by sequencing their 16S rRNA genes.

The taxonomy of all isolated bacteria largely represents the taxonomy obtained via culture-independent characterization by 16S ribotyping. With 78% of all obtained CFUs the most abundant phylum was Proteobacteria, which was comprised app. 50% of the root-inhabiting bacterial community as determined by 16S ribotyping. The second most abundant group of isolated bacteria was Actinobacteria (10%), which are slightly underrepresented in comparison to the RA abundance of Actinobacteria detected by 16S ribotyping (app. 25%). 2% of all isolated CFUs were classified as Bacteroidetes, in comparison to app. 10% RA in the 16S ribotyping survey. 9.5%, however, of all cultivated bacteria were classified as Firmicutes, a phylum that was found to make up only 0.4% of all detected sequences in 16S ribotyping (Figure 3 and Figure 18). Since the root material for the culture-dependent and culture-independent approaches was prepared in the same way this might reflect a bias introduced by the utilized primer pair 799F2-1193R which seems to fail to detect bacteria of the phylum Firmicutes. This is supported by the comparison of the primer pair 799F2-1193R with its parental primer pair 799F-1193R that was modified to exclude plant derived sequences from the amplicon pools. Characterizing soil microbial communities with 799F-1193R detected Firmicutes at a RA of app. 10%, while Firmicutes were hardly detected using the primer pair 799F2-1193R. Nevertheless, also the parental unmodified primer pair did not detect Firmicutes in root-associated bacterial communities (see Appendix), therefore it cannot be excluded that the growth media used for isolation of bacteria are enriching for Firmicutes. At the same time the cultivation approach is biased towards Proteobacteria, since in the goal of isolation shifted in the course of the study from isolating a broad diversity of bacterial species towards isolating additional exemplars of Arabidopsis-associated Rhizobia and therefore media that support growth of Rhizobia (growth media YEM, M408, M715) were applied unproportionally often (Figure 19).

66% of all isolated CFUs could be mapped to an OTU representative sequence of the 16S ribotyping survey (Figure 20), while 564 CFUs could not be assigned to any OTU. For 315 CFUs of these no 16S rRNA gene sequence could be obtained. In the course of the study the methodology was refined leading to an increased proportion of positively amplified 16S rRNA genes. CFUs that

could be classified based on their 16S rRNA gene sequence but not mapped to an OTU in the ribotyping dataset predominately belong to the phylum of Proteobacteria (155 CFUs), Firmicutes (33 CFUs) and Actinobacteria (23 CFUs). This could be either explained by the use of growth media enriching for Proteobacteria or by a PCR primer bias. In the primer comparison of the modified version 799F2 with its parental primer 799F the parental primer detects Proteobacteria more abundantly than the modified primer.

Of in total 43 OTUs that were identified by 16S ribotyping to be significantly enriched in Arabidopsis bacterial communities in comparison to unplanted soil and wood stick bacterial communities and therefore members of the rOTUs subcommunity, 12 OTUs (28%) could be cultivated. Similarly, of 34 lOTUs that were shared by Arabidopsis and wood stick bacterial communities and significantly enriched against unplanted soil, 13 OTUs (38%) could be isolated (Figure 10 and Figure 21). Compared to the estimated overall culturability of a maximum of 5% of all soil bacteria with the current standard growth media (Keller & Zengler, 2004) this represents a significant proportion. The 13 rOTUs classified as Proteobacteria that could not be cultivated so far comprise a diverse range of α-, β-, γ- and δ-Proteobacteria. Obviously excluded from cultivation was a group of four OTUs classified as *Myxococcales* which may be cultivated using a more targeted approach. Since *Myxococcales* have been described to lyse other bacteria and feed on these an enriching growth medium containing lysed *E. coli* or yeast cells could be applied (Dawid, 2000).

Another 10 rOTUs that were not isolated so far are classified as Actinobacteria. Half of them are Streptomycetales, of which other representatives could be successfully isolated using growth media that were described as enriching for Actinobacteria (Flour1, Flour2, YEDC YPDA (Coombs & Franco, 2003)). Screening more CFUs showing the characteristic growth phenotypes of Streptomycetales might therefore lead to the cultivation of the missing representatives. Streptomycetes are known to produce and being resistant against a vast array of secondary metabolites like antibiotics (Watve *et al.*, 2001). These could be incorporated into isolation media to prevent growth of other bacterial species and increase the chance of isolation of Streptomycetes.

Of six Bacteroidetes rOTUs only one, a Flavobacterium, could be isolated, while other four Flavobacteria remain uncultured. Flavobacteria are not known to be difficult to culture from environmental samples and have been isolated by a number of common growth media (Jooste & Hugo, 1999). Screening more colonies might lead to the cultivation of these missing members of the rOTUs subcommunity.

Since the lOTUs subcommunity comprises almost solemnly Proteobacteria also the fraction that could not be cultured so far consists mostly of Proteobacteria. Of the 19 uncultured

Proteobacteria eight are classified as Burkholderiales. The app. 30 species within the genus Burkholderia occupy remarkably diverse niches and have been described both as human and plant pathogens, but also been shown to promote plant growth and would be therefore interesting to study (Coenye & Vandamme, 2003). Since a number of Burkholderia species are described to fix molecular nitrogen N_2 a nitrogen-free isolation medium might be worthwhile to test (Estrada-De Los Santos *et al.*, 2001).

The remaining uncultured lOTUs of the phylum Proteobacteria belong to the Myxococcales, Rhizobiales, Xanthomonadales and Pseudomonadales and with the exception of Myxococcales other members of these groups have been successfully cultivated. The only two uncultured lOTUs members belonging to the Bacteroidetes are classified as Flavobacteria.

Generally a problem in cultivation of these missing members of the rOTUs and lOTUs subcommunity is the number of colonies that have to be picked from the initial isolation growth plate and to screen them by sequencing their 16S rRNA gene to obtain not only very abundant but also low abundant bacteria. At the same time fast growing bacteria will overgrow bacteria with a longer replication time, making it difficult to capture the full bacterial diversity. This can be circumvented by creating species collections by a limited dilution approach. Instead of plating macerated root material on solid growth medium, the buffer is diluted into 96well plates filled with liquid growth medium. At a certain dilution the bacterial density is so low, that at a high chance individual wells receive single bacterial cells that will develop clonal populations that are not overgrown by fast growing bacteria. The identity of these clonal cultures can then be identified by high-throughput barcoded pyrosequencing of their 16S rRNA gene as done for culture collections of the human gut microbiota (Goodman *et al.*, 2011).

Another way to identify target bacteria within all CFUs on solid growth media is using colony hybridization on replica membranes of the growth medium plate. Colonies are hybridized with a probe specifically binding only the 16S rRNA of the targeted bacterial species. Probes can then be detected by fluorescence or antibodies and display the targeted bacteria without the need to sequence the 16S rRNA genes of all CFUs growing on the growth medium (Vartoukian *et al.*, 2010).

3.2 Arabidopsis root-associated Rhizobia

In the culture-independent as well as in the culture-dependent approach to characterize Arabidopsis root-associated bacterial communities Rhizobia were abundantly detected. Two of the tested strains, $Rhizobium_{129}$ and $Rhizobium_{219}$, are members of the rOTUs subcommunity and therefore colonization by these strains depends of assembly cues of living Arabidopsis roots. The third tested strain $Rhizobium_{13}$ is a member of the lOTUs subcommunity, it colonizes both living

Arabidopsis roots as well as metabolically inactive lignocellulosic matrices. To test the biological significance of this finding these three different Rhizobia strains were tested in for re-colonization and plant growth promotion capabilities in association with their natural host Arabidopsis in a gnotobiotic system.

Rhizobium$_{129}$, Rhizobium$_{219}$ and *R. leguminosarum* increased shoot fresh weight when Arabidopsis was grown presence of these strains (Figure 24 and Figure 25). Since the experiments were conducted in a gnotobiotic system and no other bacteria or fungi were added plant growth promotion does not occur indirectly via competition with pathogens about nutrients or priming of the plant's innate immune system but via a direct plant growth promotion mechanism. The plant fahraeus growth medium contains all nutrients in soluble and plant accessible form, suggesting that the induced changes of fresh weight are a result of interference with the plant's phytohormone signaling. Described main mechanisms how PGPR can tune phytohormone signaling include ethylene breakdown by the enzyme ACC deaminase and bacterial auxin production. Auxin production has been shown for both nodulating and non nodulating strains of *R. leguminosarum* (Wang *et al.*, 1982) and auxin producing *R. leguminosarum* strains are able to promote plant growth in non-leguminous plants (Bhattacharjee *et al.*, 2008). *R. leguminosarum* strains have also been shown to produce ACC deaminase which promotes nodulation in interaction with legume hosts, but might also be involved in the observed plant growth promoting effect on non-leguminous plants like rice (Ma, WB *et al.*, 2003; Bhattacharjee *et al.*, 2012). Both auxin production and ACC deaminase activity could explain the increased shoot fresh weight observed in Arabidopsis upon inoculation with *R. leguminosarum*. Enzymes necessary to produce auxin from the precursor tryptophan are not encoded in any of the genomes of Rhizobia isolated from Arabidopsis roots. The genome of Rhizobium$_{129}$ exemplar E encodes for genes of denitrification pathway necessary for the production of nitric oxide from nitrate. Nitric oxide produced by *A. brasilense* has been reported in a single publication to mimic effects of IAA on root morphology and promote plant growth in tomato if nitrate is added to the growth medium (Molina-Favero *et al.*, 2008). Rhizobobium$_{129}$ however does not induce formation of higher numbers of lateral roots or shortened primary roots (data not shown) which is shown to be a typical root response to PGPR-produced auxin (Spaepen & Vanderleyden, 2011). It is therefore unlikely that auxin production or auxin mimicking are responsible for the observed plant growth promotion effect.

An ACC deaminase can be found in the genome of Rhizobium$_{219}$ which exhibits a similar growth promotion effect as *R. leguminosarum*. Interference with the plant's ethylene levels might be a conserved mechanism of plant growth promotion between these two strains, if the ACC deaminase gene is indeed expressed and the protein is functional. No ACC deaminase gene could

be detected in the genome of Rhizobium$_{129}$, the strain exhibiting the strongest growth promotion effect. Since gaps in the genome were not closed it is however possible that the gene might be present but is not detected because it is located on a part of the genome that has not been completely assembled. It has to be noted, that the genome of Rhizobium$_{13}$, member of the lOTUs subcommunity and not able to promote plant growth of Arabidopsis under the tested conditions, encodes an ACC deaminase as well as the genes necessary to produce nitric oxide. The presence of these genes does not mean that they are expressed and the proteins functional. Dose dependent effects can also explain differences in the plant growth promotion capabilities observed for the Rhizobium strains. Differences in colonization efficiency can as well account for the differences and should be tested by re-isolating bacteria from Arabidopsis roots after plant growth promotion assays. Another possibility might be the plant growth promoting strains take up nutrients derived from rhizodeposition, metabolize them and secrete them again in a form that is bioavailable and useful for the plant. Plant roots are able to actively uptake sugars and organic nitrogen compounds like amino acids via H^+-ATPase powered transporters (Jones & Darrah, 1996; Hirner *et al.*, 2006). In this way bacteria might provide additional nutrients even in the gnotobiotic plate assay, in which all nutrients are given in a bioavailable form.

In conclusion, the mechanism how the tested Rhizobium strains promote plant growth is still elusive and cannot be retrieved from the genome sequence alone, but has to be elucidated by more targeted approaches like testing for the production of phytohormones or testing for plant growth promotion capabilities under defined plant growth limiting nutrient deficient conditions. Likewise, exploitation of genetic tools available for a model system like Arabidopsis will allow to test host mutants impaired in the perception of putative plant-growth promotion factors in the tested conditions. These combined approaches will facilitate the dissection of molecular mechanism underlying plant-PGPR interactions.

3.3 Genome analysis of Rhizobium$_{129}$

The genome of Rhizobium$_{129}$ was analyzed in detail as a representative of Arabidopsis associated Rhizobia that have been based on culture-independent 16S ribotyping classified as significantly enriched in Arabidopsis associated rhizobacterial communities.

The genome of Rhizobium$_{129}$ reflects its two-phasic lifestyle of being soil-borne and subsequent colonization of the rhizoplane of Arabidopsis. With app. 6.5 Mb Rhizobium$_{129}$ possesses a big genome in comparison to other bacteria which is in coherence with the finding that plant-associated bacteria have significantly larger genomes than nonplant-associated bacteria (Pini *et al.*, 2011). Genes encoding for beneficial traits in plant colonization are well represented in the

genome. Based on gene annotations Rhizobium$_{129}$ is motile and able to sense potentially root-derived chemical stimuli. Rhizosphere competence is gained via the capability to compete for iron by the production of siderophores and by an extensive protein machinery to deal with oxidative stress. Type IV pili might be involved in attachment to the root. Rhizobium$_{129}$ lacks the *nif* gene encoding for the nitrogenase enzyme complex and *nod* genes necessary for induction of nodulation in legumes. The genome also does not encode for a type three secretion system, Rhizobium$_{129}$ is therefore most likely not able to modulate its host's innate immune system by delivery of effector proteins in a canonical way.

To estimate the genetic variability within bacteria that would all based on their 16S rRNA gene sequence be clustered into the same OTU the genomes of six independently isolated Rhizobium$_{129}$ exemplars were sequenced. These exemplars were all isolated from plants grown in the same soil, but in individual pots. Each exemplar represents therefore an independent colonization event and exemplars are unlikely to be derived from a clonal population. Within the 16S rRNA sequence window that was analyzed via 16S ribotyping the six exemplars show four SNPs, while over the whole sequence 10 SNPs can be detected, which means that the exemplars would be clustered into the same OTU even taking into account the full length sequence information. Pairwise full genome alignments as well as bi-directional BLAST analysis of the predicted coding sequences the six Rhizobium$_{129}$ exemplars reveal a remarkable genome variation between exemplars within one OTU. The fraction of a total genome that can be aligned to another exemplar can be as low as 32% (Table 10).

In the era of whole genome sequencing bacterial species can be defined based on their genome similarities. Genomes that show an average nucleotide identity (ANI) of ≥95% of the total genomic sequence shared between two strains are proposed to belong to the same species (Konstantinidis *et al.*, 2006). Plasmids should not be considered in the definition of bacterial species since they are not stable properties of the genome (Konstantinidis & Tiedje, 2005). Since it could not be determined whether a given predicted coding sequence of the Rhizobium$_{129}$ exemplars is encoded on a plasmid or on the stable genome unambiguously species definition within the Rhizobium$_{129}$ exemplars was not possible. However, taking into account the whole genomes of the isolated Rhizobium$_{129}$ exemplars one would estimate three different species based on the 95% ANI definition: Species 1 represented by exemplars A and C, species 2 represented by exemplar B, and species 3 represented by exemplars D, E and F (Table 11). The presence of several species within one OTU based on ≥97% sequence identity over a segment of the 16S rRNA gene sequence could be expected as this OTU definition was shown to correlate with the bacterial taxon of a genus (Schloss & Handelsman, 2005).

The size of the core-genome, i.e. the number of genes conserved in all six exemplars and the size of the pan-genome, i.e. the estimated total size of the gene repertoire of Rhizobium$_{129}$ was estimated using a binomial mixture model (Snipen *et al.*, 2009) that aligns all predicted coding sequences of all exemplars against each other and defines gene families based on alignments that have least 50% identity along at least 50% of both sequences. These gene families can then be defined as belonging to the core- or the pan-genome, dependent on if they are present in the genomes of all exemplars or only in a subset. The number of genes belonging to core-genome families in Rhizobium$_{129}$ was 3405±29 while 693±484 genes could be identified that belong to gene families that are unique for a given Rhizobium$_{129}$ exemplar and cannot be detected in any other genome (Table 12). Considering the size of the analyzed Rhizobium genomes this number of genes is in the range of what has been observed before as for example for *Sinorhizobium* (Tian *et al.*, 2012). The core-genome of ten *Pseudomonas fluorescens* strains, with an average individual genome size of app. 5.5 Mb, encodes for 2,789 genes (Loper *et al.*, 2012). Similarly, the core-genome of 17 *E. coli* human pathogenic and commensal isolates with genomes around 5 Mb is predicted to contain 2,344 genes (Rasko *et al.*, 2008).

Based on the number of core- and dispensable gene families the pan-genome of Rhizobium$_{129}$ is estimated to comprise 30,000 genes families in total. Estimation of the pan-genome based on gene families instead of individual genes reduces the estimated size of the pan-genome since multiple genes are combined in one gene family. Therefore, the estimated pan-genome of Rhizobium$_{129}$ is comparatively big. In the aforementioned studies the pan-genomes of *E. coli* and *P. fluorescens* were calculated to add up to 13,000 and 13,800 genes, respectively. Based on the total ANI of the analyzed genomes Rhizobium$_{129}$ is a genus that contains three different species inflating the pan-genome size. Additionally, based on the number of *repA* homologs per Rhizobium$_{129}$ exemplar the number of plasmids varies between three and six. Plasmids are one of the factors with the highest influence on pan-genome size as reported for *Erwinia amylovora* for which the pan-genome can be considered as closed, apart from the plasmid content (Smits *et al.*, 2011).

Since the Rhizobium$_{129}$ exemplars could not unequivocally be divided into species the exemplars were separated into two larger sub-clusters of three exemplars each. All six exemplars were isolated from Arabidopsis ecotype Shakdara grown in Cologne soil, ruling out an obvious environmental factor for the separation of the exemplars into two subgroups. All six exemplars have a similar genome size and a similar number of predicted coding sequences. Part of the divergence at the whole genome level is probably explained by the presence of different types and different numbers of plasmids in the isolated exemplars.

Comparison of subgroup$_{DEF}$ to the genomes of the other exemplars reveals on average 1337 subgroup enriched genes. The genomes within subgroup$_{ABC}$ are more diverse due to the inclusion of exemplar B, which reduces the number of subgroup-enriched genes for this group on average to 679. Within both subgroup enriched gene sets (Figure 27), however, genes assigned to the same biological functions are represented. More than 25% of the subset specific genes to which a function could be assigned by GO terms are predicted to be involved in transport. This points towards an adaptation to different habitats because of the ability to take up different subsets of nutrients from the bacterial environment as the genome of exemplar E encodes for example for several subgroup$_{DEF}$-specific ABC-type sugar transport proteins. Another 20% of both subgroup-specific gene sets are defined as involved in 'oxidation reduction processes', pointing towards subgroup specific metabolic pathways. In the genomes of subgroup$_{DEF}$ for example three operons are present encoding for proteins necessary to denitrify nitrite to molecular nitrogen via nitric oxide and dinitrogen oxide. Denitrification is an alternative way of generating energy under anaerobic conditions, for example in soil (Hayatsu *et al.*, 2008). In members of the Rhizobiales denitrifying species have been reported to grow under oxygen-limited conditions in the presence of nitrate or nitrite as terminal electron acceptor (Ohara & Daniel, 1985). In *Bradyrhizobium japonicum* USDA110 mutants lacking the genes encoding the cytochrome c-containing subunit II and cytochrome b-containing subunit I of the nitric oxide reductase are not able to grow microaerobically under nitrate-respiring conditions (Mesa *et al.*, 2002). In rhizobia-legume symbiosis nitric oxide is accumulating in the nodule and might act as a signaling compound in nodule formation but also inhibits the nitrogenase leading to reduced nitrogen fixing efficiency (Meilhoc *et al.*, 2011). Rhizobium$_{129}$ exemplars of the subgroup$_{DEF}$ therefore either have the potential to grow anaerobically under nitrate-respiring conditions or are closer related to nodulating Rhizobium strains that involve denitrification genes in symbiosis than Rhizobium$_{129}$ exemplars of the subgroup$_{ABC}$.

The subgroup enriched gene sets were tested for GO term enrichments against the rest of the predicted proteins of a genome of a representative of the respective subgroups. In this way it can be determined whether specific gene functions are overrepresented in the subgroup enriched gene sets. Gene duplication/amplification and subsequent diversification could for example lead to such overrepresentations in subgroup specific gene sets. Gene amplification is a common mechanism in evolution and adaptation to a new habitat. If genes are redundant, one homolog can evolve to adapt to changing environmental factors, while the other homolog ensures fitness in the present situation. In this way gene amplification is a relatively fast way of evolution (Andersson & Hughes, 2009). GO term enrichment analysis for subgroup$_{DEF}$ enriched genes as well as for the subgroup$_{ABC}$ enriched gene-set within the genomes of Rhizobium$_{129}$ exemplars E and A,

respectively, reveals that within both subset genes encoding for proteins involved in transcription are enriched. Additionally GO term enrichment analysis reveals a subgroup specific operon of five genes for synthesizing the iron chelating siderophore rhizobactin 1021 in subgroup$_{ABC}$ which was initially identified in *Sinorhizobium meliloti* (Lynch *et al.*, 2001) and might represent an adaptation to an environment low in iron.

Transcription factors are described to be strongly selected for under complex environmental conditions. Genomes of free living non-pathogenic bacteria, including bacteria that are able to form symbiosis with plants or animals, are enriched in the number of transcription factors in comparison to bacteria with a very defined habitat like intracellular symbionts (Cases *et al.*, 2003). Rhizobium$_{129}$ is soil-borne and subsequently colonizes the rhizoplane, it therefore represents such a free-living bacterium. The enrichment of transcription factors in the two subgroup-specific gene-sets might represent adaptations to a specific environmental conditions or microhabitats. Changing the regulation of transcription by duplication or by mutation of genes of transcription factors is a powerful way to influence whole networks of responses to environmental conditions. In dynamically changing environmental conditions expanded and diverged sets of transcription factors yield fitness advantages, as shown for the transcription factor B family in archeae (Turkarslan *et al.*, 2011).

3.4 Genome comparison between rOTUs and lOTUs Rhizobia

To obtain insights into the genetic determinants for the different lifestyles observed for Rhizobia classified as rOTUs and Rhizobia classified as lOTUs the genomes of representatives of both classes were sequenced. Since the lOTUs test set comprised only a single OTU this comparison rather reveals the adaptation of Rhizobium$_{13}$ to the 'habitat wood stick' than allowing inference of first principles of the observed lOTUs/rOTUs separation. Definition of subgroup (rOTUs/lOTUs) enriched gene family members resulted in a similar pattern as also observed for subgroups within the Rhizobium$_{129}$ exemplars alone (Figure 28). The main diverged genes encoded by lOTUs and rOTUs subgroup members were assigned to function in transcription, transport and oxidation-reduction processes. This points to an adaptation to two different habitats, a living plant root and metabolically inactive lignocellulose, by the aforementioned conserved mechanisms.

The number of lOTUs enriched gene family members was higher than the number of rOTUs specific genes, most likely because in this comparison the lOTUs fraction comprised only exemplars of a single OTU while the rOTUs fraction contained isolates of two OTUs. Genetic diversity within the analyzed lOTUs subgroup is therefore lower which increases the number of detected subgroup enriched gene families. The higher number of genes increases also the number of genes with an assigned function via GO terms which is why for the lOTUs enriched genes

significant enrichments of GO terms could be detected but not for the rOTUs enriched genes. Enriched GO terms were 'sequence-specific DNA binding transcription factor activity', 'ATP-binding cassette (ABC) transporter complex', 'ATP catabolic process' and 'peptide transport', hence GO terms that were also detected in the intra-OTU comparison of Rhizobium$_{129}$ and that point to adaptation to differing habitats. The enrichment of genes involved in peptide transport might point to the saprophytic lifestyle of Rhizobium$_{13}$. Saprophytes process dead and decaying organic matter by extracellular digestion of for example proteins. Secreted proteins of saprophytes therefore contain an array of degrading enzymes (Desvaux *et al.*, 2010). Peptide transporters are then needed to take up the resulting digestion products. Interestingly, peptide transporters have been also demonstrated to function in chemotaxis in sensing peptide attractants (Abouhamad *et al.*, 1991).

Rhizobium$_{13}$ of the lOTUs sub-community abundantly colonize wooden sticks and might be able to use cellulose as a carbon source. In a first step of cellulose degradation endoglucanases hydrolyze internal bonds at random positions of less ordered regions of cellulose, generating chain ends that can be further processed by a second class of enzymes, exoglucanases. The major product of this reaction is cellobiose which is then further hydrolyzed to glucose by cellobiase enzymes (Mba Medie *et al.*, 2012). The lOTUs specific gene subset contains genes encoding endoglucanases and a β-glucosidases/cellobiases, therefore for enzymes necessary for the first and the last step in cellulose degradation. However, no exoglucanases could be detected in the genome of Rhizobium$_{13}$ exemplars. It is unlikely that this gene could not be annotated because it lies in a genetic region that has not been assembled completely since the genomes of four Rhizobium$_{13}$ exemplars have been sequenced and none encodes an exoglucanase. If Rhizobium$_{13}$ is involved in cellulose degradation it would have to rely on other cellulose degradation microbes, like fungi, to catalyze this step. It has to be noted, that endoglucanases and β-glucosidases are also encoded for in the genome of Rhizobium$_{129}$, but the number of the encoded homologs is significantly lower than in the genomes of Rhizobium$_{13}$ exemplars. Gene duplication is a proposed mechanism for adaptation to a specific environment. The number of genes encoding for the same protein can increase the amount of the gene product (Schuster-Bockler *et al.*, 2010; Zhou *et al.*, 2011) which might lead to increased cellulose degradation in the case of Rhizobium$_{13}$.

Essential for inferring genetic determinants for the rOTUs or lOTUs lifestyle, however, will be to obtain additional genomes of other Rhizobia and non-Rhizobia that are members of the lOTUs community. The undertaken approach of targeted large scale cultivation guided by a 16S ribotyping survey is the first step towards deciphering the functional significance host-microbiota interactions. But large scale full genome sequencing will have to be followed by large scale metabolic phenotyping to validate predictions made on the basis of bacterial genomes. Based on genomic and metabolomic data first principles can by hypothesized and tested with selected members of the host-associated bacterial microbiota.

4 Material and Methods

4.1 Material

4.1.1 Bacterial strains

Rhizobium leguminosarum strain 8401 (Downie *et al.*, 1983) was kindly provided by Prof. Allan Downie, John Innes Centre, Norwich, England. *Rhizobium leguminosarum* strain 8401 lacks the symbiotic plasmid and is not able to induce nodulation in legumes. *Escherichia coli* DH5α used for control experiments was purchased from Invitrogen (Karlsruhe, Germany). Bacteria isolated in the course of this study are listed in Table 16 together with the plant genotype strains were isolated from and the type of natural soil plants were grown in for isolation of root-associated bacteria. Each individual strain was isolated from an independent root, and therefore each strain represents an independent colonization event.

Table 16: Bacterial strains and exemplars isolated from Arabidopsis roots in the course of this study

OTU	exemplar	Isolated from
129	**A**	Arabidopsis Shakdara grown in Cologne soil
	B	Arabidopsis Shakdara grown in Cologne soil
	C	Arabidopsis Shakdara grown in Cologne soil
	D	Arabidopsis Shakdara grown in Cologne soil
	E	Arabidopsis Shakdara grown in Cologne soil
	F	Arabidopsis Shakdara grown in Cologne soil
219	**A***	Arabidopsis Shakdara grown in Cologne soil
13	**A**	Arabidopsis Shakdara grown in Cologne soil
	B	Arabidopsis Shakdara grown in Cologne soil
	C*	Arabidopsis Shakdara grown in Cologne soil
	D	Arabidopsis Shakdara grown in Cologne soil

4.1.2 Plant material

Arabidopsis thaliana (here referred to as Arabidopsis) seeds from the accessions Landsberg *erecta* (L*er*), Shakdara (Sha) and Columbia-0 (Col-0) were kindly provided by Dr. Maarten Koornneef, Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne (Germany).

4.1.3 Vectors

$Rhizobium_{129}$ and $Rhizobium_{219}$ were transformed with an RFP expressing version of the pA16 plasmid (Lewin *et al.*, 1990), kindly provided by Prof. Allan Downie, John Innes Centre, Norwich,

England. The selectable marker conferred by the RFP pA16 plasmid is tetracycline. Rhizobium$_{13}$ was transformed with the pLMB426 plasmid, kindly provided by Philip Poole, John Innes Centre, Norwich, England (unpublished). The plasmid encodes for the fluorophore mCherry constitutively expressed under the Ptac promotor. The selectable marker conferred by the plasmid is gentamycin (Figure 29).

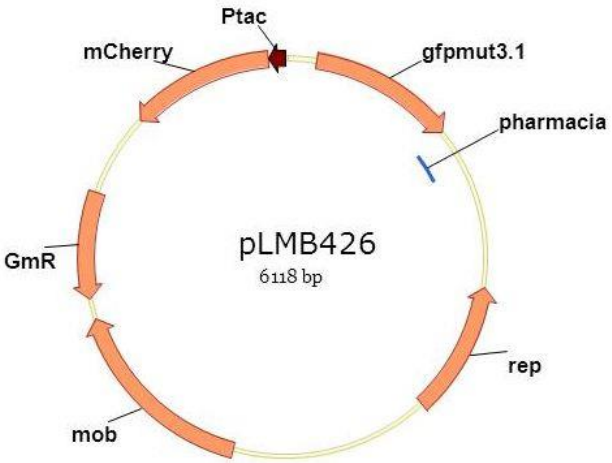


Figure 29: vector map of pLMB426

4.1.4 Antibiotics

Antibiotics used in this study are listed in Table 17. Antibiotics were prepared in 1,000 times concentrated stock solutions and stored at -20°C. Benzimidazole was added to growth media plates used for isolation of bacteria from Arabidopsis roots to prevent fungal growth.

Table 17: Antibiotics and fungicides used in this study

Antibiotic/fungicide	final concentration	source
Streptomycin	200/400 µg/ml	Sigma
Tetracycline	5 µg/ml	Sigma
Gentamycin	25 µg/ml	Roth
Benzimidazole	5 mg/ml	Sigma

4.1.5 Growth media, solutions, and buffers

Media were sterilized by autoclaving at 121 °C for 20 min. Antibiotics and other heat labile compounds were added after media have been cooled down to approximately 50°C. Heat labile compounds were sterilized using filter sterilization units prior to addition to the media. Media used for cultivation of root-associated bacteria are listed in Table 18, fahraeus plant growth medium composition is listed in Table 19. Buffers and solutions used in this study are described

within each method. If not otherwise stated, buffers were prepared in ddH_2O and aqueous solutions were sterilized by autoclaving at 121 °C for 20 min.

Table 18: Growth media used for cultivation of root-associated bacteria

Growth medium	**purpose/cultivation of**	**reference/supplier**
1/10 strength LB	basal medium	(Bertani, 1951)
1/10 strength 869	basal medium	(Mergeay *et al.*, 1978)
1/2 strength Nutrient agar (NA)	basal medium	Merck
Standard I agar	basal medium	Merck
Tryptone yeast (TY)	basal medium, rhizobia	(Beringer, 1974)
Tap water yeast extract (TWYE)	Actinobacteria	(Coombs & Franco, 2003)
Flour 1	Actinobacteria	(Coombs & Franco, 2003)
Flour 2	Actinobacteria	(Coombs & Franco, 2003)
Yeast extract casein hydrolysate (YEDC)	Actinobacteria	(Coombs & Franco, 2003)
Czapek doc agar	soil bacteria, fungi	Merck
R2A agar	slow growing bacteria	(Reasoner & Geldreich, 1985)
Yeast extract mannitol (YEM)	rhizobia	(Rao, 1995)

Table 19: Composition of fahraeus plant growth medium

Component	Concentration	
$CaCl_2$, x 2 H2O	0.1	g/L
$MgSO_4.7H_2O$,	0.1	g/L
KH_2PO_4,	0.1	g/L
Na_2HPO_4. x $12H_2O$	0.15	g/L
Fe-citrate	5	mg/L
H_3BO_3	2.86	mg/L
$MnSO_4$ x 4 H2O	2.03	mg/L
$ZnSO_4$ x 7 H2O	0.22	mg/L
$CuSO_4$ x 5 H2O	0.8	mg/L
H_2MoO_4	0.8	mg/L

4.1.6 Oligonucleotides

Oligonucleotides used for PCR amplification are listed in Table 5 and were purchased from Sigma (Hamburg, Germany) or Invitrogen (Karlsruhe, Germany).

Table 20: Oligonucleotides used for PCR amplifications

name	sequence (5' - 3')	reference
799F	AAC MGG ATT AGA TAC CCK G	Chelius & Triplett, 2001
799F2	AAC MGG ATT AGA TAC CCG G	Bulgarelli et al, 2012
1401R	CGG TGT GTA CAA GGC CC	Buchholz-Cleven et al, 1997

4.1.7 Enzymes

DNA free DFS Tag polymerase was purchased from Bioron (Ludwigshafen, Germany).

4.2 Methods

4.2.1 Seed surface sterilization

Seeds were surface-sterilized by a washing series in ethanol (2 x 70% and 96%). The ethanol was removed and seeds were dried in a sterile hood.

4.2.2 Plant growth conditions

The seeds were sown on natural soil at a density of nine plants per pot. Weeds germinating from seeds naturally present in the experimental soils were manually removed upon identification. Pots were randomly distributed in trays and placed in a greenhouse with the following conditions: 16 hours light (day) and 8 hours dark (night), 22°C during the day and 18°C during the night at a relative humidity of 70%. Plants were watered periodically every other day to maintain for each soil type a maximum water retention capacity between 50-70%. Plants were maintained in the greenhouse until early flowering stage.

4.2.3 Preparation of root material of Arabidopsis grown in natural soil for isolation of bacteria

To standardize the sampling and because in natural soils (not amended by sand) fragile root tips can be lost during preparation, we have operationally chosen to dissect 3 cm root segments just below the rosette with a sterile scalpel. Root sections of all 9 plants derived from a single pot were pooled into a 15 ml Falcon tube containing 2.5 ml sterile Silwet L-77 amended PBS buffer

(PBS-S; 130 mM NaCl, 7 mM Na_2HPO_4, 3 mM NaH_2PO_4, pH 7.0, 0.02% Silwet L-77) and washed on a shaking platform for 20 min at 180 rpm (Movie S1). Roots were transferred to a new 15 ml Falcon tube containing 2.5 ml PBS-S. After a second washing step (20 min, 180 rpm), roots were transferred to another 15 ml Falcon tube (2.5 ml PBS-S) and were sonicated for 10 cycles consisting of 30 second pulses at 160 W and breaks of 30 seconds (Bioruptor Next Gen UCD-300, diagenode, Liège, Belgium). After removing from the PBS-S buffer the washed and sonicated roots were transferred into 700 µl PBS buffer without addition of Silwet L-77 in 2 ml screw cap tubes. Roots were then homogenized utilizing sterile 2 mm metal beads and a Precellys®24 tissue lyzer (Bertin Technologies, Montigny-le-Bretonneux, France) at 5,000 rotations per minute for three cycles of 30 seconds. The homogenate was centrifuged for 5 minutes at 1,000 x g to pellet large cell debris and the supernatant was transferred into a new tube without disturbing the pellet. The debris-free homogenate was then diluted 1:10, 1:100 and 1:000 and 100 µl of each dilution was plated on a diverse range of minimal growth media (Table 18). The plates were incubated at 28°C for 4-7 days until a sufficient number of bacterial colonies could be observed.

4.2.4 PCR amplification and sequencing of the 16S rRNA gene of isolated root-associated bacteria

Wells of 96 deep well plates were filled with 400 µl of the same growth media used for initial plating and each well was inoculated with one of the bacterial colonies growing on the agar plates used for initial plating. Deep well plates were then sealed with adhesive foil and incubated for 4 days at 28°C on a rotating incubator at 180 rounds per minute. After incubation 100 µl of the liquid growth medium was transferred into a 96 well PCR plate and incubated at 99°C for 10 minutes to lyse the bacteria and extract their DNA. Cell debris was removed by centrifuging the 96well PCR plate for 10 minutes at 2,000 x g in a swing out rotor in a Heraeus multifuge 3 S-R and transferring the supernatant into a new 96 well PCR plate. 2 µl of the debris free bacterial lysate was used in subsequent PCRs to amplify the bacterial 16S rRNA gene using 799F (Chelius & Triplett, 2001) as forward and 1401R (BuchholzCleven *et al.*, 1997) as reverse primer (Table 5). PCRs were performed for each sample in a volume of 25 µl. PCR components include 1.25 U DFS-Taq DNA Polymerase (Bioron, Ludwigshafen, Germany), 1x complete reaction buffer, 2 mmol of $MgCl_2$, 200 µmol of dNTPs and 400 nmol of each primer. PCR reactions were mixed in a laminar flow hood and amplified using the following conditions: initial denaturation at 95°C for 3 minutes followed by 35 cycles of denaturation for 30 seconds at 95°C, annealing at 55°C for 30 seconds and elongation at 72°C for 30 seconds and by a final elongation step at 72°C for 5 minutes. PCR reactions were checked for positive amplification on a 1 % agarose gel stained with ethidium

bromide. PCR amplicons were then Sanger sequenced at the Max Planck Genome Centre Cologne utilizing the 1401R primer.

4.2.5 Identification of isolated bacteria and matching with the bacterial database obtained by culture-independent 16S ribotyping

The OTU representative sequences obtained by culture-independent 16S ribotyping of the root-associated bacteria (see chapter I) were used to create a local BLAST database using the Bioedit software package (http://www.mbio.ncsu.edu/bioedit/bioedit.html). 16S rRNA gene sequences obtained from the isolated bacteria were then blasted against this local database and colonies were defined as matching an OTU if the sequence identity was ≥ 97% over 200 bp of the OTU representative sequence. In case a colony 16S sequence matched several OTU representative sequences the colony was assigned to the most abundant OTU. Furthermore, colony 16S rRNA gene sequences were classified using the classifier algorithm version 2.5 of the ribosomal database project (http://rdp.cme.msu.edu/classifier/classifier.jsp) (Wang *et al.*, 2007).

4.2.6 Selection of Streptomycin resistant derivatives of bacterial isolates for generation of clonal strains

Wild-type Rhizobia strains isolated from Arabidopsis roots identified by sequencing of their 16S rRNA gene (Table 16) were used to inoculate 5 ml of TY medium and incubated at 28°C for 2 days. Cells were then concentrated by centrifuging 4 ml of the culture at 3,500 x g for 3 min and resuspending the pellet in 200 µl of 10 mM $MgCl_2$. Resuspended bacteria were diluted 1:10, 1:100 and 1:1000 and 100 µl of each dilution were plated on TY agar plates containing 20 µg/ml streptomycin (TY strep^{20}) to select for derivatives carrying a natural streptomycin resistance. Growing colonies were then transferred to a second TY agar plate, containing 400 µg/ml streptomycin (TY strep^{400}), to exclude mutations that could have pleiotropic effects, like cell wall mutations, and select for mutations either in the 16S rRNA gene or in the gene coding the 30S ribosomal subunit protein 12S (rpsL). Both mutations prevent binding of streptomycin to the translation machinery and confer high level streptomycin resistance without pleiotropic effects. Transfer of colonies was done by placing a Protran BA 85 membrane (Schleicher und Schnell, Dassel, Germany) on colonies growing on the TY strep^{20} agar plate for a few seconds and transferring the membrane to the TY strep^{400} agar plates. Single colonies growing on the TY strep^{400} agar plates and the wild-type isolate were streaked out on YEM minimal medium to check for obvious growth phenotypes induced by the selection procedure. One colony without obvious growth deficiencies per strain was chosen for further analysis.

4.2.7 Transformation of Rhizobia via triparental mating

RFP or GFP expressing plasmids (see section 8.1.3) were transferred from the donor *E. coli* strains to the recipient streptomycin resistant Rhizobium strain by using the F-plasmid carrying *E. coli* pRK2013 as a helper strain (Figurski & Helinski, 1979). By mixing all 3 strains with an inoculation loop on a TY agar plate without selective pressure plasmids can be transferred from the donor strain to the recipient strain by the F-plasmid of the helper strain. After incubation of the agar plate at 28°C for 1 day without any selective pressure bacteria were transferred to a second TY agar using a membrane as described in section 8.2.6. This plate contained the antibiotic the RFP or GFP plasmid confers resistance to, respectively, to select for positive transformants and 200 µg/ml streptomycin to select against to *E. coli* donor and helper strains. The selection procedure was successful if none of the three strains could grow alone on the selective agar plate. Transformants were streaked out to obtain single colonies which were used to inoculate liquid cultures to be stored as glycerol stocks at -80°C and used for further experiments.

4.2.8 Agarose gel electrophoresis of DNA

DNA fragments were separated by agarose gel electrophoresis in gels consisting of 1 - 3 % (w/v) agarose (Bio-Budget Technologies, Krefeld, Germany) supplied with ethidium bromide solution (1:40000) in TAE buffer (400 mM Tris; 10 mM EDTA; 200 mM acetic acid; pH 8.5).

4.2.9 Correlation of OD_{600} with number of colony forming units

$Rhizobium_{129}$ exemplar E was grown over night in 5 ml of TY medium. Cultures were diluted 1:1, 1:2, 1:4, 1:8, 1:16 and 1:32. Absorbance at 600 nm was measured twice for each dilution using a Biophotometer plus (Eppendorf, Hamburg, Germany). For each culture dilution, two dilution series (undiluted, 10^{-1} – 10^{-8}) were plated on TY agar plates containing the appropriate antibiotics. 15 µl of each dilution were plated and incubated for 2 days at 28°C. The dilution step at which single colonies could be determined was used for counting CFUs. Then the total number of bacterial was calculated and correlated against the absorbance at 600 nm. In all further experiments applied bacterial densities were based on this correlation. Since all isolated Rhizobia strains showed a high sequence similarity at the whole genome level, the same OD_{600}-CFU correlation was used for all isolated Rhizobium strains.

4.2.10 Re-colonization of Arabidopsis roots

4.2.10.1 Surface sterilization

Seeds were surface-sterilized by a washing series in ethanol (2 x 70% and 96%). The ethanol was removed and seeds were dried in a sterile hood.

4.2.10.2 Filter inoculation

Rhizobium strains were grown over night in 5 ml of TY medium supplemented with the appropriate antibiotics. Cells were washed three times by repetitive centrifugation at 3,000 x g for 3 min and resuspension in 10 mM $MgCl_2$ to avoid carryover of TY medium. 10 ml of inoculum was prepared in 10 mM $MgCl_2$ and the inoculum density, e.g. 10^5 cells/ml, was adjusted according to the OD_{600}-CFU correlation obtained in section 8.2.9. The inoculum was poured into a sterile 120 mm x 120 mm petri dish and envelope strips grade 0858 were (Whatman, Einschlagstreifen/Envelope strips grade 0858. Art. No 10 334 365. 110 x 580 mm) were soaked in the inoculum and placed on FM medium agar plates. Afterwards 5-10 Arabidopsis Col-0 seeds were sawn out on the envelope strips. Plants were then grown in a light chamber at a humidity of 65%, a day length of 10 hours and 23°C at day and 22°C at night.

4.2.10.3 Embedding bacteria in agar

Rhizobium strains were grown over night in 5 ml of TY medium supplemented with the appropriate antibiotics. Cells were washed three times by repetitive centrifugation at 3,000 x g for 3 min and resuspension in 10 mM $MgCl_2$ to avoid carryover of TY medium. The cell density of the inoculum was prepared in 10 mM $MgCl_2$ according to the OD_{600}-CFU correlation obtained in section 8.2.9. 1 ml of inoculum was mixed with 49 ml of hand-warm FM agar medium in a sterile 50 ml falcon tube resulting in the desired final cell density and poured into 120 x 120 mm petri dishes. After the agar had solidified Arabidopsis Col-0 seeds were sawn out on the agar. Plants were then grown in a light chamber at a humidity of 65%, a day length of 10 hours and 23°C at day and 22°C at night.

4.2.10.4 Visualization of bacteria on the Arabidopsis rhizoplane

Arabidopsis Col-0 was inoculated with different Rhizobium strains as described in sections 8.2.10.2 or 8.2.10.3. Visualization was performed using a CLSM 510 Meta microscope (Carl Zeiss, Oberkochen, Germany). RFP was excited with a 488 nm laser line and detected with a BP 500-530 nm filter.

4.2.10.5 Plant growth promotion assays

Arabidopsis were inoculated with the respective Rhizobium strain as described in sections 8.2.10.2 or 8.2.10.3. Additionally as controls plates were prepared where instead of 1 ml of Rhizobium inoculum, 1 ml of MgCl2, or 1 ml of heat-killed bacteria at a comparable cell density, or 1 ml of an *E. coli* culture at a comparable cell density was mixed with 49 ml of FM agar medium, respectively, before pouring the agar plates. Per strain and inoculation density or control treatment, respectively, three FM agar plates à 10 seeds were prepared. Plants were then grown in a light chamber at a humidity of 65%, a day length of 10 hours and 23°C at day and 22°C at night. 21 days after sawing shoots of all germinated plants of one plate were pooled and shoot fresh weight was measured. Average plant shoot fresh weight was then calculated as the pool fresh weight divided by the number of plants in the pool. The average fresh weight of the $MgCl_2$ control was defined as 1 and fresh weight of all other treatments are expressed relative to the $MgCl_2$ control.

4.2.11 Whole genome analysis of bacterial isolates

4.2.11.1 Extraction of genomic DNA of bacteria

Strains were grown over night in 5 ml of TY medium. 1 to 2 ml of overnight culture were used for extraction of genomic DNA. Genomic DNA was extracted using a modified protocol of the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Cells were harvested by centrifugation at 3,000 x g for 3.5 min and washed by resuspending in Tris buffer (10 mM Tris-HCl, pH 8.5). Cells were then again centrifuged and resuspended in 270 µl of ATL buffer and 30 µl of Proteinase K supplied with the DNA extraction kit. Lysis of the cells was carried out at 56°C for 1 hour. Proteinase K was inactivated by the addition of EDTA to a final concentration of 10 mM and incubation at 75°C for 10 min. Afterwards RNase was added to a final concentration of 100 ng/ml and RNA was digested at 37°C for 30 min. DNA extraction was then carried on as suggested by the kit's manufacturer. DNA concentrations were determined by Nanodrop (ThermoScientific, Wilmington, USA) and DNA was checked for RNA contamination on an agarose gel.

4.2.11.2 Whole genome sequencing

Whole genome shotgun sequencing of bacterial strains was carried out by the Max Planck Genome Centre Cologne. Draft sequences were produced by Illumina HiSeq2000 TruSeq chemistry for 100 bp single or 100 bp paired end read sequencing following the manufacturer's recommendations after fragmentation of the genomic DNA to fragments of 300 bp using a

COVARIS S2. The genome of Rhizobium$_{129}$ exemplar E was additionally sequenced using Roche's GS FLX+ Titanium system following the manufacturer's instructions.

4.2.11.3 Genome assembly

Draft genomes were assembled using the CLC workbench software package (CLC bio, Aarhus, Denmark). To obtain high quality sequences for the assembly, raw sequences were trimmed using the software's "trim sequences" function applying a quality score of 0.03 and without allowing any ambiguous nucleotides. Only reads that were longer than 50 bp after trimming were used for the assembly. Draft contigs were then assembled using the "de novo assembly" function applying default settings. Minimum contig size was set to 500 bp.

4.2.11.4 Genome analysis and annotation

Draft genomes uploaded to the "rapid annotation using subsystem technology (RAST)" server to predict coding sequences. (Aziz *et al.*, 2008). rRNA gene sequences were extracted from draft genomes using the RNammer 1.2 server (Lagesen *et al.*, 2007). Coding sequences predicted by the RAST server of representative exemplars of identified subgroups of the Arabidopsis associated Rhizobia were additionally annotated with the Blast2go software package (Conesa *et al.*, 2005).

4.2.11.5 Genomic comparison

The estimation of core- and pan-genomes of the isolated Rhizobium strains and the identification of the genes being specific for a given subgroup of isolates were conducted by Emiel Ver Loren van Themaat, MPIPZ, Cologne. The size of the core-genome, i.e. the number of genes conserved in all sequenced strains exemplars and the size of the pan-genome, i.e. the estimated total size of the gene repertoire of the tested strains, were estimated using a binomial mixture model according to Snipen et al., 2009. Briefly, in a first step an all-against-all BLASTing is performed for all predicted protein amino acid sequences of a given genome against all predicted protein amino acid sequences in all other genomes in the dataset. Only alignments with at least 50% identity along at least 50% of both sequences are considered. Two sequences are defined to belong to the same gene family if both their reciprocal alignments fulfill this 50-50-cutoff rule. This results in a gene family matrix that gives the number of genes per analyzed genome belonging to a defined gene family.

Gene families are then analyzed if they belong to the core-, or the dispensable genome, or if they show a subgroup-specific (e.g. lOTUs or rOTUs specific) behavior. Assuming a Poisson distribution

of the gene family counts a linear model is fitted to the gene family counts using group-specificity (lOTUs/rOTUs/unique for one genome/present in all genomes) as underlying assumption. A chi-square test is used to determine how likely genomes of compared subgroups have the same amount of members for each gene family. Gene families that are shown in the chi-square test to have the same amount of gene family members at likelihood of < 0.05 are defined as significant for the tested group specificity.

Genes that were assigned to these gene families by the 50-50-cutoff rule are then also defined as being specific for the tested condition (e.g. lOTUs or rOTUs specific).

4.2.11.6 Analysis of identified subgroup specific genes

Subgroup specific genes were analyzed in representative exemplars of the respective subgroups using the Blast2Go software package. GO term assignment pie charts for subgroup specific genes were plotted by creating multilevel pie charts at a term filter value of 5 from combined graphs of biological processes that take into account all GO term nodes that contain at least 4% of all sequences in the dataset. Analysis of GO term enrichments were conducted within representative exemplars for the respective subgroup by testing subgroup specific genes against the rest of the respective genome by Fisher's exact test. GO term enrichments with a FDR < 0.05 were defined as significantly enriched.

4.2.11.7 Whole genome alignment to determine overall genome similarity

Whole genome alignments were conducted by Barbara Kracher, MPIPZ Cologne. Draft genomes of all six sequenced $Rhizobium_{129}$ exemplars were aligned in all possible pairwise combinations using the nucmer function of the MUMmer software package with the default alignment settings (Delcher *et al.*, 2003). For both genomes in each pairwise comparison, the length of all aligned sequence segments was then summed up and this total alignment length was subsequently divided by the total assembly size of the respective genome, giving the fractions of each genome that could be aligned to the respective other genome. For the calculation of the total alignment length, sequence segments that could be aligned to multiple locations in the other genome were considered only once. Additionally, for each comparison the average sequence identity over all aligned sequence segments was calculated.

Abbreviations

%	percent
µl	mikroliter
ABC transporter	ATP binding cassette transporter
ACC	1-aminocyclopropane-1-carboxylate
AHL	*N*-acylhomoserine
ANI	average nucleotide identity
ARISA	Automated ribosomal intergenic spacer analysis
ATP	adenosintriphosphat
CARD-FISH	catalyzed reporter deposition - in situ fluorescence hybridization
CAS	cologne agricultural soil
CFU	colony forming unit
DAPG	2,4-diacetylphloroglucinol
DNA	Deoxyribonucleic acid
E. coli	Escherichia coli
FAO	soil classification according to the Food and Agriculture Organization of the United Nations
FDR	false discovery rate
GFP	green fluorescent protein
h	hour
HCN	hydrogen cyanide
HGT	horizontal gene transfer
HRP	horseradish peroxidase
IAA	indol-acetic-acid
IPCD	indole-3-pyruvate decarboxylase
ISR	induced systemic resistance
JA	jasmonic acid
Ler	Arabidopsis thaliana ecotype Landsberg erecta
lOTU	lignocellulose-associated OTU
MAMP	microbe associated molecular pattern
MeJA	methyl jasmonic acid
min	minute
ml	milliliter
MPIPZ	Max Planck Institute for Plant Breeding Research
NB-LRR	nucleotide binding leucine rich repeat
OD	optical density
OTU	operational taxonomic unit
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PGPR	Plant growth promoting rhizobacteria
pH	negative decimal logarithm of H+ concentration
pv.	pathovar
QS	quorum sensing
RA	relative abundance
RAST	rapid annotation using subsystems technology
RFP	red fluorescence protein

RNA	ribonucleic acid
ROS	reactive oxygen species
rOTU	root-associated OTU
rRNA	ribosomal ribonucleic acid
RT	room temperature
SA	salicylic acid
SD	standard deviation
SE	standard error
sec	second
SEM	Scanning electron microscopy
Sha	Arabidopsis thaliana ecotype Shakdara
Taq	Thermophilus aquaticus
T-RFLP	terminal restriction length polymorphism
TRIS	tris-(hydroxymethyl)-aminomethan
U	unit
UV	Ultraviolet
V	Volt
wOTU	wood-associated OTU
WT	wild type

References

Abouhamad WN, Manson M, Gibson MM, Higgins CF. 1991. Peptide transport and chemotaxis in Escherichia coli and Salmonella typhimurium: characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein. *Mol Microbiol* **5**(5): 1035-1047.

Ahmad F, Ahmad I, Khan MS. 2008. Screening of free-living rhizospheric bacteria for their multiple plant growth promoting activities. *Microbiological Research* **163**(2): 173-181.

Alonso C, Pernthaler J. 2006. Concentration-dependent patterns of leucine incorporation by coastal picoplankton. *Appl Environ Microbiol* **72**(3): 2141-2147.

Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**(6): 1919-1925.

Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**(1): 143-169.

Andersson DI, Hughes D. 2009. Gene amplification and adaptive evolution in bacteria. *Annu Rev Genet* **43**: 167-195.

Antoun H, Beauchamp CJ, Goussard N, Chabot R, Lalande R. 1998. Potential of Rhizobium and Bradyrhizobium species as plant growth promoting rhizobacteria on non-legumes: Effect on radishes (Raphanus sativus L.). *Plant and Soil* **204**(1): 57-67.

Appleby CA. 1984. Leghemoglobin and rhizobium respiration. *Annual Review of Plant Physiology and Plant Molecular Biology* **35**: 443-478.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. *Bmc Genomics* **9**: 75.

Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. 2005. Host-bacterial mutualism in the human intestine. *Science* **307**(5717): 1915-1920.

Badri DV, Chaparro JM, Manter DK, Martinoia E, Vivanco JM. 2012. Influence of ATP-Binding Cassette Transporters in Root Exudation of Phytoalexins, Signals, and in Disease Resistance. *Front Plant Sci* **3**: 149.

Badri DV, Quintana N, El Kassis EG, Kim HK, Choi YH, Sugiyama A, Verpoorte R, Martinoia E, Manter DK, Vivanco JM. 2009. An ABC transporter mutation alters root exudation of phytochemicals that provoke an overhaul of natural soil microbiota. *Plant Physiol* **151**(4): 2006-2017.

Bais HP, Walker TS, Schweizer HP, Vivanco JA. 2002. Root specific elicitation and antimicrobial activity of rosmarinic acid in hairy root cultures of Ocimum basilicum. *Plant Physiology and Biochemistry* **40**(11): 983-995.

Bak S, Olsen CE, Petersen BL, Moller BL, Halkier BA. 1999. Metabolic engineering of p-hydroxybenzylglucosinolate in Arabidopsis by expression of the cyanogenic CYP79A1 from Sorghum bicolor. *Plant J* **20**(6): 663-671.

Balder R, Hassel J, Lipski S, Lafontaine ER. 2007. Moraxella catarrhalis strain O35E expresses two filamentous hemagglutinin-like proteins that mediate adherence to human epithelial cells. *Infect Immun* **75**(6): 2765-2775.

Banerjee MR, L. Y, K. VJ. 2006. plant-growth promoting rhizobacteria as biofertilizers and biopesticides. *Handbook of microbial fertilizers*(Haworth Press, New York).

Barac T, Taghavi S, Borremans B, Provoost A, Oeyen L, Colpaert JV, Vangronsveld J, van der Lelie D. 2004. Engineered endophytic bacteria improve phytoremediation of water-soluble, volatile, organic pollutants. *Nat Biotechnol* **22**(5): 583-588.

Barret M, Morrissey JP, O'Gara F. 2011. Functional genomics analysis of plant growth-promoting rhizobacterial traits involved in rhizosphere competence. *Biology and Fertility of Soils* **47**(7): 729-743.

Bartsev AV, Deakin WJ, Boukli NM, McAlvin CB, Stacey G, Malnoe P, Broughton WJ, Staehelin C. 2004. NopL, an effector protein of Rhizobium sp. NGR234, thwarts activation of plant defense reactions. *Plant Physiol* **134**(2): 871-879.

Basilio A, Gonzalez I, Vicente MF, Gorrochategui J, Cabello A, Gonzalez A, Genilloud O. 2003. Patterns of antimicrobial activities from soil actinomycetes isolated under different conditions of pH and salinity. *J Appl Microbiol* **95**(4): 814-823.

Benson AK, Kelly SA, Legge R, Ma FR, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D, Hua KJ, Kachman SD, Moriyama EN, Walter J, Peterson DA, Pomp D. 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**(44): 18933-18938.

Berendsen RL, Pieterse CM, Bakker PA. 2012. The rhizosphere microbiome and plant health. *Trends Plant Sci*.

Berg G, Smalla K. 2009. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *Fems Microbiology Ecology* **68**(1): 1-13.

Beringer JE. 1974. R factor transfer in rhizobium leguminosarum. *Journal of General Microbiology* **84**(SEP): 188-198.

Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol* **62**(3): 293-300.

Bezemer TM, Lawson CS, Hedlund K, Edwards AR, Brook AJ, Igual JM, Mortimer SR, Van der Putten WH. 2006. Plant species and functional group effects on abiotic and microbial soil properties and plant-soil feedback responses in two grasslands. *Journal of Ecology* **94**(5): 893-904.

Bhattacharjee RB, Jourand P, Chaintreuil C, Dreyfus B, Singh A, Mukhopadhyay SN. 2012. Indole acetic acid and ACC deaminase-producing Rhizobium leguminosarum bv. trifolii SN10 promote rice growth, and in the process undergo colonization and chemotaxis. *Biology and Fertility of Soils* **48**(2): 173-182.

Bhattacharjee RB, Singh A, Mukhopadhyay SN. 2008. Use of nitrogen-fixing bacteria as biofertiliser for non-legumes: prospects and challenges. *Applied Microbiology and Biotechnology* **80**(2): 199-209.

Biswas JC, Ladha JK, Dazzo FB. 2000. Rhizobia inoculation improves nutrient uptake and growth of lowland rice. *Soil Science Society of America Journal* **64**(5): 1644-1650.

Bjornsson L, Hugenholtz P, Tyson GW, Blackall LL. 2002. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology* **148**(Pt 8): 2309-2318.

Blazejak A, Erseus C, Amann R, Dubilier N. 2005a. Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Applied and Environmental Microbiology* **71**(3): 1553-1561.

Blazejak A, Erseus C, Amann R, Dubilier N. 2005b. Coexistence of bacterial sulfide oxidizers, sulfate reducers, and spirochetes in a gutless worm (Oligochaeta) from the Peru margin. *Appl Environ Microbiol* **71**(3): 1553-1561.

Boller T, Felix G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol* **60**: 379-406.

Bonsall RF, Weller DM, Thomashow LS. 1997. Quantification of 2,4-diacetylphloroglucinol produced by fluorescent Pseudomonas spp. in vitro and in the rhizosphere of wheat. *Applied and Environmental Microbiology* **63**(3): 951-955.

Bouffaud ML, Kyselkova M, Gouesnard B, Grundmann G, Muller D, Moenne-Loccoz Y. 2012. Is diversification history of maize influencing selection of soil bacteria by roots? *Mol Ecol* **21**(1): 195-206.

Bowling SA, Guo A, Cao H, Gordon AS, Klessig DF, Dong XI. 1994. A mutation in arabidopsis that leads to constitutive expression of systemic acquired-resistance. *Plant Cell* **6**(12): 1845-1857.

Bressan M, Roncato MA, Bellvert F, Comte G, Haichar FZ, Achouak W, Berge O. 2009. Exogenous glucosinolate produced by Arabidopsis thaliana has an impact on microbes in the rhizosphere and plant roots. *ISME J* **3**(11): 1243-1257.

Brigham LA, Michaels PJ, Flores HE. 1999. Cell-specific production and antimicrobial activity of naphthoquinones in roots of lithospermum erythrorhizon. *Plant Physiol* **119**(2): 417-428.

Broek AV, Lambrecht M, Vanderleyden J. 1998. Bacterial chemotactic motility is important for the initiation of wheat root colonization by Azospirillum brasilense. *Microbiology (Reading)* **144**(9): 2599-2606.

Brühl GW. 1987. Soilborne plant pathogens. *Macmillian NY*.

Buchan A, Crombie B, Alexandre GM. 2010. Temporal dynamics and genetic diversity of chemotactic-competent microbial populations in the rhizosphere. *Environ Microbiol* **12**(12): 3171-3184.

BuchholzCleven BEE, Rattunde B, Straub KL. 1997. Screening for genetic diversity of isolates of anaerobic Fe(II)-oxidizing bacteria using DGGE and whole-cell hybridization. *Systematic and Applied Microbiology* **20**(2): 301-309.

Buddrus-Schiemann K, Schmid M, Schreiner K, Welzl G, Hartmann A. 2010. Root Colonization by Pseudomonas sp DSMZ 13134 and Impact on the Indigenous Rhizosphere Bacterial Community of Barley. *Microbial Ecology* **60**(2): 381-393.

Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, Peplies J, Gloeckner FO, Amann R, Eickhorst T, Schulze-Lefert P. 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**(7409): 91-95.

Burse A, Weingart H, Ullrich MS. 2004. The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, Erwinia amylovora. *Mol Plant Microbe Interact* **17**(1): 43-54.

Buschart A, Sachs S, Chen X, Herglotz J, Krause A, Reinhold-Hurek B. 2012. Flagella Mediate Endophytic Competence Rather Than Act as MAMPS in Rice-Azoarcus sp Strain BH72 Interactions. *Molecular Plant-Microbe Interactions* **25**(2): 191-199.

Buttner D, Bonas U. 2010. Regulation and secretion of Xanthomonas virulence factors. *FEMS Microbiol Rev* **34**(2): 107-133.

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D. 2011. Whole-genome sequencing of multiple Arabidopsis thaliana populations. *Nat Genet* **43**(10): 956-963.

Cases I, de Lorenzo V, Ouzounis CA. 2003. Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol* **11**(6): 248-253.

Chaintreuil C, Giraud E, Prin Y, Lorquin J, Ba A, Gillis M, de Lajudie P, Dreyfus B. 2000. Photosynthetic bradyrhizobia are natural endophytes of the African wild rice Oryza breviligulata. *Appl Environ Microbiol* **66**(12): 5437-5447.

Chandra S, Choure K, Dubey RC, Maheshwari DK. 2007. Rhizosphere competent Mesorhizobium loti MP6 induces root hair curling, inhibits Sclerotinia sclerotiorum and enhances growth of Indian mustard (Brassica campestris). *Brazilian Journal of Microbiology* **38**(1): 124-130.

Chelius MK, Triplett EW. 2001. The Diversity of Archaea and Bacteria in Association with the Roots of Zea mays L. *Microb Ecol* **41**(3): 252-263.

Chi F, Shen SH, Cheng HP, Jing YX, Yanni YG, Dazzo FB. 2005. Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology. *Appl Environ Microbiol* **71**(11): 7271-7278.

Coenye T, Vandamme P. 2003. Diversity and significance of Burkholderia species occupying diverse ecological niches. *Environ Microbiol* **5**(9): 719-729.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37**(Database issue): D141-145.

Compant S, Clement C, Sessitsch A. 2010. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology & Biochemistry* **42**(5): 669-678.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18): 3674-3676.

Conn VM, Franco CM. 2004. Analysis of the endophytic actinobacterial population in the roots of wheat (Triticum aestivum L.) by terminal restriction fragment length polymorphism and sequencing of 16S rRNA clones. *Appl Environ Microbiol* **70**(3): 1787-1794.

Cooley MB, Miller WG, Mandrell RE. 2003. Colonization of Arabidopsis thaliana with Salmonella enterica and enterohemorrhagic Escherichia coli O157:H7 and competition by Enterobacter asburiae. *Appl Environ Microbiol* **69**(8): 4915-4926.

Coombs JT, Franco CM. 2003. Isolation and identification of actinobacteria from surface-sterilized wheat roots. *Appl Environ Microbiol* **69**(9): 5603-5608.

Costa R, Gotz M, Mrotzek N, Lottmann J, Berg G, Smalla K. 2006. Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *FEMS Microbiol Ecol* **56**(2): 236-249.

Dakora FD, Phillips DA. 2002. Root exudates as mediators of mineral acquisition in low-nutrient environments. *Plant and Soil* **245**(1): 35-47.

Dawid W. 2000. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol Rev* **24**(4): 403-427.

de Weert S, Vermeiren H, Mulders IH, Kuiper I, Hendrickx N, Bloemberg GV, Vanderleyden J, De Mot R, Lugtenberg BJ. 2002. Flagella-driven chemotaxis towards exudate components is an important

trait for tomato root colonization by Pseudomonas fluorescens. *Mol Plant Microbe Interact* **15**(11): 1173-1180.

De Weger LA, Van Der Vlugt CIM, Wijfjes AHM, Bakker PAHM, Schippers B, Lugtenberg B. 1987. Flagella of a plant-growth-stimulating pseudomonas-fluorescens strain are required for colonization of potato roots. *Journal of Bacteriology* **169**(6): 2769-2773.

DeAngelis KM, Brodie EL, DeSantis TZ, Andersen GL, Lindow SE, Firestone MK. 2009. Selective progressive response of soil microbial community to wild oat roots. *ISME J* **3**(2): 168-178.

Delcher AL, Phillippy AM, Carlton J, Salzberg SL. 2002. Fast Algorithms for Large-scale Genome Alignment and Comparision. *Nucleic Acids Res* **30**(11): 2478-2483.

Delcher AL, Salzberg SL, Phillippy AM. 2003. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* **Chapter 10**: Unit 10 13.

Delvasto P, Valverde A, Ballester A, Munoz JA, Gonzalez F, Blazquez ML, Igual JM, Garcia-Balboa C. 2008. Diversity and activity of phosphate bioleaching bacteria from a high-phosphorus iron ore. *Hydrometallurgy* **92**(3-4): 124-129.

Denarie J, Debelle F, Prome JC. 1996. Rhizobium lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu Rev Biochem* **65**: 503-535.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**(7): 5069-5072.

Desvaux M, Dumas E, Chafsey I, Chambon C, Hebraud M. 2010. Comprehensive appraisal of the extracellular proteins from a monoderm bacterium: theoretical and empirical exoproteomes of Listeria monocytogenes EGD-e by secretomics. *J Proteome Res* **9**(10): 5076-5092.

Dimkpa C, Weinand T, Asch F. 2009. Plant-rhizobacteria interactions alleviate abiotic stress conditions. *Plant Cell Environ* **32**(12): 1682-1694.

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B. 1993. Cellular organisation of the Arabidopsis thaliana root. *Development* **119**(1): 71-84.

Domenech J, Ramos SB, Probanza A, Lucas GJA, Gutierrez MFJ. 2007. Elicitation of systemic resistance and growth promotion of Arabidopsis thaliana by PGPRs from Nicotiana glauca: a study of the putative induction pathway. *Plant and Soil* **290**(1-2): 43-50.

Dong X. 2004. NPR1, all things considered. *Curr Opin Plant Biol* **7**(5): 547-552.

Doornbos RF, Geraats BP, Kuramae EE, Van Loon LC, Bakker PA. 2011. Effects of jasmonic acid, ethylene, and salicylic acid signaling on the rhizosphere bacterial community of Arabidopsis thaliana. *Mol Plant Microbe Interact* **24**(4): 395-407.

Downie JA. 2010. The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *Fems Microbiology Reviews* **34**(2): 150-170.

Downie JA, Ma QS, Knight CD, Hombrecher G, Johnston AWB. 1983. Cloning of the symbiotic region of Rhizobium-Leguminosarum - the nodulation genes are between the nitrogenase genes and a nifA-like gene. *Embo Journal* **2**(6): 947-952.

Duan J, Müller KM, Charles TC, Vesely S, Glick BR. 2009. 1-aminocyclopropane-1-carboxylate (ACC) deaminase genes in rhizobia from southern Saskatchewan. *Microb Ecol* **57**(3): 423-436.

Duijff BJ, Recorbet G, Bakker PA, Loper JE, Lemanceau P. 1999. Microbial Antagonism at the Root Level Is Involved in the Suppression of Fusarium Wilt by the Combination of Nonpathogenic Fusarium oxysporum Fo47 and Pseudomonas putida WCS358. *Phytopathology* **89**(11): 1073-1079.

Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM, Saar MO, Alexander S, Alexander EC, Jr., Rohwer F. 2006. Using pyrosequencing to shed light on deep mine microbial ecology. *Bmc Genomics* **7**: 57.

Eickhorst T, Tippkotter R. 2008. Improved detection of soil microorganisms using fluorescence in situ hybridization (FISH) and catalyzed reporter deposition (CARD-FISH). *Soil Biology & Biochemistry* **40**(7): 1883-1891.

Eilers KG, Lauber CL, Knight R, Fierer N. 2010. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. *Soil Biology & Biochemistry* **42**(6): 896-903.

Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P. 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* **4**(5): 642-647.

Esseling JJ, Lhuissier FG, Emons AM. 2003. Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. *Plant Physiol* **132**(4): 1982-1988.

Estrada-De Los Santos P, Bustillos-Cristales R, Caballero-Mellado J. 2001. Burkholderia, a genus rich in plant-associated nitrogen fixers with wide environmental and geographic distribution. *Appl Environ Microbiol* **67**(6): 2790-2798.

Fierer N, Bradford MA, Jackson RB. 2007. Toward an ecological classification of soil bacteria. *Ecology* **88**(6): 1354-1364.

Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* **103**(3): 626-631.

Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R. 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* **6**(5): 1007-1017.

Figurski DH, Helinski DR. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**(4): 1648-1652.

Fisher MM, Triplett EW. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl Environ Microbiol* **65**(10): 4630-4636.

Flemming HC, Wingender J. 2010. The biofilm matrix. *Nature Reviews Microbiology* **8**(9): 623-633.

Fujishige NA, Kapadia NN, De Hoff PL, Hirsch AM. 2006. Investigations of Rhizobium biofilm formation. *FEMS Microbiol Ecol* **56**(2): 195-206.

Gage DJ. 2002. Analysis of infection thread development using Gfp- and DsRed-expressing Sinorhizobium meliloti. *Journal of Bacteriology* **184**(24): 7042-7046.

Gage DJ. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol Biol Rev* **68**(2): 280-300.

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J. 2001. The composite genome of the legume symbiont Sinorhizobium meliloti. *Science* **293**(5530): 668-672.

Galippe V. 1887. Note sur la présence de micro-organismes dans les tissus végétaux. *Comptes Rendus Hebdomadaires des Séances et Mémoires de la Société de Biologie et des ses Filiales et Associées* **39**: 410-416.

Gans J, Wolinsky M, Dunbar J. 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**(5739): 1387-1390.

Glick BR, Cheng Z, Czarny J, Duan J. 2007a. Promotion of plant growth by ACC deaminase-producing soil bacteria. *European Journal of Plant Pathology* **119**(3): 329-339.

Glick BR, Todorovic B, Czarny J, Cheng Z, Duan J, McConkey B. 2007b. Promotion of plant growth by bacterial ACC deaminase. *Critical Reviews in Plant Sciences* **26**(5-6): 227-242.

Goedhart J, Hink MA, Visser A, Bisseling T, Gadella TWJ. 2000. In vivo fluorescence correlation microscopy (FCM) reveals accumulation and immobilization of Nod factors in root hair cell walls. *Plant Journal* **21**(1): 109-119.

Gonzalez JE, Reuhs BL, Walker GC. 1996. Low molecular weight EPS II of Rhizobium meliloti allows nodule invasion in Medicago sativa. *Proc Natl Acad Sci U S A* **93**(16): 8636-8641.

Gonzalez V, Santamaria RI, Bustos P, Hernandez-Gonzalez I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramirez MA, Jimenez-Jacinto V, Collado-Vides J, Davila G. 2006. The partitioned Rhizobium etli genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci U S A* **103**(10): 3834-3839.

Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI. 2011. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A* **108**(15): 6252-6257.

Gottel NR, Castro HF, Kerley M, Yang Z, Pelletier DA, Podar M, Karpinets T, Uberbacher E, Tuskan GA, Vilgalys R, Doktycz MJ, Schadt CW. 2011. Distinct microbial communities within the endosphere and rhizosphere of Populus deltoides roots across contrasting soil types. *Appl Environ Microbiol* **77**(17): 5934-5944.

Gottig N, Garavaglia BS, Garofalo CG, Orellano EG, Ottado J. 2009. A Filamentous Hemagglutinin-Like Protein of Xanthomonas axonopodis pv. citri, the Phytopathogen Responsible for Citrus Canker, Is Involved in Bacterial Virulence. *Plos One* **4**(2).

Gough C, Galera C, Vasse J, Webster G, Cocking EC, Denarie J. 1997. Specific flavonoids promote intercellular root colonization of Arabidopsis thaliana by Azorhizobium caulinodans ORS571. *Mol Plant Microbe Interact* **10**(5): 560-570.

Graner G, Persson P, Meijer J, Alstrom S. 2003. A study on microbial diversity in different cultivars of Brassica napus in relation to its wilt pathogen, Verticillium longisporum. *Fems Microbiology Letters* **224**(2): 269-276.

Haas D, Keel C. 2003. Regulation of antibiotic production in root-colonizing Pseudomonas spp. and relevance for biological control of plant disease. *Annu Rev Phytopathol* **41**: 117-153.

Haichar FZ, Marol C, Berge O, Rangel-Castro JI, Prosser JI, Balesdent J, Heulin T, Achouak W. 2008. Plant host habitat and root exudates shape soil bacterial community structure. *ISME J* **2**(12): 1221-1230.

Halder AK, Chakrabartty PK. 1993. Solubilization of inorganic-phosphate by Rhizobium. *Folia Microbiologica* **38**(4): 325-330.

Halder AK, Mishra AK, Bhattacharyya P, Chakrabartty PK. 1990. Solubilization of rock phosphate by Rhizobium and Bradyrhizobium. *Journal of General and Applied Microbiology* **36**(2): 81-92.

Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* **57**: 303-333.

Hallmann J, QuadtHallmann A, Mahaffee WF, Kloepper JW. 1997. Bacterial endophytes in agricultural crops. *Canadian Journal of Microbiology* **43**(10): 895-914.

Han JI, Choi HK, Lee SW, Orwin PM, Kim J, Laroe SL, Kim TG, O'Neil J, Leadbetter JR, Lee SY, Hur CG, Spain JC, Ovchinnikova G, Goodwin L, Han C. 2011. Complete genome sequence of the metabolically versatile plant growth-promoting endophyte Variovorax paradoxus S110. *J Bacteriol* **193**(5): 1183-1190.

Hardoim PR, Andreote FD, Reinhold-Hurek B, Sessitsch A, van Overbeek LS, van Elsas JD. 2011. Rice root-associated bacteria: insights into community structures across 10 cultivars. *FEMS Microbiol Ecol* **77**(1): 154-164.

Hardoim PR, van Overbeek LS, van Elsas JD. 2008. Properties of bacterial endophytes and their proposed role in plant growth. *Trends in Microbiology* **16**(10): 463-471.

Hassan S, Mathesius U. 2012. The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany* **63**(9): 3429-3444.

Hayatsu M, Tago K, Saito M. 2008. Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification. *Soil Science and Plant Nutrition* **54**(1): 33-45.

Hein JW, Wolfe GV, Blee KA. 2008. Comparison of rhizosphere bacterial communities in Arabidopsis thaliana mutants for systemic acquired resistance. *Microb Ecol* **55**(2): 333-343.

Heuer H, Krsek M, Baker P, Smalla K, Wellington EM. 1997. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl Environ Microbiol* **63**(8): 3233-3241.

Hider RC, Kong X. 2010. Chemistry and biology of siderophores. *Nat Prod Rep* **27**(5): 637-657.

Hiltner J. 1904. Über neuere Erfahrungen und Probleme auf dem Gebiet der Bodenbakteriologie und unter besonderer Berücksichtigung der Gründüngung und Brache. *Arbeiten der Deutschen Landwirtschaftlichen Gesellschaft* **98**: 59-78.

Hirner A, Ladwig F, Stransky H, Okumoto S, Keinath M, Harms A, Frommer WB, Koch W. 2006. Arabidopsis LHT1 is a high-affinity transporter for cellular amino acid uptake in both root epidermis and leaf mesophyll. *Plant Cell* **18**(8): 1931-1946.

Hooper LV, Midtvedt T, Gordon JI. 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annual Review of Nutrition* **22**: 283-307.

Hubber A, Vergunst AC, Sullivan JT, Hooykaas PJ, Ronson CW. 2004. Symbiotic phenotypes and translocated effector proteins of the Mesorhizobium loti strain R7A VirB/D4 type IV secretion system. *Mol Microbiol* **54**(2): 561-574.

Hurek T, Handley LL, Reinhold-Hurek B, Piche Y. 2002. Azoarcus grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Mol Plant Microbe Interact* **15**(3): 233-242.

Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* **4**(11): e1000255.

Huys G, Vanhoutte T, Joossens M, Mahious AS, De Brandt E, Vermeire S, Swings J. 2008. Coamplification of eukaryotic DNA with 16S rRNA gene-based PCR primers: possible consequences for population fingerprinting of complex microbial communities. *Current Microbiology* **56**(6): 553-557.

Inceoglu O, Al-Soud WA, Salles JF, Semenov AV, van Elsas JD. 2011. Comparative analysis of bacterial communities in a potato field as determined by pyrosequencing. *Plos One* **6**(8): e23321.

Iniguez AL, Dong Y, Carter HD, Ahmer BM, Stone JM, Triplett EW. 2005. Regulation of enteric endophytic bacterial colonization by plant defenses. *Mol Plant Microbe Interact* **18**(2): 169-178.

Iniguez AL, Dong YM, Triplett EW. 2004. Nitrogen fixation in wheat provided by Klebsiella pneumoniae 342. *Molecular Plant-Microbe Interactions* **17**(10): 1078-1085.

Janssen PH. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* **72**(3): 1719-1728.

Jetiyanon K, Kloepper JW. 2002. Mixtures of plant growth-promoting rhizobacteria for induction of systemic resistance against multiple plant diseases. *Biological Control* **24**(3): 285-291.

Johnson DC, Dean DR, Smith AD, Johnson MK. 2005. Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* **74**: 247-281.

Johnston-Monje D, Raizada MN. 2011. Conservation and diversity of seed associated endophytes in Zea across boundaries of evolution, ethnography and ecology. *Plos One* **6**(6): e20396.

Jones DL, Darrah PR. 1996. Re-sorption of organic compounds by roots of Zea mays L and its consequences in the rhizosphere .3. Characteristics of sugar influx and efflux. *Plant and Soil* **178**(1): 153-160.

Jones DL, Nguyen C, Finlay RD. 2009. Carbon flow in the rhizosphere: carbon trading at the soil-root interface. *Plant and Soil* **321**(1-2): 5-33.

Jones JD, Dangl JL. 2006. The plant immune system. *Nature* **444**(7117): 323-329.

Jooste PJ, Hugo CJ. 1999. The taxonomy, ecology and cultivation of bacterial genera belonging to the family Flavobacteriaceae. *Int J Food Microbiol* **53**(2-3): 81-94.

Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *Fems Microbiology Reviews* **33**(2): 376-393.

Kachlany SC, Planet PJ, Bhattacharjee MK, Kollia E, DeSalle R, Fine DH, Figurski DH. 2000. Nonspecific adherence by Actinobacillus actinomycetemcomitans requires genes widespread in bacteria and archaea. *J Bacteriol* **182**(21): 6169-6176.

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. *Mol Plant Microbe Interact* **19**(10): 1121-1126.

Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS. 2009. Transcriptomic Analysis of Rhizobium leguminosarum Biovar viciae in Symbiosis with Host Plants Pisum sativum and Vicia cracca. *Journal of Bacteriology* **191**(12): 4002-4014.

Keller M, Zengler K. 2004. Tapping into microbial diversity. *Nature Reviews Microbiology* **2**(2): 141-150.

Kim YC, Miller CD, Anderson AJ. 2000. Superoxide dismutase activity in Pseudomonas putida affects utilization of sugars and growth on root surfaces. *Appl Environ Microbiol* **66**(4): 1460-1467.

Kloepper JW, Leong J, Teintze M, Schroth MN. 1980. Enhanced plant-growth by siderophores produced by plant growth-promoting rhizobacteria. *Nature* **286**(5776): 885-886.

Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, von Mering C, Vorholt JA. 2012. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* **6**(7): 1378-1390.

Koch R. 1893. Ueber den augenblicklichen Stand der bakteriologischen Choleradiagnose. *Medical Microbiology and Immunology* **14**(1): 319-338.

Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B-Biological Sciences* **361**(1475): 1929-1940.

Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**(7): 2567-2572.

Krause A, Ramakumar A, Bartels D, Battistoni F, Bekel T, Boch J, Bohm M, Friedrich F, Hurek T, Krause L, Linke B, McHardy AC, Sarkar A, Schneiker S, Syed AA, Thauer R, Vorholter FJ, Weidner S, Puhler A, Reinhold-Hurek B, Kaiser O, Goesmann A. 2006. Complete genome of the mutualistic, N2-fixing grass endophyte Azoarcus sp. strain BH72. *Nat Biotechnol* **24**(11): 1385-1391.

Kravchenko LV, Azarova TS, Makarova NM, Tikhonovich IA. 2004. The effect of tryptophan present in plant root exudates on the phytostimulating activity of rhizobacteria. *Microbiology* **73**(2): 156-158.

Kristiansson E, Hugenholtz P, Dalevi D. 2009. ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* **25**(20): 2737-2738.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**(1): 118-123.

Kunin V, Hugenholtz P. 2010. PyroTagger: A fast, accurate pipeline for rRNA amplicon pyrosequence data. *The Open Journal* **Article 1**.

Kwak YS, Bonsall RF, Okubara PA, Paulitz TC, Thomashow LS, Weller DM. 2012. Factors impacting the activity of 2,4-diacetylphloroglucinol-producing Pseudomonas fluorescens against take-all of wheat. *Soil Biology & Biochemistry* **54**: 48-56.

Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**(9): 3100-3108.

Lewin A, Cervantes E, Wong CH, Broughton WJ. 1990. NodSU, 2 new nod genes of the broad host range Rhizobium strain NGR234 encode host-specific nodulation of the tropical tree leucaena-leucocephala. *Molecular Plant-Microbe Interactions* **3**(5): 317-326.

Li Y, Lee CG, Watanabe T, Murase J, Asakawa S, Kimura M. 2011. Identification of microbial communities that assimilate substrate from root cap cells in an aerobic soil using a DNA-SIP approach. *Soil Biology & Biochemistry* **43**(9): 1928-1935.

Liu WT, Marsh TL, Cheng H, Forney LJ. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology* **63**(11): 4516-4522.

Long HH, Sonntag DG, Schmidt DD, Baldwin IT. 2010. The structure of the culturable root bacterial endophyte community of Nicotiana attenuata is organized by soil composition and host plant ethylene production and perception. *New Phytologist* **185**(2): 554-567.

Loper JE, Hassan KA, Mavrodi DV, Davis EW, 2nd, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS, 3rd, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT. 2012. Comparative genomics of plant-associated Pseudomonas spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* **8**(7): e1002784.

Loper JE, Henkels MD. 1997. Availability of iron to Pseudomonas fluorescens in rhizosphere and bulk soil evaluated with an ice nucleation reporter gene. *Appl Environ Microbiol* **63**(1): 99-105.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH. 2004. ARB: a software environment for sequence data. *Nucleic Acids Research* **32**(4): 1363-1371.

Lugtenberg B, Kamilova F. 2009. Plant-growth-promoting rhizobacteria. *Annu Rev Microbiol* **63**: 541-556.

Lundberg DS, Lebeis SL, Herrera Paredes S, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Glavina del Rio T, Edgar R, Eickhorst T, Ley RE, Hugenholtz P, Green Tringe S, Dangl JL. 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature*.

Lynch D, O'Brien J, Welch T, Clarke P, Cuiv PO, Crosa JH, O'Connell M. 2001. Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by Sinorhizobium meliloti. *Journal of Bacteriology* **183**(8): 2576-2585.

Lynch JM, Whipps JM. 1990. Substrate flow in the rhizosphere. *Plant and Soil* **129**(1): 1-10.

Ma W, Sebestianova SB, Sebestian J, Burd GI, Guinel FC, Glick BR. 2003. Prevalence of 1-aminocyclopropane-1-carboxylate deaminase in Rhizobium spp. *Antonie Van Leeuwenhoek* **83**(3): 285-291.

Ma WB, Guinel FC, Glick BR. 2003. Rhizobium leguminosarum biovar viciae 1-aminocyclopropane-1-carboxylate deaminase promotes nodulation of pea plants. *Applied and Environmental Microbiology* **69**(8): 4396-4402.

Madhaiyan M, Poonguzhali S, Ryu J, Sa T. 2006. Regulation of ethylene levels in canola (Brassica campestris) by 1-aminocyclopropane-1-carboxylate deaminase-containing Methylobacterium fujisawaense. *Planta* **224**(2): 268-278.

Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G.

2005. Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* **309**(5731): 148-150.

Manter DK, Delgado JA, Holm DG, Stong RA. 2010. Pyrosequencing reveals a highly diverse and cultivar-specific bacterial endophyte community in potato roots. *Microb Ecol* **60**(1): 157-166.

Manz W, Amann R, Ludwig W, Vancanneyt M, Schleifer KH. 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. *Microbiology* **142 (Pt 5)**: 1097-1106.

Manz W, Amann R, Ludwig W, Wagner M, Schleifer KH. 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria - problems and solutions. *Systematic and Applied Microbiology* **15**(4): 593-600.

Mark GL, Dow JM, Kiely PD, Higgins H, Haynes J, Baysse C, Abbas A, Foley T, Franks A, Morrissey J, O'Gara F. 2005. Transcriptome profiling of bacterial responses to root exudates identifies genes involved in microbe-plant interactions. *Proc Natl Acad Sci U S A* **102**(48): 17454-17459.

Marschner P. 2004. Development of specific rhizosphere bacterial communities in relation to plant species, nutrition and soil type. *Plant Soil* **261**: 199-208.

Mba Medie F, Davies GJ, Drancourt M, Henrissat B. 2012. Genome analyses highlight the different biological roles of cellulases. *Nat Rev Microbiol* **10**(3): 227-234.

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. *Curr Opin Genet Dev* **15**(6): 589-594.

Meilhoc E, Boscari A, Bruand C, Puppo A, Brouquisse R. 2011. Nitric oxide in legume-rhizobium symbiosis. *Plant Science* **181**(5): 573-581.

Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M. 1998. Arabidopsis thaliana: a model plant for genome analysis. *Science* **282**(5389): 662, 679-682.

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JH, Piceno YM, DeSantis TZ, Andersen GL, Bakker PA, Raaijmakers JM. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**(6033): 1097-1100.

Mergaert P, Uchiumi T, Alunni B, Evanno G, Cheron A, Catrice O, Mausset AE, Barloy-Hubler F, Galibert F, Kondorosi A, Kondorosi E. 2006. Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc Natl Acad Sci U S A* **103**(13): 5230-5235.

Mergeay M, Houba C, Gerits J. 1978. Extrachromosomal-inheritance controlling resistance to cadmium, cobalt, copper and zinc - evidence from curing in a pseudomonas. *Archives Internationales De Physiologie De Biochimie Et De Biophysique* **86**(2): 440-442.

Merzaeva OV, Shirokikh IG. 2010. Production of auxins by the endophytic bacteria of winter rye. *Prikl Biokhim Mikrobiol* **46**(1): 51-57.

Mesa S, Velasco L, Manzanera ME, Delgado MJ, Bedmar EJ. 2002. Characterization of the norCBQD genes, encoding nitric oxide reductase, in the nitrogen fixing bacterium Bradyrhizobium japonicum. *Microbiology* **148**(Pt 11): 3553-3560.

Micallef SA, Shiaris MP, Colon-Carmona A. 2009. Influence of Arabidopsis thaliana accessions on rhizobacterial communities and natural variation in root exudates. *J Exp Bot* **60**(6): 1729-1742.

Miche L, Battistoni F, Gernmer S, Belghazi M, Reinhold-Hurek B. 2006. Upregulation of jasmonate-inducible defense proteins and differential colonization of roots of Oryza sativa cultivars with the endophyte Azoarcus sp. *Molecular Plant-Microbe Interactions* **19**(5): 502-511.

Millet YA, Danna CH, Clay NK, Songnuan W, Simon MD, Werck-Reichhart D, Ausubel FM. 2010. Innate Immune Responses Activated in Arabidopsis Roots by Microbe-Associated Molecular Patterns. *Plant Cell* **22**(3): 973-990.

Mishra RPN, Singh RK, Jaiswal HK, Kumar V, Maurya S. 2006. Rhizobium-mediated induction of phenolics and plant growth promotion in rice (Oryza sativa L.). *Current Microbiology* **52**(5): 383-389.

Misk A, Franco C. 2011. Biocontrol of chickpea root rot using endophytic actinobacteria. *Biocontrol* **56**(5): 811-822.

Molina-Favero C, Creus CM, Simontacchi M, Puntarulo S, Lamattina L. 2008. Aerobic nitric oxide production by Azospirillum brasilense Sp245 and its influence on root architecture in tomato. *Molecular Plant-Microbe Interactions* **21**(7): 1001-1009.

Morgan JL, Darling AE, Eisen JA. 2010. Metagenomic sequencing of an in vitro-simulated microbial community. *Plos One* **5**(4): e10209.

Muyzer G, de Waal EC, Uitterlinden AG. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**(3): 695-700.

Nacke H, Thurmer A, Wollherr A, Will C, Hodac L, Herold N, Schoning I, Schrumpf M, Daniel R. 2011. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *Plos One* **6**(2): e17000.

Neilands JB. 1995. Siderophores: structure and function of microbial iron transport compounds. *Journal of Biological Chemistry* **270**(45): 26723-26726.

O'Callaghan KJ, Stone PJ, Hu X, Griffiths DW, Davey MR, Cocking EC. 2000. Effects of glucosinolates and flavonoids on colonization of the roots of Brassica napus by Azorhizobium caulinodans ORS571. *Appl Environ Microbiol* **66**(5): 2185-2191.

Ohara GW, Daniel RM. 1985. RHIZOBIAL DENITRIFICATION - A REVIEW. *Soil Biology & Biochemistry* **17**(1): 1-9.

Okazaki S, Sugawara M, Minamisawa K. 2004. Bradyrhizobium elkanii rtxC gene is required for expression of symbiotic phenotypes in the final step of rhizobitoxine biosynthesis. *Applied and Environmental Microbiology* **70**(1): 535-541.

Oldroyd GE, Downie JA. 2008. Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* **59**: 519-546.

Oldroyd GED, Murray JD, Poole PS, Downie JA 2011. The Rules of Engagement in the Legume-Rhizobial Symbiosis. In: Bassler BL, Lichten M, Schupbach G eds. *Annual Review Genetics, Vol 45*. Palo Alto: Annual Reviews, 119-144.

Ostle N, Whiteley AS, Bailey MJ, Sleep D, Ineson P, Manefield M. 2003. Active microbial RNA turnover in a grassland soil estimated using a 13CO2 spike. *Soil Biology & Biochemistry* **35**: 877-885.

Patten CL, Glick BR. 2002. Regulation of indoleacetic acid production in Pseudomonas putida GR12-2 by tryptophan and the stationary-phase sigma factor RpoS. *Can J Microbiol* **48**(7): 635-642.

Peix A, Rivas-Boyero AA, Mateos PF, Rodriguez-Barrueco C, Martinez-Molina E, Velazquez E. 2001. Growth promotion of chickpea and barley by a phosphate solubilizing strain of Mesorhizobium mediterraneum under growth chamber conditions. *Soil Biology & Biochemistry* **33**(1): 103-110.

Pernthaler A, Pernthaler J 2005. Simultaneous fluorescence in situ hybridization of mRNA and rRNA for the detection of gene expression in environmental microbes. *Environmental Microbiology*. San Diego: Elsevier Academic Press Inc, 352-+.

Pernthaler A, Pernthaler J, Amann R. 2002. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Applied and Environmental Microbiology* **68**(6): 3094-3101.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* **8**(10): 785-786.

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, Baker CC, Di Francesco V, Howcroft TK, Karp RW, Lunsford RD, Wellington CR, Belachew T, Wright M, Giblin C, David H, Mills M, Salomon R, Mullins C, Akolkar B, Begg L, Davis C, Grandison L, Humble M, Khalsa J, Little AR, Peavy H, Pontzer C, Portnoy M, Sayre MH, Starke-Reed P, Zakhari S, Read J, Watson B, Guyer M. 2009. The NIH Human Microbiome Project. *Genome Res* **19**(12): 2317-2323.

Pieterse CMJ, Van Pelt JA, Verhagen BWM, Ton J, Van Wees SCM, Leon-Kloosterziel KM, Van Loon LC. 2003. Induced systemic resistance by plant growth-promoting rhizobacteria. *Symbiosis* **35**(1-3): 39-54.

Pini F, Galardini M, Bazzicalupo M, Mengoni A. 2011. Plant-Bacteria Association and Symbiosis: Are There Common Genomic Traits in Alphaproteobacteria? *Genes* **2**.

Prosser JI, Rangel-Castro JI, Killham K. 2006. Studying plant-microbe interactions using stable isotope technologies. *Curr Opin Biotechnol* **17**(1): 98-102.

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

Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvette C, Moenne-Loccoz Y. 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and Soil* **321**(1-2): 341-361.

Ran LX, Li ZN, Wu GJ, van Loon LC, Bakker P. 2005. Induction of systemic resistance against bacterial wilt in Eucalyptus urophylla by fluorescent Pseudomonas spp. *European Journal of Plant Pathology* **113**(1): 59-70.

Rao NSS. 1995. *Soil microorganisms and plant growth*: Science Publishers, Inc.

Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. *J Bacteriol* **190**(20): 6881-6893.

Rea PA 2007. Plant ATP-Binding cassette transporters. *Annual Review of Plant Biology*. Palo Alto: Annual Reviews, 347-375.

Reasoner DJ, Geldreich EE. 1985. A new medium for the enumeration and subculture of bacteria from potable water. *Applied and Environmental Microbiology* **49**(1): 1-7.

Reinhold-Hurek B, Hurek T. 1998. Life in grasses: diazotrophic endophytes. *Trends Microbiol* **6**(4): 139-144.

Reinhold-Hurek B, Hurek T. 2011. Living inside plants: bacterial endophytes. *Curr Opin Plant Biol* **14**(4): 435-443.

Reintanz B, Lehnen M, Reichelt M, Gershenzon J, Kowalczyk M, Sandberg G, Godde M, Uhl R, Palme K. 2001. Bus, a bushy Arabidopsis CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *Plant Cell* **13**(2): 351-367.

Reitz M, Rudolph K, Schroder I, Hoffmann-Hergarten S, Hallmann J, Sikora RA. 2000. Lipopolysaccharides of Rhizobium etli strain G12 act in potato roots as an inducing agent of systemic resistance to infection by the cyst nematode Globodera pallida. *Applied and Environmental Microbiology* **66**(8): 3515-3518.

Robert-Seilaniantz A, Grant M, Jones JD. 2011. Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism. *Annu Rev Phytopathol* **49**: 317-343.

Rodriguez H, Fraga R. 1999. Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnology Advances* **17**(4-5): 319-339.

Rodriguez H, Fraga R, Gonzalez T, Bashan Y. 2006. Genetics of phosphate solubilization and its potential applications for improving plant growth-promoting bacteria. *Plant and Soil* **287**(1-2): 15-21.

Roller C, Wagner M, Amann R, Ludwig W, Schleifer KH. 1994. In situ probing of gram-positive bacteria with high DNA G + C content using 23S rRNA-targeted oligonucleotides. *Microbiology* **140 (Pt 10)**: 2849-2858.

Rosenblueth M, Martinez-Romero E. 2004. Rhizobium etli maize populations and their competitiveness for root colonization. *Arch Microbiol* **181**(5): 337-344.

Rosenblueth M, Martinez-Romero E. 2006. Bacterial endophytes and their interactions with hosts. *Mol Plant Microbe Interact* **19**(8): 827-837.

Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**(10): 1340-1351.

Rudrappa T, Czymmek KJ, Pare PW, Bais HP. 2008. Root-secreted malic acid recruits beneficial soil bacteria. *Plant Physiol* **148**(3): 1547-1556.

Ryu CM, Murphy JF, Mysore KS, Kloepper JW. 2004. Plant growth-promoting rhizobacteria systemically protect Arabidopsis thaliana against Cucumber mosaic virus by a salicylic acid and NPR1-independent and jasmonic acid-dependent signaling pathway. *Plant Journal* **39**(3): 381-392.

Santaella C, Schue M, Berge O, Heulin T, Achouak W. 2008. The exopolysaccharide of Rhizobium sp YAS34 is not necessary for biofilm formation on Arabidopsis thaliana and Brassica napus roots but contributes to root colonization. *Environmental Microbiology* **10**(8): 2150-2163.

Savage DC. 1977. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* **31**: 107-133.

Schloss PD, Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**(3): 1501-1506.

Schuster-Bockler B, Conrad D, Bateman A. 2010. Dosage sensitivity shapes the evolution of copy-number varied regions. *Plos One* **5**(3): e9474.

Seefeldt LC, Hoffman BM, Dean DR. 2009. Mechanism of Mo-dependent nitrogenase. *Annu Rev Biochem* **78**: 701-722.

Sekine M, Ichikawa T, Kuga N, Kobayashi M, Sakurai A, Syono K. 1988. Detection of the iaa biosynthetic-pathway from tryptophan via indole-3-acetamide in bradyrhizobium spp. *Plant and Cell Physiology* **29**(5): 867-874.

Semenov AM, van Bruggen AHC, Zelenev VV. 1999. Moving Waves of Bacterial Populations and Total Organic Carbon along Roots of Wheat. *Microb Ecol* **37**(2): 116-128.

Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E. 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. *Appl Environ Microbiol* **67**(9): 4215-4224.

Shanmugam KT, Ogara F, Andersen K, Valentine RC. 1978. Biological nitrogen fixation. *Annual Review of Plant Physiology and Plant Molecular Biology* **29**: 263-276.

Shibistova O, Lloyd J, Evgrafova S, Savushkina N, Zrazhevskaya G, Arneth A, Knohl A, Kolle O, Schulze ED. 2002. Seasonal and spatial variability in soil CO2 efflux rates for a central Siberian Pinus sylvestris forest. *Tellus Series B-Chemical and Physical Meteorology* **54**(5): 552-567.

Shipton PJ, Cook RJ, Sitton JW. 1973. Occurrence and transfer of a biological factor in soil that suppresses take all of wheat in eastern Washington. *Phytopathology* **63**: 511-517.

Siyambalapitiya N, Blackall LL. 2005. Discrepancies in the widely applied GAM42a fluorescence in situ hybridisation probe for Gammaproteobacteria. *FEMS Microbiol Lett* **242**(2): 367-373.

Smits TH, Rezzonico F, Duffy B. 2011. Evolutionary insights from Erwinia amylovora genomics. *J Biotechnol* **155**(1): 34-39.

Smyth GK. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**: Article3.

Snipen L, Almoy T, Ussery DW. 2009. Microbial comparative pan-genomics using binomial mixture models. *Bmc Genomics* **10**.

Spaepen S, Vanderleyden J. 2011. Auxin and plant-microbe interactions. *Cold Spring Harb Perspect Biol* **3**(4).

Spiegelman D, Whissell G, Greer CW. 2005. A survey of the methods for the characterization of microbial consortia and communities. *Can J Microbiol* **51**(5): 355-386.

Staddon PL. 2004. Carbon isotopes in functional soil ecology. *Trends Ecol Evol* **19**(3): 148-154.

Steinkellner S, Lendzemo V, Langer I, Schweiger P, Khaosaad T, Toussaint JP, Vierheilig H. 2007. Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules* **12**(7): 1290-1306.

Sullivan JT, Trzebiatowski JR, Cruickshank RW, Gouzy J, Brown SD, Elliot RM, Fleetwood DJ, McCallum NG, Rossbach U, Stuart GS, Weaver JE, Webby RJ, De Bruijn FJ, Ronson CW. 2002. Comparative sequence analysis of the symbiosis island of Mesorhizobium loti strain R7A. *J Bacteriol* **184**(11): 3086-3095.

Taghavi S, Garafola C, Monchy S, Newman L, Hoffman A, Weyens N, Barac T, Vangronsveld J, van der Lelie D. 2009. Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Environ Microbiol* **75**(3): 748-757.

Tan Z, Hurek T, Reinhold-Hurek B. 2003. Effect of N-fertilization, plant genotype and environmental conditions on nifH gene pools in roots of rice. *Environ Microbiol* **5**(10): 1009-1015.

Tateda K, Comte R, Pechere JC, Kohler T, Yamaguchi K, Van Delden C. 2001. Azithromycin inhibits quorum sensing in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy* **45**(6): 1930-1933.

Teale WD, Paponov IA, Palme K. 2006. Auxin in action: signalling, transport and the control of plant growth and development. *Nat Rev Mol Cell Biol* **7**(11): 847-859.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A* **102**(39): 13950-13955.

TheArabidopsisGenomeInitiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**(6814): 796-815.

Thrall PH, Watkin ELJ, Burdon JJ 2001. Coevolution: plant–Microorganism. *eLS*: John Wiley & Sons, Ltd.

Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR, Chen WX. 2012. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc Natl Acad Sci U S A* **109**(22): 8629-8634.

Torres MA, Dangl JL, Jones JD. 2002. Arabidopsis gp91phox homologues AtrbohD and AtrbohF are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc Natl Acad Sci U S A* **99**(1): 517-522.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. 2005. Comparative metagenomics of microbial communities. *Science* **308**(5721): 554-557.

Tsuda K, Sato M, Stoddard T, Glazebrook J, Katagiri F. 2009. Network properties of robust immunity in plants. *PLoS Genet* **5**(12): e1000772.

Turkarslan S, Reiss DJ, Gibbins G, Su WL, Pan M, Bare JC, Plaisier CL, Baliga NS. 2011. Niche adaptation by expansion and reprogramming of general transcription factors. *Mol Syst Biol* **7**: 554.

Uroz S, Buée M, Murat C, Frey-Klett P, Martin F. 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environmental Microbiology Reports* **2**(2): 281-288.

Uroz S, Dessaux Y, Oger P. 2009. Quorum sensing and quorum quenching: the yin and yang of bacterial communication. *Chembiochem* **10**(2): 205-216.

Urquiaga S, Xavier RP, de Morais RF, Batista RB, Schultz N, Leite JM, Maia e Sa J, Barbosa KP, de Resende AS, Alves BJR, Boddey RM. 2012. Evidence from field nitrogen balance and N-15 natural abundance data for the contribution of biological N-2 fixation to Brazilian sugarcane varieties. *Plant and Soil* **356**(1-2): 5-21.

Van der Ent S, Verhagen BWM, Van Doorn R, Bakker D, Verlaan MG, Pel MJC, Joosten RG, Proveniers MCG, Van Loon LC, Ton J, Pieterse CMJ. 2008. MYB72 is required in early signaling steps of rhizobacteria-induced systemic resistance in arabidopsis. *Plant Physiology* **146**(3): 1293-1304.

van Overbeek L, van Elsas JD. 2008. Effects of plant genotype and growth stage on the structure of bacterial communities associated with potato (Solanum tuberosum L.). *FEMS Microbiol Ecol* **64**(2): 283-296.

Vande Broek A, Gysegom P, Ona O, Hendrickx N, Prinsen E, Van Impe J, Vanderleyden J. 2005. Transcriptional analysis of the Azospirillum brasilense indole-3-pyruvate decarboxylase gene and identification of a cis-acting sequence involved in auxin responsive expression. *Mol Plant Microbe Interact* **18**(4): 311-323.

Vartoukian SR, Palmer RM, Wade WG. 2010. Cultivation of a Synergistetes strain representing a previously uncultivated lineage. *Environ Microbiol* **12**(4): 916-928.

Vassilev N, Vassileva M, Nikolaeva I. 2006. Simultaneous P-solubilizing and biocontrol activity of microorganisms: potentials and future trends. *Appl Microbiol Biotechnol* **71**(2): 137-144.

Vial L, Groleau MC, Dekimpe V, Deziel E. 2007. Burkholderia diversity and versatility: an inventory of the extracellular products. *J Microbiol Biotechnol* **17**(9): 1407-1429.

Vicre M, Santaella C, Blanchet S, Gateau A, Driouich A. 2005. Root border-like cells of Arabidopsis. Microscopical characterization and role in the interaction with rhizobacteria. *Plant Physiol* **138**(2): 998-1008.

Voisard C, Keel C, Haas D, Defago G. 1989. Cyanide production by Pseudomonas fluorescens helps suppress black root rot of tobacco under gnotobiotic conditions. *EMBO J* **8**(2): 351-358.

Wairuri CK, van der Waals JE, van Schalkwyk A, Theron J. 2012. Ralstonia solanacearum needs Flp pili for virulence on potato. *Mol Plant Microbe Interact* **25**(4): 546-556.

Wallner G, Amann R, Beisker W. 1993. Optimizing fluorescent in situ hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. *Cytometry* **14**(2): 136-143.

Wang KL, Li H, Ecker JR. 2002. Ethylene biosynthesis and signaling networks. *Plant Cell* **14 Suppl**: S131-151.

Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**(16): 5261-5267.

Wang TL, Wood EA, Brewin NJ. 1982. Growth-regulators, Rhizobium and nodulation in peas - indole-3-acetic-acid from the culture-medium of nodulating and non-nodulating strains of R. leguminosarum. *Planta* **155**(4): 345-349.

Wang Y, Qian PY. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *Plos One* **4**(10): e7401.

Wang YQ, Ohara Y, Nakayashiki H, Tosa Y, Mayama S. 2005. Microarray analysis of the gene expression profile induced by the endophytic plant growth-promoting rhizobacteria, Pseudomonas fluorescens FPT9601-T5 in Arabidopsis. *Molecular Plant-Microbe Interactions* **18**(5): 385-396.

Watve MG, Tickoo R, Jog MM, Bhole BD. 2001. How many antibiotics are produced by the genus Streptomyces? *Arch Microbiol* **176**(5): 386-390.

Weigel D, Mott R. 2009. The 1001 genomes project for Arabidopsis thaliana. *Genome Biol* **10**(5): 107.

Weinert N, Piceno Y, Ding GC, Meincke R, Heuer H, Berg G, Schloter M, Andersen G, Smalla K. 2011. PhyloChip hybridization uncovered an enormous bacterial diversity in the rhizosphere of different potato cultivars: many common and few cultivar-dependent taxa. *FEMS Microbiol Ecol* **75**(3): 497-506.

Weller DM. 2007. Pseudomonas biocontrol agents of soilborne pathogens: looking back over 30 years. *Phytopathology* **97**(2): 250-256.

Weller DM, Raaijmakers JM, Gardener BB, Thomashow LS. 2002. Microbial populations responsible for specific soil suppressiveness to plant pathogens. *Annu Rev Phytopathol* **40**: 309-348.

White JR, Nagarajan N, Pop M. 2009. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5**(4): e1000352.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**(11): 5088-5090.

Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI. 2003. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**(5615): 2074-2076.

Yanni YG, Rizk RY, Abd El-Fattah FK, Squartini A, Corich V, Giacomini A, de Bruijn F, Rademaker J, Maya-Flores J, Ostrom P, Vega-Hernandez M, Hollingsworth RI, Martinez-Molina E, Mateos P, Velazquez E, Wopereis J, Triplett E, Umali-Garcia M, Anarna JA, Rolfe BG, Ladha JK, Hill J, Mujoo R, Ng PK, Dazzo FB. 2001. The beneficial plant growth-promoting association of Rhizobium leguminosarum bv. trifolii with rice roots. *Australian Journal of Plant Physiology* **28**(9): 845-870.

Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J. 2006. The genome of Rhizobium leguminosarum has recognizable core and accessory components. *Genome Biol* **7**(4): R34.

Zhou J, Lemos B, Dopman EB, Hartl DL. 2011. Copy-Number Variation: The Balance between Gene Dosage and Expression in Drosophila melanogaster. *Genome Biology and Evolution* **3**: 1014-1024.

Appendix

Primer comparison 799F-1193R vs. 799F2-1193R

In the course of the study sequencing reads classified as Chloroflexi strongly dominated soil- and root-inhabiting bacterial communities. Since such an overrepresentation of Choroflexi was never described before (Janssen, 2006; Nacke *et al.*, 2011; Fierer *et al.*, 2012), it was likely that it was due to a bias introduced by the applied PCR primers 799F2 – 1993R. The applied forward primer 799F2 is a derivative of the previously published primer 799F (Chelius & Triplett, 2001). When using this parental primer organellar 18S rDNA was massively co-amplified together with bacterial 16S rDNA when employing the PCR primer 799F on DNA templates of Arabidopsis root compartments (Figure 30a). Co-amplification might affect the 16S rRNA ribotype profile of the root-inhabiting bacterial communities, as previously demonstrated for other host-associated prokaryotic communities (Huys *et al.*, 2008). A single nucleotide change was introduced in the original 799F PCR primer. This modified PCR primer, designated 799F2, does not compete with Arabidopsis organellar 18S RNA gene templates and allows the preferential amplification of the bacterial 16s rRNA gene. Pyrosequencing of '799F2-1193R amplicons' demonstrated that this PCR primer pair in combination with the touchdown PCR cycling program results in <0.5% plant sequences. 799F2 derived bacterial community profiles were compared with community structures of the parental PCR primer 799F (Figure 30). Bacterial community profiles of '799F-1193R amplicons' from 3 unplanted soil samples, 3 root samples from the accession Ler and 2 root samples from the accession Sha and were generated and compared to independent samples of the presented study that share the same experimental variables (soil type and compartment). All sequences of 799F samples were co-clustered with all 799F2 sequence reads and defined OTUs in a common dataset (Definition of OTUs see 2.3)

The key difference between the two PCR primers is the over-representation of Chloroflexi in 799F2 derived bacterial community profiles in both tested compartments (Figure 30b). The dataset of soil and root samples of Cologne soil as determined by 799F2 comprises 162 dominant OTUs (>5 ‰ RA) of which most (145) are also displayed with the PCR primer 799F (Figure 30c). The PCR primer 799F detects additional dominant OTUs that mostly belong to Proteobacteria. Importantly, of the 145 commonly captured OTUs most ribotypes are similarly displayed with both PCR primers except for OTUs belonging to the phylum of Chloroflexi (Figure 30d). Likewise, the differential accumulation of OTUs in the root compartment compared to soil is confirmed with the 799F PCR primer (Figure 30e). Together this demonstrates that the PCR primer 799F2 is

biased towards Chloroflexi and importantly, despite this bias the other bacterial phyla are reliably quantified.

This conclusion is supported by comparative PCR primer analysis and PCR independent *catalyzed reporter deposition – in situ hybridization* (CARD-FISH) analysis on soil samples, which is described in (2.4.3). Nevertheless, because a satisfying exclusion of amplification of the plant organelle 18S rRNA gene was only achieved using the aforementioned primer combination, it was used throughout the study and reads assigned to the phylum Choroflexi were removed *in silico*.

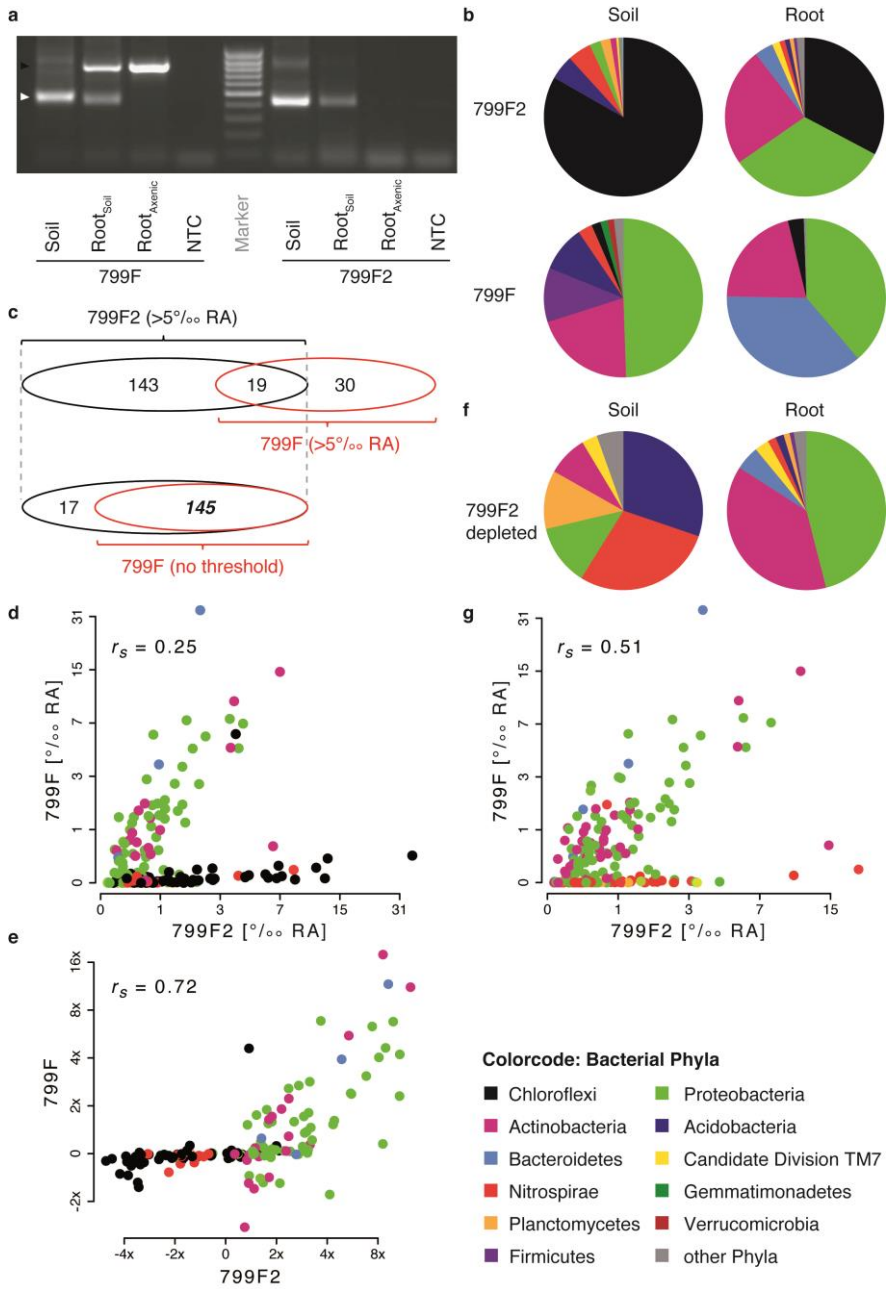
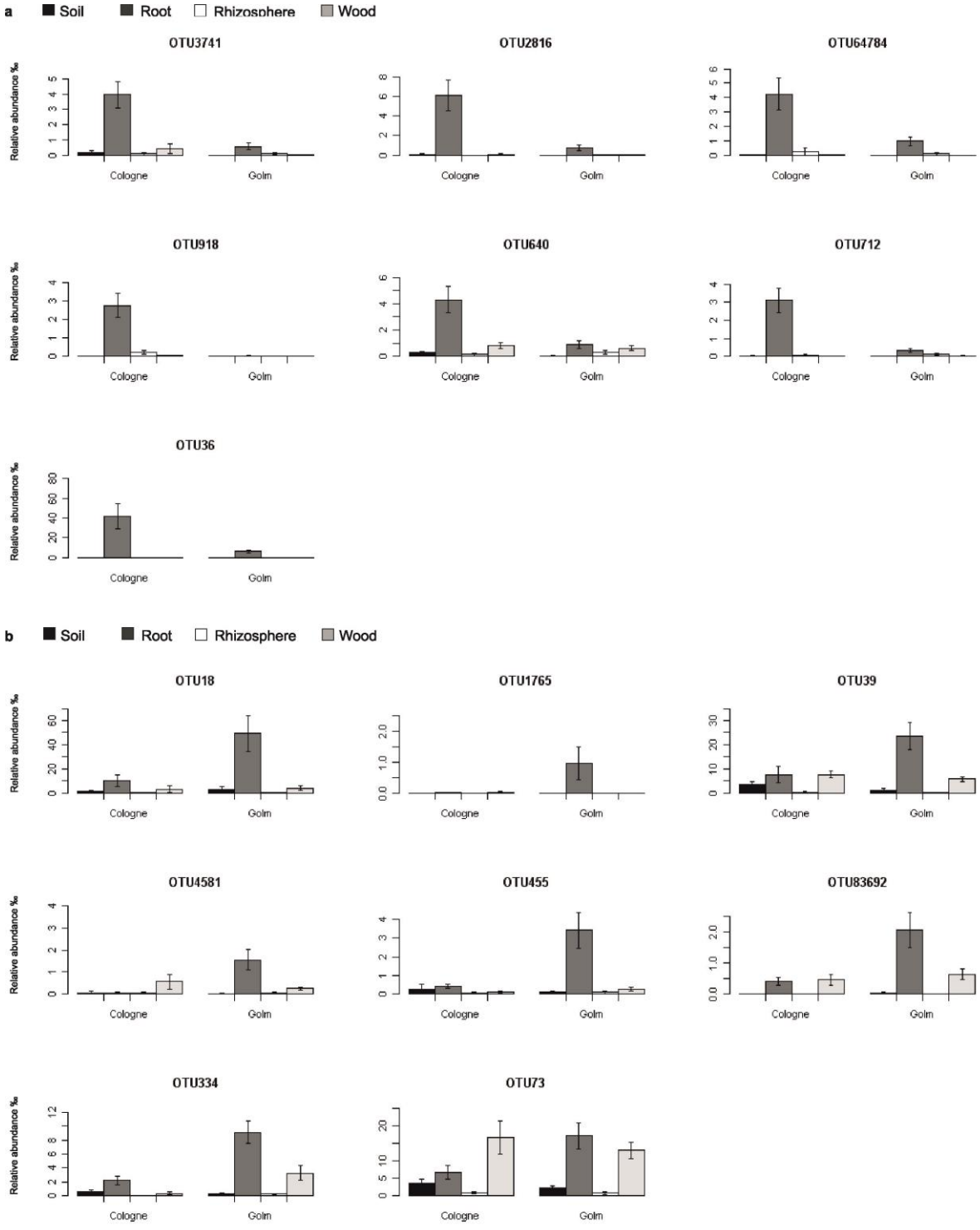


Figure 30: PCR primer comparison and depletion of Chloroflexi.

(a) End-point PCR was performed on the indicated DNA templates with the PCR primer combinations 799F - 1193R (left) and 799F2 - 1193R (right). Bacterial 16S rDNA (~ 400 bp) and Arabidopsis 18S rDNA (~ 760 bp) amplicons are indicated with white and black arrowheads, respectively. No bacterial 16S rDNA amplification product is detected using both PCR

primer combinations on DNA extracted from roots of axenically grown plants. NTC, non-template negative control. **(b-e)** Comparison of OTU based bacterial community profiles of soil and root samples from Cologne soil generated with PCR primers 799F (3 soil and 5 root samples) and 799F2. **(b)** Cumulative average RA of the OTUs per bacterial phylum (color code bottom right) as detected by the indicated PCR primers in soil and root samples. OTUs with less than two counts were removed for this analysis. **(c)** Soil and root samples profiled with the PCR primer 799F2 display 162 dominant OTUs (>5 ‰ RA), which are compared to the number of dominant (>5 ‰ RA, upper Venn diagram) and detected (no threshold, lower Venn diagram) OTUs by the PCR primer 799F. The 145 OTUs, which are dominant in the 799F2 dataset and also detected by 799F, are plotted in subfigures d, e and g. **(d)** Average RA of OTUs in soil and root compartments as detected by the indicated PCR primers. OTUs are color-coded according to their taxonomic assignment (bottom right). Spearman´s rank correlation coefficient, rs. **(e)** Fold enrichment of OTUs in roots compared to soil as detected by the indicated PCR primers. OTUs are color-coded according to their taxonomic assignment (bottom right). Spearman´s rank correlation coefficient, rs. **(f)** similar as (b) but taxonomic composition after depletion of OTUs assigned to Chloroflexi in the indicated samples displayed by the PCR primer 799F2. **(g)** similar as (d) but average RA of OTUs in soil and root compartments were calculated after depletion of OTUs assigned to Chloroflexi

Supplementary Figures



Supplementary Figure 1: Root-inhabiting OTUs differentially enriched in the tested soil types. Bar graphs depict RA (‰ normalized data ± s.e.m) of individual OTUs in the indicated compartments and soil types. **(a)** OTUs significantly enriched to higher levels in Cologne soil or absent in Golm soil. **(b)** OTUs significantly enriched to higher levels in Golm soil or absent in Cologne soil. Significance was tested by the interaction terms between soil type and compartments ($P<0.05$, t-test, FDR corrected).

Supplementary tables

Supplementary table 1: All bacterial and archeal taxa identified by 16S ribotyping

Phylum	Domain	average relative abundance [‰]
Crenarchaeota	Archaea	7.51
Euryarchaeota	Archaea	0.11
Acidobacteria	Bacteria	24.86
Actinobacteria	Bacteria	63.19
Aquificae	Bacteria	0.01
Armatimonadetes	Bacteria	0.36
BD1-5	Bacteria	0.55
Bacteroidetes	Bacteria	32.40
Caldiserica	Bacteria	0.01
Candidate division BRC1	Bacteria	0.05
Candidate division OD1	Bacteria	0.12
Candidate division OP11	Bacteria	0.00
Candidate division OP3	Bacteria	0.01
Candidate division OP9	Bacteria	0.98
Candidate division TM7	Bacteria	10.17
Candidate division WS3	Bacteria	0.27
Chlamydiae	Bacteria	0.02
Chlorobi	Bacteria	0.12
Chloroflexi	Bacteria	559.97
Cyanobacteria	Bacteria	0.26
Deferribacteres	Bacteria	0.01
Deinococcus-Thermus	Bacteria	0.73
Dictyoglomi	Bacteria	0.00
Elusimicrobia	Bacteria	0.07
Fibrobacteres	Bacteria	0.01
Firmicutes	Bacteria	4.00
Fusobacteria	Bacteria	0.01
GAL08	Bacteria	0.05
Gemmatimonadetes	Bacteria	0.88
JL-ETNP-Z39	Bacteria	0.01
Kazan-3B-28	Bacteria	0.00
Lentisphaerae	Bacteria	0.01
MVP-21	Bacteria	0.01
NPL-UPA2	Bacteria	0.00
Nitrospirae	Bacteria	27.99
Planctomycetes	Bacteria	35.27
Proteobacteria	Bacteria	226.89
Spirochaetes	Bacteria	0.31
Synergistetes	Bacteria	0.00
TA06	Bacteria	0.01
TM6	Bacteria	0.16
Tenericutes	Bacteria	0.01
Thermodesulfobacteria	Bacteria	0.02
Verrucomicrobia	Bacteria	2.53
WCHB1-60	Bacteria	0.04

Danksagung

Vor allem danke ich Prof. Schulze-Lefert für das Betreuen meiner Arbeit als gestrenger Doktorvater und dafür, dass ich meine Füße unter seinen MPI Tisch stellen durfte.

I thank Allan Downie for being part of my thesis committee and teaching me some *real* microbiology.

Für das Begutachten meiner Arbeit danke ich weiterhin Prof. Hülskamp.

Prof. Flügge danke ich für das Übernehmen des Prüfungsvorsizes wärend meiner Disputation.

Allerherzlichste Bedankerisierung an den Wurzel-Trupp Davide, Klaus, Yang, Nina und Federica. Mei, war das schön, mit euch im Sandkasten zu matschen! Außerdem danke ich Klaus und Davide noch einmal für das Korrekturlesen dieser Arbeit. Ein oder auch zwei Rehctschreibfeler habt ihr ja dann doch noch entdeckt…

Danke an Emiel und Barbara für den Bioinformatik-Part, ohne den es diese Arbeit nicht gäbe, an Hannes und Thilo für den Angelkurs und an Elmon und Rainer für Hilfe an allen möglichen Mikroskopen. An dieser Stelle möchte ich mich noch kurz vor dem anonymen Forscher verneigen, der die Lautsprecher am Confocal gesponsort hat, das hat mir manch dunkle Kellerstunde versüßt.
Ah, und Elke, kannst Du mir vielleicht noch mal die Federstahlpinzette leihen?

Weil ich hier am MPI außerdem gelernt habe, dass sich alles um's wording and phrasing dreht, möchte ich noch kurz folgende Perlen der menschlichen Kommunikation teilen:

asap, mental masturbation, should work, result-oriented, gentle

Und jetzt mal, ohne Spaß, ein DANKE an meine Mitbewohnerin Gretel und den P-Fisch Moritz. In diesem Laden hat man's nicht leicht, und ein paar Freunde, die einen wirklich verstehen, die helfen einem schon.

Ein Herz aus Hack an Florian, dem Kessel und dem Wirjes für Erdung und dringenst benötigte intellektuelle Auszeiten. Wenns knallt, dann spritzt es auch!

Ich danke Kollege Kaffee dafür mich wach zu halten, wenn nötig, und Peter-Joseph, Heinrich und Rainer dafür, mich in einen traumlosen Schlaf zu wiegen. Vorallem während des Zusammenschreibens.

Dann gibt es noch einen Haufen Leute mit denen ich verwandt sein darf, und die zweifelsohne zu diesem Machwerk beigetragen haben, evtl. sogar ohne zu wissen dass und wie. Aber ohne euch wär's jedenfalls nicht gegangen. Auf keinen Fall. Ich danke, und es berührt mich in diesem Moment tief im Herzen drinne:

meinem Mütterlein

dem Vadder

Stephan

Thomas

Andreas

Markus

Soweit,
Matthes, Matze, No.5

PS: Paul, Du schuldest mir noch ´n Sixpack Bier für meinen Posterpreis beim Annual Day 2011

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegt worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.
Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Paul Schulze-Lefert betreut worden.

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem Dekanat unverzüglich mitzuteilen

Datum

Unterschrift

Curriculum Vitae

Persönliche Daten

Name:	Matthias Rott
Anschrift:	Thielenstraße 8 50825 Köln
Geburtsdatum, -ort:	17.12.1984, Troisdorf
Familienstand:	Ledig

Promotion

Seit 10.2009	Promotionsarbeit am Max Planck Institut für Pflanzenzüchtungsforschung in der Arbeitsgruppe Paul Schulze-Lefert mit dem Thema „Structure and assembly cues of Arabidopsis root-inhabiting bacterial communities and comparative genomics of selected Rhizobium members"

Studium

11.2008-09.2009	Diplomarbeit am Max Planck Institut für chemische Ökologie in der Arbeitsgruppe Axel Mithöfer mit dem Thema „Untersuchungen des Verdauungssekretes und Klonierung eines PR1 Proteins aus *Nepenthes* Kannen."
08.2007-09.2007	Praktikum am am Max Planck Institut für chemische Ökologie in der Arbeitsgruppe Axel Mithöfer
10.2004-09.2009	Studium der Biologie an der Rheinischen Friedrich Wilhelm Universität Bonn

Schulbildung

08.1995-06.2004	Albert Einstein Gymnasium Sankt Augustin

Köln, 15.10.2012 _____

Matthias Rott