Universität Bremen

Study on the major diazotrophs associated with Norway spruce (*Picea abies* L. Karst) roots

Yuan Liu

Dissertation submitted in partial fulfillment of the

requirements for the degree

Dr. rer. nat.

Bremen, May 2011

The experiments of the presented work have been carried out from April 2006 until December 2010 at the Department of Biology/Chemistry of Bremen University, Germany, under the guidance of Prof. Dr. Barbara Reinhold-Hurek and Dr. Thomas Hurek.

First reviewer: Prof. Dr. Barbara Reinhold-Hurek

Second reviewer: Prof. Dr. Uwe Nehls

Parts of the presented results have been published:

Burbano CS, **Liu Y, Rösner K, Reis V, Caballero-Mellado J, Reinhold-Hurek B, Hurek T** (2011) Predominant *nifH* transcript phylotypes related to *Rhizobium rosettiformans* in field grown sugarcane plants and in Norway spruce. *Environmental Microbiology Reports* 3 : 383-389

Table of Contents

Summary

Conifers are of great ecological and commercial importance in the northern hemisphere temperate forests and in similar climates all over the world. Like in most temperate forests, nitrogen is an essential and often limiting nutrient. It is known that forest ecosystem may accumulate significant, albeit low amounts of nitrogen without the establishment of a symbiosis with known nitrogen-fixing bacteria. However, there is no agreement between researchers on the ecological role of biological nitrogen fixation in forest ecosystems. While mycorrhizas have been repeatedly implicated with nitrogen fixation in conifers, studies remain inconclusive since they did not reveal the identity of the responsible diazotrophs.

In this study major diazotrophs associated with medium-coarse roots of Norway spruce (*Picea abies* L. Karst) in a typical northern hemisphere temperate forest were investigated by applying RNA-based and protein-based methods. *NifH* transcripts identical to the partial *nifH* sequence of the grass endophyte *Azoarcus* sp. BH72 were repeatedly detected in RNA extracts from spruce sampled in different months using broad-range *nifH* primer sets with both single-stage and nested RT-PCR approaches. The *nifH* expression level was evaluated with real-time reverse transcriptase-polymerase chain reaction (RT-PCR) using *Azoarcus* sp. BH72-specific and universal *nifH* primers. *NifH* expression profiles showed that the *nifH* transcript levels changed with season and were highest in summer.

Real-time RT-PCR with *Azoarcus* sp. BH72-specific *nifH* and 16S rRNA gene primer sets were carried out to compare the ratios of the levels of 16S rRNA to *nifH* transcripts and of 16S rRNA genes to *nifH* genes in pure cultures with those *in planta*. Statistically significant differences were found in expression levels and presence of 16S rRNA and *nifH* genes from *A*. sp. BH72 *in planta* and in pure culture. Transcript levels of *A*. sp. BH72 16S rRNA genes relative to those of *A*. sp. BH72 *nifH* genes decreased by half when this

bacterium was associated with spruce. Even more pronounced differences were found at the DNA level. Here, the ratio changed by a factor >60: surprisingly, 16S rRNA genes from *Azoarcus* sp. BH72 were not detectable in spruce root DNA (detection limit: 6 copies), while 3×10^4 copies of *nifH* / g roots were found in the same DNA extract; while in pure cultures the ratio between 16S rRNA and *nifH* gene copies is four, in spruce it was <0.07.These results suggested that *in planta, nifH* from *Azoarcus* sp. BH72 is a non-chromosomal gene.

Apart from *A*. sp. BH72 *nifH* genes, also *nifH* sequences identical to the partial *nifH* from *Pseudomonas stutzeri* A1501 were detected in a particular spruce sample. Additionally, also *nifH* sequences with 98.7% DNA sequence identity to the partial *nifH* from *Rhizobium rosettiformans* were detected from spruce RNA extracted with an alternative RNA extraction method. However real-time qPCR analysis with specific primers indicated that the copy numbers of *P. stutzeri nifH* and 16S rRNA genes and of *R*. *rosettiformans nifH* genes in spruce roots were all below the limit of detection. Based on these data it was concluded that *Azoarcus* sp. BH72 is the major active nitrogen-fixing bacterium associated with spruce roots, while nitrogen fixation in *P. stutzeri* or *R. rosettiformans* is probably not important.

Apart from PCR dependent methods also a protein-based, cultivation independent approach was developed for detection and identification of N_2-fixing bacteria in plants. For this purpose spruce root proteins were extracted and first analyzed for presence of the NifH protein by Western blot analysis using a polyclonal anti-NifH antibody raised against the nitrogenase iron protein of *A*. sp. BH72 and a newly designed universal polyclonal anti-NifH peptide antibody. Subsequently, the immunoreactive protein spots were excised from two-dimensional (2-D) gels, analyzed by liquid chromatography Orbitrap mass spectrometry (LC-Orbitrap MS), and identified by comparative sequence analyses against public databases. All NifH peptide sequences obtained matched with the corresponding sequence of the NifH protein from *A*. sp. BH72. Because of the high specificity of two peptides, NifH from *Azoarcus* sp. BH72 could be identified in spruce roots with certainty.

This is the first study detecting NifH in naturally growing plants, which are not known to form any nitrogen fixing symbiosis or symbiotic structures with diazotrophic bacteria.

In another methodological development it was observed that *nifH* RT-single stage PCR products representing *nifH* sequences from *A*. sp. BH72 were obtained from spruce samples only by using LNA-substituted primers, but the PCR products were hardly clonable with the TOPO TA cloning kit. Because of the importance of this observation for this study, effects of LNA-substituted primers on topoisomerase- cloning efficiencies were examined. The results obtained suggested that the use of LNA-substituted primers in PCR with *Taq* DNA polymerase may inhibit topoisomerase-cloning by the base composition in the primer target region and by the prevention of the formation of a 3'-overhang.

In summary, NifH from the endophyte *Azoarcus* sp. BH72, which is central to nitrogen fixation in Kallar grass, is also central to nitrogen fixation in the root bark of spruce. This conclusion was drawn from results obtained by applying proteomic and PCR-based methods to the study of proteins and nucleic acids derived from naturally grown Norway spruce plants.

Zusammenfassung

In der gemäßigten Zone der nördlichen Hemisphäre sowie in anderen Regionen mit ähnlichen klimatischen Bedingungen spielen Nadelwälder eine ökologisch und kommerziell wichtige Rolle. Stickstoff ist ein essentieller und in Nadelwäldern ein das Pflanzenwachstum oft limitierender Nährstoff. Zwar sind Nadelwälder dafür bekannt, dass sie eine gewisse Menge an Stickstoff ohne die Etablierung einer Symbiose mit bekannten N_2-bindenden Bakterien akkumulieren können. Es ist aber umstritten, welche Rolle biologische Stickstofffixierung dort für das Pflanzenwachstum spielt. Während Mykorrhiza schon mehrfach mit der Stickstofffixierung bei Koniferen in Verbindung gebracht geworden ist, blieben die Untersuchungen zur Identität der verantwortlichen Diazotrophen bisher ergebnislos.

In dieser Studie wurden anhand von RNA- und Protein-basierten Methoden die wichtigsten Stickstofffixierer identifiziert, die in Assoziation mit der Rinde von etwa fingerstarken Wurzeln der norwegischen Fichte leben. In einem für die nördliche Hemisphäre typischen Nadelwald konnten in Fichtenwurzeln *nifH* Transkripte, die mit der Teilsequenz des *nifH* Gens von dem Grasendophyten *Azoarcus* sp. BH72 identisch sind, reproduzierbar über mehrere Monate hin detektiert werden. Diese Ergebnisse wurden mit Hilfe von klassischer, wie auch verschachtelter (‚nested') reverser Transkriptase-Polymerasekettenreaktion (RT-PCR) zur Erhöhung der Nachweisempfindlichkeit erhalten. Der Spiegel von *nifH* Transkripten in der Fichtenwurzel wurde anhand von Echtzeit RT-PCR beurteilt. Hierfür wurden sowohl *Azoarcus* sp. BH72-spezifische-, als auch universelle Primer verwendet. Die damit erhaltenen Ergebnisse zeigten, dass sich der *nifH* mRNA-Spiegel von *Azoarcus* sp. BH72 Transkripten in der Fichte mit den Jahreszeiten veränderte und im Sommer am höchsten war.

Zum Vergleich von kultiviertem mit pflanzenassoziiertem *Azoarcus* sp. BH72, wurden die 16S rRNA und *nifH* Gene sowie 16S rRNA und *nifH* Transkripte von diesem Bakterien in Reinkulturen und in Fichtenwurzeln mit Hilfe von Echtzeit RT-PCR quantifiziert und miteinander verglichen. Hierbei wurden statistisch signifikante Unterschiede gefunden. So reduzierte sich das Verhältnis zwischen 16S rRNA und *nifH* Transkripten in *A*. sp. BH72 im

Vergleich zur Reinkultur um die Hälfte, wenn das Bakterium mit der Fichte assoziiert war. Die Unterschiede waren noch viel größer auf DNA Ebene. Hier verschob sich das Verhältnis um mehr als den Faktor 60: Überraschenderweise wurden keine 16S rRNA Gene von *Azoarcus* sp. BH72 in Fichtenwurzeln entdeckt (Detektionsgrenze: sechs Kopien), während in denselben Wurzeln (der selben DNA Extraktion) 3×10^4 *nifH* Genkopien pro g Wurzel gefunden wurden; während in Reinkulturen auf eine *nifH* Genkopie vier 16S rRNA Genkopien kommen, waren es in Fichtenwurzeln weniger als 0,07 16S rRNA Genkopien. Dieses Ergebnis ließ darauf schließen, dass es sich bei *nifH* von *Azoarcus* sp. BH72 *in planta* um ein nichtchromosomales Gen handelte.

Abgesehen von den *nifH* Genen von *A*. sp. BH72 wurden auch in einer einzelnen Fichtenprobe *nifH* Transkripte gefunden, deren DNA Sequenz identisch zu dem entsprechenden Teil der *nifH* Sequenz von *Pseudomonas stutzeri* A1501 ist. Zusätzlich wurden *nifH* Transkripte mit 98,7% Sequenzidentität auf DNA Ebene zu Teilen der *nifH* Sequenz von *Rhizobium rosettiformans* detektiert, wenn die RNA mit einer alternativen RNA Extraktions Methode aus Fichtenwurzeln extrahiert wurde. Jedoch zeigten Ergebnisse, die mit Hilfe von Echtzeit RT-PCR erhalten wurden, dass die Kopienzahl von *nifH* und 16S rRNA Genen dieser Bakterien in Fichtenwurzeln unterhalb der Nachweisgrenze lagen. Anhand dieser Daten wurde geschlußfolgert, daß *Azoarcus* sp. BH72 der wichtigste aktive Stickstofffixierer in Fichtenwurzeln ist, und dort *Pseudomonas stutzeri* sowie *Rhizobium rosettiformans* in der N_2-Fixierung nur eine Nebenrolle spielen.

Neben PCR-abhängigen Methoden wurde in dieser Untersuchung auch ein Protein-basiertes, kultivierungsunabhängiges Verfahren zur Detektion und Identifikation von N_2-fixierenden Bakterien in Pflanzen entwickelt. Zu diesem Zweck wurden die Gesamtproteine von Fichtenwurzeln nach gelelektrophoretischer Auftrennung einer Western Blot Analyse unterzogen. Hierfür wurden polyklonale anti-NifH Antikörper gegen NifH von *A*. sp. BH72, sowie ein neu entworfener universeller polyklonaler anti-NifH Peptid Antikörper verwendet. Anschließend wurden die immunoreaktiven Proteine aus zweidimensionalen (2D) Gelen ausgeschnitten und über Flüssigkeits-Chromatographie mit Orbitrap Massenspektrometrie

Kopplung (LC-Orbitrap MS) analysiert und mit Hilfe von Sequenzvergleichen identifiziert. Alle NifH Peptid Sequenzen stimmten mit der Sequenz des Eisenproteins der Nitrogenase aus *A*. sp. BH72 überein. Auf grund der hohen Spezifität zweier Peptide konnte NifH von *Azoarcus* sp. BH72 in Fichtenwurzeln eindeutig identifiziert werden. Dies ist die erste Studie, in der es gelang, NifH in natürlich gewachsenen Pflanzen nachzuweisen, von denen bisher nicht bekannt ist, dass sie Symbiosen oder symbiotische Strukturen mit N_2-bindenden Bakterien ausbilden können.

Zu den methodischen Entwicklungen gehört, dass bei der Topoisomerase-abhängigen Klonierung von RT-PCR Produkten aus Fichtenwurzeln, deren Sequenzen identisch zu dem entsprechenden Abschnitt des *nifH* Gens von *A*. sp. BH72 sind, eine außerordentliche Verzerrung der Ergebnisse festgestellt wurde. Diese PCR Produkte waren kaum klonierbar, wenn sie mit Hilfe von Primern mit verbrückten Nukleinsäuren (engl. locked nucleic acid, LNA) hergestellt wurden. Die Topoisomerase-abhängige Klonierung von PCR Produkten mit anderen *nifH* Sequenzen erwies sich dagegen als unproblematisch. Aufgrund der Bedeutung dieser Beobachtung für diese Arbeit wurden die Effekte von LNA-Primern auf die Effizienz von Topoisomerase-abhängigen Klonierungen genauer untersucht. Hierbei wurden Anhaltspunkte dafür erhalten, daß die Verwendung von LNA-substituierten Primern in PCR Reaktionen mit *Taq* DNA Polymerase die Topoisomerase-abhängige Klonierung durch einen fehlenden 3‘-Überhang und durch die Basenzusammensetzung im Primerbereich behindern kann.

Zusammenfassend konnte in dieser Doktorarbeit erstmalig auf Protein- und Transkriptebene gezeigt werden, dass NifH von *Azoarcus* sp. BH72 für die Stickstofffixierung in der Rinde von Fichtenwurzeln von zentraler Bedeutung ist, ähnlich wie dies schon vorher für *nifH* mRNA in Kallargras beschrieben war. Diese Schlußfolgerung konnte aus Ergebnissen abgeleitet werden, die durch den erstmaligen Einsatz von proteomischen und PCR-basierten Methoden an Proteinen und Nukleinsäuren aus natürlich gewachsenen Fichten erhalten wurden.

1 Introduction

1.1 Nitrogen cycle and biological nitrogen fixation

The nitrogen cycle represents one of the most important nutrient cycles in ecosystems. It is a set of biogeochemical processes by which nitrogen undergoes chemical reactions, changes forms and moves through different reservoirs. Nitrogen can exist in both inorganic and organic forms as well as many different oxidation states. The conversions between the different forms and states of nitrogen are shown in Fig. 1.

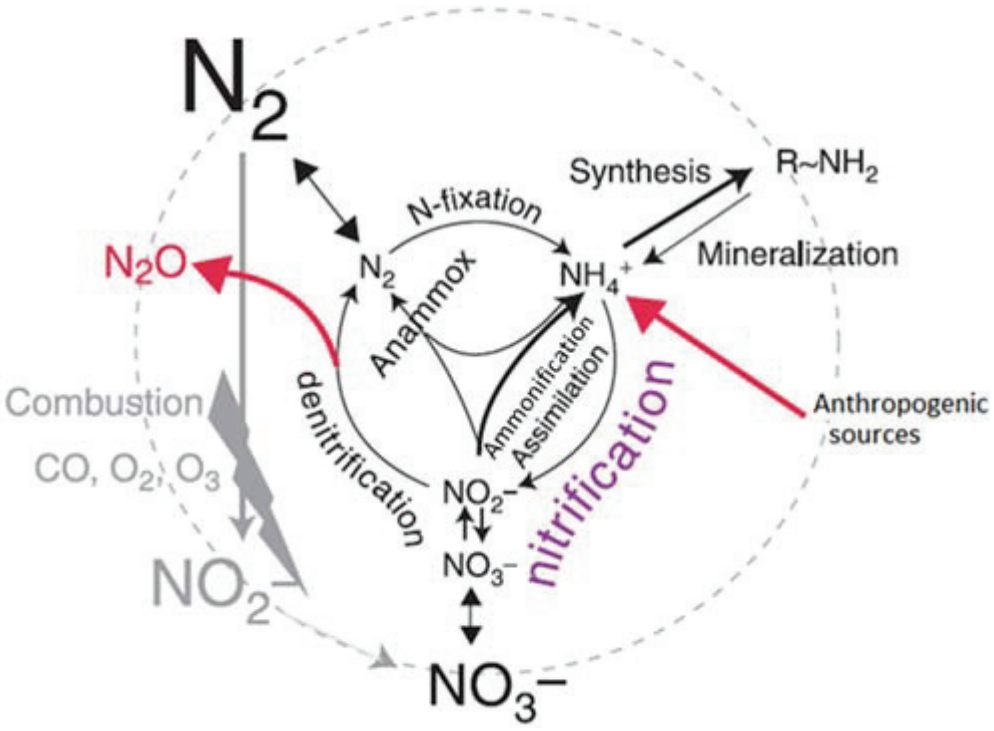


Fig. 1 The nitrogen cycle [compiled from (Klotz & Stein, 2008)].

Main processes in the nitrogen cycle include nitrogen fixation, assimilation (nitrogen uptake by organisms for growth), mineralization (decay), nitrification, denitrification, and the two less well characterized pathways, anammox (anaerobic ammonium

oxidation) and ammonification (dissimilatory nitrate reduction to ammonium). Microorganisms, particularly bacteria, play major roles in all of the principal nitrogen transformations.

Nitrogen fixation is a process to convert the atmospheric nitrogen (primarily N_2) to nitrogen compounds. Most elemental nitrogen is fixed through biological processes, which are exclusively carried out by prokaryotes (Burns & Hardy, 1975). Smaller, but significant amounts of nitrogen are fixed through high-energy non-biological events such as lightning, industrial manufacture of fertilizer, and fuel combustion (Tripathi*, et al.*, 2007). Nitrogen is an essential but most commonly deficient plant nutrient, since nitrogen in the atmosphere, constituting the largest reservoir, is metabolically not directly available to plants. Nitrogen applied in fertilizers usually provides benefit to plants, but the excessive applied nitrogenous fertilizers may be lost from the soil-plant system through leaching, volatilization, and denitrification (Granstedt, 2000), and may cause adverse environmental effects (Ayoub, 1999). Therefore, biological nitrogen fixation offers an economically and ecologically attractive way of reducing external nitrogen input (Peoples & Craswell, 1992).

1.2 Diazotrophic endophytes in seed plants

There are four established nitrogen fixing plant symbioses in seed plants (Table 1): (i) rhizobia symbiosis with legume or non-legume plants. The legume-rhizobial symbioses have been well examined due to their economic importance. *Parasponia* is the only non-legume plant which can establish a nodule symbiosis with rhizobia. (ii) *Frankia* symbiosis with actinorhizal plants. Actinorhizal plants are only nodulated by strains of actinomycetes from the genus *Frankia*. (iii) *Nostoc* symbiosis with *Gunnera*. *Gunnera*-*Nostoc* symbiosis is the only known symbiosis where a cyanobacterium resides intracellularly in an angiosperm. *Nostoc* infects specialized glands which are located on the stems of *Gunnera*, but in contrast to nodular symbioses the two symbionts show no coordinate differentiation (Seckbach, 2002), and the glands are formed before the symbiosis initiated. (iv) Cyanobacteria symbiosis with cycads.

Cycads represent the only gymnosperms known to establish a symbiosis with diazotrophs. Cyanobacteria occupy a zone in the outer part of the root cortex of the coralloid roots, the highly specialized lateral roots of cycads, but remain extracellular (Vessey, *et al.*, 2005).

Table 1. Established nitrogen fixing plant symbioses in seed plants (spermatophyta) [a]

Plant symbionts	Bacterial symbionts	Location of bacterial symbionts	Symbiotic structures induced
Gymnosperms			
-Cycads	Cyanobacteria (generally *Nostoc*)	extracellular	no
Angiosperms			
Core eudicot -*Gunnera* (Haloragaceae)	*Nostoc*	intracellular	no
-Rosid cluster I (Rosidae)	*Frankia, Rhizobia* [b]	intracellular	yes, nodules

[a], compiled from (Stacey, *et al.*, 1992, Soltis, *et al.*, 1995, Gualtieri & Bisseling, 2000, Rai, *et al.*, 2000, Moulin, *et al.*, 2001, Yates, *et al.*, 2007, Marchetti, *et al.*, 2011)

[b], α-rhizobia: *Brady-*, *Sino-*, *Meso-*, *Azorhizobium*, *Rhizobium, Methylobacterium*; ß-rhizobia: *Cupriavidus*, *Burkholderia*

For these four types of nitrogen fixing symbioses in seed plants, reconstitution of the functional symbioses was achieved. The microsymbionts were confirmed to have the capacity to fix nitrogen within plant and transfer the fixed nitrogen to the host plant. Symbiotic plants were shown to grow completely independent of other combined-nitrogen sources (Silvester & Smith, 1969, Callaham, *et al.*, 1978, Torrey, *et al.*, 1981, Yoneyama, *et al.*, 1986, Bergersen & Postgate, 1987, Lindblad, *et al.*, 1991, Bothe, *et al.*, 2007). Other diazotrophs closely associated with plants were defined as diazotrophic endophytes. Plant endophytes refer to microorganisms (i) colonizing the internal tissues of health plants, but are not residing in living plant cells or surrounded by a host membrane compartment; (ii) they typically live in a tight association with their plant host and, unlike *Rhizobium*, cannot be isolated from root-free soil, since they don't survive well without the plant host; (iii) they spend most

of their life cycle inside plant without causing symptoms of plant damage (Reinhold-Hurek & Hurek, 1998). Diazotrohic endophytes such as *Azoarcus* sp. BH72 (Reinhold*, et al.*, 1986), *Herbaspirillum* seropedicae (James & Olivares, 1998) and *Gluconacetobacter* diazotrophicus (James*, et al.*, 1994) can infect the interior of graminaceous plants without causing symptoms of plant disease. The contribution of nitrogen fixation by these endophytes to growth of certain graminaceous plants was confirmed by ^{15}N procedures (Sevilla*, et al.*, 2001, Hurek*, et al.*, 2002, Oliveira*, et al.*, 2002). However, *Azoarcus* sp. BH72 is the only cultivated diazotrophic grass endophyte, for which molecular evidence is suggesting, that the capability of this bacterium to provide nitrogen derived from N_2-fixation to inoculated Kallar grass plants may be of relevance in uninoculated, naturally growing plants. Only this bacterium has been shown to be the most active nitrogen fixing bacterium within the natural population of diazotrophic bacteria in roots of such plants (Hurek*, et al.*, 2002, Reinhold-Hurek & Hurek, 2011).

Nevertheless, it is still not known whether any flowering plant except those eudicots, which are able to establish an intracellular symbiosis with certain nitrogen fixing bacteria (Table 1), can grow completely independent of any combined-nitrogen source (Boddey*, et al.*, 2003, Giller & Merckx, 2003). Much less is known in this respect on non-flowering seed plants, apart from cycads. In this study, culture-independent methods were used to detect the major diazotrophs in a natural stand of Norway spruce, a gymnosperm, which is not known to form any nitrogen fixing symbiosis or symbiotic structures with diazotrophic bacteria.

1.3 Nitrogen fixation research on forest ecosystem

Nitrogen is a critical, limiting nutrient in most temperate forests (Vitousek & Howarth, 1991). It was reported that N_2-fixation by free-living and associative bacteria occurs in forest soils (Dawson, 1983, Limmer & Drake, 1996), mainly in forest litter and in the plant rhizosphere (Granhall & Lindberg, 1978). Several reports showed different rates of N_2-fixation, depending on forest type (Roskoski, 1980, Dawson, 1983, Baker &

Attiwill, 1984, Jones & Bangs, 1985, Jurgensen*, et al.*, 1987, Heath*, et al.*, 1988). Average rates documented on the forest floor of northern temperate forests are nearly 1 kg N ha^{-1} yr^{-1} and most rates were estimated to be less than 10 kg N ha^{-1} yr^{-1} (Pérez*, et al.*, 2003), but much higher estimates have been reported as well (Bormann*, et al.*, 1993). It was reported that nitrogen fixation by diazotrophic microorganisms is the primary source of N addition to undisturbed, unfertilized forest ecosystems (Cleveland*, et al.*, 1999). However, there is no agreement between researchers on the ecological role of nitrogen fixation in forest ecosystems (Różycki*, et al.*, 1999). Bormann *et al.* (1993) claimed that pine forest ecosystem can be nitrogen self-sufficient due to non-symbiotic N_2-fixation, while Barkmann and Schwintzer (Barkmann & Schwintzer, 1998) argued that N_2-fixation rates associated with pine are low and have little impact on the N budget of pine forest ecosystem.

Conifers are of great ecological and commercial importance in the northern hemisphere temperate forests and in similar climates all over the world, where they often form the dominant vegetation cover. With a scarcity of known endophytic interactions in coniferous forests and considering that conifer communities normally carry ectomycorrhizas (Bewege et al., 1978), associative N_2-fixing bacteria, occurring within the fungal sporocarps and mycelia as well as within the mycorrhizal roots (Florence & Cook, 1984, Li & Hung, 1987), were assumed to be more important than free-living diazotrophs, because of their direct impact on the mycorrhizal symbiont and the host plant (Garbaye, 1994). Mycorrhizal tips and the vegetative mycelium represent a niche that is ideally suited for the enrichment of associative bacteria (Rygiewicz*, et al.*, 1984, Li*, et al.*, 1992, Różycki*, et al.*, 1999, Sen, 2000, Hawkes, 2003). Although the role of ectomycorrhiza-associated bacteria is not fully understood, some reports suggest that they may act as “mycorrhization helper bacteria” (MHB) that promote mycorrhizal formation and improve both mycorrhizal and plant growth (Garbaye, 1994, Enebak*, et al.*, 1998, Poole*, et al.*, 2001). Recent studies (Li et al., 1992; Perez-Moreno and Read, 2000; Sen, 2000) have suggested that nitrogen fixers may be associated with mycorrhizas, more specifically in the mycorrhizosphere of

forest trees, and that a close association between mycorrhizas and N_2-fixing bacteria could contribute significantly to soil, fungal and plant nitrogen in nutrient-poor ecosystems (Burke*, et al.*, 2006). Mycorrhiza associated diazotrophs were mostly studied using culturing techniques. Different species of mycorrhiza-associated diazotrophs were isolated, e.g. *Azospririllum* spp. from sporocarps of ectomycorrhizal fungi (Li & Hung, 1987), *Bacillus* sp. from the tuberculate ectomycorrhizas of Douglas-fir (Li*, et al.*, 1992), from Scots pine ectomycorrhizas (Pachlewski*, et al.*, 1991) and *Paenibacillus amylolyticus* from the inner tissue of *Suillus tomentosus* tuberculates ectomycorrhizae (Paul*, et al.*, 2007).

Studies of free-living diazotrophs in the coniferous forest soil were carried out with traditional culture-dependent methods as well as molecular techniques. Culture-independent studies on the diversity and phylogeny of diazotrophs have been carried out by targeting the *nifH* gene, encoding the iron protein of nitrogenase. Researchers (Richards, 1973, Grayston*, et al.*, 1997, Różycki*, et al.*, 1999) reported that diazotrophic bacteria were found to be less numerous in root-free than in rhizosphere soil, as bacteria are more frequent and metabolically active in the root zone due to higher carbon inputs in the form of root exudates than in root-distant soil. Several studies showed that N_2-fixing bacteria occur abundantly in the upper soil layer (5 cm depth) and their abundance decreases with soil depth (Mergel*, et al.*, 2001, Rosch*, et al.*, 2002). Low levels of diversity among N_2-fixing bacteria were found in a soil of a Douglas-fir forest (Widmer*, et al.*, 1999) and a soil of loblolly pine plantation (Burke*, et al.*, 2006), and also in an acid soil of oak-hornbeam forest, which is in contrast to the apparent richness of 16S rRNA gene sequences in the same soil sample (Rosch*, et al.*, 2002).

Data on the composition of the bacterial community obtained with cultivation-dependent methods do not necessarily reflect the abundance of the diazotrophs in the samples analyzed. Though several studies have successfully targeted *nifH* genes within conifer forest samples using nested PCR approach

(Widmer*, et al.*, 1999, Shaffer*, et al.*, 2000, Levy-Booth & Winder, 2010), little is known about the diversity of bacterial expression of *nifH* by examining *nifH* transcripts. Up to now there is only a single report on *nifH* transcript diversity in conifer forest samples. In this work on ectomycorrhizas of Corsican pine (Izumi*, et al.*, 2006), it is reported that mainly methanotrophic alphaproteobacteria such as *Methylocella* spp. and *Methylocapsa* spp. as well as betaproteobacteria of the genus *Burkholderia* spp. are transcribing *nifH*. However, nitrogenase gene expression was only rarely detected in the samples examined, suggesting that nitrogen fixation did not play an important role in providing nitrogen to the ectomycorrhizal roots investigated. Taken together, studies of diazotrophs in conifer forest by molecular techniques are rare and the limited data from either culture-dependent or culture-independent studies are not sufficient to draw conclusions on important nitrogen fixing bacteria in conifer forest ecosystems.

The application of mRNA-based methods to conifer forest samples has been hampered by several methodological challenges. It is often difficult to obtain high-quality RNA from conifer tissues which is rich in polyphenols, polysaccharides and RNases leading to RNA degradation and contaminating RNA with reverse transcription (RT)-PCR inhibitors (Kolosova*, et al.*, 2004). Furthermore, due to the degeneracy of the genetic code and the divergence of *nifH* genes, (universal) *nifH* primer sets with broad–range coverage usually have a high degree of DNA sequence degeneracy, which is compromising the specificity and efficiency of the PCR reaction. Accordingly, the application of highly degenerate *nifH* primers in RT-PCR reactions to analyze complex environmental samples often results in an insufficient amplification of *nifH* transcripts.

In this study, data on diversity and abundance of *nifH* transcripts associated with coarse Norway spruce roots are presented. To facilitate the isolation of suitable RNA for RT-PCR applications from such recalcitrant materials, the well known RNA extraction protocol from Chang *et al*. (1993), had to be slightly modified by including a

proteinase K purification step. In this protocol the RNA extraction buffer was developed to include reagents overcoming several common problems associated with this type of tissue, for example, PVP and ß-mercaptoethanol as reducing reagents to avoid oxidation of nucleic acids by phenolic compounds, CTAB as detergent combined with chloroform to remove proteins instead of using phenol, high concentration of sodium chloride to remove polysaccharides. RNA extracts were purified by proteinase K and DNase I treatment to further remove proteins and remove DNA contamination, respectively. With the newly introduced proteinase K step the resulting RNA was suitable for the subsequent enzymatic reactions. Broad-range *nifH* primers (Zani*, et al.*, 2000) were used in RT-PCR to amplify *nifH* transcripts. This approach is widely used for the detection of nitrogen-fixing microorganisms in environment samples. Because of the tight relationship between nitrogenase activity and expression of *nifH* (Egener*, et al.*, 1998, Dixon & Kahn, 2004), communities that are actively transcribing *nifH* are also likely to fix nitrogen. Moreover, the degenerate primers used here have been shown to amplify phylogenetically diverse *nifH* without bias in an artificial mixture of *nifH* templates (Tan*, et al.*, 2003). The compromised specificity and sensitivity of PCR amplification caused by the high degeneracy of primers could be to some extent complemented by using a nested PCR approach or by substituting locked nucleic acid (LNA) for DNA at selected positions within these primers (Burbano*, et al.*, 2010). RT-nested PCR and single stage PCR using LNA-substituted primers were applied to the same spruce RNA extracts to address the potential bias in the RT-nested PCR approach caused by the increased number of PCR cycles. Finally, also data on 16S rRNA / rDNA amplicons from spruce samples were included to relate the community structure of the diazotrophs with that of the total bacterial community. This is the first report on *nifH* transcript diversity in coniferous medium-coarse roots.

1.4 Quantitative PCR and its application to this study

Quantitative PCR or qPCR is now widely used in microbial ecology to determine gene

numbers present within environmental samples. Furthermore, reverse transcription (RT) analyses are increasingly combined with qPCR in RT-qPCR assays, offering a powerful tool for quantification of gene expression and relating biological activity to ecological function (Smith & Osborn, 2009). Compared to other methods for the study of gene expression e.g. microarray hybridization, quantitative PCR analysis is the most sensitive method for detection of low abundance RNA. The highest sensitivity, accuracy and specificity are all achieved by qPCR (Bustin, 2000), while microarray hybridizations offer the advantage of a high-throughput technique which allows to simultaneously analyze many genes (Cho & Tiedje, 2002).

Quantitative PCR can be performed using three different methods: most probable number (MPN)-PCR, competitive PCR and real-time PCR. The advantages and limitations of each method were evaluated by multiple researchers. MPN-PCR consists of serial dilutions of DNA / cDNA samples until extinction and replicated PCR reactions for each dilution. The result of each reaction is scored positive or negative after gel-electrophoresis analysis. The number of gene copies is calculated using MPN statistics (Philippot, 2006). Competitive PCR is based on the simultaneous amplification of the target and a competitor DNA in a single tube. The competitor molecule must have the same primer binding sites, but differs in size from the target gene, making a co-amplification possible during the PCR and allowing the distinction of the amplicons by electrophoresis. Data analysis is usually performed at the plateau phase assuming that both templates are amplified with the same efficiency (Sharma*, et al.*, 2007). Both, the competitive PCR and MPN-PCR require a post-PCR processing, which is very labour-intensive and requires a large number of PCR reactions per sample in quantitative MPN PCR applications. Real-time PCR is based on detection of fluorescence generated during synthesis of PCR amplicons by *Taq* polymerase. The quantification of amplicons in the exponential phase of the PCR, when the efficiency is recognized to be the highest, is one of the greatest advantages of the real-time PCR in comparison to other quantification PCR methods (Sharma*, et al.*, 2007). Other advantages of real-time PCR are an increased dynamic range and

low detection limit, no post-PCR manipulation, a reduced risk of cross contamination and therefore all in all a quick, reproducible and less labour-intensive procedure. Although real-time PCR requires expensive equipment and reagents, this method is most widely used.

Briefly, the real-time RT-PCR technique uses the same enzymatic reaction procedures as conventional RT-PCR. It involves applying reverse transcriptase to total isolated RNA to produce cDNA that is used as template in the subsequent PCR along with a primer set designed to amplify the genes of interest. The real-time PCR integrates the amplification and analysis steps by monitoring the DNA produced during each PCR cycle. Fluorescent values are documented at the end of each qPCR cycle and the amount of fluorescence is directly related to the amount of product. The three most used real-time PCR fluorescent chemistries are SYBR Green (Higuchi*, et al.*, 1992), TaqMan (Heid*, et al.*, 1996) and molecular beacons (Tyagi & Kramer, 1996). SYBR Green is the simplest and most affordable method, and consists of a fluorescent dye that binds to double-stranded DNA. The unbound dye exhibits little fluorescence in solution. The SYBR Green assay is very sensitive but has diminished specificity, as the dye binds to all double-stranded DNA present, and primer dimers may result in a false reading (Bustin, 2000).

In order to obtain meaningful data on gene expression, it is necessary to extract high-quality and intact RNA from environmental samples, upon which all of the subsequent manipulations and analyses are dependent. Purity and integrity of RNA are critical elements for real-time RT-PCR, as well as other RNA-based analyses. Starting with low quality RNA may strongly compromise the results of downstream applications. The assessment of RNA quantity and quality can be done by various methods such as, micro-volume spectrophotometry (NanoDrop, Thermo Scientific), microfluidic analysis (Agilent Technologies' Bioanalyzer, Bio-Rad Laboratories' Experion), and RiboGreen (Bustin*, et al.*, 2009). These methods produce different total RNA quantification results, making it advisable to measure all samples with a

single method (Bustin*, et al.*, 2009).

In this study, three *nifH* transcript phylotypes were described, which were either identical or highly similar (98.7% DNA sequence similarity) to the corresponding *nifH* sequence from *Azoarcus* sp. BH72 and *Pseudomonas stutzeri* A1501 or *Rhizobium rosettiformans*, respectively (Results section 3.1). Only *nifH* transcripts identical to the partial *nifH* sequence of *A.* sp. BH72 were repeatedly detected and identified to be predominant in several months' samples, but the expression levels of *nifH* gene could not be represented by the end-point RT-PCR. Therefore real-time (RT)-PCRs were carried out (i) using strain BH72 specific *nifH* primer set to quantify the BH72 specific *nifH* transcripts in the approximately same amount of spruce total RNA extract from different samples; (ii) using universal *nifH* primers sets to quantify the total *nifH* transcripts in different samples and to compare them with the data obtained from (i); (iii) using BH72 specific *nifH* and 16S rRNA gene primer sets to compare the ratios of 16S rRNA transcripts to *nifH* transcripts and of 16S rRNA genes to *nifH* genes in pure cultures with those *in planta*. Since the RNA extracts containing *P. stutzeri* and *R. rosettiformans nifH* transcripts had been used up in earlier experiments, the quantification of *nifH* transcripts was only performed with primer set specifically targeting *nifH* from *A*. sp. BH72.

1.5 Proteomics and its application to environmental samples

Compared to nucleic acids, proteins are promising alternative markers of biological function, since they reflect the actual activity with respect to metabolic reactions and regulatory cascades and provide more direct information about microbial activity than functional genes and even their corresponding messenger RNAs (Wilmes & Bond, 2006). The term "proteomics" was first coined in 1995 (Wilkins*, et al.*, 1996) and was defined as the large-scale characterization of the entire protein complement of a cell line, tissue, or organism. Today the definition of proteomics does not only include the complete protein complement of the expressed genome, but also encompasses diverse techniques that provide a macroscopic view of what is expressed and present

under different growth conditions. Complementary to the information of genome and transcriptome, proteomics not only characterizes the final gene products, but also provides detailed information about protein abundances, stabilities, turnover rates, posttranslational modifications and protein-protein interactions (Keller & Hettich, 2009). Proteomics is a rapidly developing and technology-driven discipline with high potential. As in the early emerging stage of proteomics, the 2-dimensional (2-D) gel-based protein identification was time consuming and tedious due to the lack of sensitive and fast sequencing technologies for protein analysis, and thus information could be provided on only the most abundant proteins. With the advent of more sophisticated and higher-throughput chromatographic-mass spectrometric instrumentation, proteomics has greatly advanced with respect to coverage. In parallel with the progress made in bioinformatics, it is now possible to identify proteins by database searching with mass spectrometry (MS) information and to characterize the corresponding genes by reverse genetics (Wilmes & Bond, 2006). At present it is possible to use proteomic approaches to identify at least 50 to 70 % of the predicted proteome for most bacteria grown under a single growth condition, while estimations suggested that most bacteria may deploy only 50 to 80 % of their predicted genes under a single growth condition (Keller & Hettich, 2009). Thus the current level of proteome measurement is far beyond “only the most abundant” (proteins could be studied) stage.

Wilmes and Bond proposed the term “metaproteomics” for the large-scale characterization of the entire protein complement of environmental microbiota at a given point in time (Wilmes & Bond, 2004). Environmental samples are often highly complex, which makes proteome studies in this field especially challenging. Some of the challenges are the lack of genome sequences for the vast majority of environmental bacteria, difficulties in isolating bacteria and proteins from certain environments, and the presence of complex microbial communities (Lacerda & Reardon, 2009). Despite these challenges, environmental proteomics offers a unique view to further our knowledge on e. g. function, cellular location, posttranslational

modification and source of proteins in environmental samples and complements the nucleic acid-based approaches to study activity and diversity in microorganisms.

Environmental proteomic analysis implies the development of different technical steps, from the extraction of microbial proteins from the environmental matrix to the resolution of their diversity and identification. There are two major approaches for environmental protein extraction: (i) direct extraction of bulk protein from environmental samples and (ii) separation of the organisms from environmental matrix and subsequent extraction. For the first approach, organisms are lysed directly in the environmental sample and the resulted protein extracts appear to be very complex and vary according to the target environment and the surrounding local conditions (Taylor & Williams, 2010). *In situ* lysis facilitates a thorough protein recovery from indigenous bacteria, fungi, protozoa and multicellular organisms, this mixture being likely to introduce difficulties in the taxonomic delineation of the detected proteins. Furthermore, a direct lysis approach applied to environmental samples makes it technically difficult for subsequent protein characterization due to the interfering compounds contaminated in the protein extracts (Maron*, et al.*, 2007). As with *in situ* nucleic acid-based studies, the most crucial step is ensuring that the quality and quantity of the proteins extracted is representative of the samples. This is particularly true for environmental proteomics studies because of the complexity of indigenous microbial communities, the heterogeneity of the natural environment, and the presence of interfering compounds (e.g. phenolic compounds, humic acids), which make it difficult to extract a suitable protein fraction for analysis (Maron*, et al.*, 2007). In contrast with the direct lysis approach, in the indirect lysis approach, proteins are extracted, purified from organisms that have been previously separated from the environmental matrix (Maron*, et al.*, 2004, Taylor & Williams, 2010). This strategy allows the precise targeting of the bacterial fraction and to obtain a cellular fraction only slightly contaminated by interfering compounds. However, the efficiency of bacterial separation is dependent on the physicochemical characteristics of the environmental sample, which may introduce further bias in the quantitative and

qualitative recovery of environmental bacterial proteins (Maron*, et al.*, 2007). In a pioneer study, Schulze *et al.* (Schulze*, et al.*, 2005) analyzed proteins isolated from dissolved organic matter in different environments and showed that the relative proportion of the proteins originating from bacteria varied from 78 % in lake water to less than 50 % in a forest soil solution.

Once the protein extracts are obtained, the subsequent characterization and identification of proteins is done by mass spectrometry (MS)-based proteomics experiments can be separated into two major approaches: (i) 2-D gel electrophoresis with either peptide mass fingerprinting or tandem MS (MS/MS) and (ii) liquid chromatography (LC)-MS/MS (Keller & Hettich, 2009). Despite the technical limitations of 2-D gel electrophoresis, such as a reliably monitoring low abundant, very hydrophobic, very acidic or very basic proteins (Lee, 2001), this gel-based analysis still provides a vast amount of data for proteomics studies. Wilmes and Bond analyzed the metaproteome in a laboratory-scale activated sludge by combining 2D-PAGE with MS. Highly expressed protein spots were excised and could be identified with MS (Wilmes & Bond, 2004). This demonstrated the possibility to isolate and identify microbial proteins from a complex environment. This work focused on interesting spots from the entire proteome of a laboratory-scale activated sludge system. Likewise, dinitrogenase reductase (NifH protein) associated with the roots of Norway spruce was intended to be detected, here.

As nitrogen fixation is performed exclusively by prokaryotes, and data on nitrogen fixation associated with conifers are contradictory (Richards, 1964, Bormann*, et al.*, 1993, Barkmann & Schwintzer, 1998, Binkley, 2002), only low amounts of dinitrogenase reductase protein were to be expected in spruce roots. Here, studies on detection and identification of NifH in spruce roots are reported. This is the first time that Western blot analysis in combination with two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was applied to environmental samples and the immunoreactive protein spots were subsequently identified by MS analysis. There

were three tasks in this study which needed to be addressed: (i) a protein extraction method had to be established which allowed obtaining proteins from spruce roots suitable for 2D gel electrophoresis analysis; (ii) the performance of anti NifH antibody and anti NifH peptide antibodies in immunodetection experiments on environmental samples had to be evaluated. A polyclonal antibody raised against the nitrogenase Fe protein from *Azoarcus* sp. BH72 (Oetjen*, et al.*, 2009) was used. Nitrogenase is a conserved protein and all polyclonal antibodies to the Fe protein of nitrogenase tested so far cross-react with heterologous nitrogenase proteins (from other species) to some extent (Zehr*, et al.*, 1990). Universal polyclonal antibodies are commercially available from Agrisera, but application data are rare and only available for marine diazotrophs (Kupper*, et al.*, 2008, Levitan*, et al.*, 2010). For this purpose a universal polyclonal anti-peptide antibody directed against a conserved peptide of the nitrogenase Fe protein was designed based on sequence information available in public databases. Performance of the peptide antibody in Western blot assays for specific detection of divergent nitrogenase iron proteins was tested using nitrogen-fixing pure cultures and environmental samples; (iii) Last but not least, predominant iron proteins synthesized *in situ* had to be identified by MS analysis of immunoreactive spots from 2D-gels. Results obtained from the nucleic acid-based studies suggested that *nifH* transcripts from *Azoarcus* sp. BH72 were predominant in most of the samples evaluated. However, since this bacterium is widely used in this laboratory, random contamination of PCR reactions with genomic templates or templates stemming from PCR reactions from this bacterium cannot be completely excluded. If protein data would be consistent with *nifH* expression data, contamination was not likely to play a considerable role.

1.6 LNA-substituted primer and potential biases of TOPO cloning

Locked nucleic acids (LNAs), first described by Singh *et al*. (Singh*, et al.*, 1998), are a novel class of conformational restricted nucleotide analogs. LNA is a bicyclic nucleic acid where a ribonucleoside is linked between the 2′-oxygen and the 4′-carbon atom

to a methylene unit, which results in a locked *3'*-endo conformation that reduces the conformational flexibility of the ribose. The incorporation of LNA into oligonucleotide primers has been shown to increase template binding strength and specificity for DNA amplification (Ballantyne*, et al.*, 2008).

Over the past two decades, culture-independent molecular techniques have improved our view of microbial community to an unprecedented level (DeLong, 2002). Apart from 16S rRNA genes, which are now commonly used for the inventory of prokaryotic diversity, several key genes for certain metabolic processes are often employed as target genes for the functional analysis of microbial communities. For the process of nitrogen fixation, the Fe protein gene (*nifH*) of the evolutionary conserved nitrogenase protein complex is particularly suitable for phylogenetic analysis. The specificity and sensitivity of *nifH* (RT)-PCR was improved by using LNA-substituted primers instead of corresponding DNA primers (Burbano*, et al.*, 2010).

To analyze the heterogeneous pool of PCR products, clone libraries are superior to fingerprinting based methods, when explicit data on abundance and diversity of phylotypes are required. Making use of the non-templated addition of adenosine to the 3' end of PCR products by *Taq* polymerase, the commercial TA cloning system from Invitrogen has been widely adopted. Cloning biases are introduced through mainly three aspects (Taylor*, et al.*, 2007): (i) the template-independent terminal transferase activity of *Taq* polymerase is probably juxtaposed sequence-related. Brownstein *et al.* (1996) reported that non-templated A addition adjacent to a 3' terminal C was favored and to a 3' terminal A was not. In addition to the 3' terminal base, the adjacent nucleotides may also affect the fraction of adenylated PCR product. Incomplete addition of A to a majority of PCR products potentially decreases the cloning efficiency of such products; (ii) certain insertions transcribed under the promoter of the T vector may interfere with the growth of transformed *Escherichia coli* cells, eliminating these clones from the library; (iii) it is known that reaction kinetics favor insertion of shorter DNA fragments over longer ones. Huber *et al.* (2009)

constructed libraries with different sizes of amplicons of the 16S rRNA gene, and found that the library with the smallest amplicons contained more unique sequences, a higher diversity, and a different community structure than the library with larger amplicons. Nevertheless, these surveys were all based on cloning PCR products obtained with DNA primers. So far there was no report on possible effects caused by TA cloning of PCR fragments obtained with LNA-substituted primers.

In this study, BH72 *nifH* amplicons could be obtained by single stage RT-PCR from spruce samples only by using LNA-substituted primers, but the amplicons were almost not clonable with the TOPO TA cloning kit (data not shown). It is shown here, that this inefficient TOPO TA cloning was caused by the failure of *Taq* DNA polymerase to produce 3' overhangs on PCR fragments obtained with LNA-substituted primers. The effect of certain LNA substitutions on *Vaccinia* topoisomerase I ligation was estimated as well.

1.7 Research outline

This dissertation aims to identify the major diazotrophs associated with Norway spruce roots. For this purpose the diversity and abundance of *nifH* transcripts were studied in spruce roots (3.1). The most abundant *nifH* transcripts were quantified by real-time RT-PCR and monitored over several months. *In planta* and *ex planta nifH* transcription of the most active nitrogen fixer was compared by normalization with 16S rRNA transcript levels which were estimated by real-time PCR as well (3.2). Furthermore, *nifH* expression was studied at the protein level by two-dimensional Western blot analysis using NIfH antibodies and mass spectrometry (3.3). Finally, a PCR cloning bias caused by using LNA-substituted primers was investigated (3.4).

2 Materials and Methods

2.1 RT-PCR studies

2.1.1 Sampling

The root samples of Norway spruce (*Picea abies* L. Karst) were collected from the Solling plateau in Central Germany (51°46'N, 9°35'E; elevation 500 m), where the Norway spruce plantation is located on a strongly acidified, low-fertile loam-silt (Martinson*, et al.*, 2005). Sampling was conducted several times during a two-year period, from July 2006 to October 2007. For this purpose roots, about 2 cm in diameter, were dug out from the organic layer of soil, 5 to 10 cm in depth and adhering soil and the cortex remnants were carefully removed. After removing most of the stone-cork cells of the outer cork layer (phellem) the remaining bark consisting of the phellem, phellogen (bark cambium) and phelloderm was peeled off, and wrapped in aluminum foil. After snap freezing in a liquid nitrogen dewar, samples were transported back to the lab, and stored in liquid nitrogen until further processed. Root-free soil samples were collected from the rhizosphere of spruce roots in Sept. 2007.

2.1.2 RNA and DNA extraction

Protocol I: RNA and DNA extraction followed Chang's protocol (Chang*, et al.*, 1993), as described (Knauth*, et al.*, 2005) with slight modifications. Instead of 700 mg plant material 300 mg root bark (fresh weight) were used and after LiCl precipitation a proteinase K treatment followed by a phenol / chloroform extraction was included. LiCl-precipitated nucleic acids, consisting mainly of RNA, were treated with proteinase K (final concentration 1mg / ml, from Merck, Darmstadt, Germany) at 60℃ for 60 min, while the supernatant was precipitated with isopropanol to retrieve the portion of nucleic acids containing most of the DNA of the sample. After proteinase K treatment,

the RNA was purified by extraction with acid phenol / chloroform and then precipitated with ethanol and sodium acetate from the water phase. The DNA-free RNA was dissolved in 20 µl RNA secure solution (Ambion, Austin, TX, USA) and stored in liquid nitrogen. The isopropanol-precipitated DNA was spun down, washed twice with 70 % ethanol, dissolved in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and stored at -20℃. The NanoDrop (Thermo scientific, Wilmington, USA) was used to quantify the extracted RNA and DNA. RNA and DNA extractions from soil samples were performed with the same protocol, replacing 300 mg spruce roots by 300 mg soil. Protocol II (only for RNA extraction): RNA extraction was done as described in protocol I until the LiCl precipitation step. Thereafter the RNA preparation was subjected to the RNeasy Plus Mini kit (Qiagen, Hilden, Germany) to further purify and remove the co-precipitated DNA according to the manufacturer's instruction.

2.1.3 Primers

Primers for RT-PCR, RT-nested PCR and clone library screening were listed in Table 2. All DNA oligonucleotides were synthesized by Eurofin MWG GmbH (Ebersberg, Germany) and LNA-substituted oligonucleotides were synthesized by Eurogentec (Seraing, Belgium).

Table 2. Oligonucleotide primer sequences

Primer name	Sequence (5'-3') [a]	Reference
nifH1	tgygayccnaargcnga	(Zehr & McReynolds, 1989)
LNA-Zehr-F	tGygAyCcnaargcnga	(Burbano*, et al.*, 2010)
nifH2	adngccatcatytcncc	(Zehr & McReynolds, 1989)
LNA-Zehr-R	aDngCcAtcatytcncc	(Burbano*, et al.*, 2010)
nifH3	atrttrttngcngcrta	(Zani*, et al.*, 2000)
LNA-nifH3	aTrtTrtTngcngcrta	This study
nifH4	ttytayggnaarggngg	(Zani*, et al.*, 2000)
R518	attaccgcggctgctgg	(Muyzer & Smalla, 1998)
F357	ctacgggaggcagcag	(Muyzer & Smalla, 1998)
GC-clamp	cgcccgccgcgccccgcgccc ggcccgccgcccccgccccc	(Muyzer & Smalla, 1998)
BH-118F	tgcgacccgaaggctga	This study
BH-214R	ggtcttcaaccgagccg	This study

[a], y = t or c, r = a or g, d = a, g or t and n = a, c, g, or t; LNA bases in capitals.

2.1.4 Reverse transcription (RT)-PCR

To amplify *nifH* transcripts, RT-PCR (Burbano*, et al.*, 2010) and RT-nested PCR (Zani*, et al.*, 2000) were carried out as described previously. Reverse transcription reactions were performed in a 20 µl system with SuperScript III (Invitrogen, Darmstadt, Germany) according to the manufacturer's instruction, using 200 ng RNA as template and LNA-nifH3 as primer for cDNA synthesis. PCR was carried out with LNA-Zehr-F and R, while nested PCR was carried out with DNA oligonucleotide primers, primer set nifH3 and nifH4 for the first round PCR and primer set nifH1 and nifH2 for the second round PCR. When the RT- nested PCR products needed to be analyzed with DGGE, the GC-clamp-nifH1 was used in the second-round PCR in the nested PCR procedure. RT-PCR amplification of 16S rRNA fragments was performed with Ready-to-Go RT-PCR beads (Amersham, Munich, Germany) according to the manufacturer's instruction, using R518 as primer at the RT step. PCR was carried out as previously described (Diallo*, et al.*, 2004). RT-PCR products were separated on a 2% agarose gel in Tris-Acetate-EDTA (TAE) buffer (1 L of 50 × TAE buffer including: 242 g Tris base, 0.5 M pH 8.0 EDTA and 57.2 ml glacial acetic acid). After electrophoresis, the gel was stained with ethidium bromide (0.5 µg / ml) and the band pattern was documented by taking photos with the Image Master VDS camera (GE Healthcare, Fairfield, CT, USA) or by scanning with the Typhoon 8600 Variable Mode Imager (Amersham Pharmacia Biotech, Sunnyvale, CA, USA).

2.1.5 Denaturing gradient gel electrophoresis (DGGE)

The DGGE analysis of RT-PCR products was carried out with the D-Code system (Bio-Rad, Munich, Germany) as described by Demba Diallo (2004). Approximately 500 ng of PCR products for each lane were loaded onto 8% (w / v) polyacrylamide gels, 1 mm thick, running in 1 × TAE buffer. The denaturing gradient contained 35 - 80% denaturant (100% denaturant corresponded to 7 M urea and 40% deionized

formamide). Electrophoresis was performed at a constant voltage of 75 V for 16 h and at 60℃. After electrophoresis, the gels were stained with 1 × SYBR Gold (Invitrogen, Darmstadt, Germany) for 30 min with gentle agitation. The DGGE bands were recorded with the Typhoon 8600 Variable Mode Imager (Amersham Pharmacia Biotech, Sunnyvale, CA, USA) and the profiles were analyzed with ImageQuant software (Amersham, Uppsala, Sweden). Three lanes of standards, a mixture of 16S amplicons with a gradient of the GC-content from nine environmental clones (Demba Diallo*, et al.*, 2004), were run in parallel with the samples on each DGGE gel. Since the standards were always run together with the samples at the same denaturant concentration in the gel, their positions could be used to compare the patterns formed in different gels.

2.1.6 Southern blot

The *nifH* RT-PCR / RT-nested PCR products were hybridized with the Ps (*Pseudomonas stutzeri*) nifH DIG4 probe (Eurofins MWG, Ebersberg, Germany). Blotting of an agarose gel onto a nylon membrane (Hybond-N; GE Healthcare, Fairfield, CT, USA) was performed as described previously (Southern, 1975, Evans*, et al.*, 1994) by capillary transfer. Ps nifH DIG4 is a 60-mer oligonucleotide probe labeled at 4 positions with digoxigenin (DIG). This probe sequence was derived from *P. stutzeri* A1501 *nifH* (CP000304), and is corresponding to positions 261 to 320 of its *nifH* gene. These nucleotide positions are conserved, but not identical, in other known *nifH* sequences. In order to balance the stringency and detection coverage, eight *nifH* fragments with different identity in the probe binding region to *P. stutzeri* were used as references to optimize the hybridization conditions (details described in the Results section). Under the optimized conditions, hybridization was carried out with a moderate stringency, a hybridization at 60°C overnight and with 2 washes in 2 × SSC + 0.1% SDS at 60°C. The 16S RT-PCR products were separated on a DGGE gel and hybridized with a 16S rRNA gene targeted *Pseudomonas*-specific digoxigenin-labeled probe, which corresponds to positions 348 to 374 of the *P. stutzeri* 16S rRNA

sequence (AF143245). The DGGE gel was washed several times with TAE buffer after staining and was blotted onto a N+ nylon membrane (Hybond-N+, GE Healthcare, Fairfield, CT, USA) with a semi-dry electroblotter (Bio-Rad, Munich, Germany). Blotting was carried out at 40 mA (about 5-6 volts) for 1 h using TAE as transfer buffer. Denaturation and neutralization were performed by placing the blot membrane on top of a filter paper soaked with denaturation or neutralization buffer. Denaturation was carried out for 30 min and neutralization for 5 min. The air-dried membrane was exposed to 312 nm UV light (Biometra, Goettingen, Germany) for 2 min to covalently crosslink the DNA fragments. For hybridization the SET system was used, as described previously (Hurek*, et al.*, 1993). Hybridization was done at 68°C overnight. Chemiluminescent detection was carried out using CDP-star (Roche Applied Science, Mannheim, Germany) and images were recorded with the LAS-3000 system (Fujifilm, Japan).

2.1.7 Clone library

Clone libraries of *nifH* RT-PCR products were constructed using the TOPO TA cloning kit (Invitrogen, Darmstadt, Germany) or the CloneJET cloning kit (Fermentas, St. Leon-Rot, Germany) in accordance with the manufacturer's instructions. Transformants were analyzed for *nifH* inserts by colony PCR followed by dot hybridization with Ps nifH DIG4. Colony PCR was performed with the universal forward and reverse primers of the clone vector, M13 F+R for pCR2.1-TOPO vector and pJET1.2 F+R for pJET1.2 / blunt vector. One microliter colony PCR products (usually containing the same amount of DNA) were spotted onto a N+ nylon membrane (Hybond-N+; GE Healthcare, Fairfield, CT, USA). Denaturation, neutralization and UV cross linking were carried out following the same protocol as described before for DGGE gel blotting, but denaturation time was decreased to 5 min. Membranes were hybridized with the Ps nifH DIG4 probe at 60°C overnight. Primer set BH-118F and 214R, corresponding to positions 118 to 134 and 198 to 214 in the *Azoarcus* sp. BH72 *nifH* sequence (NC008702.1), was designed to specifically

amplify a 97-bp fragment of *A.* sp. BH72 *nifH* and used to screen the clone library for the *A.* sp. BH72 *nifH* fragment.

2.1.8 Sequencing

PCR products were directly sequenced by Barcode sequencing (LGC Genomics, Berlin, Germany) with universal primers of the clone vector or with primers used for individual PCR reactions.

2.2 Real-time PCR studies

2.2.1 Sampling

Sampling was done as described in section 2.1.1. Only samples were analyzed, where *nifH* transcript phylotypes affiliated with *Azoarcus* sp. BH72 were predominant.

2.2.2 RNA and DNA extraction

RNA and DNA extraction was done by protocol I as described in section 2.1.2. The same protocol was applied to nucleic acid extraction from 5×10^{10} nitrogen-fixing cells of an *A*. sp. BH72 culture, grown in a fermentor at 2 µM dissolved O_2 according to standard conditions.

2.2.3 Primers

The primer set HF and HDR, used for *nifH* cRNA synthesis, were designed based on the *nifHDK* operon sequence (GenBank accession number AF200742) of *A.* sp. BH72. These primers cover the coding sequence of *nifH* and part of the intergenic region between the *nifH* and *nifD* genes of *A*. sp. BH72 and generate a 1054 bp fragment in PCR reactions. For 16S cRNA synthesis, primers 8F and 1492R (Table 3) were used. Two universal primers, nifH3 for *nifH* and R518 for 16S rRNA genes (Table 3), were used for reverse transcription. For nested RT-qPCR the universal *nifH* primer sets

were used as described before (2.1.4). The *nifH* primers specific for *A.* sp. BH72 were developed on the basis of multiple-sequence alignments and validated by PCR. The 16S rRNA primers specific for *A.* sp. BH72 were designed and validated in the same way. For this purpose several phylogenetically closely related bacterial taxa, *Azoarcus communis* SWuB3 and *Azovibrio restrictus* S5b2 (Reinhold-Hurek & Hurek, 2000) were used as references and for optimization of PCR conditions (data not shown). Based on these results, *nifH* primers 390F, 479R and 16S primers 320F, 417R (Table 3) were regarded to be specific at the species level. For universal quantification of *nifH* transcripts a nested RT-qPCR, whereas for specific detection of *nifH* transcripts from *A*. sp. BH72 a single stage RT-qPCR were used. The primer sets specific for *Psudomonas stutzeri*, Ps390F+Ps467R for *nifH* and Ps60F+Ps220R for 16S, and for *Rhizobium rosettiformans*, RrHF+RrHR for *nifH* and RrSF+RrSR for 16S, were developed in the same way as for the specific primer sets of *A.* sp. BH72 (Table 3).

Table 3. Primers used for reverse transcription, PCR amplification and cRNA synthesis

Gene target	Name	Sequence (5'-3')	Reference
*nifH*DK operon of *Azoarcus* sp. BH72	HF	atggcaaagctgcgtc	This study
	HDR	atgagggcgtcggtttcttcacg	This study
16S rRNA gene-universal (Archaea not included)	8F	agagtttgatccttggctcag	(Dojka, *et al.*, 1998)
	1492R	gcytaccttgttacgactt	(Dojka, *et al.*, 1998)
16S rRNA gene-universal (Archaea not included)	R518	attaccgcggctgctgg	(Muyzer & Smalla, 1998)
nifH of *Azoarcus* sp. BH72	390F	cgacgtggtgtgtggtg	This study
	479R	atggccatcatttcgcc	This study
16S rRNA gene of *Azoarcus* sp. BH72	320F	ccagtcrtgggggataactacgc	This study
	417R	gccgctccaatcacgtgag	This study
nifH-universal	nifH3	atrttrttngcngcrta	(Zani, *et al.*, 2000)
	nifH4	ttytayggnaarggngg	(Zani, *et al.*, 2000)
	nifH1	tgygayccnaargcnga	(Zehr & McReynolds, 1989)
	nifH2	adngccatcatytcncc	(Zehr & McReynolds, 1989)
nifH of *P. stutzeri*	Ps390F	cgacgtggtctgtggcg	This study
	Ps462R	gccggagcagaccacgtagat	This study
16S rRNA gene of *P. stutzeri*	Ps60CF	agcggatgaagggagcttgc	This study
	Ps220R	gctaatccgacctaggctcatc	This study
nifH of *R. rosettiformans*	RrHF	acgcaaaggctcaggacac	This study
	RrHR	atgtcgcgatagcccacc	This study
16S rRNA gene of *R.*	RrF	gacatccgggtcgcggac	This study

rosettiformans	RrR	gctcgacgtcaccgtctc	This study

2.2.4 Synthesis of cRNAs

For *nifH* cRNA synthesis specific for *A.* sp. BH72, PCR was carried out with primer set HF and HDR, using *A.* sp. BH72 genomic DNA as template. The PCR was carried out as follows: initial DNA denaturation for 5 min at 95℃, 35 cycles of denaturation for 1 min at 95℃, annealing for 1 min at 65℃, extension for 1 min at 72℃, final extension for 10 min at 72℃. The 1054 bp PCR product was cloned into the plasmid vector pCR2.1-TOPO using the TOPO TA cloning kit (Invitrogen, Darmstadt, Germany). Positive clones were further analyzed with primer combinations HF+M13R and M13F+HDR to confirm that the fragment was inserted in the correct orientation. Plasmid were extracted from the correct clone using QIAprep Miniprep kit (QIAGEN, Hilden, Germany) and were used as templates for *in vitro* transcription under the T7 promoter after linearization with *Sac*I (T7 transcription kit, Fermentas, St. Leon-Rot, Germany). *In vitro* transcription was done according to the manufacturer's instruction. The linearized plasmid DNA in the *in vitro* transcriptional reaction system was removed by DNase I treatment (Roche Applied Science, Mannheim, Germany) by incubation for 30 min at 37℃. The cRNA was purified by phenol / chloroform extraction and ethanol / sodium acetate (3 M, pH 5.2) precipitation. The cRNA pellet was collected by centrifugation, dissolved in RNA secure solution (Ambion, Darmstadt, Germany) and stored in liquid nitrogen until further use in RT-qPCR applications. Synthesis of *A.* sp. BH72 specific 16S cRNA was done following the same procedures, using primer set 8F and 1492R and the PCR program as described previously (Dojka*, et al.*, 1998). Cloning of PCR products, linearization of plasmids, and *in vitro* transcription were done as described above for *A.* sp. BH72 *nifH* cRNA. The quality of the cRNA was documented with the Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA).

2.2.5 Quantification and dilution of cRNA and DNA standards

The cRNA concentration was estimated by NanoDrop and the calculation of RNA copy number was performed by using the online tool EndMemo (http://www.endmemo.com). Serial tenfold dilutions of *A.* sp. BH72 specific *nifH* cRNA (4.7×10^7 to 4.7×10^0 copies) and 16S cRNA (8.7×10^9 to 8.7×10^2 copies) were performed in the presence of 20 ng / µl carrier RNA (QIAGEN, Hilden Germany). Linearized plasmids used as templates for *in vitro* transcription were used as DNA standards. The serial dilutions of DNA standards were done in the same way as of cRNA standards, using seriall tenfold dilutions of *A.* sp. BH72 specific *nifH* gene (8.0×10^6 to 8.0×10^0 copies) and 16S rRNA gene (6.2×10^6 to 6.2×10^0 copies). As a *R. rosettiformans*-specific *nifH* gene standard a linearized plasmid from the clone library (Burbano *et al*., 2011) was used. This plasmid carries the *R. rosettiformans* partial *nifH* sequence corresponding to *A.* sp. BH72 *nifH* position 118 to 479. Serial 10-fold dilutions of the *R. rosettiformans*-specific *nifH* gene standard were done (6.5×10^6 to 6.5×10^0 copies). For the *P. stutzeri*-specific *nifH* gene standard a linearized plasmid from the clone library (details described in Results section) was used. This plasmid carries the *P. stutzeri nifH* sequence corresponding to *A.* sp. BH72 *nifH* position 118 to 479. For the *P. stutzeri*-specific 16S rRNA gene standard PCR products were used which had been obtained by PCR amplification with primer set 8F+1492R from genomic DNA of *P. stutzeri* (DSM No. 4166). Serial dilutions of *P. stutzeri*-specific DNA standards were done using 6.7×10^6 to 6.7×10^0 copies for the 16S rRNA gene and 3.3×10^6 to 3.3×10^1 copies for the *nifH* gene.

2.2.6 Reverse transcription of samples and cRNA standards

Four hundred nanograms of total RNA from spruce samples or 200 ng of total RNA from *A.* sp. BH72 were reverse transcribed using SuperScript III (Invitrogen, Darmstadt, Germany). All reverse transcriptions were done in a 20 µl reaction system. Serial dilutions of cRNA standards were reverse transcribed. The universal primers

nifH3 and R518 were used for the reverse transcription of *nifH* and 16S cDNA respectively. Aliquots of cDNAs were stored at -80℃ prior to use in qPCR.

2.2.7 Optimizing real-time PCR using SYBR Green I

In order to optimize the qPCR conditions for quantification, the concentration of SYBR Green I was determined. The 10,000 × SYBR Green I stock (Invitrogen, Karlsruhe, Germany) was 100 fold diluted in PCR enhancer (Molzym, Bremen, Germany). In PCR reactions 0.6 x SYBR Green I was used. All PCRs were carried out in a final volume of 25 µl, containing 1 × reaction buffer, 1.5 mM $MgCl_2$, each deoxynucleoside triphosphate (dNTP) at a concentration of 200 µM, each oligonucleotide primer at a concentration of 0.3 µM, and 1.25 U of *Taq* DNA polymerase (MolTaq, Molzym, Bremen, Germany). As template for PCR reactions 1 / 20 volume (1 µl) of the RT reaction was used.

The first round of the universal *nifH* nested PCR was performed as follows: initial denaturation for 5 min at 95℃, 25 cycles of denaturation for 20 sec at 95℃, annealing for 1.5 min at 55℃, extension for 30 sec at 72℃, and a final extension for 5 min at 72℃. An aliquot of 2.5 µl from the first round were used as template for the second round qPCR using the same PCR program, increasing the cycle number to 45 and omitting a final extension step. For the first round of PCR amplification, DNA primer set nifH3 and nifH4 and for the second round DNA primer set nifH1 and nifH2 were used.

For *A.* sp. BH72-specific *nifH* qPCR, the reaction conditions for primer set BH-118F and 214R consisted of an initial denaturation of 5 min, followed by 45 cycles of 20 sec at 95°C, 30 sec at 56°C, and 30 sec at 72°C. Fluorescence was monitored after each extension step. Melting curves on real-time PCR reactions were obtained upon increasing the temperature from 55°C to 96°C at a rate of 1.0℃ in 10 sec. The qPCRs with other primer sets were performed with the same protocol, except that the annealing temperatures were changed to 68℃ for *A.* sp. BH72-specific 16S PCR, to

58 ℃ for *R. rosettiformans*-specific *nifH* PCR and to 60 ℃ or 64 ℃ for *P. stutzeri*-specific *nifH* or 16S PCRs, respectively.

All real-time PCR programs were carried out with the Chromo 4 PTC-200 real-time PCR cycler (MJ Research, Waltham, MA, USA). Replicated reactions and independently diluted standard curves were run under identical conditions. To quantify unknown samples a calibration curve was run routinely with each sample set and compared with previous standard curves. Results showed that variability of amplification efficiency was < 5 % between runs (data not shown). Opticon Monitor Analysis software (version 2.03, MJ Research) was used for data analysis. Template concentrations were determined in duplicates and were calculated from the calibration curve. For all (RT)-qPCR assays, there was a linear relationship between the log of the standard cRNA or DNA copy number and the measured threshold cycle (C_T) value across the specified concentration range. The correlation (r^2) of the linear relationship was strong (> 0.99, in all cases, except for the 16S rRNA gene of *P. stutzeri*, which was 0.98). PCR amplification efficiencies of both standards and samples were calculated by the equation, PCR efficiency = $10^{-1/slope}$ -1 using the qPCR calculator from Agilent Technologies (http://www.genomics.agilent.com). The calculated efficiencies were 73 to 135% for all qPCRs using specific primer sets, while the efficiency was 55% for the nested qPCR using universal primer sets. Analysis of the melting curve was performed to determine the specificity of qPCR amplification. PCR product size was verified by agarose gel electrophoresis. If amplicons of the proper size were detected by agarose gel electrophoresis in the no-template control or the DNA-contamination control, the real-time (RT)-PCR data obtained from the same batch of PCR reaction were rejected.

2.3 Protein studies

2.3.1 Sampling

The sampling location was described in section 2.1.1. Sampling was conducted in

fifteen months during a four-year period, from July 2006 to September 2009. In September 2006, different types of roots (differentiation by the diameter of roots) were collected from different soil depths (from 0 to 40 cm). Based on the preliminary results of this study, the root samples collected from the later sampling dates consisted mainly of medium-coarse roots, with a diameter of about 2 cm from the organic layer at 5 to 10 cm soil depths. Transport and storage of samples were carried out as described in section 2.1.1.

2.3.2 Protein extraction from root samples

Protocol I: 200 mg (fresh weight) root material was ground in liquid nitrogen into fine powder. The following protein extraction steps were done as described previously (Hurek*, et al.*, 1995, Hauberg*, et al.*, 2010). The air-dried pellet was dissolved in 80 µl protein buffer (5 M urea, 2 M thiourea, 2% CHAPS, 2% Sulfobetaine 3-10, 1% DTT and 2% Bio-Rad ampholytes 3-10) and stored at -80℃ until further processed. Protein concentration was estimated with the RC/DC protein assay kit (Bio-Rad, Munich, Germany) using BSA to generate a standard curve. *A.* sp. BH72 was used as a reference to evaluate the specificity of the polyclonal NifH antibodies and to monitor the whole process of immunoblotting. The N_2-fixing cells of *A.* sp. BH72 were obtained as described by Oetjen and Reinhold-Hurek (Oetjen & Reinhold-Hurek, 2009). Proteins were extracted from the N_2-fixing cells using the same protocol. Protocol II: Protein extraction was done as described in protocol I, until the step following the first centrifugation. The whole supernatant, both the water and phenol phases, were transferred into a new tube. One volume of chloroform was added and the tube was kept shaking at 4℃ for 15 min. Proteins were pelleted between the water and organic phases after centrifugation. The protein layer was washed several times with chloroform and finally two times with methanol. The air-dried pellet was dissolved in 80 µl of protein buffer.

2.3.3 SDS-PAGE with mini Protean II system

Protein samples were mixed with loading buffer (4 × loading buffer containing: 60% glycerol, 0.25 M Tris-HCl, pH 6.9, 9.2% SDS, 5% ß – mercaptoethanol and 0.012% bromphenolblue) and centrifuged at 16,000 × g for 1 min before loading. A prestained protein ladder (10-170 kDa, Fermentas, St. Leon-Rot, Germany) was used to size proteins on gels and blots, to monitor protein migration during electrophoresis and to monitor protein transfer onto membranes. Proteins were separated by standard SDS-PAGE in a mini Protean II cell (Bio-Rad, Munich, Germany) using the discountinuous buffer system of Laemmli (Laemmli, 1970). The stacking or separation gel contained 5% or 12% acrylamide (37.5:1 acrylamide/bis-acrylamide) respectively.

2.3.4 Blotting

The gel-separated proteins were blotted to a membrane using the mini Trans-Blot electrophoretic transfer cell (Bio-Rad, Munich, Germany) according to the manufacturer's instructions. The nitrocellulose membrane (Whatman, Dassel, Germany) or the polyvinylidene difluoride (PVDF) membrane (Amersham, Munich, Germany) were used and described in detail below (Results section 3.3). Semi-dry blotting was carried out at 8 volts for 50 min, using transfer buffer (1 L containing 14.4 g glycine, 3 g Tris base, 380 µl 20% SDS and 200 ml methanol). To confirm the efficiency of protein transfer, MemCode reversible protein stain kit (Thermo Fisher Scientific, Bonn, Germany) or Ponceau S (0.2% Ponceau S in 3% acetic acid) were used for detecting proteins on nitrocellulose or PVDF membrane, respectively. The blot was either used immediately or was air-dried and stored at -20℃ until further processing.

2.3.5 Retrieve and purify proteins from blot membrane

Proteins extracted from roots were separated in a SDS-PAGE gel with the mini Protean II system and blotted to a nitrocellulose membrane as described before. The

region between 30-40 kDa was cut out of the membrane with a scalpel and divided into small pieces. The excised pieces of membrane were incubated in phenol for 30 min to extract the proteins from the membrane into the phenol phase. After removal of the membrane, proteins were purified using ProteoExtract Protein Precipitation kit (Merck, Darmstadt, Germany) in accordance with the manufacturer's instruction.

2.3.6 Immunological detection

The membrane was blocked for 1 hour in Tris buffered saline (TBS) (20 mM Tris, pH 7.5, 500 mM NaCl) containing 5% non-fat dry milk followed by three brief washes with TBS. Subsequently the membrane was incubated with the primary antibody or pre-immune serum in TBS containing 0.5% bovine serum albumin (BSA) at 4°C overnight. Two types of primary antibodies were used in this study, a polyclonal NifH antibody raised against the Fe protein from *A.* sp. BH72 (Oetjen*, et al.*, 2009), and a polyclonal universal peptide antibody (Pacific Immunology Corp., Ramona, USA) raised against a custom synthesized synthetic peptide (data not shown), which was designed by Dr. T. Hurek from comparisons of available *nifH* gene sequences. For immunoblotting the affinity-purified peptide antibody was used. For obtaining a monospecific peptide antibody, the antiserum was custom affinity-purified against the synthesized peptide. The *A*. sp. BH72 NifH antibody was stored at -80°C in PBS at a concentration of 15 mg / ml and was further diluted 1:5000 for Western blot analyses. The affinity-purified NifH peptide antibody was stored at 4°C at a concentration of 0.5 mg / ml and was further diluted 1:500 for immunoblot analyses. After three washes with TBS, the membrane was incubated at room temperature for 1 h in TBS containing 0.5% BSA and polyclonal swine anti-rabbit immunoglobulins conjugated to horseradish peroxidase (Dakocytomation, Glostrup, Denmark) as secondary antibody, which was diluted 1:10,000. After three washes with TBS, chemiluminescence detection was carried out, unless stated otherwise, with the Roche Lumi-Light Plus Western blotting substrate (Roche applied science, Mannheim, Germany). Chemiluminescent images were recorded with the LAS-3000 system (Fujifilm, Japan).

2.3.7 Two-dimensional gel electrophoresis and gel staining with colloidal Coomassie

Isoeletric focusing (IEF) was carried out with the Model 175 Tube cell (Bio-Rad, Munich, Germany). The IEF gel contained 3.5% (w/v) acryl-bisacrylmide (30:1), 5.8% ampholytes (Serva, Heidelberg, Germany) in a ratio of 2:1:1 at pH ranges of 3 to 10, 4 to 6 and 5 to 8, 9 M urea, 2% CHAPS, 0.3% TEMED and 0.05% ammoniumpersulfate (APS). The IEF gel was casted in glass tube and was pre-run for 15 min at 200 V, 30 min at 300 V and 60 min at 400 V using 20 mM NaOH as the cathode solution and 10 mM phosphoric acid as the anode solution. The protein sample was loaded on top of the gel, covered with 20 µl overlay buffer (6 M urea, 1% ampholines pH 3.5 to 10, 100 mM DTT and 2% CHAPS) and the excess space of the glass tube was filled with the cathode buffer. The IEF gel was run at 100 V for 10 min, 200 V for 10 min, 300 V for 2 h and 400 V over night. After running, the IEF gels were stored at -20℃ inside the glass tubes prior to further use. The second-dimensional SDS-PAGE was performed with the Scie-Plas TV 400Y vertical electrophoresis system (Biostep, Jahnsdorf, Germany). Proteins were separated in a 12% polyacrylamide gel overlaid with a 5% polyacrylamide stacking gel. The IEF gel was removed from the glass tube and equilibrated for 20 min in equilibration buffer (60 mM Tris·Cl, pH 6.8, 1% SDS, 20% glycerol, 50 mM DTT and trace of bromphenolblue). The equilibrated IEF gel was loaded on top of the slab gel and a standard SDS-PAGE gel electrophoresis was carried out with the discountinuous buffer system of Laemmli (Laemmli, 1970). After electrophoresis, the gel was stained with colloidal Coomassie G-250 in accordance with Candiano's method (Candiano*, et al.*, 2004). Briefly, the gel was incubated overnight with gentle shaking in the freshly prepared staining solution (0.15% Coomassie brilliant blue G-250, 10% ammonium sulfate, 10% phosphoric acid and 20% ethanol). Usually a destaining step was not necessary. The gel was stored at 4℃ in storage solution (18% ethanol and 3% glycerol).

2.3.8 Mini 2-D gel electrophoresis and gel staining with SYPRO Ruby for protein quantification

The IEF was carried out with the mini Protean tube cell (Bio-Rad, Munich, Germany) The IEF gel containing the same components as described before was casted in the capillary tube according to the manufacturer's instruction. Five to 10 µg protein was loaded without a pre-run. The protein sample was overlaid with 40 µl overlay buffer and the remaining space of the sample reservior was filled up with cathode buffer. The gel was run at 500 V for 10 min, and then the voltage was increased to 750 V for 3.5 h. After the first dimensional run was complete, the IEF gel was removed from the capillary tube, loaded directly on top of the slab gel, overlaid with 1 ml equilibration buffer and incubated for 10 min. The second dimensional SDS-PAGE gel was carried out with the mini Protean II cell (Bio-Rad, Munich, Germany) as described before. The slab gel was stained with SYPRO Ruby (Sigma-Aldrich, Munich, Germany) in accordance with the manufacturer's instruction. Briefly, the slab gel was fixed for 30 min in 50% (v/v) ethanol with 3% acetic acid. After fixation the gel was incubated overnight with gently shaking in the staining solution. After staining, the gel was washed for 30 min in 10% ethanol with 7% acetic acid. SYPRO Ruby images were obtained using the Typhoon 8600 laser scanner with the following settings: emission filter Rox 610, PMTV 500 V and laser green 532 nm. Quantitative analysis of SDS-PAGE images was carried out using ImageQuant software. Quantification of the protein spots separated on 2-D gel was done using a standard curve. In this study, BSA and yeast carboxypeptidase Y (glycosilated) were used as standards. The two standards were serially diluted in protein buffer to the following protein amounts per gel: 50 ng, 250 ng, 500 ng and 1000 ng. Equal amount of the two standards were loaded together on one IEF capillary gel and the four SDS-PAGE gels were run in parallel. After electrophoresis, SYPRO Ruby staining and Typhoon imaging were carried out under the same conditions as described above. The standard curve was plotted using the volume in densitometry units of scanned standard spots against the various standard protein loadings in the gels.

2.3.9 Protein identification

The immunoreactive proteins were excised from the slab gels and stored in LoBind tubes (Eppendorf, Hamburg, Germany) at -80℃ prior to further use. Trypsin digestion and MS (mass spectrometry) analysis were performed by Dr. F. Schmidt at the Interfaculty Institute for Genetics and Functional Genomics in Greifswald, Germany. The MS data were converted to the protXML format, showing the reference protein ID and statistics, and were visualized with the protXML Viewer. One mini 2-D gel was carried out by loading 25 µl protein buffer with the same sample loading syringe used in this study as a blank control. The blank control slab gel was stained with SYPRO Ruby and no protein spot was detected in the scanned image. Nevertheless, a spot was excised, from an area corresponding to the target spots validated by MS analysis to contain NifH from *A*. sp. BH72 and was sent for MS analysis to evaluate the inter-sample contamination.

2.3.10 Glycoprotein detection and differentiation and deglycosylation

The DIG glycan detection kit (Roche applied science, Mannheim, Germany) was used for the detection of sugars in glycoconjugates by an enzyme immunoassay. The DIG glycan differentiation kit (Roche applied science, Mannheim, Germany) was used for the characterization of carbohydrate chains from glycoproteins bound to the nitrocellulose membrane. An enzymatic protein deglycosylation kit (Sigma-Aldrich, Munich, Germany) was used for removal of all N-linked and simple O-linked carbohydrates from glycoproteins. All experiments were done in accordance with the manufacturer's instructions.

2.4 Studies on cloning bias

2.4.1 DNA templates

Plasmids pA72 or pRW3 were used as DNA templates, carrying a partial *nifH* gene

fragment from *A.* sp. BH72 or from *R. rosettiformans* W3, respectively. Both *nifH* fragments had been obtained by PCR from genomic DNA of *A*. sp. BH72 or from sugarcane (Burbano et al., 2011) with LNA-substituted Zehr forward and reverse primers (Table 4) before they were cloned into the TOPO TA vector (Invitrogen, Darmstadt, Germany). Nucleotide sequencing of inserts from both plasmids had revealed that the Zehr primer target regions in both plasmids were identical (data not shown).

2.4.2 Primer sequences

For universal *nifH* targeting the DNA primers Zehr-F and -R designed by Zehr and McReynolds (1989) were used, which have been shown to amplify highly divergent *nifH* genes with equal efficiencies and to cover the majority of *nifH* genes without mismatches in the primer target regions (Demba Diallo*, et al.*, 2008). LNA-Zehr primers were designed by Burbano *et al.* (2010). In these primers three DNA bases are substituted by LNA bases (Table 4). Based on previous Barcode sequencing information (unpublished data), the degenerate positions in LNA primer of Zehr-F and R were substituted by specific DNA or LNA bases to generate non-wobble primers NW-F, AR-R and BHNW-R. Primer pair NW-F and AR-R had no mismatch to the Zehr primer-target regions in plasmids pA72 and pRW3, while primer pair NW-F and BHNW-R had no mismatch to the Zehr primer-target regions in the *nifH* gene from *A.* sp. BH72. Primers were synthesized by Eurogentec (Seraing, Belgium). All LNA primers were RP-cartridge purified and all DNA primers were salt-free.

Table 4. Oligonucleotide primer sequences. LNA bases in capitals.

Primer name	Sequence (5'-3')	Reference
DNA-Zehr-F	tgygayccnaargcnga	Zehr and McReynolds (1989)
LNA-Zehr-F	tGygAyCcnaargcnga	Burbano *et al*. (2010)
DNA-Zehr-R	adngccatcatytcncc	Zehr and McReynolds (1989)
LNA-Zehr-R	aDngCcAtcatytcncc	Burbano *et al*. (2010)
DNA-NW-F	tgtgatccgaaggcaga	This study
LNA-NW-F	tGtgAtCcgaaggcaga	This study
DNA-AR-R	agagccatcatttcgcc	This study
LNA-AR-R	aGagCcAtcatttcgcc	This study
DNA-BHNW-R	atggccatcatttcgcc	This study
LNA-BHNW-R	aTggCcAtcatttcgcc	This study
KP2S	tcgcgaccgtagtcc	This study

2.4.3 PCR amplification

PCR reactions were performed in a 20 µl system, containing 1 × reaction buffer, 2.0 mM $MgCl_2$, 50 µM dNTPs, 0.5 µM primers, 1 U DreamTaq (Fermentas, St. Leon-Rot, Germany) or *Pfu* DNA polymerase (Fermentas, St. Leon-Rot, Germany) and 20 ng *A.* sp. BH72 genomic DNA or 0.5 ng plasmid DNA (pA72 or pRW3). The PCR program was as follows: 30 cycles of 95℃ for 20 s, 56℃ for 30 s and 72℃ for 30 s, and a final extension for 10 min at 72℃. For a colony PCR, an initial step of 10 min at 95℃ was included to lyse the cells and inactivate nucleases.

2.4.4 Cloning and screening

Amplicons obtained by PCR with the DreamTaq DNA polymerase were cloned using the TOPO TA cloning kit (Invitrogen, Darmstadt, Germany), while amplicons obtained with *Pfu* DNA polymerase were cloned using the Zero Blunt TOPO cloning kit (Invitrogen, Darmstadt, Germany) according to the manufacturer's instructions. Clone libraries were screened by colony PCR using the vector primers M13 forward and reverse, supplied in the kit. Some confirmed plasmid inserts were directly sequenced by Barcode sequencing (LGC Genomics, Berlin, Germany) using M13 primers.

2.4.5 Single strand DNA ligation

In order to investigate whether *Taq* DNA polymerase can add a single deoxynucleotide to the 3'-ends of PCR products obtained with LNA-substituted primers, a ligation to a single-stranded DNA adaptor (kindly provided by J. Doering, University of Bremen) was carried out. This adaptor consists of 43 nucleotides. It is 5'-phosphorylated and modified by a 3'-ddC (dideoxycytidine, which prevents extension by DNA polymerases and degradation by 3' exonucleases) to permit ligation only to the free 3'-ends of PCR products. For this purpose genomic DNA from *A.* sp. BH72 was used as template for PCR reactions with LNA- or DNA-Zehr primers. Ligation reactions were done in a 20 µl system, containing 1 × ligase buffer, 20 pmol adaptor, approx. 1 pmol purified PCR products, 10 µl 50% polyethylene glycerol (PEG) 4000 and 10 U T4 RNA ligase I (New England Biolabs, Frankfurt am Main, Germany). Ligation was carried out at 22℃ for at least 1 hour. To obtain the sequence information of the 3'-ends from the PCR products, 1.0 µl of the ligation mixture was used as template for PCR reactions with primer KP2S (Table 4), which anneals to the 5'-end of the adaptor. PCR products were cloned and sequenced by clone vector primers. The adaptor sequence is as follows (5' to 3'):
GGACTACGGTCGCGATGGCACGATACGCAGCATCACAAGAACT.

3 Results

3.1 Study of the diversity and abundance of *nifH* transcripts in Norway spruce roots

3.1.1 Optimization of hybridization conditions for Ps nifH DIG4 probe

In this study broad range, highly degenerate primers (Zani*, et al.*, 2000) were used in reverse transcription (RT)-PCR reactions to amplify *nifH* transcripts from spruce root samples. It is well known that degenerate primers bind imperfectly to the target region, which typically leads to low amplification efficiency, sensitivity, and specificity of the PCR reaction (Martin*, et al.*, 1985, Watkins & SantaLucia, 2005). PCR performance may further decrease when complex environmental samples, such as plants or soil are analyzed, which often contain low template concentrations and inhibitors (Poussier*, et al.*, 2002). Therefore it was not surprising that with the high degeneracy of the Zehr primers, non-specific amplification from environmental samples was sometimes observed. For example, a non-specific PCR product was occasionaly obtained which had a size similar to that of the desired PCR product. A 330-bp fragment with (90%) nucleotide sequence identity to the *P. stutzeri* A1501 penicillin-binding protein-2 gene (CP000304.1), amplified with the reverse primer, was frequently detected by sequencing (data not shown) and its migration in agarose gel electrophoresis (2% agarose gel) was almost indistinguishable from the 362-bp *nifH* fragment. To allow an initial screening for presence of *nifH* fragments, PCR fragments were hybridized with a *nifH* probe. For this purpose the Ps nifH DIG4 probe was used. This probe is *nifH*-specific and covers divergent *nifH* sequences. The hybridization conditions for Ps nifH DIG4 probe were optimized by using eight *nifH* fragments with different nucleotide sequence identities (from 70 - 93%, Fig. 2) to the probe binding region of the *P. stutzeri nifH* fragment. Hybridizations were carried out in SET (Varshney*, et al.*, 1991) and SSC (Hurek*, et al.*, 1997) hybridization solutions for

comparison (Fig. 2).

When the SSC system was used, hybridization was carried out at 50℃, 60℃, or 65℃ overnight, and blots were washed in either 6 x SSC or 2 x SSC with 0.1% SDS (Fig. 2). In case of hybridization in SET buffer, hybridization was carried out at 25℃. Under these conditions, phylogenetically distant *nifH* fragments, with a nucleotide sequence identity as low as 70%, were still detectable with the *P. stutzeri nifH* gene-based probe allowing to differentiate non-specific amplificons from *nifH* fragments by oligonucleotide hybridization.

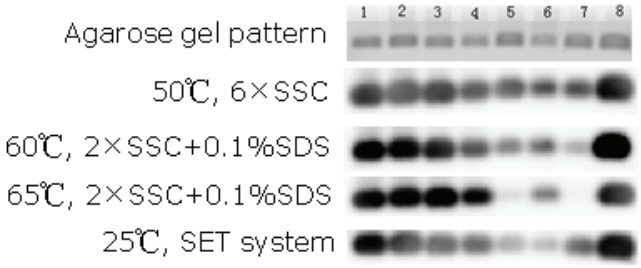


Fig. 2 Southern blot analysis with the Ps nifH DIG4 probe in different conditions of DNA hybridization. The hybridization buffer system, temperature, and stringency of the washing solution were varied. For details see text. The eight reference *nifH* fragments and their% nucleotide sequence identity to the *nifH* fragment from Ps nifH DIG4 as follows: 1. γ-proteobacteria, 81%; 2. Gram positive bacteria, 81%; 3. α-proteobacteria, 85%; 4. γ-proteobacteria, 93%; 5. Gram positive bacteria, 70%; 6. β-proteobacteria, 81%; 7. *anfH*, 73%; 8. Gram positive bacteria, 80%.

3.1.2 Detection of *nifH* transcripts identical to *P. stutzeri nifH*

Isolation of high-quality RNA from spruce samples was challenging, since the tissues were rich in phenolic compounds, polysaccharides and RNases (Schneiderbauer*, et al.*, 1991). Since the spruce root samples were collected directly from the forest, it was possible that some samples were contaminated to some degree by humic acids from soil, which are known to be strong PCR inhibitors. Although, soil was carefully removed from roots, this possibility cannot be excluded. To avoid RNA degradation during collection and transport, root samples were snap-frozen and stored at the

temperature of liquid nitrogen in a gaseous-phase dewar for a safe transport. Once back to the laboratory, samples were stored in liquid nitrogen prior to molecular analyses. Different methods, including several commercial kits, were evaluated for RNA extraction. Methods involving phenol in the early steps of extraction, e.g. Trizol extraction and hot phenol extraction, resulted in extremely low yields and very poor quality of RNA. This result is consistent with Schneiderbauer's report (Schneiderbauer*, et al.*, 1991) and might be due to the insufficient dissociation of nucleic acids from contaminants and the co-extraction of phenolic compounds associated with RNA into the phenolic phase. The modified Chang's method used in this study performed best. With this extraction protocol a moderate yield and a relatively high-quality of RNA for subsequent enzymatic manipulations were consistently obtained. Nevertheless, amplification of *nifH* transcripts were successfully performed from only few samples and only with a LNA-substituted instead of the standard DNA primer set, using RT-single stage PCR. This indicated that contaminants were not completely removed during the extraction procedure.

Besides to the interference of contaminants, successful RT-PCR amplification of the target sequences was also dependent on the primers used. To get successful amplification from a wide range of diazotrophs, all the primers used here had to be highly degenerate, whereby the degeneracy represents the number of unique sequence combinations each primer contains. With the degeneracy increasing, the amplification efficiency usually decreases, as mentioned above. This drawback could be avoided to some extent by the use of LNA-substituted primers. A 362-bp fragment was amplified from a September 2006 root RNA extract (Fig. 3A) using LNA-substituted primers. The desired band was excised from the agarose gel and purified by gel purification kit. Purified RT-PCR products were directly sequenced with the LNA-Zehr-F primer. Additionally a clone library was constructed from these PCR products using the CloneJET cloning kit. Finally, *nifH* fragments were analyzed by DGGE (Denaturing gradient gel electrophoresis) after addition of a GC-clamp by a ten-cycle re-amplification using GC-clamp-nifH1 + nifH2.

Direct sequencing resulted in 362 bp of unambiguously readable sequence, which was identical to the partial *nifH* gene sequence from *P. stutzeri* (CP000304). In direct sequencing, one sequencing reaction constitutes the consensus of initial amplification events from many template molecules (Hoss*, et al.*, 1994). Therefore, the fact that the nucleotide sequence could be unambiguously read with no indels indicated that the diazotrophic bacterial community associated with spruce roots was not very complex and that *P. stutzeri* probably was the most abundant N_2-fixing bacterium in this sample.

This was further confirmed by sequencing of the clone library and by the DGGE analysis. The clone library was pre-screened by colony PCR with the universal primers of the pJET clone vector and positive clones were distinguished from the false ones by the fragment size of the colony PCR products. Five positive clones were randomly picked and sequenced with the universal primer of the clone vector. All five sequences were identical to the corresponding *P. stutzeri nifH* sequence, suggesting that only a single phylotype was present. The sequencing effort was sufficient, since the Chao1 estimator reached an asymptotic maximum at one (Fig. 3B). The distinct band obtained by DGGE analysis indicated that the particular *nifH* gene phylotype was predominantly transcribed in this root sample. This was the first time in this study that *nifH* transcripts were detected in spruce root RNA extracts using a single stage RT-PCR. This successful RNA extraction protocol was used in all subsequent experiments.

However, the messenger RNA with a 100% DNA sequence identity to the *nifH* transcript from *P. stutzeri* along 362 bp could only be detected in this particular RNA extract. This transcript could be not detected again in any other sample, even not in those collected at the same date.

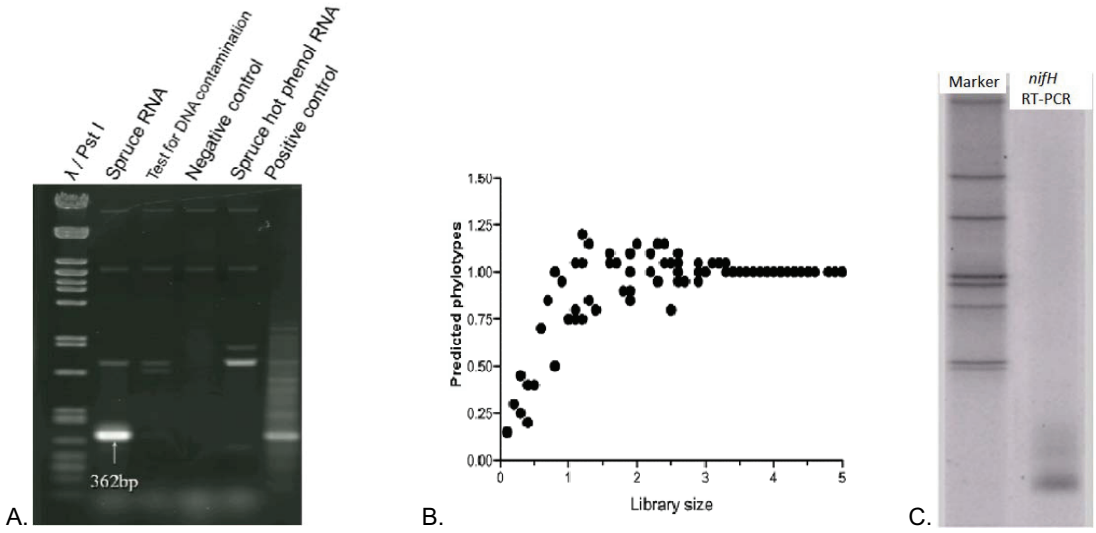


Fig. 3 (A) Representative agarose gel pattern of the 362 bp *nifH* RT-PCR product from a spruce root sampled in September 2006 showing the *Pseudomonas stutzeri*-affiliated-phylotype. *A*. sp. BH72 was used as positive control. Further controls are detailed in the text; (B) Predicted number of phylotypes based on S_{Chao1} versus number of clones analysed. S_{Chao1} was estimated using an online tool (http://www.aslo.org/lomethods/free/2004/0114a.html); (C) DGGE gel pattern of the *nifH* RT-PCR product. Molecular weight marker according to Demba Diallo *et al*. (2004)

3.1.3 Detection of *nifH* transcripts identical to *Azoarcus* sp. BH72 *nifH*

When the RT-nested PCR approach was applied to RNA extracts from root samples collected in different months, *nifH* transcripts were efficiently amplified from all the tested samples after two rounds of amplification. As to be expected from application of nested PCR approaches, the amplification sensitivity was significantly increased by almost doubling the PCR cycle numbers. PCR inhibitors were further diluted in the second round PCR than in the single stage PCR, and the specificity was increased by using a primer set in the second round PCR targeting the inner sequence positions of the initial amplicons. It was also possible to apply the GC-clamp nifH1 primer instead of the nifH1 primer in the second-round, allowing for the use of PCR products with GC clamp in subsequent DGGE analysis without re-amplification (Fig. 4A). Because of the high cycle numbers of nested PCR and an increased risk for false positive PCR

products, it is very important to perform the proper controls for each set of PCR reaction. Each RT-nested PCR experiment included a no-template control for the RT-step and for PCR, as well as a DNA control. The DNA control tested for DNA contamination in the RNA extract. In that way the possibility was largely ruled out that the amplicons in RT-nested PCR reactions were stemming from contaminating DNA (either from the working ambience or due to insufficient DNase I treatment) and not from reverse transcribed sample RNA. When a PCR fragment was obtained in any of the no-template controls or the DNA control, the RT-nested PCR products from root RNA amplified in the same batch were discarded. This happened in less than 5% of all RT-PCR reactions, which were analyzed by agarose gel electrophoresis.

A uniform one-band DGGE pattern was obtained for all samples: all pattern had only one particular *nifH* phylotype in common, which was predominant in all the samples tested (Fig. 4B). Two PCR products from no-template controls were also analyzed on DGGE. Bands migrated to different positions (Fig. 4B), indicating that products obtained from root samples were not contaminated by these fragments. PCR products of Sept. 2006 and Jul. 2007 were directly sequenced and revealed that they were identical to the partial *nifH* gene sequence of *A.* sp. BH72 (NC008702). These results suggested that in all spruce extracts examined but one (see above) *nifH* transcripts from *A.* sp. BH72 were most abundant and accordingly this bacterium was probably the most active nitrogen fixing bacterium there. *A.* sp. BH72 is widely used in this laboratory and is investigated by mass cultivation in pure cultures as well as by molecular tools including PCRs of its *nifH* gene. This inevitably increases the risk of random contamination by *A*. sp. BH72 *nifH* PCR products or genomic DNA generated in the laboratory. However, since the level of PCR contamination in control reactions was generally low (< 5%), and contamination of RT-reactions or samples by *A*. sp. BH72 mRNA can practically be excluded, the introduction of RNA or DNA templates into all spruce samples collected at the six time points present in Fig. 4B appears to be rather unlikely.

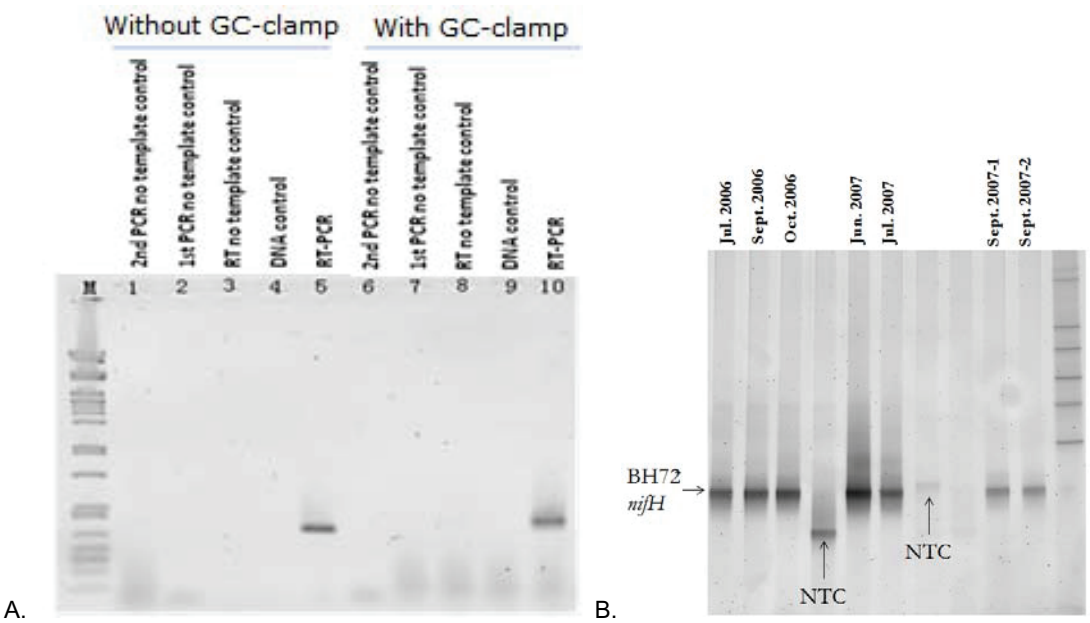


Fig. 4 (A) Representative agarose gel pattern showing ethidium bromide stained RT-nested PCR products including controls from a sample collected in July 2007. M, size marker (*Pst*I-digested lambda DNA); (B) DGGE fingerprint analysis of expressed nitrogenase genes in spruce roots sampled between July 2006 and September 2007 showing *Azoarcus* sp. BH72-affiliated-phylotypes (marked by an arrow). The DGGE pattern of two RT-nested PCR products from no-template controls is also shown. Molecular marker as in Fig. 3C

Because of the high cycle numbers of nested PCR and an increased risk for false positive PCR products, a RT-single stage PCR was applied to the same batches of cDNAs synthesized in the RT step for nested PCR reactions shown in (Fig. 4). To improve the low PCR sensitivity of the RT-single stage PCR, LNA-substituted RT and PCR primers were used. Yet, from the 7 spruce root samples investigated, RT-PCR bands were only obtained from roots collected in Oct. 2006 and Jul. 2007 (Fig. 5A) suggesting that the use of LNA-substituted primers alone was generally not sufficient to make up for the high PCR sensitivity and specificity of the nested PCR approach in spruce samples. Furthermore, for both samples multiple bands were obtained, which indicated in accordance with Burbano *et al*. (2010) that the specificity decreased drastically, when all primers in RT-PCR reactions were LNA-substituted. To identify possible *nifH* amplicons among the PCR fragments, a Southern blot analysis (Fig. 5A) with the Ps nifH DIG4 probe was carried out. This analysis identified two 362 bp

bands of the expected size which hybridized to the probe. The 362-bp bands were excised from the agarose gel and purified with a gel purification kit. The purified RT-PCR products could not be efficiently directly sequenced, because of a high degree of contamination of *nifH* fragments with other PCR products. To obtain the sequence information, clone libraries were initially constructed using the TOPO TA cloning kit. Because of the cloning bias, associated with LNA-substituted primers (details described in Results section 3.4), the purified RT-PCR fragments were subsequently cloned in the pJET1.2 / blunt vector with the CloneJET cloning kit. The two libraries were screened by colony PCR and dot-blot hybridization with the Ps nifH DIG4 probe. Fifty colony PCR products from the Oct. 2006 library and 100 colony PCR products from the Jul. 2007 library were analyzed. This analysis showed that 54% of the clones in each library represented *nifH* sequences (Fig. 5B and 5C). Clone libraries were further screened by PCR using the *A.* sp. BH72 specific primer set 118F and 214R, targeting the *nifH* gene of this bacterium to detect BH72 identical *nifH* sequences from the libraries. Twenty-two positive colony PCR products were randomly chosen as templates from each library. Results showed that all hybridization-positive clones with correct-size inserts carried *nifH* sequences identical to *A.* sp. BH72 *nifH* (Fig. 6A and 6B). These results eliminated the possibility of contamination of the 2nd round PCR in the two samples collected in Oct. 2006 and Jul. 2007 and most probably also for samples from other months.

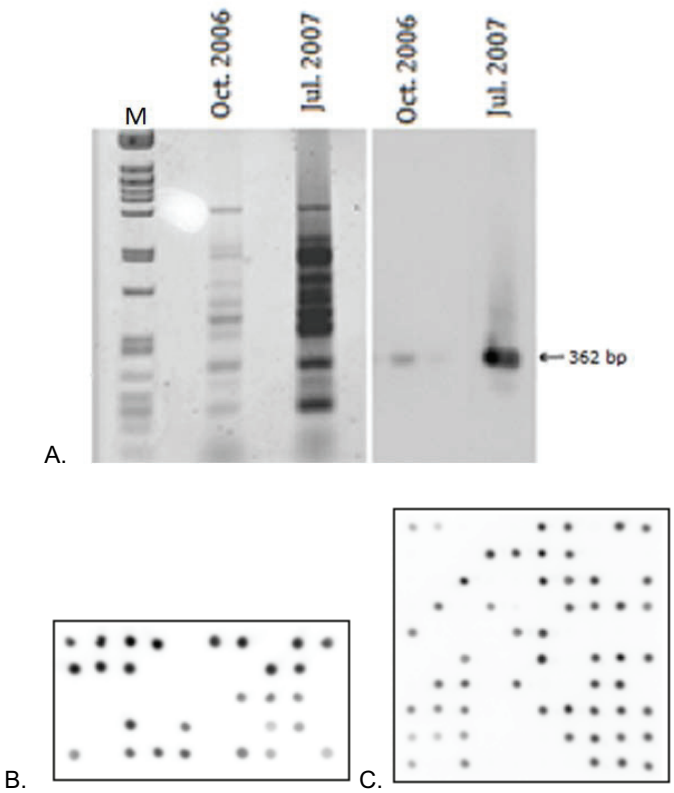


Fig. 5 (A) Agarose gel pattern showing the single stage RT-PCR products amplified from RNA extracts of spruce roots sampled in October 2006 and July 2007 (left) and corresponding Southern blot drawn to scale using the Ps nifH DIG4 probe for hybridization (right); M, size marker (*Pst*I-digested lambda DNA); (B) Dot blot of 50 colony PCR products of the clone library from the sample collected in October 2006 using the Ps nifH DIG4 probe for hybridization; (C) Dot blot of 100 colony PCR products of the clone library from the sample collected in July 2007, using the Ps nifH DIG4 probe for hybridization

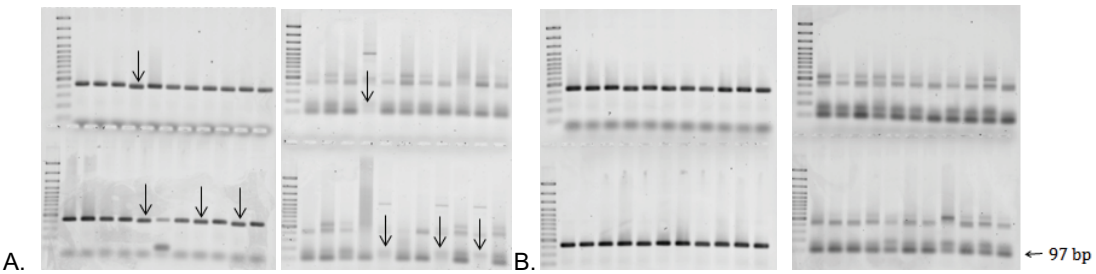


Fig. 6 Agarose gel patterns showing PCR products using clone vector primers (pJET1.2 F+R) on colonies (left) and *A*. sp. BH72 *nifH* specific primers (118F and 214R) on colony PCR products (right). Colony PCR products were obtained from clones of the October 2006 (A) and July 2007 (B) libraries. For BH72 specific *nifH* PCR, only clones were selected as templates which were positive in dot blot hybridization with the with Ps nifH DIG4 probe (Fig. 5B and 5C). Arrows pointing to the clones which

showed positive signals in dot blot hybridization but carried incorrect size of inserts which could not get the amplicon with BH72 *nifH* specific primers.

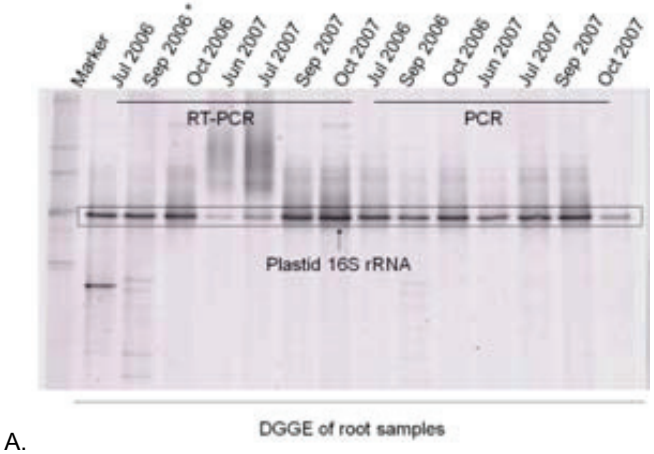
3.1.4 DGGE analysis of 16S rRNA and 16S rDNA

In order to relate the community structure of the diazotrophs with that of the total bacterial community, root samples were also subjected to a 16S rRNA (RT-) PCR-DGGE analyses. Three soil samples were also analyzed in parallel to obtain the information on the 16S gene pool in root-free rhizosphere soil. The 16S primer set from Muyzer and Smalla (Muyzer & Smalla, 1998) which targets 16S rRNA gene sequences of bacteria and plant organelles was used. DGGE gel patterns showed that the soil community structure was complex and no phylotype appeared to be particularly predominant or active. In contrast, the DGGE pattern of spruce root samples, very consistent within the two-year period and not unexpectedly, was dominated by a single band (Fig. 7) which was identified as plastid 16S rRNA or rDNA. Furthermore a Southern blot analysis of PCR fragments separated by DGGE was carried out with a digoxigenin-labeled probe specific for 16S rRNA genes of *P. stutzeri*. Hybridization signals obtained by hybridization with this probe were weak for soil (Fig. 7) or hardly detectable for root samples, even not after prolonged exposure of the blot (data not shown). DGGE in combination with Southern blot analysis indicated that *P. stutzeri* was neither abundant in rhizosphere soil nor in the roots of spruce.

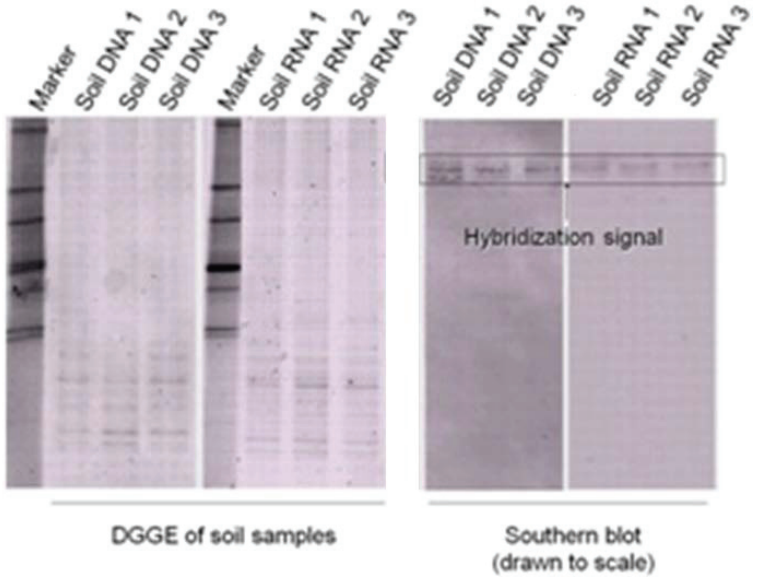
This result is not unexpected, since it is well known from 16S rRNA-based cultivation independent PCR-DGGE and clone library approaches that although diazotrophs may be highly adapted to the soil physical and chemical properties, they are rarely dominant (Demba Diallo*, et al.*, 2004). As mentionted before, the 16S primer set used here, and most of the other universal 16S rRNA gene primer sets commonly used (Weidner*, et al.*, 2000, Sakai*, et al.*, 2004), co-amplify 16S rRNA genes from plastids and plant mitochondria because of a high sequence conservation of the primer target sites. Accordingly, 16S rRNA gene clone libraries obtained from plant DNA may consist only to 10% of sequences stemming from bacteria, while the overwhelming

majority is represented by ribosomal sequences from plant organelles (Kretzer*, et al.*, 2009). Likewise, Burke *et al*. (Burke*, et al.*, 2006) reported that as many as 28% of cloned 16S rRNA gene sequences represented *Pinus* spp. chloroplast ribosomal RNA when a clone library obtained from DNA extracted from rhizosphere soil in a loblolly plantation was analyzed. Consistent with these observations, in this study less than 15 % of all 16S rRNA sequences analyzed originated from bacteria.

Other universal primer sets targeting 16S rRNA genes have been designed which avoid co-amplification of plastid or mitochondrial sequences by either specifically amplifying only bacterial 16S sequences (Chelius & Triplett, 2001) or by allowing the separation of bacterial from eukaryotic amplicons by size fractionation (Sun*, et al.*, 2008). However, the bacterial PCR fragments obtained with these primer sets are all larger than 500 bp which hardly suitable for DGGE analysis.



A.



B.

Fig. 7 (A) DGGE fingerprint analysis of *nifH* genes and *nifH* transcripts in spruce roots; (B) root-free soil with corresponding Southern blot showing the specific hybridization of a probe targeting 16S rRNA genes of *P. stutzeri*. DGGE gels and Southern blot are drawn to scale. *, *P. stutzeri nifH* transcripts were detected in roots collected on this sampling date. Molecular markers as in Fig. 3C

3.1.5 Detection of *nifH* transcripts most similar to *Rhizobium rosettiformans nifH*

To evaluate the effect of RNA extraction on the apparent diversity of *nifH* transcripts RNA extraction was carried out with additional spruce root material sampled in July 2009 by protocol II, combining the modified Chang's method with a commercial kit which replaced the proteinase K and DNase I treatment steps. With this extraction

protocol highly pure RNA was obtained from less starting material (about 100 mg roots per extraction), but much less RNA was extracted. Single stage RT-PCR was applied to this RNA extract using LNA-substituted primers to increase the sensitivity. The desired 362 bp PCR product was retrieved from agarose gel and directly sequenced. A 276 bp *nifH* sequence was obtained (Burbano*, et al.*, 2011) which has a 98.7 % DNA sequence identity to the partial *nifH* sequence from *Rhizobium rosettiformans* (Kaur*, et al.*, 2010).

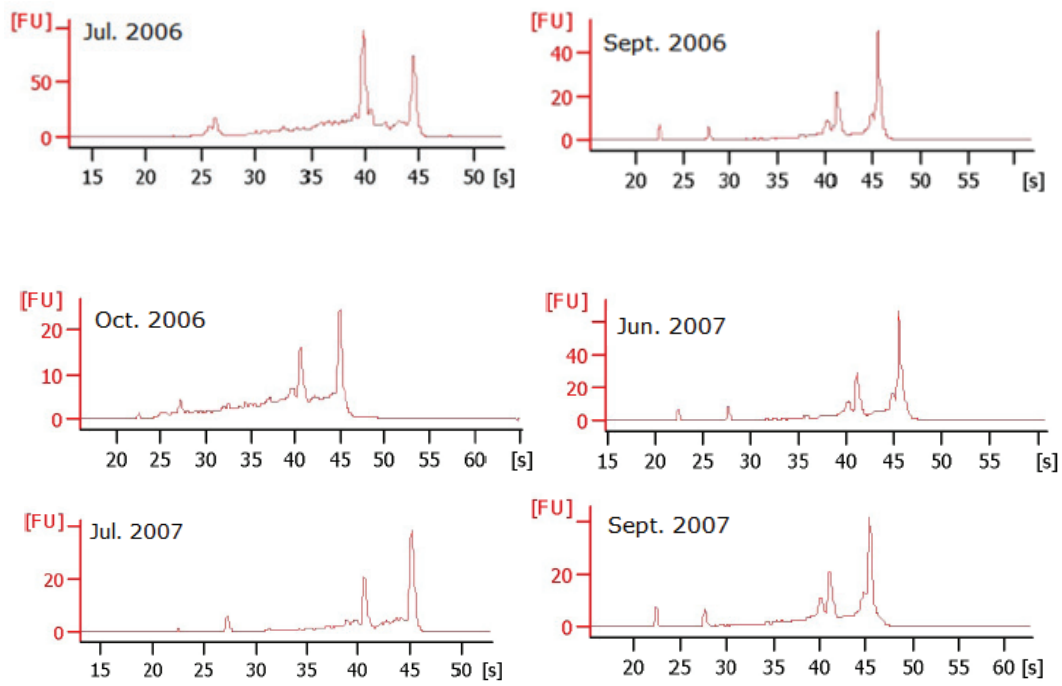
So far by comparative sequence analysis of RT-(nested) PCR products *P. stutzeri nifH* transcripts from one particular sample, *A.* sp. BH72 *nifH* transcripts from most tested samples, and *R. rosettiformans nifH* transcripts from a RNA sample extracted with an alternative method had been detected. Therefore, the diversity of active diazotrophs associated with the spruce coarse roots was apparently extremely low, the three detected nitrogen-fixing phylotypes interestingly excluding each other in the individual samples. These results suggested that apart from sample heterogeneity and variability, the RNA extraction protocol may have a drastic effect on the apparent composition of *nifH* phylotypes in spruce roots.

3.2 Quantification of *nifH* transcripts in spruce roots by RT-real-time PCR

3.2.1 Quality control of RNA extracts and synthesized cRNAs

Determining the integrity of RNA is a critical step in gene expression analysis. Spruce RNA extracts were tested with the Eukaryote Total RNA Nano assay on the Agilent 2100 bioanalyzer. Using electrophoretic separation on microfabicated chips, RNA samples were separated and subsequently detected by laser induced fluorescence detection. The bioanalyzer software generated an electropherogram as shown in Fig. 8A and the RNA integrity number (RIN) was estimated as well. The synthesized *Azoarcus* sp. BH72 specific *nifH* cRNA with expected size 1174 bp and 16S cRNA with expected size 1605 bp were tested in the same assay. The electropherogram and electrophoresis profiles were shown in Fig. 8B.

A.



B.

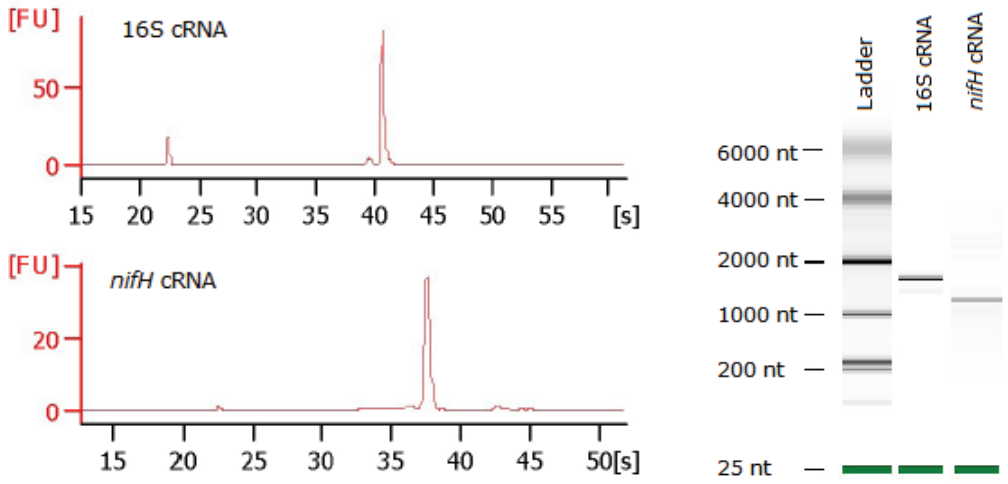


Fig. 8 (A) The electropherograms of spruce RNA extracted from different months. The RIN of the samples were 4.8, 4.2, 6.6, 3.9, 4.4 or 3.9 for Jul. 2006, Sept. 2006, Oct. 2006, Jun. 2007, Jul. 2007, or Sept. 2007, respectively. (B) The electropherogram (left) and electrophoresis profiles (right) of *Azoarcus* sp. BH72 specific *nifH* and 16S cRNAs, using the Agilent RNA 6000 Nano ladder as a marker.

3.2.2 Quantification of *nifH* transcripts in different months' samples

As reported in section 3.1, *nifH* transcripts from bacteria affilated with *Azoarcus* sp.

BH72 represented the predominant phylotypes in spruce roots which were sampled in several months in 2006 and 2007. However, no data were available on the total amount of *nifH* transcript there. To quantify total *nifH* expression in these samples, nested real-time RT-PCR with universal primer sets was used. *A.* sp. BH72 specific *nifH* transcripts in spruce roots were quantified by real-time RT-PCR with specific primers. Universal primer nifH3 was used for reverse transcription and the cDNA generated was used for *nifH* quantification with universal and specific primer sets. This did not only allow a comparison of the *nifH* transcript level among the samples, but also between the two amplification approaches. Since *nifH* of *A.* sp. BH72 was the predominant phylotype in the spruce root samples studied (details in section 3.1), the *nifH* quantification using universal and specific primer sets should result in similar values. Using BH72 *nifH* cRNA as standard, calibration curves were constructed spanning 8 orders of magnitude (4.7×10^7 to 4.7×10^0 copies) for RT-qPCR (Fig. 9A) and 6 orders of magnitude (4.7×10^7 to 4.7×10^2 copies) for RT-nested qPCR applications (Fig. 9B). Since the amplification efficiency slightly varied from batch to batch, several standards were run routinely with each sample set and were used as for quantification. A strong linear relationship with a correlation coefficient of $r^2 > 0.99$ between the fractional cycle number and the starting copy number was found for cRNA standard curves for specific and universal primer sets. However, the standard curve obtained with universal primers resulted in a narrower linear range. Furthermore, the PCR efficiency was only 55.2%, compared to 95.7% in the case of specific primers

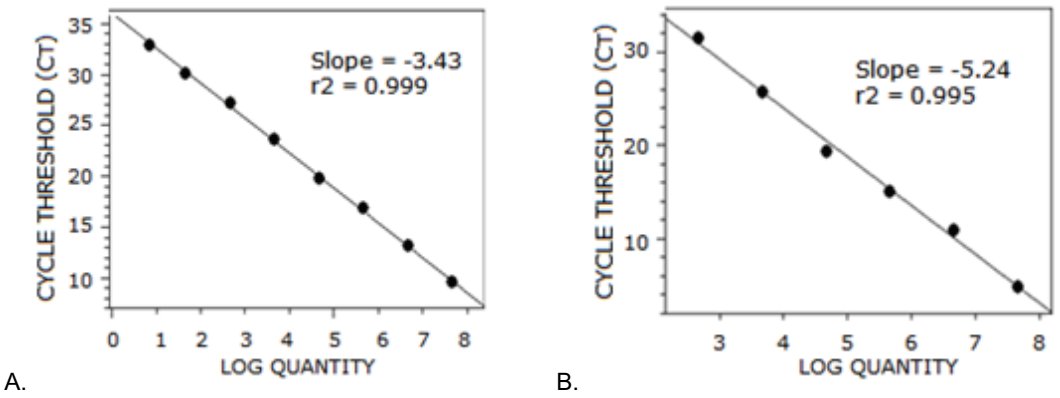


Fig. 9 Standard curves for *Azoarcus* sp. BH72 *nifH* cRNA templates used in RT-qPCR (A) or RT-nested qPCR (B) applications applying primersets which are either suitable for specific apmlification of *nifH* from *A*. sp. BH72 or for broad-range *nifH* amplification, respectively.

Results on quantification of *nifH* expression in spruce roots in 6 months of 2006 and 2007 are shown in Fig. 10. Cloning and sequencing of RT-qPCR products obtained with *A*. sp. BH72 specific *nifH* primers revealed that the nucleotid sequence of the PCR fragment was 100% identical to the partial *nifH* gene from *A.* sp. BH72 (data not shown). This result confirmed that the primers used were specific for *nifH* from *A.* sp. BH72. The copy number of *nifH* transcripts varied 100-fold among the 6 samples. The highest expression level was detected in the September 2006 and July 2007 samples. Surprisingly, copy numbers detected by using universal primers were all lower than using specific primers. To determine if these differences are significant, an unpaired *t* test was carried out using normalized qPCR data from samples collected in June 2006, September 2006, October 2006, and July 2007 with an online tool (QuickCalcs, www.graphpad.com). The copy numbers calculated from the June and September 2007 samples were far beyond the linear range of the standard curve and therefore too low to be considered. This analysis showed that the RT-nested qPCR approach resulted in an estimate which is 64.7 % ± 3.3 % (mean ± SD) of the estimate obtained with the qPCR using specific primers and that this difference between these two values is extremely statistically significant ($p<0.0001$). The result of the diversity analysis described in section 3.1, namely that *nifH* transcripts from *A.* sp. BH72

represented the predominant expressed *nifH* phylotype in the spruce root RNA extracts was confirmed by RT-qPCR. The lower estimate obtained with universal than with specific primers is intriguing and requires further investigations including comparative studies on pure cultures.

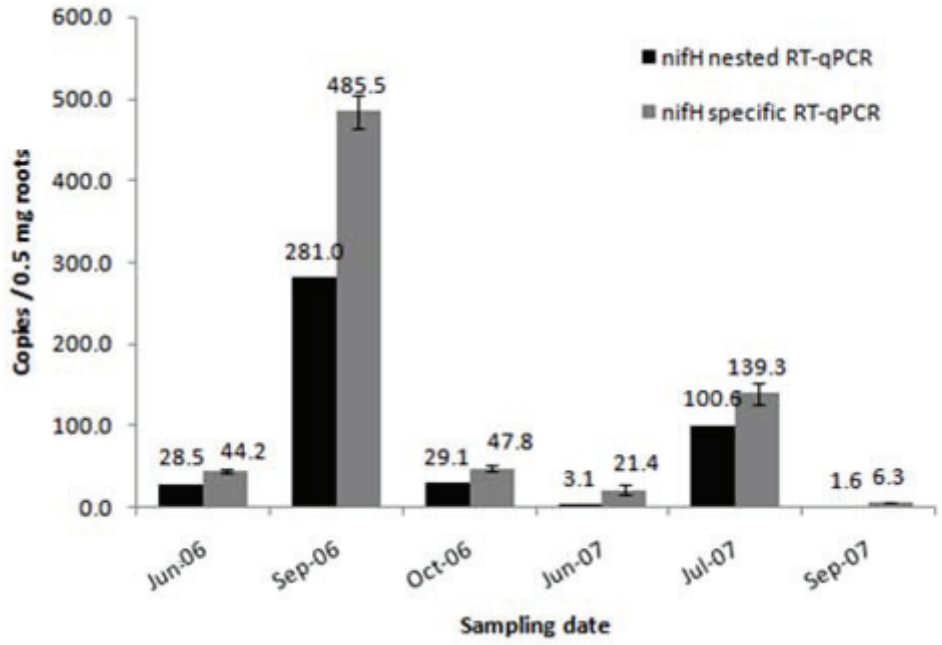


Fig. 10 Comparison of *nifH* transcript levels in spruce roots by *A*. sp. BH72 specific RT-qPCR (gray column) and universal RT-nested qPCR (black column). Roots were collected in 2006 and 2007. Columns with error bars (SEM, standard error of the mean) were averages from two replicates. RT-nested qPCR with universal primers was not replicated.

The observation that the copy numbers determined using universal primers were below or close to the detection limit requires improving the whole nested PCR system. Especially, it might be helpful to perform more cycles in the first round PCR to increase the detection limit. Nevertheless, the *nifH* expression levels evaluated by using universal primers showed the same tendency as those obtained with specific primers and this indicated that the highest *nifH* expression was detected in summer. This result is consistent with Thaweenut's study (Thaweenut*, et al.*, 2011) on *nifH* gene expression in sugarcane. In this study it was shown by RT-PCR that *nifH*

expression was well detectable in the high-temperature (June to September) and negligible in the low-temperature season (September to December). Therefore, the Sept. 2006 sample, which showed the highest *nifH* expression level was used for further studies. Although performing poorly with the experimental setup used here, nested qPCR with universal primers has successfully been applied for quantification of *nifH* genes in Douglas-fir forest soil (Levy-Booth & Winder, 2010). The poor performance of the nested RT-PCR was probably caused by the inclusion of the reverse transcription step, since it is well known that reverse transcriptase may inhibit PCR at low concentration of template (Chandler*, et al.*, 1998). Chandler reported that the RT-PCR efficiency was reduced, due to the reverse transcriptase inhibition of the PCR, in uncoupled RT-PCR systems below 10^5 to 10^6 copies of starting RNA. This result matched with the observation in this study that template concentrations were at the low end of the linear detection range of the *nifH* cRNA standard curve when universal primers were used. The fact that the same template concentrations were not at the low end of the linear detection range of the *nifH* cRNA standard curve when specific primers were used suggests that PCR reactions involving highly degenerate, universal primers might be more affected by reverse transcriptase inhibition than those using primers without degenerate positions.

3.2.3 Comparison of *Azoarcus* sp. BH72 specific 16S rRNA and *nifH* transcript abundances in spruce and pure culture

While the *nifH* transcription level is generally used to indicate the activity of *nifH* expression, the amount of ribosomal RNA (rRNA) serves as an indicator for the growth status of cells (Nomura*, et al.*, 1984, Condon*, et al.*, 1995, Perez-Osorio*, et al.*, 2010). The following experiments were carried out to investigate whether the *nifH* and 16S rRNA transcript levels of *A.* sp. BH72 cells *in planta* were comparable to those in pure culture, where cells were grown under optimum N_2-fixing conditions. For this purpose real-time quantitative RT-PCR with *A*. sp. BH72 specific *nifH* and 16S primer sets was used on the spruce sample collected in September 2006 and on nitrogen

fixing cells of strain BH72. *NifH* and 16S rRNA transcripts were quantified using *nifH* (Fig. 9A) and 16S cRNA standard curves (Fig. 11).

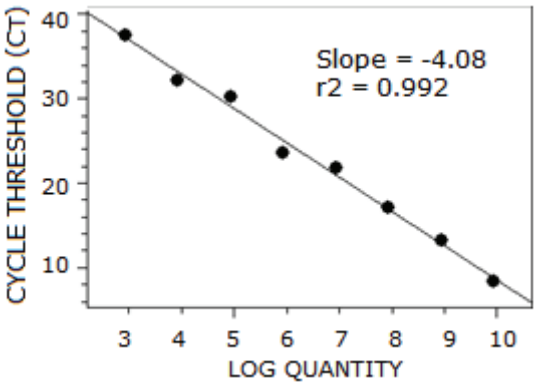


Fig. 11 Standard curve for *Azoarcus* sp. BH72 *in vitro*-transcribed 16S rRNA templates (8.7×10^9 to 8.7×10^2 copies). The RT-qPCR efficiency was 76%.

The identity of RT-qPCR products amplified from spruce roots was verified by sequencing. RT-PCR products showed a perfect nucleotide sequence match to the corresponding partial *nifH* and 16S rRNA sequences from *A*. sp. BH72. Presence of inhibitors in RNA extract obtained from spruce roots was tested by spiking the RT reaction with RNA obtained from nitrogen-fixing cells of strain BH72. When the C_T values of spiking RNA alone and spiking RNA in the root extract were compared, the difference obtained was <0.5 (data not shown). This indicated that, if at all, only negligible amounts of inhibitors were present in RNA extracts from spruce (Nolan*, et al.*, 2006). The transcript abundances calculated for spruce and *A*. sp. BH72 are shown in Table 5.

Table 5. Abundance of 16S rRNA and *nifH* transcripts from *Azoarcus* sp. BH72 in spruce roots (collected in September 2006) and in nitrogen fixing cells of a pure culture [a]

	16S rRNA transcripts (copy number)	*nifH* transcripts (copy number)	16S rRNA : *nifH*
Azoarcus sp. BH72 [b] (pure culture)	(2.82 ± 0.28) × 10^9	(1.2 ± 0.03) × 10^6	2352 ± 233 [d]
Azoarcus sp.BH72 [c] (*in planta*)	(6.25 ± 0.36) × 10^5	485.5 ± 20.3	1290 ± 75 [d]

[a], means from two replicates with standard error of the mean are given

[b], estimated from 10 ng of *A*. sp. BH72 RNA

[c], estimated from 20 ng of spruce RNA

[d], means are significantly different from each other (p = 0.049), as estimated by an unpaired t-test.

These results showed that the 16S rRNA transcript normalized *nifH* expression level of *A*. sp. BH72 *in planta* was 2 times (2552/1290) higher than under optimal conditions in pure culture and that this increase was statistically significant (Table 5).

In order to obtain normalized gene abundances for A. sp. BH72 too, the 16S rRNA : *nifH* gene ratios of *A.* sp. BH72 were determined *in planta* and in pure culture as well. For this purpose the genomic DNA of BH72 cells was taken as reference, since the genome sequence is available (Krause*, et al.*, 2006), and under ideal qPCR conditions the 16S rRNA gene / *nifH* ratio should be 4 to 1. Standard curves were generated from 7 tenfold serial dilutions (8.0×10^6 to 8.0×10^0 copies) of strain BH72 *nifH* linearized plasmid (Fig. 12A) and from 7 tenfold dilutions (6.2×10^6 to 6.2×10^0 copies) of BH72 16S rRNA gene linearized plasmid standard (Fig. 12B) as templates. Using 10 ng BH72 DNA as reference, 16S rRNA gene / *nifH* was calculated as 6.03 ± 0.95 (SD derived from two repetitions).

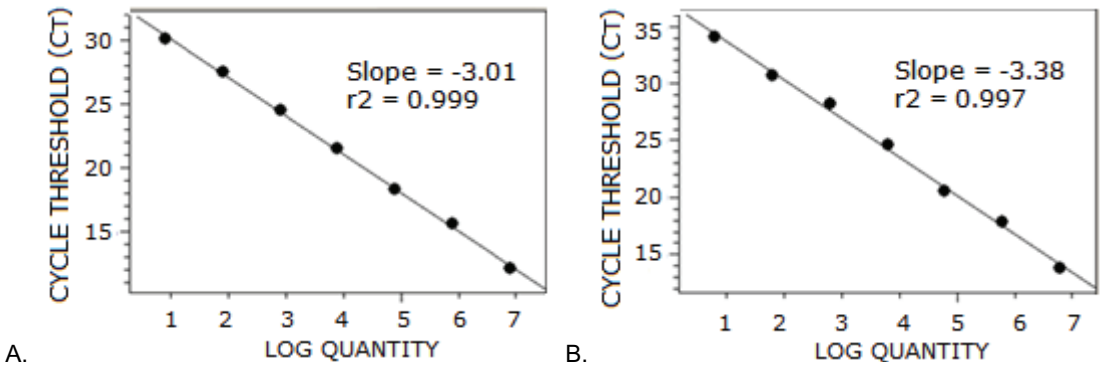


Fig. 12 Standard curves for BH72 *nifH* (A) and 16S rRNA gene templates (B). The PCR efficiency was 115% and 97.6%, respectively.

Quantification of *nifH* gene in spruce roots resulted in 81 ± 1.4 copies / 300 ng DNA, corresponding to approximately 3×10^4 copies / g roots, values which are higher than those reported for *Azotobacter vinelandii* (2.1 to 3.7 log *nifH* gene copies / g soil) in a Douglas-fir forest (Levy-Booth & Winder, 2010). However, quantification of the BH72 16S rRNA gene copy number in the same DNA extracts revealed that the 16S rRNA gene was under the detection limit of 6.2 copies in 300 ng spruce root DNA.

Presence of inhibitors in the spruce DNA extract could be excluded by diluting DNA templates from 300 ng to 100 ng and 50 ng per *nifH* qPCR reaction and observing a corresponding decrease of *nifH* copies from 81 ± 1.4 to 24.4 ± 3.3 and 6.6 ± 0.8 (SD derived from two repetitions), respectively. Furthermore, to test for the possibility that the 16S rRNA gene was more difficult to be retrieved than the *nifH* gene because of contaminants in the root sample, or that the amplification was inhibited by substances co-purified with DNA during extraction (Bustin*, et al.*, 2009), an extraction was performed by spiking spruce root samples with *A*. sp. BH72 cells. The extracted DNA, from the spiked spruce root sample, was subsequently used for BH72-specific *nifH* and 16S qPCR. For a better comparison with previous results the DNA extraction from the BH72-spiked spruce root sample was carried out by mixing 300 mg roots and 5×10^{10} BH72 cells, representing the same amount of material used for obtaining spruce and BH72 DNA before. Using as template each 8.5 ng of DNA from the spiked spruce root sample for *nifH* and 16S qPCR, a 16S rRNA gene / *nifH* gene ratio of 8.12

± 1.0 (SD derived from two repetitions) was obtained from the copy numbers calculated. Therefore, inhibition was most probably not the reason why 16S rRNA genes from *A*. sp BH72 could not be quantified, and their abundance was at the limit of detection in spruce roots. Clearly, the 16S rRNA gene copy number normalized *nifH* abundances from *Azoarcus* sp BH72 *in planta* was drastically increased in comparison to pure cultures.

3.2.4 *NifH* and 16S rRNA gene-based-quantitation of *Pseudomonas stutzeri* and *Rhizobium rosettiformans* in roots of Norway spruce by real-time qPCR

Besides the *A*. sp. BH72 *nifH* transcripts, which were repeatedly found in different spruce root samples, in two distinct samples also *P. stutzeri* and *R. rosettiformans nifH* transcripts (using an alternative RNA extraction protocol) had been detected (Results section 3.1). In order to obtain normalized gene abundances, the 16S rRNA : *nifH* gene ratios of *P. stutzeri* and *R. rosettiformans* were determined *in planta*. Standard curves were generated from 6 tenfold serial dilutions (3.3×10^6 to 3.3×10^1 copies) of *P. stutzeri nifH* linearized plasmid (Fig. 13A), 7 tenfold dilutions (6.7×10^6 to 6.7×10^0 copies) of *P. stutzeri* 16S rRNA gene PCR product (Fig. 13B) and 5 tenfold serial dilutions (2.5×10^5 to 2.5×10^1 copies) of *R. rosettiformans nifH* linearized plasmid standard (Fig. 13C).

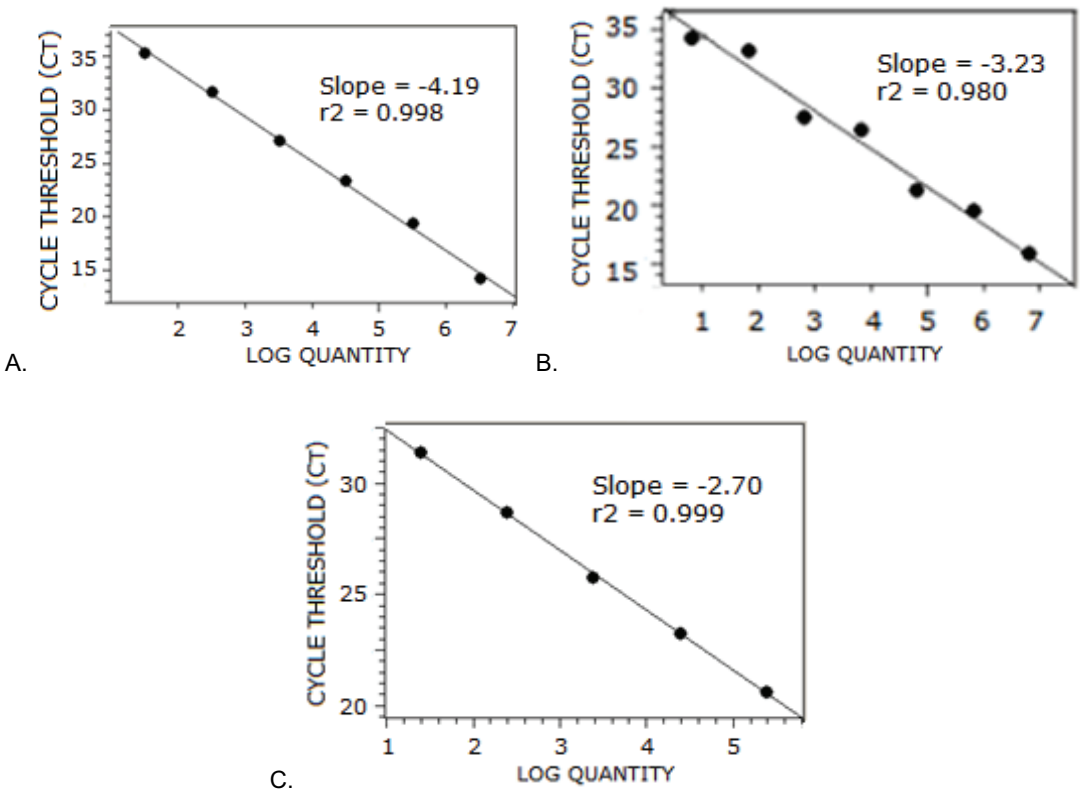


Fig. 13 Standard curves for *P. stutzeri nifH* (A) and 16S rRNA gene templates (B), or for *R. rosettiformans nifH* templates. The PCR efficiency was 73.2% and 104%, or 135%, respectively.

The real-time qPCR analysis of *P. stutzeri nifH* and 16S rRNA genes and *R. rosettiformans nifH* genes in spruce suggested that the copy numbers of the respective genes were all below the low end of the linear range of the standard curves. Moreover, when the qPCR products were analyzed by agarose gel electrophoresis, no PCR products of the predicted size were obtained (data not shown). This indicated that the abundances of these genes in spruce was below the limit of detection and that probably primer-dimers contributed to the fluorescence signal recorded. These data suggested that *A.* sp. BH72 is the major active nitrogen-fixing bacterium associated with spruce roots, while nitrogen fixation in *P. stutzeri* or *R. rosettiformans* is probably not important and or confined to particular roots.

3.3 Detecting NifH protein in spruce roots

3.3.1 Specificity of the primary antibodies

Two polyclonal antibodies were used, which are termed here “NifH antibody” (raised against dinitrogenase reductase of *Azoarcus* sp. BH72) and “NifH peptide antibody” (a universal polyclonal anti-dinitrogenase reductase peptide antibody). To assess the specificity of the NifH antibody, protein extracts from N_2-fixing *Azoarcus* sp. BH72 were separated by SDS-PAGE and immunoblotted. As to be expected (Oetjen*, et al.*, 2009), two bands representing the post-translationally modified and unmodified Fe proteins (Hurek*, et al.*, 1995, Martin & Reinhold-Hurek, 2002, Oetjen & Reinhold-Hurek, 2009), were detected (Fig. 14A). No band was detected with pre-immune serum (Fig. 15A), a control which had not been included previously (Oetjen*, et al.*, 2009).

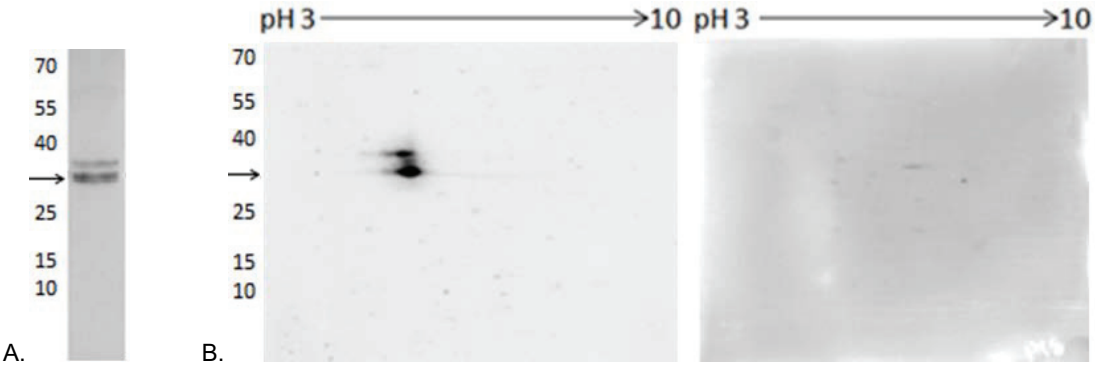


Fig. 14 Protein extracts from N_2-fixing *Azoarcus* sp. BH72 cells which were separated on (A) SDS-PAGE gel and blotted on nitrocellulose membrane. A representative immunoblot with antiserum to dinitrogenase reductase from *A*. sp. BH72 is shown. Representative results were derived from more than three independent experiments; (B) mini 2-DE gel blotted on PVDF membrane. Immunoblot with universal anti-dinitrogenase reductase peptide antibody is shown. The two bands in (A) or two spots in (B left) showed the post-translationally modified (upper) and unmodified (lower, 33-kDa, arrows) forms of NifH subunits. The blots of 1-D and 2-D gel slabs were drawn to the same scale. Positions of molecular weight markers (in kDa) are indicated on the left.

In the case of the NifH peptide antibody, BH72 proteins were separated on a mini 2-D gel and immunoblotted. Two spots, the modified and unmodified Fe proteins, were detected (Fig. 14B left) and no cross reaction was detected with pre-immune serum (Fig. 14B right). Antibodies were not applied to other reference nitrogen-fixing strains, since the focus was to detect NifH of strain BH72 in environmental samples. Though the antibodies showed no cross-reactivity to non-related proteins in *A*. sp. BH72, the possibility of cross-reactivity to non-related proteins present in environmental samples cannot be excluded, considering the high complexity of environmental samples. To obtain universal antibodies to the nitrogenase Fe protein, anti-peptide antibodies were raised against a 13-amino acid peptide which is highly conserved among iron proteins from different species. This peptide is predicted to be located on the surface of the native nitrogenase and has some hydrophilic residues (data not shown), which facilitate recognition of the peptide sequence within the protein. However, the bioinformatically predicted specificity for NifH and broad-range recognition of the anti-peptide antibody of NifH proteins from various nitrogen fixing bacteria needs to be further verified in practice.

3.3.2 Spruce root proteins reacted with NifH antibody

Spruce root samples were collected in 15 different months during a four-year period. Proteins were extracted from samples from each of the 15 months, separated on SDS-PAGE gels and putative NifH proteins were screened by Western blot analysis using the NifH antibody. The results of this analysis are shown in Fig. 15.

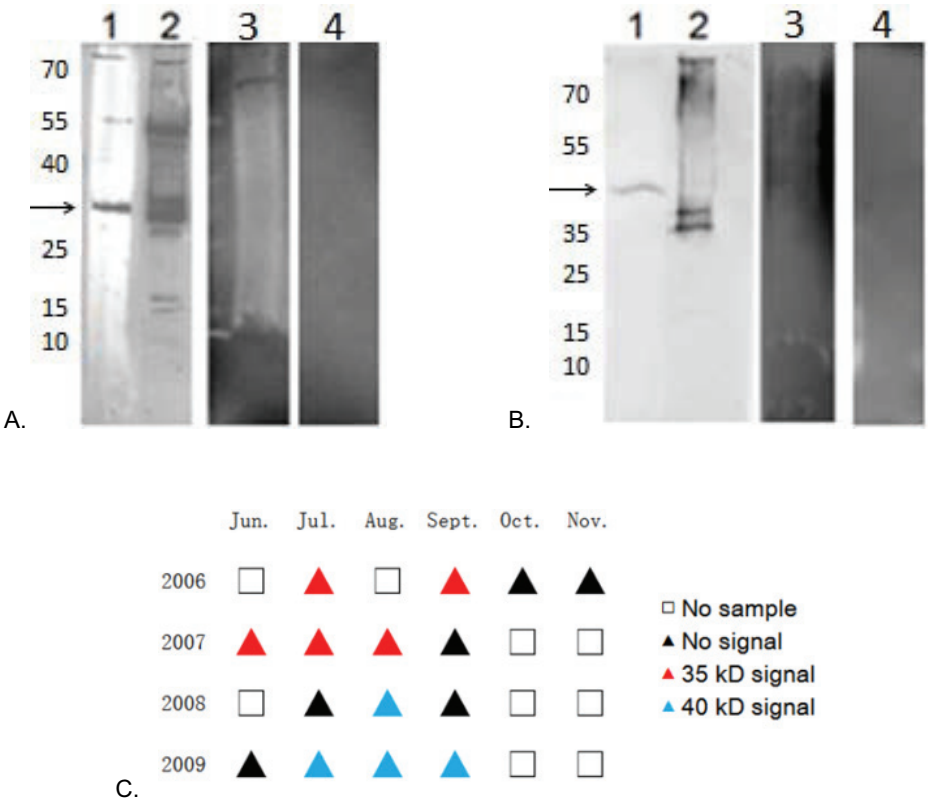


Fig. 15 Western blot analysis of protein extracts from spruce (1, 3) and *A*. sp. BH72 (2, 4) as control. Blots were incubated with antiserum to dinitrogenase reductase from *A*. sp. BH72 (1, 2) or with pre-immune serum (3, 4). The antiserum shows strong reactions with bands (arrows) of the apparent molecular masses of 35-kDa (A) and 40-kDa (B) in spruce protein extracts sampled between June and November in 2006 to 2009 (C). Representative results were derived from more than three independent experiments. Positions of molecular weight markers (in kDa) are indicated on the left.

In six of the months, samples were negative. In nine of the months samples showed positive reactions with the NifH antibody and no reactions with pre-immune serum, but the signals had two different apparent molecular masses of 35-kDa (Fig. 15A) and 40-kDa (Fig. 15B) in four and five months of the 9 months sampled, respectively (Fig. 15C). NifH is a conserved protein (Zehr*, et al.*, 1990), and the apparent or the predicted molecular weight of unmodified NifH from different diazotrophs are close to 32-kDa (Mevarech*, et al.*, 1980, Torok & Kondorosi, 1981, Fuhrmann & Hennecke, 1984, Murphy*, et al.*, 1993, Beesley*, et al.*, 1994). As judged by the apparent molecular masses, the two signals obtained probably neither represented the

unmodified NifH subunit nor the ADP-ribosylated larger one (Oetjen & Reinhold-Hurek, 2009). However, since the proteins were recognized by the NifH antibody raised against dinitrogenase reductase from *Azoarcus* sp. BH72 and no cross reactions were observed between the samples and pre-immune serum, the 35 and 40 kDa proteins may represent NifH inspite of the different sizes compared to *Azoarcus* sp. BH72 modified and unmodified NifH. Possibly, other modifications of NifH, as described for cyanobacterium *Gloeothece* (Gallon*, et al.*, 2000) and *Azospirillum amazonense* (Song*, et al.*, 1985), than ADP-ribosylation might exist *in planta*, which have been not detected yet. Furthermore, it also appears to be possible that the protein extracts from spruce roots were not as pure as the bacterial cell proteins and the associated interfering compounds affected protein migration during SDS-PAGE. It is known that protein extracted from mature plant tissue typically contains higher levels of interfering compounds, such as polysaccharides, organic acids, lipids, phenolic compounds, and a broad array of other secondary metabolites (Karppinen*, et al.*, 2010). Subsequently, a 2-D gel system was applied to the root protein samples to improve resolution and to eventually excise the immunoreactive spots from 2-D gels for MS analyses.

3.3.3 Identification of the 40-kDa protein

As shown in Fig. 15C, the 35-kDa signal could only be detected until August 2007. Therefore, from August 2007 on protocol II was used for protein extraction and the 40-kDa signal has been detected ever since. Two-dimensional gel electrophoresis was performed using the protein extracted from Aug. 2008 samples. Multiple IEF gels were run in parallel and subsequently the slab gels were blotted onto nitrocellulose membranes, which were incubated with NifH antibody or pre-immune serum. For staining, gels were incubated with colloidal Coomassie. The 40-kDa signal was detected by 2-D gel immunoblot analysis, but instead of an expected spot the signal appeared as an unfocused, noncontinuous band in an acidic pH range (Fig. 16B). Possible reasons for this horizontal streaking were that the 40-kDa protein was a

glycoprotein and that protein extracts contained substances interfering with isoelectric focusing.

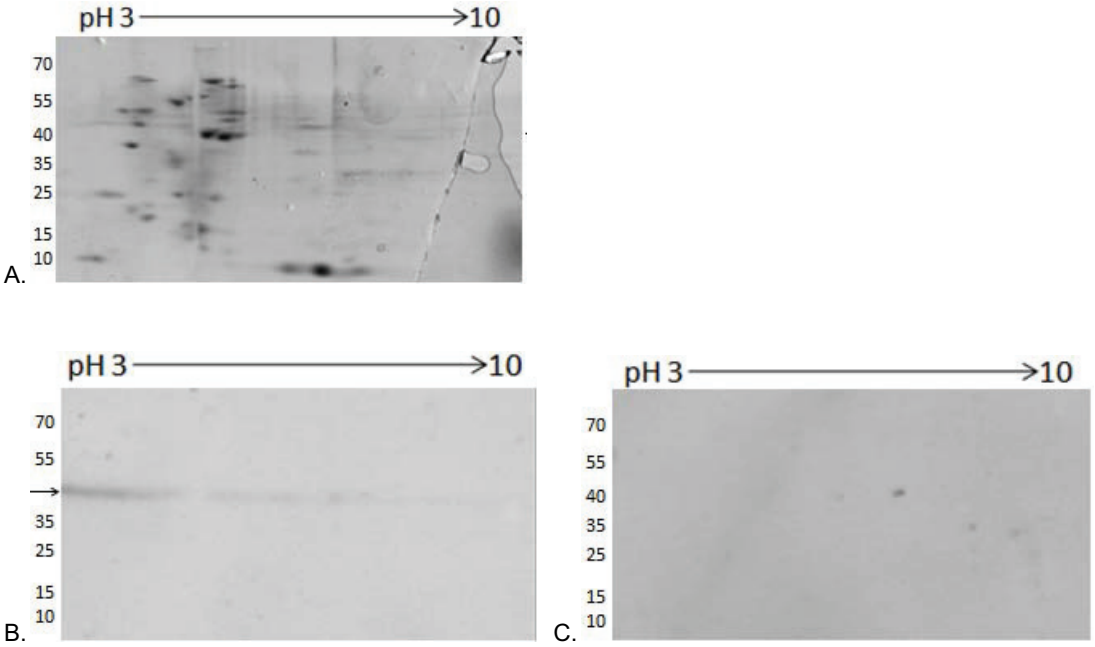


Fig. 16 2-DE gel pattern of a protein extract from a August 2008 spruce sample stained with colloidal Coomassie (A); Immunoblots incubated with antiserum to NifH of *A*. sp. BH72 (B) or with pre-immune serum (C). The signal obtained on the immunoblot with antiserum showed horizonzal streaking and had an apparent molecular mass of 40-kDa (arrow). Positions of molecular weight markers (in kDa) are indicated on the left. The 2-D gel and blots are drawn to scale. Representative results are shown which were derived from three independent experiments.

Taylor *et al.* (2004) reported that post translational modification of a particular protein could result in multiple immunoreactive spots. Furthermore it was known from the 2-DE pattern of glycoproteins that different glycoforms of the same protein can form trains of spots caused by different isoelectric points (Nawarak*, et al.*, 2004). Therefore, it appeared to be possible that a glycosylation modification of the NifH subunit might have been responsible for the form of the immunological signal obtained on the blot (Fig. 16B). However, so far there is no report on dinitrogenase reductase with an apparent molecular weight of 40 kDa with such a modification yet.

The Digoxigenin (DIG) glycan detection and DIG glycan differentiation kits were used to test whether the 40-kDa protein could be a glycoprotein and which type of glycosylation it might be. Due to the limitation of sample, these experiments were carried out by SDS-PAGE instead of 2-DE system. Using the DIG glycan detection kit for the detection of sugars in glycoconjugates, signals were observed in the range of 40-kDa. There are five types of protein glycosylations: N- and O- glycosylation, C-mannosylation, phospoglycation and glypiation (Spiro, 2002). Applying the five DIG-labeled lectins available in the DIG glycan differentiation kit, signals could only be observed with DIG-labeled GNA (*Galanthus nivalis* agglutinin) (Fig. 17).

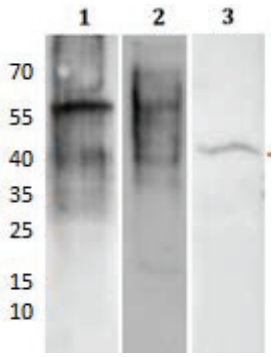


Fig. 17 Glycan detection (1) and glycan differentiation (2) blots using digoxigenin hydrazide and DIG-labeled GNA, respectively, for detection of carbohydrate moities. Immunoblot incubated with antiserum to NifH from *A*. sp. BH72 (3) for comparison. Positions of molecular weight markers (in kDa) are indicated on the left. Blots are drawn to scale. Representative results were derived from two independent experiments.

The lectin GNA recognizes terminal mannose (Shibuya*, et al.*, 1988) , indicating the presence of N- or O-glycosidically linked mannose residues on the blot. To remove the sugar moieties an enzymatic deglycosylation kit was used. However, deglycosylation was not successful with root protein samples. Possibly, the reaction conditions used were not suitable for enzymatic deglycosylation or resolubilization. In order to retain the activity of the deglycosylation enzymes, the root proteins needed to be precipitated out of the protein buffer (containing detergents) and dissolved in the deglycosylation reaction buffer (no detergent), in which deglycosylation or resolubilization might have not occured because of presence or absence of detergents, respectively. Since enzymatic deglycosylation could not be achieved and

because of the limitation of sample, analyses on the signals obtained with the glycan detection and differentiatin kits were not further pursued. Clearly, the resolution obtained with 1-D SDS-PAGE was not sufficient to match the three signals obtained in Figure 17 with certainty.

Apart from glycosylation there are also other factors known that can contribute to low-quality or multiple spots from single proteins in 2-DE gels. Nonprotein impurities with a net ionic charge such as phenolic compounds can interfere with isoelectric focus and lead to horizontal streaking and insufficient focusing. When several commercial kits (e.g. clean-up column and filtration) were applied to further purify the protein extracts and to overcome this effect, the immunoreactive signal dramatically decreased after purification (data not shown). Finally, the cleanup was done by using the ProteoExtract Protein Precipitation kit (commercially available from Merck) on proteins which have been eluted from the nitrocellulose membrane by a phenol treatment (section 2.3.5). With this method a protein extract with improved quality (no brown color in the protein sample) could be obtained with a minimal loss of protein (70% recovery, comparing the immunoreactive signal before and after purification).

This time, immunoreactive signals of good quality were obtained on Western blots after 2-D gel electrophoresis of the purified extracts. Signals were separated into a train of three discrete spots with an apparent molecular mass of 40-kDa (Fig. 18 A-C).By comparing the blot to the 2-D Coomassie-stained gel three gel spots were located and excised (Fig. 18A and B) and analyzed by mass spectrometry (MS). Homology-based protein identification using LC-Orbitrap-MS revealed that none of the peptides detected were related to NifH.

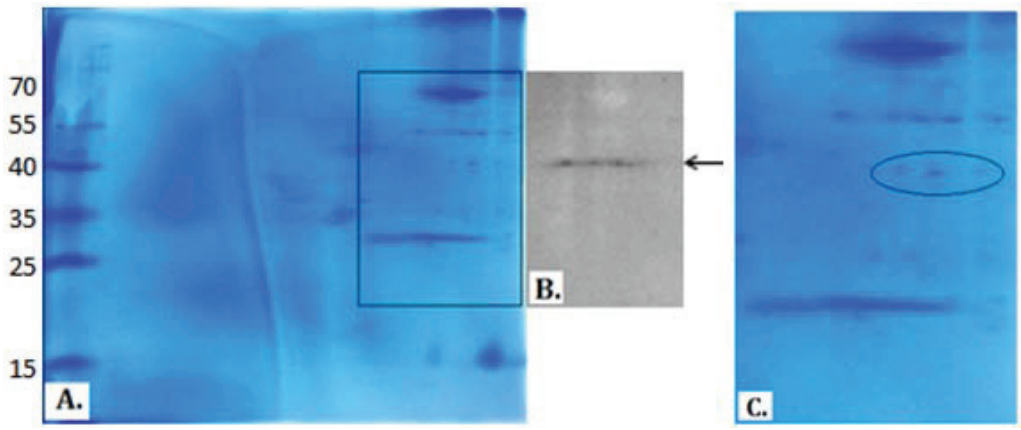


Fig. 18 2-DE gel pattern of a purified protein extract from an August 2008 spruce sample stained with colloidal Coomassie (A); Partial immunoblot from (A) incubated with antiserum to NifH of *A*. sp. BH72 (B); The boxed region in (A) was zoomed in and is shown in (C). The encircled region in (C) pointed out the three spots (arrow), which were excised from the 2-DE gel. The boxed part of the 2-D gel and partial blot are drawn to scale. Positions of molecular weight markers (in kDa) are indicated on the left. Representative results were derived from two repetitions.

3.3.4 Identification of the 35-kDa protein

The 35-kDa immunoreactive signal was detected in spruce samples from July 2006 to August 2007 by immunoblots on nitrocellulose membranes which had been incubated with antiserum to NifH of *A*. sp. BH72. This signal could not be detected with the same procedure if the extracted proteins were stored at -80℃ for more than one year. However, the 35-kDa signal could be detected in protein extracts after long-term storage at -80℃ when a PVDF membrane was used for immunoblotting. Compared to the nitrocellulose membrane the PVDF membrane offers better protein retention, physical strength, broad chemical compatibility (Pluskal*, et al.*, 1986), and allows protein binding in the presence of SDS. When protein was extracted from environmental samples, some interfering compounds co-precipitated with proteins. Possibly, multiple freeze-thaw cycles changed the solubility of compounds in the protein extract, which effect protein binding to the nitrocellulose but not to the PVDF membrane, similar to SDS.

Due to the sample limitation, a mini 2-D gel system was applied. In order to get a satisfactory separation of proteins in a mini 2-DE gel, the amount of protein loaded was reduced to about 1 / 10 of a standard 2-DE gel. SYPRO Ruby was used for gel staining. This protein gel stain has a similar sensitivity to that of silver staining, but greater sensitivity than colloidal Coomassie staining. Besides the substantial reduction of sample size, compared to the Bio-Rad model 175 tube cell system, the mini 2-D gel system reduced more than half of the time required to complete the entire run. Western blot analysis was carried out using NifH antibody. Two immunoreactive signals were detected with molecular masses of 35-kDa and 50-kDa (Fig. 19A). The 50-kDa signal was also present in blots incubated with pre-immune serum (Fig. 19B). Therefore the 35-kDa signal was regarded as the specific immunoreactive signal for NifH which could be easily identified on the SYPRO Ruby stained mini 2-D gel as a discrete spot (Fig. 19C).

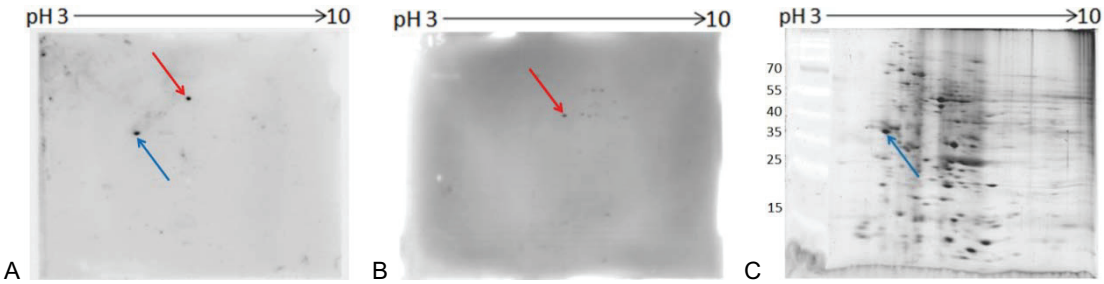
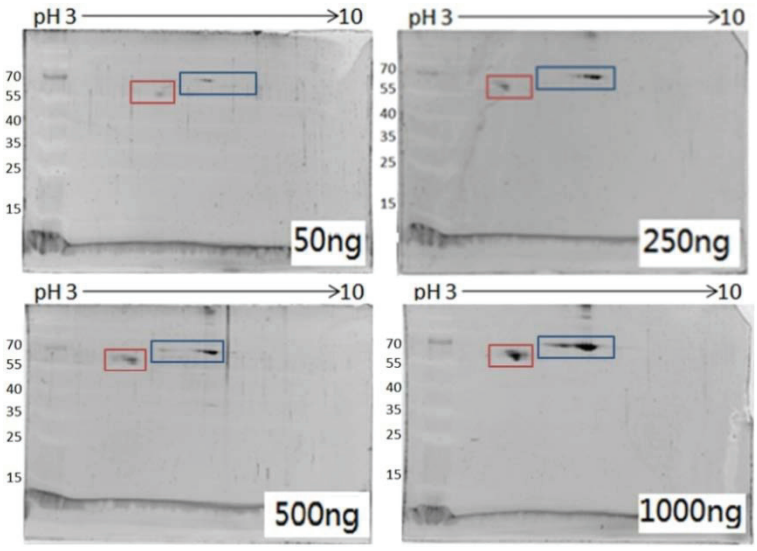


Fig. 19 Representative 2-DE slab-gel immunoblots of September 2006 samples incubated with antiserum to NifH of *A*. sp. BH72 (A) or with pre-immune serum (B). The immunoblot treated with antiserum showed two immunoreactive spots with apparent molecular masses of 35-kDa (blue arrow) and 50-kDa (red arrow). The immunoblot with pre-immune serum showed an immunoreactive spot with an apparent molecular mass of 50-kDa (red arrow). The corresponding 35-kDa protein in the SYPRO Ruby stained mini 2-DE gel is labeled with a blue arrow (C). Positions of molecular weight markers (in kDa) are indicated on the left (C). Representative results were derived from more than three independent experiments. Blots and mini 2-DE gel are drawn to scale.

One of the advantages of SYPRO Ruby protein staining, is its broader linear dynamic range compared to silver staining (Lopez*, et al.*, 2000). Making use of this advantage,

SYPRO Ruby staining was used to quantify the protein spots in 2-DE gels. A standard curve was constructed using the densitometric units of scanned standard spots against the amount of proteins loaded in gels. To explore how the SYPRO Ruby stain performed for difficult to stain proteins, a glycoprotein, yeast carboxypeptidase Y, was used as a reference together with a BSA standard (Fig. 20A). Standard curves (Fig. 20B) showed at each dilution level, with the same protein loading, that the glycoprotein relative to BSA resulted in about half of the densitometric volume. This indicated that quantification of glycoprotein according to a BSA standard curve leads to an underestimation. SYPRO Ruby could detect glycoprotein, but not with the same sensitivity as BSA. In order to accurately quantify a target protein, it would be necessary to use a standard with characteristics similar to the target to construct the standard curve.

A.



B.

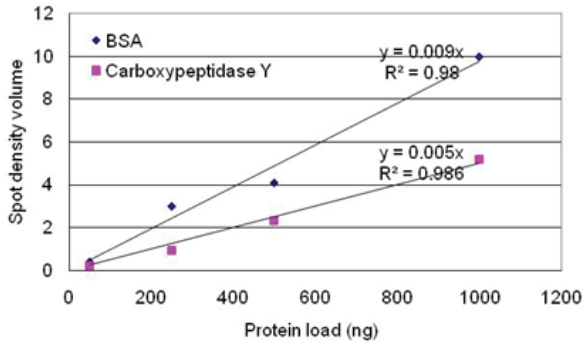


Fig. 20 (A) Protein pattern of four dilutions of the two standards, BSA (blue rectangle) and carboxypeptidase Y (red rectangle), on mini 2-DE gel stained with SYPRO Ruby, Positions of molecular weight markers (in kDa) are indicated on the left; (B) Standard curve constructed with BSA and carboxypeptidase Y. Experiments were not repeated.

The quantities of the 35-kDa protein spot (Fig. 19C) or the NifH proteins (including modified and unmodified NifH) of BH72 (Fig. 21) were determined by interpolation onto BSA standard curves to be 125 ng or 60 ng, respectively. Protein concentrations in protein extracts were determined using Bio-Rad RC/DC kit and a BSA standard curve (Fig. 22). The total amounts of protein loaded on 2-DE gels were estimated as 1600 ng for spruce root proteins (Fig. 19C) and 1300 ng for BH72 proteins (Fig. 21). According to these estimates the putative NifH protein was calculated to represent 7.8% of the total protein in spruce root, while it made up only 4.6% of the total protein in

nitrogen fixing cells of a pure culture of *A*. sp. BH72. Since it is implausible that NifH protein levels in spruce roots and nitrogen fixing cells were in the same range, the 35-kDa immunoreactive signal, like the 40-kDa signal before, probably resulted from unspecific cross-reactions and did not represent NIfH. Consistent with this conclusion no NifH related peptide sequences could be identified when the 35-kDa spot was excised from the 2-DE gel and was analyzed by MS.

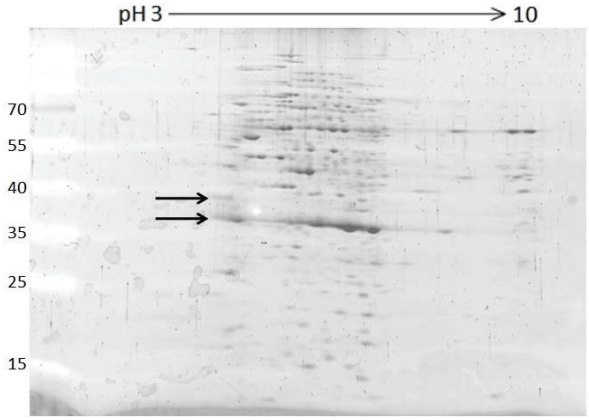


Fig. 21 Representative protein pattern of *Azoarcus* sp. BH72 N_2-fixing cells in mini 2-DE gel, stained with SYPRO Ruby. The modified (upper) and unmodified (lower) forms of NifH subunits are labeled by an arrow,. Positions of molecular weight markers (in kDa) are indicated on the left. Representative results from three repetitions are shown.

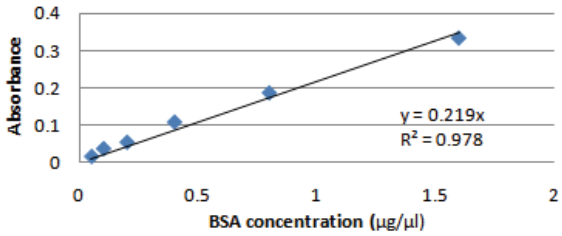


Fig. 22 Standard curve constructed with BSA using Bio-Rad RC/DC protein assay kit.

Nevertheless, the 35-kDa spot was excised from a 2-D gel and analyzed by MS. From the MS data, no NifH related peptide sequence could be identified.

3.3.5 Immunoblot with anti-NifH peptide antibody

In order to eliminate the interfering cross reaction signals, an affinity purified anti-NifH peptide primary antibody, with 2 predicted active epitopes, was applied for immunodetection. Twenty new protein extracts (obtained with extraction protocol I) from different months' samples were analyzed. The nonspecific 40-kDa and 35-kDa signals were not present in immunoblots incubated with the anti-peptide antibody, but there were other immunoreactive signals observed in the higher pH range (Fig. 23A). Since these signals were not detectable using either antiserum to NifH from *A*. sp. BH72 or the pre-immune serum of NifH peptide antibody (Fig. 23B and C), these signals were either nonspecific or represented NifH proteins which did not react with antiserum to NifH from *A*. sp. BH72. The NifH peptide sequence tag used in this study as antigen to generate NifH peptide antibody is perfectly or near-perfectly conserved in NifH proteins from all known organisms, but it is near-perfectly conserved in nitrogenase reductase-like protein e.g. light-independent protochlorophyllide reductase (ZP_07685416) as well. Nevertheless, neither the known nitrogenase reductase nor the known NifH-like protein have an isoelectric point (p*I*) at high pH range as shown in Fig. 23A. Therefore it is not likely that the immunoreactive signals boxed in red (Fig. 23 A) represent NifH proteins.

Most of the 20 extracts showed no signal other than those at the higher pH range, except for two specific immunoreactive signals that were detected in a September 2006 sample. In comparison to the immunoblot of BH72 cell proteins, these two signals showed a very similar pattern in molecular masses and isoelectric point. Subsequently, the antiserum to NifH was applied to this protein extract and these two signals were also detected along with the known 35-kDa cross reactive signal (Fig. 23C). The corresponding gel pieces were excised from the SYPRO Ruby stained gels (Fig. 23D) for MS analysis.

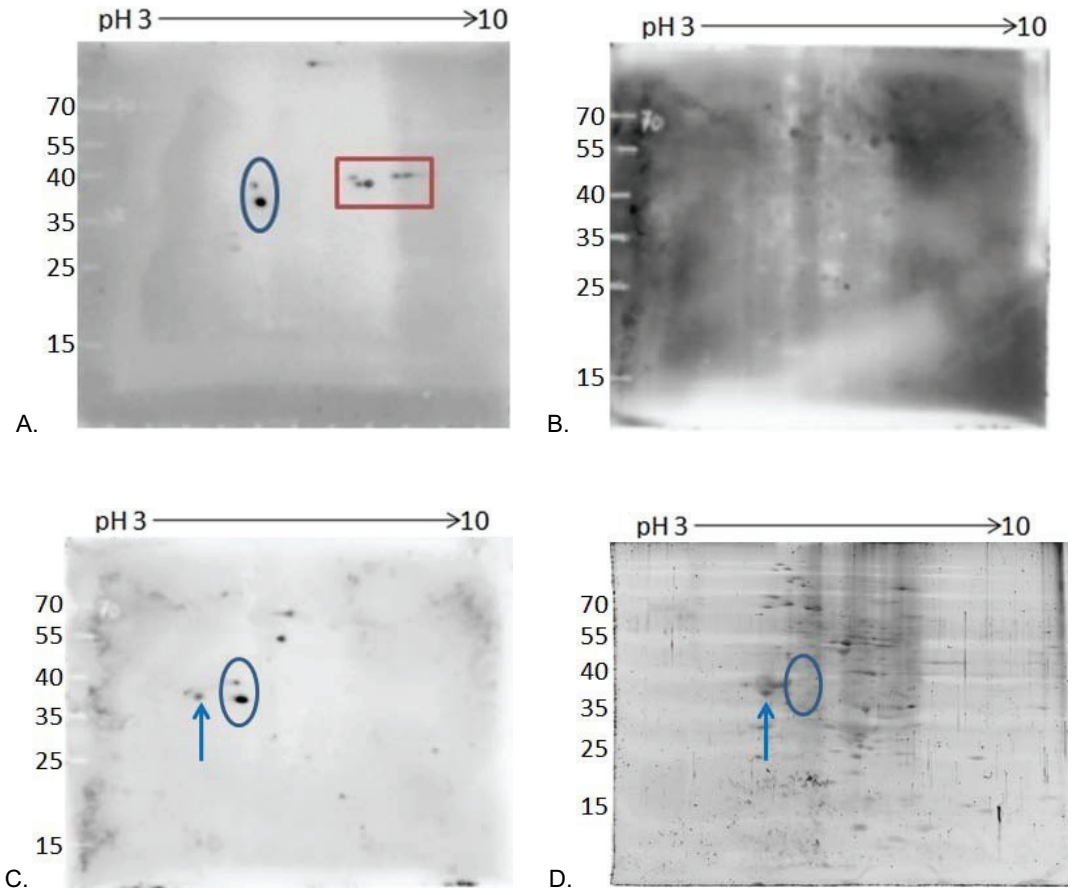


Fig. 23 2-D Western blot analysis of protein extracts obtained from spruce roots sampled in September 2006. Immunoblot incubated with universal anti-dinitrogenase reductase peptide antibody (A) or the corresponding pre-immune serum (B) or antiserum to dinitrogenase reductase from *A*. sp. BH72 (C). Representative protein profile in mini 2-DE gel stained with SYPRO Ruby (D). Matching immunoreactive signals from two antibodies and corresponding proteins in mini 2-D gel are encircled in blue(A, C and D). Immunoreactive signals detected only with universal anti-dinitrogenase reductase peptide antibody (not detected with its pre-immune serum) are boxed in red (A). Immunoreactive signals detected only with antiserum to dinitrogenase reductase from *A*. sp. BH72 (not detected with its pre-immune serum) are marked with a blue arrow (C and D). Positions of molecular weight markers (in kDa) are indicated on the left. Representative 2-D gel profiles were derived from four repetitions. Western blot experiments could not be repeated because of sample limitation.

In the Mini 2-DE gel experiments, all protein samples were loaded using the same long narrow needle. This needle was rinsed three times with distilled autoclaved water between protein loadings. To evaluate the possibility of inter-sample contamination, a

blank control gel was used to evaluate the possibility of inter-sample contamination. For this purpose 25 µl bidest water instead of protein extract was loaded. Gel electrophoresis and staining were performed as usual. Gel spots from the blank control were excised and analyzed by MS. No peptide matched proteins of *A*. sp. BH72. This indicated that the inter-sample contamination need not be considered.

Peptide sequences matching *Azoarcus* sp. BH72 NifH protein were identified in spruce roots sampled in September 2006. A total of four gel spots, duplicates of the upper and lower spots encircled in blue (Fig. 23 A, B), were analyzed by MS and from three of the spots peptides were obtained (Table 6) with 100% sequence identity to the NifH protein from *A*. sp. BH72 (Genbank accession No. YP_932042). Therefore, the two immunoreactive signals, most likely represented the modified and unmodified forms of NifH proteins in BH72 nitrogen-fixing cells. From the fourth spot peptides were not obtained, probably because the protein content was too low. Analysis of the two lower spots also identified peptides (Table 6) with 100% sequences identity (all peptides are unique, except “VQASFVR”) to the outer membrane porin protein precursor of *Azoarcus* sp. BH72 (Genbank accession No.YP_934792), which is specific to *A*. sp. BH72 and is the most abundant protein in the *A.* sp. BH72 proteome (Hauberg*, et al.*, 2010).

Among the NifH peptides detected in spruce roots, the peptide "MTVIEYDPTHK" has 100% identity to NifH from *Sideroxydans lithotrophicus* ES-1 (YP_003523506) and from *A*. sp. BH72, while all other NifH proteins from other bacteria have less % sequence identity to this peptide. On the other hand, the peptide "STTTQNLVAALAEAGKK" detected in spruce roots has one mismatch to the corresponding peptide sequence from *S. lithotrophicus* (“STTTQNLVAALAESGKK”), while there is no mismatch to the corresponding NifH peptide from BH72. Therefore the NifH sequences detected, most probably stem from BH72 (all other detected NifH peptides in spruce are identical to the corresponding sequences from BH72 NifH). This interpretation is consistent with the finding that the highly specific peptides of the

outer membrane porin protein precursor of *A.* sp. BH72 were detected in spruce too.

The two immunoreactive signals, probably representing the signals formed by modified and unmodified forms of NifH proteins as in BH72 nitrogen-fixing cells, were the overlap of the reactive signals obtained with the two primary antibodies.

Table 6. Peptide sequences match to proteins of *Azoarcus* sp. BH72

Analyzed spot	Annotation	Protein ID	Matched peptides [a]	% Sequence coverage	Organism
Lower protein spot-1 (3.3.5)	Dinitrogenase reductase subunit	YP_932042	13 / 13	15.8	*Azoarcus* sp. BH72
Upper protein spot-1 (3.3.5)	Dinitrogenase reductase subunit	YP_932042	8 / 8	9.1	*Azoarcus* sp. BH72
Lower protein spot-2 (3.3.5)	Dinitrogenase reductase subunit	YP_932042	8 / 8	13.1	*Azoarcus* sp. BH72
Protein spot (3.3.6)	Dinitrogenase reductase subunit	YP_932042	6 / 6	4.7	*Azoarcus* sp. BH72
Lower protein spot-1 (3.3.5)	Outer membrane porin protein precursor	YP_934792	24 / 24	62.7	*Azoarcus* sp. BH72
Lower protein spot-2 (3.3.5)	Outer membrane porin protein precursor	YP_934792	6 / 6	21.4	*Azoarcus* sp. BH72

[a], absolute proportions of all peptides matching to peptides from A. sp. BH72

3.3.6 Mass spectrometry analyses of 2-D gel spots

As described before, the matching immunoreactive spots obtained with the two antibodies were only detected in one protein extract from the September 2006 sample. To explore the possibility that putative NifH proteins existed in more spruce samples but could not be detected by immunoblotting, two additional protein extracts from the September 2006 sample in which no immunoreactive signals had been detected were

further analyzed. These two protein extracts were separated on mini 2-DE gels visualized by SYPRO Ruby staining and gel scanning. Corresponding gel spots were excised from 2-D gels and analyzed by MS. Peptides with 100% sequences identity to BH72 NifH were identified in one sample (Table 6).

3.4 Cloning bias caused by the use of LNA-substituted primers in PCR reactions

3.4.1 Effects related to the formation of 3' overhangs

Preliminary experiments had revealed by direct sequencing of PCR products that single stage RT-PCR on spruce samples with LNA-substituted *nifH* Zehr primers produced amplicons representing the *nifH* sequence from *A*. sp. BH72. However, PCR products obtained by amplification with *Taq* DNA polymerase were practically not clonable using TOPO TA cloning. TOPO TA cloning relies on the efficient topoisomerase ligation of PCR products with vector DNA with the help of single-base 3' overhang added to PCR products by the template independent transferase activity of Taq DNA polymerase (Clark, 1998). Since LNA nucleotides at the 3' end of the primer can inhibit PCR extension (Latorra, *et al*., 2003) presumably through conversion of B-form DNA into A-form DNA (Petersen *et al*., 2002) it was assumed that LNA nucleotides at the 5' end of the primer might have a similar effect on the addition of single-base 3' overhangs as well. To test this hypothesis the PCR product amplified from *Azoarcus* sp. BH72 genomic DNA by LNA primer set Zehr-F + R was examined for the 3' overhang using the method described before. Sequencing results showed clearly that the 3' overhangs were only formed when the DNA primers were used, while in the case of LNA-substituted primers the overhangs were completely missing (Fig. 24).

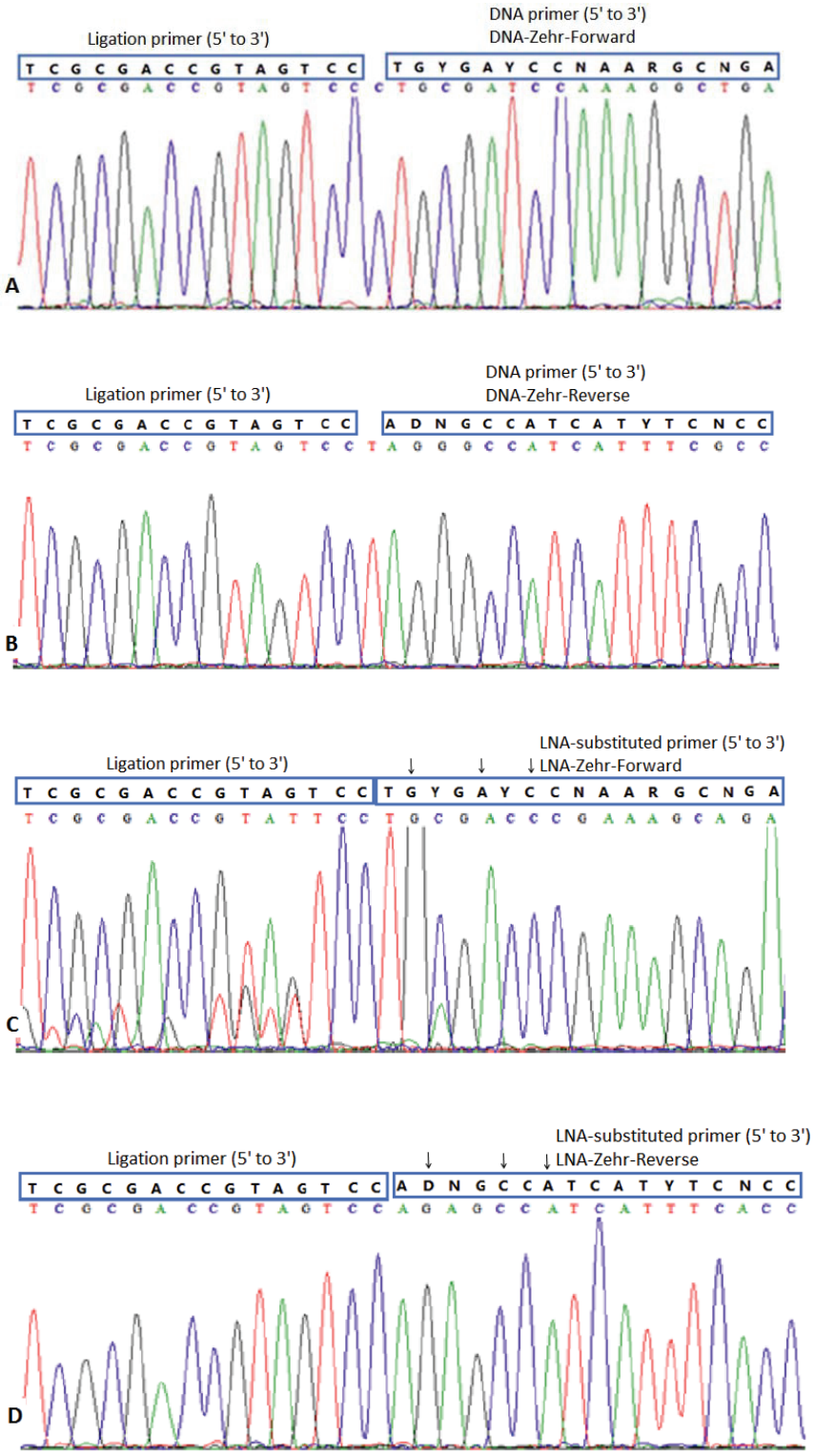


Fig. 24 DNA sequence chromatograms obtained by sequencing of PCR products with (A, B) and without single-base 3' overhangs (C, D), depending on whether DNA (A, B) or LNA-substituted primers (C, D) were used in PCR reactions with *Taq* DNA polymerase. Genomic DNA of *Azoarcus* sp. BH72 was used as template. LNA bases are headed by arrows.

3.4.2 The effect of LNA substitution on TOPO cloning efficiency

Furthermore, the cloning efficiency of PCR products obtained with LNA primer set Zehr-F + R using pRW3 as template was higher than the one obtained from *A.* sp. BH72 genomic DNA (data not shown). The results detailed in Fig. 24 had demonstrated that PCR products amplified by LNA-Zehr primers had no 3' overhang, but this could not explain the higher TOPO TA cloning efficiency when pRW3 instead of *A*. sp. BH72 genomic DNA was used as template. According to Gen *et al*., (2006), the TOPO TA vector can be used to clone blunt-end PCR products (generated with ordinary DNA primers) with a lower cloning efficiency (about 20% reduction) compared to the PCR product with a 3'-end A overhang, although the insert and vector have incompatible ends in this case. Sequence comparisons showed that primer target regions in *A.* sp. BH72 genomic DNA and in plasmids pA72 and pRW3 deviated from each other by two nucleotides complementary to two degenerate positions in the LNA-Zehr reverse primer (Fig. 25).

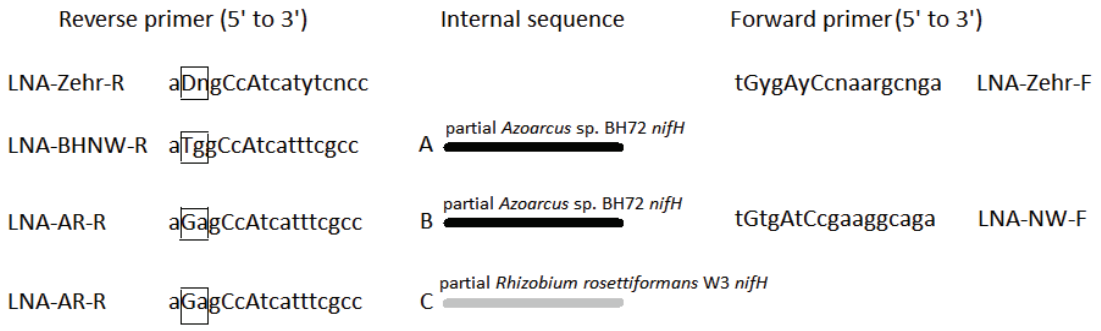


Fig. 25 Scheme showing the sequences of PCR products amplified using the primers indicated. Three non-degenerate LNA-substituted primers, LNA-BHNW-R, LNA-AR-R and LNA-NW-F, were designed based on the available sequence information. The two potentially important nucleotide positions in the target region of the reverse primers are boxed (see text for explanation). While PCR product A was hardly clonable, cloning of PCR products B and C did not pose any problem. For C, pRW3 was used as template.

Because of the reported effects of LNA-substitution on primer extension (Latorra, *et al*.,

2003) and on formation of single-base 3' overhangs as described in 3.4.1, it appeared to be most likely that the sequence composition in the primer target region of the LNA-Zehr reverse primer was responsible for the reduction of cloning efficiency. In order to prove this hypothesis, plasmids pA72 and pRW3 were used as templates to generate PCR products with the two non-degenerate LNA primer sets, NW-F+AR-R and NW-F+BHNW-R (Fig. 25). Since PCR products obtained with LNA-substituted primers had no 3' overhang, the *Pfu* polymerase and TOPO ZeroBlunt kit were used for cloning of the four types of PCR products obtained with the two LNA-substituted primer sets from the two plasmids. The results are shown in Table 7.

Table 7. The effect of sequence composition within LNA-substituted primers on TOPO clonining efficiency of four different PCR fragments

Template	Primer pair	Replication 1[a]		Replication 2[a]		Replication 3[a]		Cloning efficiency[b] (%)	P value (*t* test result)
		Tested	Positive	Tested	Positive	Tested	Positive		
pA72[b]	NW-F + AR-R	50	40	50	45	ND	ND	85±5	0.0034
	NW-F+BHNW-R	10	0	6	0	ND	ND	0	
pRW3[c]	NW-F + AR-R	50	40	50	42	50	35	78±4	0.0004
	NW-F+BHNW-R	13	2	5	0	6	0	5.±5	

[a], tested, colonies selected from the plate; positive,screened by colony PCR for inserts of appropriate size as described in 2.4.4

[b], cloning efficiency, defined as the percentage of positive clones among the clones tested.

[c], plasmid pA72, representing the TOPO TA vector, harboring the partial *nifH* sequence from *Azoarcus* sp. BH72.

[d], plasmid pRW3, representing the TOPO TA vector, harboring the partial *nifH* sequence from *Rhizobium rosettiformans* sp. W3.

Results clearly showed that, the sequence within the primer target region of the LNA-substituted BHNW-R primer was responsible for the drastic reduction of the cloning efficiency and that the two *nifH* sequences serving as templates were not involved. The efficiency of cloning PCR products obtained with LNA primer set NW-F + BHNW-R was significant lower than the efficiency of cloning PCR products obtained

with LNA primer set NW-F + AR-R and this effect was independent from the sequences of the two templates examined, which had a nucleotide sequence identity of 78% in an alignment consisting of 328 positions (without primers). When the experiment shown in Table 7 was repeated twice with DNA primers, the cloning efficiency was not affected. In case of PCR products obtained with DNA primer sets NW-F + BHNW-R and NW-F + AR-R, the cloning efficiencies obtained for both templates were >80%.

These results suggested that the LNA substitution at the penultimate position of the 5'-terminal sequence of the reverse primer is important for efficient cloning of PCR products obtained with LNA-substituted Zehr primers. Consistent with this observation more than 95% of *nifH* clones in a TOPO TA clone library had the penultimate LNA-G in their nucleotide sequence when LNA-substituted Zehr primers (Table 2) were used, which have an LNA-A, LNA-G, or LNA-T at this position (data not shown).

4 Discussion

4.1 Cloning bias caused by the use of LNA-substituted primers

This study was prompted by the preliminary observation that certain PCR products amplified from environmental samples using LNA-substituted primers were hardly clonable with TOPO TA cloning kit (data not shown). When LNA-substituted primers were used the cloning efficiency decreased, probably because the activity of the terminal transferase of *Taq* polymerase and the activity of the topoisomerase was inhibited.

In 3.4 it was shown that PCR amplification using LNA-substituted primers resulted in blunt-end PCR products. Probably, the terminal transferase activity of *Taq* DNA polymerase was affected. Di Giusto and King (Di Giusto & King, 2004) reported that LNA positioning at or near the 3' terminus of a primer could significantly affect polymerase-mediated extension. According to this report a 3'-end LNA nucleotide does not substantially slow primer extension, but the polymerization rate dropped dramatically when an additional LNA was present at the penultimate position of the 3'-end. The penultimate LNA base caused a larger conformational change than a terminal LNA base and repositioned the terminal nucleotide. Di Giusto and King (Di Giusto & King, 2004) argued that this repositioning of the terminal nucleotide within the active site of *Taq* polymerase caused reduction of the polymerization rate. The inhibition of a single-base addition of 3' overhangs by *Taq* polymerase may be explained likewise (Fig. 26A and B). LNA nucleotides in double stranded DNA cause a conversion from the B into the A conformation (Petersen*, et al.*, 2002); and inhibit PCR extension when located at the 3' end of the primer (Latorra*, et al.*, 2003). A poor extension of primers, which are LNA substituted at their 3' end, is consistent with a lack of incorporation of LNA nucleotides by *Taq* DNA polymerase (Veedu*, et al.*, 2007). Before single base 3'-overhangs are formed in a nontemplate dependent manner, blunt-end PCR products are dissociated from the polymerase after the

completion of template synthesis, and subsequently the polymerase reassociates with the blunt-end substrate to perform the non-templated synthesis (Clark, 1988). Therefore, it is likely that apart from the template-dependent polymerase activity also the template-independent terminal transferase-like activity of the *Taq* polymerase is sensitive to the A-form of duplex DNA, leading to blunt ends as shown in Fig. 26.

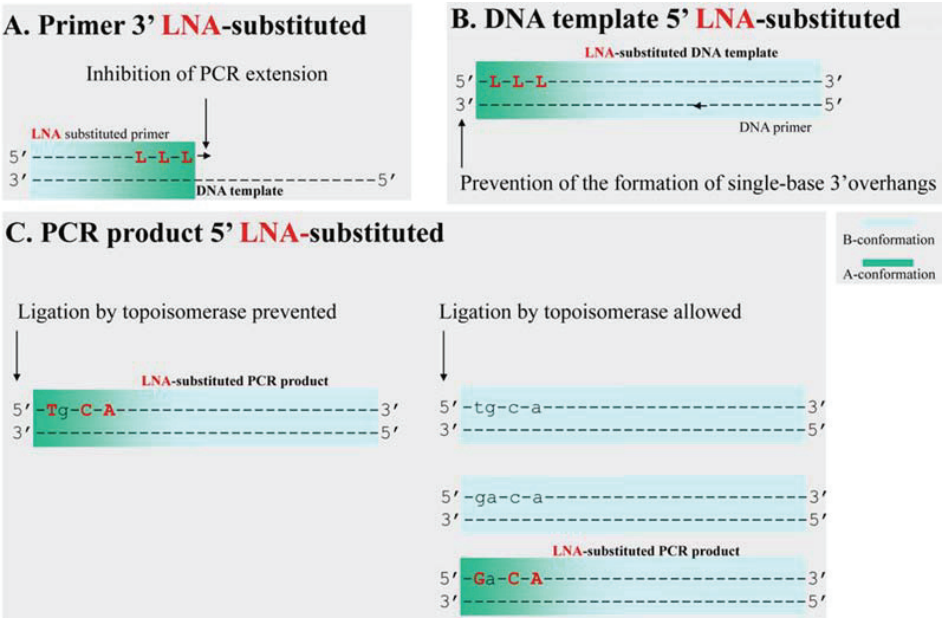


Fig. 26 Effects of conformational changes in duplex DNA caused by 3' or 5' LNA substitution on polymerase (A), terminal transferase activity (B) of Taq DNA polymerase or topoisomerase (C) respectively.

The working principle of topoisomerase mediated ligation with the TOPO vector (Shuman, 1992, Cheng & Shuman, 2000) is as follows: the plasmid vector is supplied linearized with *Vaccinia* virus DNA topoisomerase I covalently bound to the 3'-end of each DNA strand with the phosphor-tyrosyl bond. The phosphor-tyrosyl bond is subsequently attacked by the 5' terminal hydroxyl group of the PCR product, resulting in topoisomerase release and generation of recombinant DNA molecules. The close contact of topoisomerase with the 5' terminal hydroxyl group of the PCR product is essential for the ligation reaction, as the energy from the phosphor-tyrosyl bond is transferred to the religated phosphodiester bond. In case of the TOPO TA vector, a covalent intermediate between PCR product and vector is formed due to the

3'-overhangs of vector and PCR product, which helps to form a close contact between the topoisomerase and the 5' terminal hydroxyl group of the PCR product. The topoisomerase-mediated ligation can still have a high efficiency, when the PCR product contains a 3' G, T or C overhang (Cheng & Shuman, 2000) which doesn't perfectly match to the 3' T overhang of the vector. However when the blunt-end PCR product is cloned with TOPO TA vector, the covalent intermediate between PCR product and vector is not formed since the 3' overhang is missing from the PCR product, leading to lower ligation efficiency and lower cloning efficiency.

The results presented in section 3.4 also showed that the cloning efficiency using the TOPO ZeroBlunt vector was dramatically decreased when the blunt-end PCR products were generated with certain LNA-substituted primers. Hypothetically the specific LNA substitutions at the near 5'-end of the PCR products affected the cloning efficiency by interfering with the topoisomerase mediated ligation (Fig. 26C). Shuman (1992) showed that dsDNA containing one covalently topoisomerase activated terminus could ligate to a heterologous blunt-end duplex DNA regardless of the sequence of the acceptor molecule. However, Tian *et al*. (Tian*, et al.*, 2004) suggested that continuous base stacking of the scissile strand on both sides of the cleavage site is very important for the topoisomerase catalyzed religation. Since LNA modifications in duplex DNA might cause discontinuous base stacking (You*, et al.*, 2006), LNA base substitution may prevent ligation by topoisomerase. In this study Tg-C-A in LNA-substituted primers prevented ligation but Ga-C-A not (Fig. 26C). This may be related to the higher binding strength of T-LNA than G-LNA (McTigue*, et al.*, 2004) which will lead to a greater base stacking in the case of the T-LNA substituted primer compared to the G-LNA substituted primer. Accordingly, a less efficient ligation would result in a lower cloning efficiency.

4.2 Sampling

This study aimed to identify the major diazotrophs associated with Norway spruce roots. Most previous studies (reviewed in the introduction section), if not all, were

focused on the mycorrhizal roots of the coniferous root system and mostly were based on culture-dependent methods. For this study, the bark of the coarse roots of Norway spruce was used. The sampling strategy was based on preliminary experimental results. Initially, multiple types of roots, including coarse, fine and mycorrhizal roots, were sampled from different soil horizons, comprising the organic soil layer (top soil, 0-20 cm depth) to the mineral soil layer (sub-soil, 20-40 cm depth), in September 2006. Since with the RNA extraction and RT-PCR protocol used at that time no PCR amplification from any sample could be obtained, protein-extracted samples were first screened by Western blot analysis using the antiserum against NifH from *A*. sp. BH72. Immunoreactive signals were only detected in bark from coarse roots from the organic soil layer (~ 2 cm diameter roots in 5-10 cm of topsoil layer). Roots were collected according to these criteria on all following sampling dates, so that only the barks of coarse roots from the same depth of soil layer were collected.

The 35-kDa immunoreactive signal, used as an indicator for initial screening of different roots, was eventually verified to be an unspecific cross-reaction (details in Results section 3.3) between the NifH antibody and the 35-kDa plant protein from spruce. Nevertheless, in the bark of coarse roots selected by the immunodetection approach, nitrogen fixation activity was confirmed at both transcriptional and translational levels. Unfortunately, the *nifH* transcript and NifH protein information from mycorrhizal root materials were lacking in this study, since this type of materials was excluded for not giving the 35-kDa signal during the initial screening by immunodetection. It would be interesting to investigate the mycorrhizal roots for *nifH* expression in future and to compare the results with the data from the current study.

4.3 Heterogeneity of root materials

Besides that the identified *nifH* transcript was identical to a partial *nifH* sequence from *Azoarcus* sp. BH72, from one RNA extract the *nifH* transcript was found to be identical to the *nifH* partial sequence from *Pseudomonas stutzeri* A1501. However, this result could be not reproduced with any other RNA extract, suggesting that this phylotype

was not very abundant. Another *nifH* phylotype, highly similar to the partial *nifH* sequence of *Rhizobium rosettiformans*, was identified in one spruce sample from which RNA had been extracted by an alternative method. *Pseudomonas* species were reported as free-living bacteria in forest soil (Elo*, et al.*, 2000, Filion*, et al.*, 2004), while phylogenetically diverse *Rhizobium* species were isolated from pine ectomycorrhizae (Tanaka & Nara, 2009). Probably, these two phylotypes were present in the root samples at very low abundance. The *nifH* genes from these two phylotypes were not detectable in the spruce root DNA extracts according to the quantitative PCR data (in 3.2.4). However, since the root material was heterogonous, these phylotypes might be randomly distributed in niches of the samples. Therefore it was necessary to perform multiple independent experiments to identify potentially important phylotypes which were predominant in the majority of samples analyzed. In this study, at each sampling date roots from at least three trees were collected, more than 3 pieces of roots from root system of each tree, and more than 5 subdivisions from each piece of root, to ensure that sufficient material was collected to conduct multiple independent experiments.

With respect to the study on protein, NifH proteins were detected by immunoblot and / or mass spectrometry analysis in only two samples which were collected in September 2006. This might indicate that NifH proteins were heterogeneously distributed *in planta* and that they are only sufficiently enriched in certain roots at a particular time. On the other hand, co-purified compounds might interfere with protein extraction from certain samples rendering NifH proteins undetectable and leading to false negative results in Western blot and / or mass spectrometry analysis.

4.4 Extra *nifH* gene copies of *Azoarcus*. sp. BH72 in spruce

Quantitative RT-PCR using *A*. sp. BH72 specific *nifH* and 16S primer sets showed that the ratio of *nifH* transcripts to 16S rRNA transcripts *in planta* was about 2 times higher than the ratio in *A*. sp. BH72 pure cultures which had been incubated under optimal nitrogen fixing conditions. This difference was statistically significant.

Strikingly, when quantitative PCR was applied to quantify *A*. sp. BH72 specific *nifH* and 16S rRNA genes in spruce DNA samples, a ratio < 0.075 was calculated from less than 6 copies of the 16S rRNA gene and 80 copies of *nifH* gene detected in 300 ng of spruce DNA. In pure culture DNA samples of strain BH72 a ratio of 16S rRNA gene to *nifH* gene copies of 6.3 was determined which is in reasonable agreement with the theoretical ratio of 4:1 according to the genome sequence information. The disproportional high copy numbers of *nifH* genes relative to 16S rRNA genes from *A*. sp. BH72 in spruce might have several reasons:

(i) The 16S rRNA primers used are too specific and do not cover the whole diversity of *Azoarcus* sp. BH72 *in planta*. This would cause an underestimation of 16S rRNA genes / transcripts relative to *nifH* from these bacteria *in planta*, which was observed. It was intended to test for this possibility by amplification of the 16S rRNA gene with *Azoarcus* spp. specific 16S primers and subsequent sequences analyses of the amplicons. Primer sets TH1 and TH5, or TH5 and TH3, which have been reported previously to cover the genera *Azoarcus*, *Azonexus*, *Azovibrio*, *Azospira* or *Azoarcus* spp. (Hurek*, et al.*, 1993), respectively, were not successful to obtain amplicons from spruce DNA (data not shown). Most probably the gene copy numbers were under the limit of detection, since established PCR protocols were used and positive controls amplified.

(ii) The copy numbers of 16S rRNA genes / transcripts from *A*. sp. BH72 *in planta* relative to *nifH* genes / transcripts are diminished because of lateral gene transfer of *nifH*. Zehr (Zehr*, et al.*, 2003) concluded in 2003 on the issue of lateral gene transfer that "…there is not strong support or evidence for lateral gene transfer, by plasmid transfer or other mechanisms, and if such transfers occurred, they must have occurred early in evolution…". Since the sequences of the respective *nifH* genes / transcripts obtained from spruce roots were identical to *nifH* from *A*. sp. BH72 the transfer must have occurred recently, which is unlikely at the current state of knowledge.

(iii) The extra copy numbers of *nifH* genes could be due to multiple copies (on plasmids or on the chromosome), although, as known from the genome sequence of *A*. sp. BH72 the *nifH* gene is only present as a single copy in this bacterium. Several copies of *nifH* on the chromosome are known, among others, for *Azorhizobium caulinodans* (Denefle*, et al.*, 1987, Norel & Elmerich, 1987). Plasmid-borne *nifH* was demonstrated in *Sinorhizobium* spp. (Haukka*, et al.*, 1998), although no evidence indicated that the *nifH* gene may be present on non-chromosomal genetic elements in *Azoarcus* spp. so far. Isolation of bacteria from spruce roots into pure culture harboring the *nifH* gene from *Azoarcus* sp. strain BH72 would provide more information, but it was not successful yet. Isolation of *in planta* diazotrophs can be challenging, as reported for *Azoarcus* sp. BH72, nitrogen-fixing endophytes in Kallar grass (Hurek*, et al.*, 2002). Clearly, further experiments are required to solve this issue.

The observed different ratio of *nifH* transcripts : 16S rRNA copies of *A*. sp. BH72 *in planta* compared to the ratio of cells cultured under nitrogen fixing conditions in pure culture implies different physiological activities between these two stages at the time of sampling. The different ratio of *nifH* gene : 16S rRNA gene copies of *A*. sp. BH72 *in planta* suggests the presence of multiple copies of the *nifH* gene in *A*. sp. BH72 strain associated with spruce. However, since *nifH* gene expression is controlled by complex regulatory cascades, and the different copy of *nifH* genes may optimally express under different physiological conditions (Norel & Elmerich, 1987), namely not all copies of *nifH* genes are actively expressing under certain physiological conditions, it is inappropriate to normalize the copy number of *nifH* transcripts with the copy number of *nifH* genes to compare the nitrogen fixing activity of *A*. sp. BH72 *in planta* with that in pure culture.

4.5 Immunodetection of the NifH proteins of *A*. sp. BH72 in spruce roots

NifH had been previously identified by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF-MS) analysis in environmental samples, such as in biofilms

collected from acid mine drainage (Ram*, et al.*, 2005), and in cyanosymbionts collected from symbiotic organs of the nitrogen-fixing water fern, *Azolla* (Ekman, et al., 2008). This is the first study detecting NifH in naturally growing seed plants, Norway spruce, which are not known to form any nitrogen fixing symbiosis or symbiotic structures with diazotrophic bacteria (see Table 1). Because there are no symbiotic structures haboring N_2-fixing microsymbionts in spruce, diazotrophic bacteria could be not separated or enriched from plant tissues, which made the application of highly sensitive detection methods necessary. For this purpose two-dimensional Western blot analysis in combination with LC-Orbitrap MS were used, which have a detection limit in the pg range (Holmes & Pollenz, 1997, Dong*, et al.*, 2010). This allowed a homology-based protein identification of *de novo* sequenced peptides obtained from enzymaticaly degraded proteins. Applying these methods, *Azoarcus* sp. BH72, a bacterium not involved in the established symbioses of seed plants (Table 1), was found to be central to nitrogen fixation in Norway spruce, which was consistent with the *nifH* transcript analysis.

As discussed before (section 1.5), only low amounts of NifH proteins were to be expected in spruce roots. Therefore, it seemed to be unlikely that without an additional immunodetection step this protein would be detected *in situ* by separating the proteins on 2D-PAGE and subsequent identification of excised spots, alone. There are several reports describing detection and quantification of specific proteins in complex environments, such as soil, using immunological methods (Maron*, et al.*, 2003, Maron*, et al.*, 2004) to provide quantitative information on the level of expression of the genetic potential at a given point of the microbial community. However, since suitable antibodies are scarce, such a strategy is only rarely applied. Antibodies used for immunodetection need to be specific to exclude cross-reactions with other proteins in the complex environmental protein mixture. On the other hand antibodies have to be able to target functionally equivalent proteins produced by a wide range of bacteria belonging to different species. Therefore this strategy is probably also limited by the small number of enzymes catalyzing biological functions

that have been identified (Maron*, et al.*, 2007). The published protocols using immunological methods only applied one dimensional Western blot analysis for the analysis of environmental samples (Maron*, et al.*, 2003, Maron*, et al.*, 2004). Therefore, the likely occurrence of false positive cross-reactions could possibly not be excluded, although treatments of blots with pre-immune serum as a immunological control were included. Similarly, in this study the two immunoreactive 40-kDa and 35-kDa proteins (Results section 3.3), were analyzed and no NifH related peptides were identified from MS data, although the two proteins reproducibly reacted with the antiserum to NifH of *A*. sp. BH72 and showed no cross reaction with pre-immune serum from the respective rabbit. Polyclonal antibodies contain a mixture of monoclonal antibodies with different binding sites to the antigenic molecule, in this case NifH. These results suggested that immunodetection carried out with polyclonal antibodies might not be specific enough for detection of antigens in complex environmental samples. A high complexity of environmental samples makes the occurrence of cross reactions between non-target proteins and primary antibody due to serological similarities likely. If cross-reactive non-target proteins with a similar apparent molecular weight as the antigen are present in a sample, a one-dimensional Western blot analysis will always cause misleading results, independent of the negative control with pre-immune serum. Peptide sequences matching *A.* sp. BH72 NifH protein were identified based on the overlapping signals obtained with the antibodies raised against NifH from *A*. sp. BH72 and the monospecific NifH peptide antibody. Therefore it may be necessary to use at least two primary antibodies in immunoblots when analyzing environmental samples. Overlapping signals obtained with multiple primary antibodies would increase confidence in the identity of proteins detected and would help to identify cross reactions with non-target proteins. However, direct MS analysis of protein extracts or of protein extracts after separation by gel electrophoresis is probably more straightforward in this case.

Peptide sequences matching to *A.* sp. BH72 NifH protein were also identified from one sample which was negative in immunodetection. This result implied that under

current experimental conditions, the MS analysis was more sensitive than Western blot analysis. However, according to the manufacturer's instructions, the chemiluminscent substrate used in this study has a detection limit of 1 to 5 pg, while the mass spectrometry analysis has a detection limit of around 1 fmol. Considering the size of NifH (33 kDa, 1 fmol 33 kDa protein = 33 pg), immunoblot and MS theoretically have about the same sensitivity. Practically, the unsuccessful immunodetection could be ascribed to the interference by the co-extracted compounds or the efficiency of the immunoreactions on the blot. Nevertheless, for a systemic study of NifH in spruce roots, in which the proteins originating from the plant vastly outnumbered the target protein, immunoblotting is still a feasible way for pre-screening.

4.6 Is the conclusion that *A.* sp. BH72 is the major active diazotroph associated with spruce root, based on false positive results?

In this study *nifH* transcripts identical to the partial *nifH* gene sequence from *A*. sp. BH72 were repeatedly identified from spruce root samples by RT-PCR. *A.* sp. BH72, isolated from Kallar grass roots in Pakistan (Reinhold*, et al.*, 1986), is commonly used in this laboratory, which includes PCR investigation of its *nifH* gene. In order to obtain trustworthy data and not a false positive result due to a contamination of *A*. sp. BH72 DNA or PCR products from the working ambient air, no template controls and a DNA control (to examine the potential contribution of carryover of genomic DNA on the RT-PCR amplification of cDNA) were run routinely together with each batch of RT-PCR reaction. Despite of the low probability (< 5 %) of contamination in control reactions, namely approximately 1 out of 20 control reactions was usually contaminated, RT-PCR products were not further analyzed if in a control from the same batch bands without template were present. Since the *nifH* transcripts identical to the partial *A*. sp. BH72 *nifH* gene sequence were detected in samples from different months, it was unlikely that PCR amplicons obtained from all of the six time points were due to random contaminations.

Quantification studies using *A*. sp. BH72 specific *nifH* and 16S primer sets showed that the ratio of *nifH* transcripts to 16S rRNA copies *in planta* was about 2 times higher than the ratio of *A*. sp. BH72 cells incubated under optimal nitrogen fixing conditions in pure culture. This observation speaks in favor for no contamination of strain BH72 cells or RNA during RNA extractions or RT-qPCR procedures. Quantification studies on the DNA level revealed that the 16S rRNA gene was under the detection limit (6 copies) in spruce root DNAs, while about 80 *nifH* gene copies were detected within the same start DNA templates. The disproportional high *nifH* gene copy number indicated that no contamination of strain BH72 cells or DNA during DNA extraction or qPCR procedures has occurred.

To circumvent a possible PCR contamination or PCR bias in the amplification reactions, investigations were performed at the protein level as well. Peptide sequences which matched to the NifH protein of *A*. sp. BH72 were identified by mass spectrometry analysis from two spruce samples, but not in a blank, which was included as a control. The blank control aimed to test whether the routine rinsing of the loading needle was sufficient to eliminate cross-contamination between samples and especially to check for the carryover of *A*. sp. BH72 proteins in the needle after rinsing. These studies on the protein level provided additional evidence that the *A*. sp. BH72 identical *nifH* gene was not only actively transcribed, but also translated *in planta*. Moreover, the immunoreactive signals in spruce roots detected by immunoblotting (with both the anti strain BH72 NifH and anti peptide antibodies) showed similar patterns to that of pure cultures of *A*. sp. BH72, suggesting that NifH *in planta* underwent the same post-translational modification (Oetjen*, et al.*, 2009) as in pure cultures .

Nucleotide BLAST of the partial, 906 bp, 16S rRNA sequence (assembled from two fragments amplified from spruce RNA using primer sets 320F + 417R and TH3 + TH5, data not shown) obtained from spruce root sample showed 100 % sequence identity to the consensus16S rRNA gene partial sequence of *A*. sp. BH72 (AF011344). It

showed 99 % sequence identity (with 3 to 7 nucleotides differences respectively) to the four other 16S rRNA gene sequences available in NCBI's Genbank. Three out of the four are uncultured bacterial 16S rRNA gene partial sequences obtained from environmental samples, including cotton rhizosphere in China (FJ444718), FD oil contaminated soil in Taiwan (DQ984532) and deep-sea water in the Gulf of Mexico Sea (EF659431). The other one is a partial 16S rRNA gene sequence of *Azoarcus*. sp. DS30 isolated from hexachlorocyclohexane contaminated soil in India (EF494194). This implies that *Azoarcus* sp. has a wide geographical and ecological distribution and a low diversity of its 16S rRNA genes.

In conclusion, although an element of risk remains, the evidence is overwhelming that *A*. sp. BH72 is the major active diazotroph associated with medium-coarse roots of Norway spruce. In this study, multiple methods were explored and applied to a recalcitrant environmental sample. Sufficient data were obtained from nucleic acids and proteins using culture-independent approaches gaining new insights into the novel association between *A*. sp. BH72 and coniferous coarse roots which are poorly studied by previous investigations (reviewed in 1.3). It suggested that *A*. sp. BH72 may play important roles in associative nitrogen fixation not only in graminaceous plants, but also in gymnosperm plant. In future, isolating the responsible bacterium would lead to more advances in understanding its nitrogen fixation activity and the plant-bacteria interactions. Additionally ^{15}N studies would be helpful to fully estimate the nitrogen-fixing capacity of the responsible bacterium and the ecological significance of the spruce root-diazotrophic association.

References

Ayoub AT (1999) Fertilizer and the environment. *Nutr Cycl Agroecosys* **55**: 117-121.

Baker TG & Attiwill PM (1984) Acetylene reduction in soil and litter from pine and eucalypt forests in south-eastern Australia. *Soil Biol Biochem* **16**: 241-245.

Ballantyne KN, van Oorschot RA & Mitchell RJ (2008) Locked nucleic acids in PCR primers increase sensitivity and performance. *Genomics* **91**: 301-305.

Barkmann J & Schwintzer CR (1998) Rapid N_2 fixation in pines? Results of a maine field study. *Ecology* **79**: 1453-1457.

Beesley CE, Smith RJ, Temple SJ & Lea PJ (1994) Cloning and nucleotide sequence of the gene encoding dinitrogenase reductase (*nifH*) from the cyanobacterium *Nostoc* 6720. *Biochim Biophys Acta* **1219**: 548-550.

Bergersen FJ & Postgate JR (1987) *A century of nitrogen fixation research : present status and future prospects*. Royal Society, London.

Binkley D (2002) Response to "Lessons from the sandbox: Is unexplained nitrogen real?". *Ecosystems* **5**: 734-735.

Boddey RM, Urquiaga S, Alves BJR & Reis V (2003) Endophytic nitrogen fixation in sugarcane: present knowledge and future applications *Plant and Soil* **252**: 139-149.

Bormann BT, Bormann FH, Bowden WB*, et al.* (1993) Rapid N_2 fixation in pines, alder, and locust: evidence from the sandbox ecosystem study. *Ecology* **74**: 583-598.

Bothe H, Ferguson SJ & Newton WE (2007) *Biology of the nitrogen cycle*. Elsevier, Amsterdam ; Boston.

Brownstein MJ, Carpten JD & Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**: 1004-1006, 1008-1010.

Burbano CS, Reinhold-Hurek B & Hurek T (2010) LNA-substituted degenerate primers improve detection of nitrogenase gene transcription in environmental samples. *Environ Microbiol Rep* **2**: 251-257.

Burbano CS, Liu Y, Rösner K, Reis V, Caballero-Mellado J, Reinhold-Hurek B & Hurek T (2011) Predominant *nifH* transcript phylotypes related to *Rhizobium rosettiformans* in field grown sugarcane plants and in Norway spruce. *Environ Microbiol Rep* **3**: 383-389.

Burke DJ, Kretzer AM, Rygiewicz PT & Topa MA (2006) Soil bacterial diversity in a loblolly pine plantation: influence of ectomycorrhizas and fertilization. *FEMS Microbiol Ecol* **57**: 409-419.

Burns RC & Hardy RW (1975) Nitrogen fixation in bacteria and higher plants. *Mol Biol Biochem Biophys* **21**: 1-189.

Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**: 169-193.

Bustin SA, Benes V, Garson JA*, et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**: 611-622.

Callaham D, Deltredici P & Torrey JG (1978) Isolation and cultivation *in vitro* of the actinomycete causing root nodulation in *Comptonia*. *Science* **199**: 899-902.

Candiano G, Bruschi M, Musante L*, et al.* (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**: 1327-1333.

Chandler DP, Wagnon CA & Bolton H, Jr. (1998) Reverse transcriptase (RT) inhibition of PCR at low

concentrations of template and its implications for quantitative RT-PCR. *Appl Environ Microbiol* **64**: 669-677.

Chang S, Puryear J & Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* **11**: 113-116.

Chelius MK & Triplett EW (2001) The diversity of archaea and bacteria in association with the roots of *Zea mays* L. *Microb Ecol* **41**: 252-263.

Cheng C & Shuman S (2000) DNA strand transfer catalyzed by vaccinia topoisomerase: ligation of DNAs containing a 3' mononucleotide overhang. *Nucleic Acids Res* **28**: 1893-1898.

Cho JC & Tiedje JM (2002) Quantitative detection of microbial genes by using DNA microarrays. *Appl Environ Microbiol* **68**: 1425-1430.

Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* **16**: 9677-9686.

Cleveland CC, Townsend AR, Schimel DS*, et al.* (1999) Global patterns of terrestrial biological nitrogen (N_2) fixation in natural ecosystems. *Glob Biogeochem Cycle* **13**: 623-645.

Condon C, Squires C & Squires CL (1995) Control of rRNA transcription in *Escherichia coli*. *Microbiol Rev* **59**: 623-645.

Dawson JO (1983) Dinitrogen fixation in forest ecosystems. *Can J Microbiol* **28**: 979-992.

DeLong EF (2002) Microbial population genomics and ecology. *Curr Opin Microbiol* **5**: 520-524.

Demba Diallo M, Reinhold-Hurek B & Hurek T (2008) Evaluation of PCR primers for universal *nifH* gene targeting and for assessment of transcribed *nifH* pools in roots of *Oryza longistaminata* with and without low nitrogen input. *FEMS Microbiol Ecol* **65**: 220-228.

Demba Diallo M, Willems A, Vloemans N*, et al.* (2004) Polymerase chain reaction denaturing gradient gel electrophoresis analysis of the N_2-fixing bacterial diversity in soil under *Acacia tortilis* ssp. raddiana and *Balanites aegyptiaca* in the dryland part of Senegal. *Environ Microbiol* **6**: 400-415.

Denefle P, Kush A, Norel F, Paquelin A & Elmerich C (1987) Biochemical and genetic analysis of the *nifHDKE* region of *Rhizobium* ORS571. *Mol Gen Genet* **207**: 280-287.

Di Giusto DA & King GC (2004) Strong positional preference in the interaction of LNA oligonucleotides with DNA polymerase and proofreading exonuclease activities: implications for genotyping assays. *Nucleic Acids Res* **32**: e32.

Diallo MD, Martens M, Vloemans N*, et al.* (2004) Phylogenetic analysis of partial bacterial 16S rDNA sequences of tropical grass pasture soil under *Acacia tortilis* subsp. raddiana in Senegal. *Syst Appl Microbiol* **27**: 238-252.

Dixon R & Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* **2**: 621-631.

Dojka MA, Hugenholtz P, Haack SK & Pace NR (1998) Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl Environ Microbiol* **64**: 3869-3877.

Dong M, Wu M, Wang F*, et al.* (2010) Coupling strong anion-exchange monolithic capillary with MALDI-TOF MS for sensitive detection of phosphopeptides in protein digest. *Anal Chem* **82**: 2907-2915.

Egener T, Hurek T & Reinhold-Hurek B (1998) Use of green fluorescent protein to detect expression of nif genes of *Azoarcus* sp. BH72, a grass-associated diazotroph, on rice roots. *Mol Plant Microbe Interact* **11**: 71-75.

Elo S, Maunuksela L, Salkinoja-Salonen M, Smolander A & Haahtela K (2000) Humus bacteria of

Norway spruce stands: plant growth promoting properties and birch, red fescue and alder colonizing capacity. *FEMS Microbiol Ecol* **31**: 143-152.

Enebak SA, Wei G & Kloepper JW (1998) Effects of plant growth-promoting rhizobacteria on loblolly and slash pine seedlings. *For Sci* **44**: 139-144.

Evans MR, Bertera AL & Harris DW (1994) The Southern blot. An update. *Mol Biotechnol* **1**: 1-12.

Filion M, Hamelin RC, Bernier L & St-Arnaud M (2004) Molecular profiling of rhizosphere microbial communities associated with healthy and diseased black spruce (*Picea mariana*) seedlings grown in a nursery. *Appl Environ Microbiol* **70**: 3541-3551.

Florence LZ & Cook FD (1984) Asymbiotic N_2-fixing bacteria associated with three boreal conifers. *Can J For Res* **14**: 595-597.

Fuhrmann M & Hennecke H (1984) *Rhizobium japonicum* nitrogenase Fe protein gene (*nifH*). *J Bacteriol* **158**: 1005-1011.

Gallon JR, Cheng J, Dougherty LJ*, et al.* (2000) A novel covalent modification of nitrogenase in a cyanobacterium. *FEBS Lett* **468**: 231-233.

Garbaye J (1994) Helper bacteria: a new dimension to the mycorrhizal symbiosis. *New Phytol* **128**: 197-210.

Giller KE & Merckx R (2003) Exploring the boundaries of N_2-fixation in cereals and grasses: An hypothetical and experimental framework. *Symbiosis* **35**: 3-17.

Granhall U & Lindberg T (1978) Nitrogen fixation in some coniferous forest ecosystems. *Ecol Bull* **26**: 178-192.

Granstedt A (2000) Increasing the efficiency of plant nutrient recycling within the agricultural system as a way of reducing the load to the environment - experience from Sweden and Finland. *Agric Ecosyst Environ* **80**: 169-185.

Grayston SJ, Vaughan D & Jones D (1997) Rhizosphere carbon flow in trees, in comparison with annual plants: The importance of root exudation and its impact on microbial activity and nutrient availability. *Appl Soil Ecol* **5**: 29-56.

Gualtieri G & Bisseling T (2000) The evolution of nodulation. *Plant Mol Biol* **42**: 181-194.

Hauberg L, Schmidt F, Scharf C, Dorr J, Volker U & Reinhold-Hurek B (2010) Proteomic characterization of a *pilR* regulatory mutant of *Azoarcus* sp. strain BH72 with the aid of gel-based and gel-free approaches. *Proteomics* **10**: 458-469.

Haukka K, Lindstrom K & Young JP (1998) Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. *Appl Environ Microbiol* **64**: 419-426.

Hawkes CV (2003) Nitrogen cycling mediated by biological soil crusts and arbuscular mycorrhizal fungi. *Ecology* **84**: 1553-1562.

Heath B, Sollins P, Perry DA & Croomack K (1988) Asymbiotic nitrogen fixation in litter from Pacific North-west forests. *Can J For Res* **18**: 68-74.

Heid CA, Stevens J, Livak KJ & Williams PM (1996) Real time quantitative PCR. *Genome Res* **6**: 986-994.

Higuchi R, Dollinger G, Walsh PS & Griffith R (1992) Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (N Y)* **10**: 413-417.

Holmes JL & Pollenz RS (1997) Determination of aryl hydrocarbon receptor nuclear translocator protein concentration and subcellular localization in hepatic and nonhepatic cell culture lines: development of quantitative Western blotting protocols for calculation of aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator protein in total cell lysates. *Mol Pharmacol* **52**:

202-211.

Hoss M, Handt O & Paabo S (1994) Recreating the past by PCR. *The Polymerase chain reaction,*(Mullis KB, Ferré F & Gibbs RA, eds.), pp. 257-264. Birkhaeuser, Boston.

Huber JA, Morrison HG, Huse SM, Neal PR, Sogin ML & Mark Welch DB (2009) Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environ Microbiol* **11**: 1292-1302.

Hurek T, Egener T & Reinhold-Hurek B (1997) Divergence in nitrogenases of *Azoarcus* spp., Proteobacteria of the beta subclass. *J Bacteriol* **179**: 4172-4178.

Hurek T, Burggraf S, Woese CR & Reinhold-Hurek B (1993) 16S rRNA-targeted polymerase chain reaction and oligonucleotide hybridization to screen for *Azoarcus* spp., grass-associated diazotrophs. *Appl Environ Microbiol* **59**: 3816-3824.

Hurek T, Van Montagu M, Kellenberger E & Reinhold-Hurek B (1995) Induction of complex intracytoplasmic membranes related to nitrogen fixation in *Azoarcus* sp. BH72. *Mol Microbiol* **18**: 225-236.

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**: 233-242.

Izumi H, Anderson IC, Alexander IJ, Killham K & Moore ER (2006) Endobacteria in some ectomycorrhiza of Scots pine (*Pinus sylvestris*). *FEMS Microbiol Ecol* **56**: 34-43.

James EK & Olivares FL (1998) Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. *Crit Rev Plant Sci* **17**: 77-119.

James EK, Reis VM, Olivares FL, Baldani JI & Dobereiner J (1994) Infection of sugar cane by the nitrogen-fixing bacterium *Acetobacter diazotrophicus*. *J Exp Bot* **45**: 757-766.

Jones K & Bangs D (1985) Nitrogen fixation by free-living heterotrophic bacteria in an oak forest: The effect of liming. *Soil Biol Biochem* **17**: 705-709.

Jurgensen MF, Larsen MJ, Graham RT & Harvey AE (1987) Nitrogen fixation in woody residue of northern Rocky Mountain conifer forests. *Can J For Res* **17**: 1283-1288.

Karppinen K, Taulavuori E & Hohtola A (2010) Optimization of protein extraction from *Hypericum perforatum* tissues and immunoblotting detection of Hyp-1 at different stages of leaf development. *Mol Biotechnol* **46**: 219-226.

Kaur J, Verma M & Lal R (2010) *Rhizobium rosettiformans* sp. nov., isolated from hexachlorocyclohexane (HCH) dump site in India, and reclassification of *Blastobacter aggregatus* Hirsch and Muller (1985) as *Rhizobium aggregatum* comb. nov. *Int J Syst Evol Microbiol*.

Keller M & Hettich R (2009) Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol Mol Biol Rev* **73**: 62-70.

Klotz MG & Stein LY (2008) Nitrifier genomics and evolution of the nitrogen cycle. *FEMS Microbiol Lett* **278**: 146-156.

Knauth S, Hurek T, Brar D & Reinhold-Hurek B (2005) Influence of different *Oryza* cultivars on expression of *nifH* gene pools in roots of rice. *Environ Microbiol* **7**: 1725-1733.

Kolosova N, Miller B, Ralph S, Ellis BE, Douglas C, Ritland K & Bohlmann J (2004) Isolation of high-quality RNA from gymnosperm and angiosperm trees. *Biotechniques* **36**: 821-824.

Krause A, Ramakumar A, Bartels D*, et al.* (2006) Complete genome of the mutualistic, N2-fixing grass endophyte *Azoarcus* sp. strain BH72. *Nat Biotechnol* **24**: 1385-1391.

Kretzer AM, King ZR & Bai S (2009) Bacterial communities associated with tuberculate ectomycorrhizae of *Rhizopogon* spp. *Mycorrhiza* **19**: 277-282.

Kupper H, Setlik I, Seibert S*, et al.* (2008) Iron limitation in the marine cyanobacterium *Trichodesmium* reveals new insights into regulation of photosynthesis and nitrogen fixation. *New Phytol* **179**: 784-798.

Lacerda CM & Reardon KF (2009) Environmental proteomics: applications of proteome profiling in environmental microbiology and biotechnology. *Brief Funct Genomic Proteomic* **8**: 75-87.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Latorra D, Arar K & Hurley JM (2003) Design considerations and effects of LNA in PCR primers. *Mol Cell Probes* **17**: 253-259.

Lee KH (2001) Proteomics: a technology-driven and technology-limited discovery science. *Trends Biotechnol* **19**: 217-222.

Levitan O, Brown CM, Sudhaus S, Campbell D, LaRoche J & Berman-Frank I (2010) Regulation of nitrogen metabolism in the marine diazotroph *Trichodesmium* IMS101 under varying temperatures and atmospheric CO_2 concentrations. *Environ Microbiol* **12**: 1899-1912.

Levy-Booth DJ & Winder RS (2010) Quantification of nitrogen reductase and nitrite reductase genes in soil of thinned and clear-cut Douglas-fir stands by using real-time PCR. *Appl Environ Microbiol* **76**: 7116-7125.

Li CY & Hung LL (1987) Nitrogen fixing (acetylene-reducing) bacteria associated with ectomycorrhizae. *Plant Soil* **98**: 425-428.

Li CY, Massicotte HB & Moore LVH (1992) Nitrogen-fixing *Bacillus* sp. associated with Douglas-fir tuberculate ectomycorrhizae. *Plant Soil* **140**: 35-40.

Limmer C & Drake HL (1996) Non-symbiotic N_2-fixation in acidic and pH-neutral forest soils: aerobic and anaerobic differentials. *Soil Biol Biochem* **28**: 177-183.

Lindblad P, Atkins CA & Pate JS (1991) N_2-fixation by freshly isolated *Nostoc* from coralloid roots of the cycad *Macrozamia riedlei* (Fisch. ex Gaud.) gardn. *Plant Physiol* **95**: 753-759.

Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M & Patton WF (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21**: 3673-3683.

Marchetti M, Catrice O, Batut J & Masson-Boivin C (2011) *Cupriavidus taiwanensis* bacteroids in *Mimosa pudica* indeterminate nodules are not terminally differentiated. *Appl Environ Microbiol* **77**: 2161-2164.

Maron PA, Ranjard L, Mougel C & Lemanceau P (2007) Metaproteomics: a new approach for studying functional microbial ecology. *Microb Ecol* **53**: 486-493.

Maron PA, Richaume A, Potier P, Lata JC & Lensi R (2004) Immunological method for direct assessment of the functionality of a denitrifying strain of *Pseudomonas fluorescens* in soil. *J Microbiol Methods* **58**: 13-21.

Maron PA, Mougel C, Siblot S, Abbas H, Lemanceau P & Ranjard L (2007) Protein extraction and fingerprinting optimization of bacterial communities in natural environment. *Microb Ecol* **53**: 426-434.

Maron PA, Coeur C, Pink C, Clays-Josserand A, Lensi R, Richaume A & Potier P (2003) Use of polyclonal antibodies to detect and quantify the NOR protein of nitrite oxidizers in complex environments. *J Microbiol Methods* **53**: 87-95.

Martin DE & Reinhold-Hurek B (2002) Distinct roles of P(II)-like signal transmitter proteins and *amtB* in regulation of nif gene expression, nitrogenase activity, and posttranslational modification of NifH in *Azoarcus* sp. strain BH72. *J Bacteriol* **184**: 2251-2259.

Martin FH, Castro MM, Aboul-ela F & Tinoco I, Jr. (1985) Base pairing involving deoxyinosine:

implications for probe design. *Nucleic Acids Res* **13**: 8927-8938.

Martinson L, Lamersdorf N & Warfvinge P (2005) The Solling roof revisited--slow recovery from acidification observed and modeled despite a decade of "clean-rain" treatment. *Environ Pollut* **135**: 293-302.

McTigue PM, Peterson RJ & Kahn JD (2004) Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochemistry* **43**: 5388-5405.

Mergel A, Schmitz O, Mallmann T & Bothe H (2001) Relative abundance of denitrifying and dinitrogen-fixing bacteria in layers of a forest soil. *FEMS Microbiol Ecol* **36**: 33-42.

Mevarech M, Rice D & Haselkorn R (1980) Nucleotide sequence of a cyanobacterial *nifH* gene coding for nitrogenase reductase. *Proc Natl Acad Sci U S A* **77**: 6476-6480.

Moulin L, Munive A, Dreyfus B & Boivin-Masson C (2001) Nodulation of legumes by members of the beta-subclass of Proteobacteria. *Nature* **411**: 948-950.

Murphy ST, Jackman DM & Mulligan ME (1993) Cloning and nucleotide sequence of the gene for dinitrogenase reductase (*nifH*) from the heterocyst-forming cyanobacterium *Anabaena* sp. L31. *Biochim Biophys Acta* **1171**: 337-340.

Muyzer G & Smalla K (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* **73**: 127-141.

Nawarak J, Phutrakul S & Chen ST (2004) Analysis of lectin-bound glycoproteins in snake venom from the *Elapidae* and *Viperidae* families. *J Proteome Res* **3**: 383-392.

Nolan T, Hands RE, Ogunkolade W & Bustin SA (2006) SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Anal Biochem* **351**: 308-310.

Nomura M, Gourse R & Baughman G (1984) Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **53**: 75-117.

Norel F & Elmerich C (1987) Nucleotide sequence and functional analysis of the two *nifH* copies of *Rhizobium* ORS571. *J Gen Microbiol* **133**: 1563-1576.

Oetjen J & Reinhold-Hurek B (2009) Characterization of the DraT/DraG system for posttranslational regulation of nitrogenase in the endophytic betaproteobacterium *Azoarcus* sp. strain BH72. *J Bacteriol* **191**: 3726-3735.

Oetjen J, Rexroth S & Reinhold-Hurek B (2009) Mass spectrometric characterization of the covalent modification of the nitrogenase Fe-protein in *Azoarcus* sp. BH72. *FEBS J* **276**: 3618-3627.

Oliveira ALM, Urquiaga S, Dobereiner J & Baldani JI (2002) The effect of inoculating endophytic N_2-fixing bacteria on micropropageted sugarcane. *Plant Soil* **242**: 205-215.

Pérez CA, Carmona MR & Armesto JJ (2003) Non-symbiotic nitrogen fixation during leaf litter decomposition in an old-growth temperate rain forest of Chiloé Island, southern Chile: Effects of single versus mixed species litter. *Gayana Bot* **60**: 23-33.

Pachlewski R, Kermen J, Chruœciak E & Trzciñska M (1991) Studies of pine ectendomycorrhizae in nurseries. *Acta Mycol* **27**: 49-61.

Paul LR, Chapman BK & Chanway CP (2007) Nitrogen fixation associated with *Suillus tomentosus* tuberculate ectomycorrhizae on *Pinus contorta* var. latifolia. *Ann Bot* **99**: 1101-1109.

Peoples MB & Craswell ET (1992) Biological nitrogen fixation: Investments, expectations and actual contributions to agriculture. *Plant & Soil* **141**: 13-39.

Perez-Osorio AC, Williamson KS & Franklin MJ (2010) Heterogeneous *rpoS* and *rhlR* mRNA levels and 16S rRNA/rDNA (rRNA gene) ratios within *Pseudomonas aeruginosa* biofilms, sampled by laser

capture microdissection. *J Bacteriol* **192**: 2991-3000.

Petersen M, Bondensgaard K, Wengel J & Jacobsen JP (2002) Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids. *J Am Chem Soc* **124**: 5974-5982.

Philippot L (2006) Use of functional genes to quantify denitrifiers in the environment. *Biochem Soc Trans* **34**: 101-103.

Pluskal MG, Przekop B, Kavonian MR, Vecoli C & Hicks DA (1986) Immobilon PVDF transfer membrane: a new membrane substrate for western blotting of proteins. *Bio Techniques* **4**: 272-283.

Poole EJ, Bending GD, Whipps JM & Read DJ (2001) Bacteria associated with *Pinus sylvestris–Lactarius rufus* ectomycorrhizas and their effects on mycorrhiza formation *in vitro*. *New Phytol* **151**: 743-751.

Poussier S, Cheron JJ, Couteau A & Luisetti J (2002) Evaluation of procedures for reliable PCR detection of Ralstonia solanacearum in common natural substrates. *J Microbiol Methods* **51**: 349-359.

Różycki H, Dahm H, Strzelczyk E & Li CY (1999) Diazotrophic bacteria in root-free soil and in the root zone of pine (*Pinus sylvestris* L.) and oak (*Quercus robur* L.). *Appl Soil Ecol* **12**: 239-250.

Rai AN, Söderbäck E & Bergman B (2000) Cyanobacterium-plant symbioses. *New Phytol* **147**: 449-481.

Ram RJ, Verberkmoes NC, Thelen MP*, et al.* (2005) Community proteomics of a natural microbial biofilm. *Science* **308**: 1915-1920.

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

Reinhold-Hurek B & Hurek T (2000) Reassessment of the taxonomic structure of the diazotrophic genus *Azoarcus sensu lato* and description of three new genera and new species, *Azovibrio restrictus* gen. nov., sp. nov., *Azospira oryzae* gen. nov., sp. nov. and *Azonexus fungiphilus* gen. nov., sp. nov. *Int J Syst Evol Microbiol* **50**: 649-659.

Reinhold-Hurek B & Hurek T (2011) Living inside plants: bacterial endophytes. *Curr Opin Plant Biol*.

Reinhold B, Hurek T, Niemann EG & Fendrik I (1986) Close association of *Azospirillum* and diazotrophic rods with different root zones of kallar grass. *Appl Environ Microbiol* **52**: 520-526.

Richards BN (1964) Fixation of atmospheric nitrogen in coniferous forests. *Aust Forestry* **28**: 68-74.

Richards BN (1973) Nirogen fixation in the rrhizosphere of conifers. *Soil Biol Biochem* **5**: 149-152.

Rosch C, Mergel A & Bothe H (2002) Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil. *Appl Environ Microbiol* **68**: 3818-3829.

Roskoski JP (1980) Nitrogen fixation in hardwood forests of northeastern United States. *Plant & Soil* **54**: 33-44.

Rygiewicz PT, Bledsoe CS & Zasoski RJ (1984) Effects of ectomycorrhizae and solution pH on [^{15}N]ammonium uptake by coniferous seedlings. *Can J For Res* **14**: 885-892.

Sakai M, Matsuka A, Komura T & Kanazawa S (2004) Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots. *J Microbiol Methods* **59**: 81-89.

Schneiderbauer A, Sandermann H, Jr. & Ernst D (1991) Isolation of functional RNA from plant tissues rich in phenolic compounds. *Anal Biochem* **197**: 91-95.

Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M & Schulze ED (2005) A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **142**: 335-343.

Seckbach J (2002) *Symbiosis: mechanisms and model systems*. Kluwer Academic Publishers, Dordrecht.

Sen R (2000) Budgeting for the wood-wide web. *New Phytol* **145**: 161-165.

Sevilla M, Burris RH, Gunapala N & Kennedy C (2001) Comparison of benefit to sugarcane plant growth and $^{15}N_2$ incorporation following inoculation of sterile plants with *Acetobacter diazotrophicus*

wild-type and Nif- mutants strains. *Mol Plant Microbe Interact* **14**: 358-366.

Shaffer BT, Widmer F, Porteous LA & Seidler RJ (2000) Temporal and spatial distribution of the *nifH* gene of N_2 fixing bacteria in forests and clearcuts in Western Oregon. *Microb Ecol* **39**: 12-21.

Sharma S, Radl V, Hai B*, et al.* (2007) Quantification of functional genes from procaryotes in soil by PCR. *J Microbiol Methods* **68**: 445-452.

Shibuya N, Goldstein IJ, Van Damme EJ & Peumans WJ (1988) Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J Biol Chem* **263**: 728-734.

Shuman S (1992) DNA strand transfer reactions catalyzed by vaccinia topoisomerase I. *J Biol Chem* **267**: 8620-8627.

Silvester WB & Smith DR (1969) Nitrogen fixation by Gunnera-Nostoc symbiosis. *Nature* **224**: 1231.

Singh SK, Nielsen P, Koshkin AA & Wengel J (1998) LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem Commun (Camb)* **4**.

Smith CJ & Osborn AM (2009) Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol* **67**: 6-20.

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM & Martin PG (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc Natl Acad Sci U S A* **92**: 2647-2651.

Song SD, Hartmann A & Burris RH (1985) Purification and properties of the nitrogenase of *Azospirillum amazonense*. *J Bacteriol* **164**: 1271-1277.

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-517.

Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R-56R.

Stacey GS, Burris RH & Evans HJ (1992) *Biological nitrogen fixation*. Chapman & Hall, New York.

Sun L, Qiu F, Zhang X, Dai X, Dong X & Song W (2008) Endophytic bacterial diversity in rice (*Oryza sativa* L.) roots estimated by 16S rDNA sequence analysis. *Microb Ecol* **55**: 415-424.

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**: 1009-1015.

Tanaka M & Nara K (2009) Phylogenetic diversity of non-nodulating *Rhizobium* associated with pine ectomycorrhizae. *FEMS Microbiol Ecol* **69**: 329-343.

Taylor CM, Marta CB, Claycomb RJ, Han DK, Rasband MN, Coetzee T & Pfeiffer SE (2004) Proteomic mapping provides powerful insights into functional myelin biology. *Proc Natl Acad Sci U S A* **101**: 4643-4648.

Taylor DL, Herriott IC, Long J & O'Neill K (2007) TOPO TA is A-OK: a test of phylogenetic bias in fungal environmental clone library construction. *Environ Microbiol* **9**: 1329-1334.

Taylor EB & Williams MA (2010) Microbial protein in soil: influence of extraction method and C amendment on extraction and recovery. *Microb Ecol* **59**: 390-399.

Thaweenut N, Hachisuka Y, Ando S, Yanagisawa S & Yoneyama T (2011) Two seasons' study on *nifH* gene expression and nitrogen fixation by diazotrophic endophytes in sugarcane (*Saccharum* spp. hybrids): expression of *nifH* genes similar to those of rhizobia. *Plant Soil* **338**: 435-449.

Tian L, Sayer JM, Jerina DM & Shuman S (2004) Individual nucleotide bases, not base pairs, are critical for triggering site-specific DNA cleavage by vaccinia topoisomerase. *J Biol Chem* **279**: 39718-39726.

Torok I & Kondorosi A (1981) Nucleotide sequence of the *R.meliloti* nitrogenase reductase (*nifH*) gene. *Nucleic Acids Res* **9**: 5711-5723.

Torrey JG, Tjepkema JD, Turner GL, Bergersen FJ & Gibson AH (1981) Dinitrogen fixation by cultures of *Frankia* sp. CpII demonstrated by N_2 incorporation. *Plant Physiol* **68**: 983-984.

Tripathi AK, Chowdhury SP & Parashar V (2007) Biological nitrogen fixation: from fundamentals to field considerations. *Agricultural Nitrogen Use and Its Environmental Implications,*(Abrol YP, Raghuram N & Sachdev MS, eds.), pp. 369-394. I. K. International Publishing House Pvt. Ltd, New Delhi.

Tyagi S & Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **14**: 303-308.

Varshney U, Lee CP & RajBhandary UL (1991) Direct analysis of aminoacylation levels of tRNAs *in vivo*. Application to studying recognition of *Escherichia coli* initiator tRNA mutants by glutaminyl-tRNA synthetase. *J Biol Chem* **266**: 24712-24718.

Veedu RN, Vester B & Wengel J (2007) In vitro incorporation of LNA nucleotides. *Nucleosides Nucleotides Nucleic Acids* **26**: 1207-1210.

Vessey JK, Pawlowski K & Bergman B (2005) Root-based N_2-fixing symbioses: legumes, actinorhizal plants, *Parasponia* sp. and cycads. *Plant Soil* **266**: 205-230.

Vitousek PM & Howarth RW (1991) Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry* **13**: 87-115.

Watkins NE, Jr. & SantaLucia J, Jr. (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res* **33**: 6258-6267.

Weidner S, Arnold W, Stackebrandt E & Puhler A (2000) Phylogenetic analysis of bacterial communities associated with leaves of the seagrass *Halophila stipulacea* by a culture-independent small-subunit rRNA gene approach. *Microb Ecol* **39**: 22-31.

Widmer F, Shaffer BT, Porteous LA & Seidler RJ (1999) Analysis of *nifH* gene pool complexity in soil and litter at a Douglas fir forest site in the Oregon Cascade Mountain Range. *Appl Environ Microbiol* **65**: 374-380.

Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF & Williams KL (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* **13**: 19-50.

Wilmes P & Bond PL (2004) The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* **6**: 911-920.

Wilmes P & Bond PL (2006) Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* **14**: 92-97.

Yates RJ, Howieson JG, Reeve WG*, et al.* (2007) *Lotononis angolensis* forms nitrogen fixing, lupinoid nodules with phylogenetically unique, fast-growing, pink-pigmented bacteria, which do not nodulate *L. bainesii* or *L. listii*. *SOIL BIOL BIOCHEM* **39**: 1680-1688.

Yoneyama T, Fujita K, Yoshida T, Matsumoto T, Kambayashi I & Yazaki J (1986) Variation in natural abundance of ^{15}N among plant parts and in $^{15}N/^{14}N$ fractionation during N_2 fixation in the legume-rhizobia symbiotic system. *Plant Cell Physiol* **27**: 791-799.

You Y, Moreira BG, Behlke MA & Owczarzy R (2006) Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res* **34**: e60.

Zani S, Mellon MT, Collier JL & Zehr JP (2000) Expression of *nifH* genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Appl Environ Microbiol* **66**: 3119-3124.

Zehr JP & McReynolds LA (1989) Use of degenerate oligonucleotides for amplification of the *nifH* gene

from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl Environ Microbiol* **55**: 2522-2526.
Zehr JP, Limberger RJ, Ohki K & Fujita Y (1990) Antiserum to nitrogenase generated from an amplified DNA fragment from natural populations of *Trichodesmium* spp. *Appl Environ Microbiol* **56**: 3527-3531.
Zehr JP, Jenkins BD, Short SM & Steward GF (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* **5**: 539-554.

Acknowledgements

First, I would like to give my sincere thanks to my supervisor, Dr. Thomas Hurek, for guiding me to carry out this challenging project. During these five years research work here, there were many frustrating experiences and also some exciting breakthroughs. For better or for worse, he is always there for me with his remarkable patience and insightful suggestions. It would not have been possible to complete this dissertation without his support and dedication. I am also especially grateful to Prof. Dr. Barbara Reinhold-Hurek for giving me the opportunity to work in her group, offering all kinds of conveniences for lab work, and supporting me with fellowships when my DFG contract was ended.

I would also like to specifically thank Prof. Dr. Uwe Nehls, who cheerfully agreed to be the second reviewer of my thesis. My thanks also go to Prof. Dr. Friederike Koenig for serving in my dissertation committee and also for the good conversations.

Many thanks go to Andrea, Elvira, Hannah, Janina, Jessica, Theresa and Xi for their kindly help to construct the German version of summary, correct grammar, and give many valuable advices in my thesis writing. My thanks also go to Claudia, Haiyuan, Liwei and our former members Ariana, Asim, Moudjahidou and Lei. I am grateful to the intensive discussions we had on work, and thank you for your encouragements. I would like to thank all the other members of AG Reinhold-Hurek: Abhijit, Andrea, Frank, Janne, Stefanie, Teja and Thea. Thank you for making a very nice communication atmosphere for research and life in the lab.

Last, but not least, I would like to thank my parents and my friends in China for their love and support during the past few years.

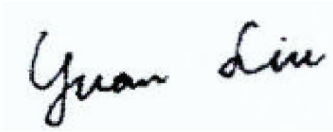
This work is financially supported by Deutsche Forschungsgemeinschaft grants to Dr. Thomas Hurek (Hu 579/7-1 and Hu 579/8-1).

Declaration

I hereby declare that I prepared my dissertation entitled “Study on the major diazotrophs associated with Norway spruce (*Picea abies* L. Karst) roots” independently. The scientific data presented in this dissertation were my own research findings.

I also admit that this dissertation, in the present form or in a similar form was never submitted to another University and has never served for other examination purposes.

Bremen, May 2011



Yuan Liu