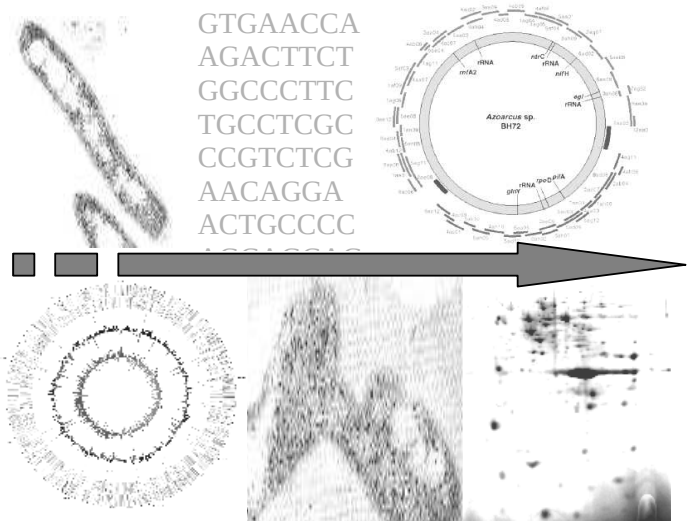
FROM GENOMICS TO PROTEOMICS IN *Azoarcus* sp. BH72, A N_2-FIXING ENDOPHYTIC BACTERIUM



Federico Battistoni

Bremen 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN (NATURAL) SCIENCE
(Dr. rer. nat.)

LABORATORY OF GENERAL MICROBIOLOGY
THE FACULTY OF BIOLOGY AND CHEMISTRY
UNIVERSITY OF BREMEN
GERMANY

Supervisor: Professor Dr. Barbara Reinhold-Hurek

VOM GENOM ZUM PROTEOM DES STICKSTOFF-FIXIERENDEN ENDOPHYTEN *Azoarcus* sp. Stamm BH72

DISSERTATION

Zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

Dem Fachbereich Biologie/Chemie der
Universität Bremen
vorgelegt von

MSc. Federico Battistoni

Bremen 2006

The studies of the presented work have been carried out from January 2003 till January 2006 at the department of Biology and Chemistry, University of Bremen, Germany, under the guidance of Prof. Dr. Barbara Reinhold-Hurek.

Die Untersuchungen zur folgenden Arbeit wurden von Januar 2003 bis Januar 2006 am Fachbereich Biologie/Chemie der Universität Bremen unter der leitung von Prof. Dr.Barbara Reinhold-Hurek durchgeführt.

Von Fachbereich Biologie/Chemie der Universität Bremen als Dissertation
angenommen am:

Datum der Disputation:
1. Erstgutachterin: Prof. Dr. Barbara Reinhold-Hurek
2. Zweitgutachterin: Prof. Dr. Annette Becker

Parts of the results presented in the following work have been submitted to or are already published in:

Battistoni, F., Reamon-Büttner, S., Bartels, D., Kaiser,O., Hurek,T. and Reinhold-Hurek,B. 2005. Physical map of the *Azoarcus* sp. strain BH72 genome based on a bacterial artificial chromosome (BAC) library as a platform for genome sequencing and functional analysis. FEMS Microbiol. Lett. **249**: 233-240.

Krause, A., Ramakumar, A., Bartels, D., Battistoni, F., Bekel, T., Boch, J., Böhm, M., Friedrich, F., Hurek, T., Krause, L., Linke, B., McHardy, A., Sarkar, A., Schneiker, S., Syed-Ali, A., Thauer, R., Vorhölter, F., Weidner, S., Pühler, A., Reinhold-Hurek, B., O. Kaiser and A. Goesmann. 2006. Life inside cereals: Genomic insights into the lifestyle of the mutualistic, N2-fixing endophyte *Azoarcus* sp. strain BH72. Nature Biotechnology. **24**: 1385-1391.

Battistoni, F., Oetjen, J. and B. Reinhold-Hurek. 2006. Proteomics analysis of *Azoarcus* sp. strain BH72 grown under different N_2-fixing conditions. Manuscript to be submitt to Journal of Bacteriology.

Miché, L., Battistoni, F., Gemmer, S., Belghazi, M. and B. Reinhold-Hurek. 2006. Up-regulation of jasmonate-inducible defense proteins and differential colonization of roots of *Oryza sativa* cultivars with the endophyte *Azoarcus* sp. Mol. Plant Microbe Interact. **19**: 502-511.

TABLE OF CONTENTS **V**

ABBREVIATIONS **1**

SUMMARY **3**

CHAPTER A: Introduction **6**

1. Microbial genome projects **7**

2. Microbial genome sequencing **8**

3. Genome-wide physical mapping using bacterial large-insert clone library **9**

3.1. Large-insert source, Bacterial Artificial Chromosome (BAC) libraries **10**

3.2. Strategies for genome physical mapping with large-insert clones **11**

3.2.1. Recurrent hybridization 13

3.2.2. Mapped DNA marker-based chromosome landing 13

3.2.3. Restriction fingerprints analysis 13

4. Genome gene finding and annotation **14**

5. From computer analysis to experimentation: Functional genomics **14**

5.1. Proteomics as a functional genomics tool **15**

5.1.1. Identification and analysis of proteins 16

5.1.2. Identification of post-translational modifications 18

5.1.3. Differential-display proteomics 18

5.2. Proteomics as a tool of plant-microbe interactions studies **20**

6. *Azoarcus* sp. BH72, a model bacterium for diazotrophic endophytes **21**

6.1. Taxonomy **21**

6.2. Plant infection and colonization **22**

6.3. Endophytic function **24**

6.4. Physiology and genetic of nitrogen fixation **25**

7. Outline of this thesis **27**

CHAPTER B: *Azoarcus* sp. BH72 genome physical map **30**

1. OBJECTIVES **31**

2. CONTRIBUTION **31**

3. RESULTS AND CONCLUSIONS **33**

4. OUTLOOKS **34**

CHAPTER C: *Azoarcus* sp. BH72 genome **42**

1. OBJECTIVES **43**

2. CONTRIBUTION **43**

3. RESULTS AND CONCLUSIONS **43**

4. OUTLOOKS **44**

CHAPTER D: Annotation and analysis of part of *Azoarcus* sp. BH72 genome sequence **53**

1. INTRODUCCION **54**

1.1. Prokaryotic genomes sequencing **54**

1.2. Genomes relevant for the environment: *Azoarcus* sp. BH72 genome project **55**

1.3. Protein classification: Clusters of Orthologous Groups of proteins (COG) database **55**

2. OBJECTIVES **56**

3. MATERIAL AND METHODS **57**

3.1. Genome sequence analysis: annotation **57**

3.2. Genome comparison **58**

3.3 Transporters classification **59**

3.4. ATP-binding cassette (ABC) transporters comparison **59**

3.5. Central metabolism in *Azoarcus* sp. BH72: Sulfate, phosphate, iron transport and metabolism related genes **60**

3.6. Siderophores biosynthesis pathways analysis **60**

3.7. Siderophore detection in the growth medium **60**

4. RESULTS AND DISCUSSION **61**

4.1. Genome comparison **61**

4.2. Proteins of the "Ion Transport and Metabolism" COG category **63**

4.2.1. Iron transport and metabolism related genes **64**

4.2.1.1. Outer membrane TonB-dependent genes *64*

4.2.1.2. Absence of siderophore biosynthetic pathways *67*

4.2.1.3. tonB-exbBD complex genes 68
4.2.1.4. ABC-iron related transporters 69
4.2.1.5. Iron regulators related genes 70
4.2.1.6. Intracellular iron homeostasis 71
4.2.1.7. Ferrous-iron transport related genes 72
4.2.2. Sulfur and phosphorus transport and metabolism related genes **72**
4.2.2.1. Sulfur assimilation 72
4.2.2.2. Phosphate assimilation 74

4.3. Proteins of the "Carbohydrate Transport and Metabolism" COG category: Central metabolism in Azoarcus sp. BH72 **74**
4.3.1. C4-dicarboxylates transport and glyoxylate shunt **74**
4.3.2. Tricarboxilic acid (TCA) cycle, glycolysis, gluconeogenesis and phosphoenolpyruvate: sugar phosphotransferase system (PTS) **77**
4.4. Transporters **80**
4.4.1. Classification of general transporters **80**
4.4.2. Comparison of number of predicted ABC-transporters among bacterial genomes **82**
4.4.3.Comparison of number of predicted ABC-transporters between different niches and between *Bacteria* and *Archaea* **84**

CHAPTER E: Functional genomic in *Azoarcus* sp. BH72, a proteomic approach **88**

1. OBJECTIVES **89**
2. CONTRIBUTION **89**
3. RESULTS AND CONCLUSIONS **90**
4. OUTLOOKS **92**

CHAPTER F: Proteomics studies of the *Azoarcus* sp. BH72-rice interaction **126**

1. OBJECTIVES **127**

2. CONTRIBUTION **128**

CHAPTER G: Discussion **139**

1. *Azoarcus* sp. BH72 genome: sequence assembly validation and physical map. **140**

1.1 Genome comparison **142**

2. From genomic to functional genomic: proteomics studies and metabolic pathways reconstruction of N_2-fixing *Azoarcus* sp. BH72 cells. **143**

2.1 Comparison of 2D-proteome pattern of strain BH72 growth in pure- or co-culture N_2-fixing conditions. **143**

2.2. Carbon metabolism **144**

2.3 N_2-fixation and nitrogen assimilation metabolism **147**

2.4 Energy production **149**

2.5. Protein synthesis and folding **149**

2.6. Iron transport and metabolism in *Azoarcus* sp. BH72 **150**

2.7. Cellular protection and detoxification **152**

2.8. Membrane proteins **153**

2.9. Motility and adhesion **153**

3.0. Diazosome related proteins **154**

3. Concluding remarks **154**

REFERENCES **155**

ABBREVIATIONS

2D-DIGE	two-dimensional difference gel electrophoresis
2D-PAGE	two-dimensional polyacrylamide gel
ABC	the ATP-binding cassette transporters superfamily
APS	ammonium persulphate
BAC	bacterial artificial chromosome
BRENDA	comprehensive enzyme information system database
CAS	chromeazurol S
COG	clusters of orthologous groups of proteins database
EDTA	ethylene diamine tetra acetic acid
ESI	electrospray ionization
EST	expressed sequence tag
GOLD	genomes online database
GUI	graphic user interface
HMW	high molecular weight
HS	hierarchical shotgun sequencing approach
IUBMB	International union of biochemistry and molecular biology
JA	jasmonic acid
KEGG	Kyoto encyclopedia of genes and genomes
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionization
MFS	major facilitator superfamily
MS	mass spectrometry
NCBI	National center for biotechnology information
OD_{578}	optical density measured at a wavelength of 578 nm
OMR	outer membrane protein
ORF	open reading frame
PCWDE	plant cell-degrading enzymes
Pfam	protein family database
PFGE	pulse-field gel electrophoresis
PMF	peptide mass fingerprint
PSD	post-source decay
PTM	post-translation modification
PTS	phosphoenolpyruvate: sugar phosphotransferase system
RFLP	restriction frame length polymorphism
RND	the resistance-nodulation-cell division transporter superfamily

SDS	sodium dodecylsulphate
SSC	standard saline citrate
TBS	tris-buffered saline
TCA	trichloroacetic acid
TC-DB	transport classification-database
TEMED	N,N,N',N'- tetramethylethylene-diamine
TIGR	The institute for genomic research
TOF	time of flight
TRAP	type C_4 dicarboxilate transporter
Tris	N-tris-(hydroxymethyl)-amino methane
WGS	whole genome shotgun sequencing approach

SUMMARY

Since the first genome sequence was completed in 1995, the number of publicly available prokaryotic genomes increased exponentially up today, when more than 300 microbial projects had been completed. Within this "genomic revolution", the complete genome analysis of environmentally and biotechnologically relevant microorganisms constitutes an emerging field of research. In this context the genome-sequencing project of *Azoarcus* sp. BH72 took place. The diazotrophic endophyte *Azoarcus* sp. BH72 posses a high biotechnological potential as it is capable of colonizing the interior of rice roots, which is one of the globally most important crops. Moreover, since the rice genome is also sequenced, the *Azoarcus*-rice system represents a great model for bacteria-grass interaction studies. In the present thesis three main topics, from genomics to functional genomic studies, were covered. The aim of the first part of this work was the characterization of an *Azoarcus* sp. BH72 bacterial artificial chromosome (BAC) library as well as the building of a physical map of strain BH72 chromosome. Both tools were used for an independent analysis of the genome structure in comparison to a shotgun library with small insert sizes, for contig assembly and gap closure of this shotgun library and for genome comparison analysis. The second main objective of this work was the annotation and analysis of a part of the genome sequence. In particular, genes which belong to the COG categories "Ion Transporters and Metabolism", and "Carbohydrates Transport and Metabolism", were studied. This analysis revealed several highlights in the genome sequence, particularly in the iron metabolism, which can be used as a starting point in future studies. Finally, in the last part of this work a functional genomic analysis of *Azoarcus* sp. BH72 grown under different conditions of N_2-fixation was performed, using a differential-display proteomic approach. Proteomic patterns of strain BH72 N_2-fixing cells, in pure- and co-culture with the endophytic fungus *Acremonium alternatum*, showed strongly significant differences. The identification of the major proteins showed that the nitrogen metabolism was very active in both conditions, as well as the carbon metabolism, which was adapted to the carbon sources available. Several membrane proteins were identified which most probably are involved in bacteria-fungus interaction, as well as in bacterial response to fungus metabolites. The results obtained contribute to a better understanding of the *Azoarcus* sp. BH72 physiology and ecology.

ZUSAMMENFASSUNG

Die Anzahl der öffentlich verfügbaren prokaryotischen Genomsequenzen ist seit 1995, dem Jahr in dem die erste Genomsequenz veröffentlicht wurde, exponentiell angestiegen. Bis heute konnten mehr als 300 Genomprojekte abgeschlossen werden. Während dieser als "genomic revolution" bezeichneten Epoche wurde die Analyse der Genome von umwelt- und biotechnologisch relevanten Mikroorganismen zu einem Schwerpunkt des sich ausdehnenden Forschungsbereiches. In diesem Zusammenhang wurde das Genomprojekt des Endophyten *Azoarcus* sp. BH72 begründet. *Azoarcus* sp. BH72 ist ein diazotropher Organismus mit einem hohen biotechnologischen Potenzial. Dieses Bakterium ist zur Kolonisierung von Reiswurzeln befähigt, wobei Reis eine der global wichtigsten Kulturpflanzen ist. Insbesondere da auch das Reisgenom sequenziert wurde, stellt die *Azoarcus* sp. BH72-Reis Interaktion ein aufschlussreiches Modell für die Bakterien-Gramineen Interaktionen dar. Die vorliegende Arbeit besteht aus drei Themenkomplexen, die sowohl Teile der Genom- als auch der funktionellen Genomanalyse umfassen. In dem ersten Teil wurde die Charakterisierung einer *Azoarcus* sp. BH72 BAC (bacterial artificial chromosome)-Genbank sowie die Herstellung einer Genkarte des *Azoarcus*-Chromosoms vorgenommen. Beide Methoden dienten zur unabhängigen Analyse der Genomstruktur im Vergleich zu einer "shotgun"-Genbank mit kleineren Inserts, zur Contig-Assemblierung, zur Schließung von Lücken in der shotgun-Genbank und zur vergleichenden Genomanalyse. In dem zweiten Teil dieser Arbeit wurde die Annotation und Analyse eines Teils der Genomsequenz behandelt. Dabei wurden im Besonderen Gene aus den COG-Kategorien "Ionen-Transporter und Metabolismus" und "Kohlenhydrat Transport und Metabolismus" studiert. Es konnten Besonderheiten in der Genomsequenz, vor allem im Eisen-Metabolismus, aufgedeckt werden, die als Startpunkt weiterer Analysen dienen könnten. Im dritten Teil stand die funktionelle Genomanalyse von *Azoarcus* sp. BH72 im Vordergrund, welcher unter unterschiedlichen Bedingungen der Stickstofffixierung kultiviert wurde. Bei einer Proteomanalyse von stickstofffixierenden Zellen in Reinkultur und in Dualkultur mit dem endhophytischen Pilz *Acremonium alternatum* ergaben sich signifikante Unterschiede. Die Aktivität des Stickstoffstoffwechsel war unter beiden genannten Bedingungen hoch; es konnten einige Schlüsselenzyme der Stickstofffixierung und der Ammonium-Assimilierung identifiziert werden. Zudem wurden viele Proteine aus dem Kohlenstoffmetabolismus detektiert, welcher an die verfügbaren Kohlenstoffquellen angepasst war. Einige der identifizierten Membranproteine könnten in der Bakterien-Pilz Interaktion eine Rolle spielen oder aber in der Einstellung des bakteriellen Metabolismus auf den Pilzstoffwechsel involviert sein. Die hier erhaltenen Ergebnisse tragen zu einem

besseren Verständnis der Physiologie und Ökologie des Endophyten *Azoarcus* sp. BH72 bei.

CHAPTER A

Introduction

"Truth is never pure, and rarely simple."
—Oscar Wilde

The Importance of Being Earnest, Act I.

1. Microbial genome projects.

The genomic approach of technology development and large-scale generation of resource data sets have introduced an important new dimension into biological research. On July 28 1995, a new era in genomics science started with the publication of the *Haemophilus influenzae* 1.8 Megabase (Mbp) genome sequence (Fleischmann *et al.*, 1995). This was followed by the publication in less than three month of the 0.58 Mbp genome sequence from *Mycoplasma genitalium* (Fraser *et al.*, 1995). Additionally, the genome sequencing project of *Escherichia coli* K-12 was started at the same period as *H. influenzae* but finished in 1997 because of its large size (4.6 Mbp). Since this period the number of publicly available prokaryotic genomes increased exponentially (Blattner *et al.*, 1997) (Figure 1).

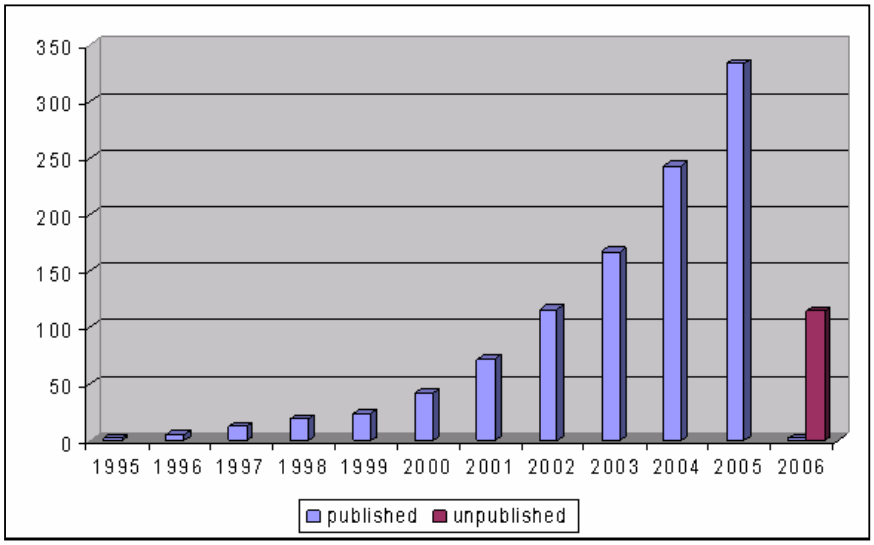


Figure 1: Number of fully sequenced organisms in the public databases since 1995.
Source: GOLD database (Bernal *et al.*, 2001). Last update January 2006.

A revolution in biological research has begun. New projects, researches, strategies and experimental technologies have generated a continues flow of large and more complex genomic data sets. This data have been transferred into public databases, and have facilitated the study of almost all life processes. Genome sequences, the bounded sets of information that guide biological development and function, lie at the heart of this revolution (Francis *et al.*, 2003).

The main driving forces behind bacterial genome sequencing have been the need to understand and control pathogenic microorganisms and to understand and manipulate

organisms of potential use in biotechnology, as well as to identify and exploit enzymes with industrial potential from extremophiles (Figure 2).

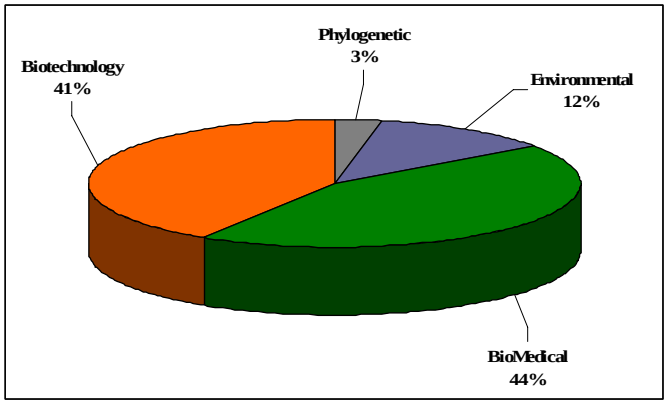


Figure 2: Funding relevance of Archaeal and Bacterial genome projects.
Source: GOLD database (Bernal *et al.*, 2001). Last update January 2006.

In the latest years additional interest in microbial genome projects have been focused in those bacteria which are relevant for the agriculture and environment (Figure 2). In this context, the present work proposes to study an environmentally and biotechnologically relevant microorganism *Azoarcus* sp. BH72, from its complete genome as a starting point, to large-scale functional genomics studies.

2. Microbial genome sequencing.

Different strategies have been developed and used in large-scale genome sequencing in the past decade. Whole genome shotgun (WGS) (Sanger *et al.*, 1977; Weber and Myers, 1997) sequencing is the most widely used strategy for microbial genome sequencing (Fleischmann *et al.*, 1995; Klenk *et al.*, 1997; Tomb *et al.*, 1997). In this technique, the primary sequence information is obtained robotically from a random library of small DNA fragments cloned in a high copy plasmid vector. The sequences of many thousands of DNA fragments (usually corresponding to a coverage between 3 and 8 times of the size of the bacterial genome), are then organized into contiguous segments (contigs). This process can be also handled automatically (Kaiser et al., 2003) (Figure 3a). On the other hand, the hierarchical shotgun (HS) approach provides an alternative more suitable for larger and more complex genomes (Venter *et al.*, 1996). In this approach a genome is decomposed into larger fragments, usually BACs (bacterial artificial chromosomes) or cosmids. A BAC-based physical map of the genome is constructed by using one or more physical map strategies. Minimal tiling path clones that completely cover the genome are selected from the physical map and used as substrates for shot gun sequencing, one clone at a time

(Zhang and Wu, 2001). The HS approach has been used for most eukaryotic genomes (Adams *et al.*, 2000; Goffeau *et al.*, 1996; Meinke *et al.*, 1998) (Figure 3b).

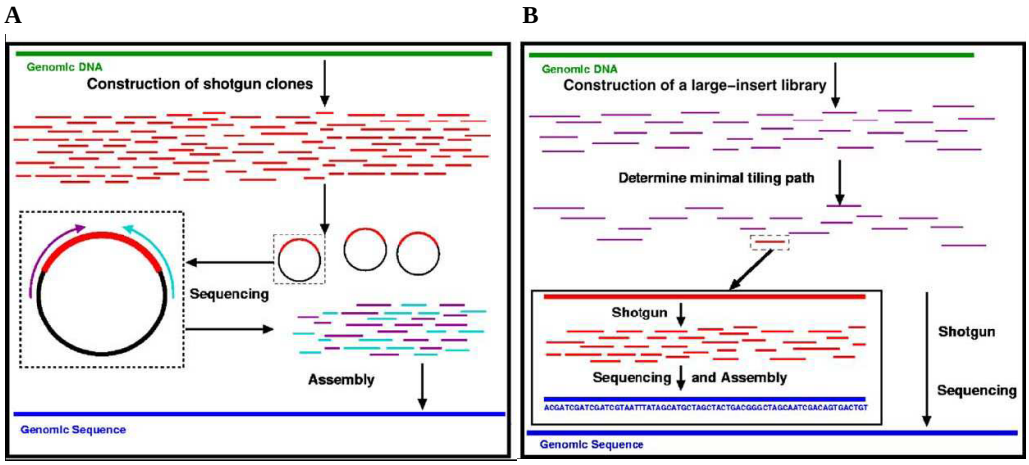


Figure 3: From genomic DNA to genomic sequence. (a) Whole genome shotgun procedure starting by fragmenting the genome into pieces (≥10 kb), sequencing the fragments and ending with the assembly based on sequence overlaps. **(b)** Hierarchical approach utilizing a large insert BAC library (up to 200 kb), a minimal set of overlapping BACs and the subsequent shotgun sequencing of the selected BACs (Kaiser et al., 2003).

3. Genome-wide physical mapping using a bacterial large-insert clone library.

Although both mentioned approaches are being used in different large-scale sequencing projects of large, complex genomes, it seems evident that bacteria-based large-insert clones are indispensable for both of them. In the HS approach the large insert-based physical map is essential for selection of minimally overlapping clones and building of a tiling clone path for genome sequencing and sequence assembly. On the other hand, for WGS, while the preliminary sequence assemblies are generated from small-insert clones of a WGS library, a large insert-based physical map is simultaneously developed, which assists in assembling the whole sequence of the genome. Several large-insert DNA clones have been developed and used to clone the DNA of large complex genomes (Table 1). This work is focused on a bacterial artificial chromosome (BAC) clone system.

Table 1: General characteristics of most common large-insert DNA clone systems.

Large-insert clone	Insert size (kb)	Host	Notes
YAC	Over 1000	yeast	High level of chimerism. Insert instability. Purification problems.
Cosmids	40-50	bacteria	Limited cloning capacity. Several copies in bacterial cell.
Fosmids	40-50	bacteria	Single copy in bacterial cell.
Bacteriophage P1	Up to 100	bacteria	Single copy in bacterial cell.
BAC	Up to 400	bacteria	Single-copy in bacterial cell. Easy to clone and purified. Stable in the host, low level of chimerism.
PAC	Up to 400	bacteria	Single-copy in bacterial cell. Easy to clone and purified. Stable in the host, low level of chimerism.
PBC	Up to 400	bacteria	Single-copy in bacterial cell. Easy to clone and purified. Stable in the host, low level of chimerism.

3.1. Large-insert source: Bacterial Artificial Chromosome (BAC) library.

A sequence-ready BAC-based physical map of high genome coverage is crucial for completely sequencing a whole genome, as well as for genome assembling validation. Successful development of a BAC-based physical map depends on several factors, including the insert sizes, genome representation in the library, the size and complexity of target genomes, as well as the methodology employed (Shizuya and Kouros-Mehr, 2001). Currently, the genome representation of a library is estimated theoretically multiplying the number of clones by the average insert size, divided by the genome size.

All BAC-libraries developed to date are generated from partial digests of high molecular weight (HMW) DNA with restriction enzymes (Zhang and Wu, 2001). Studies show that the distribution of the sites of a restriction enzyme within a genome is irregular. Therefore, the genomic regions with a particularly high or low density of the restriction sites of a restriction enzyme are difficult to clone. This is because the small (<40 kb) and the big (>400 kb) DNA fragments generated by partial digestion are removed during the size selection of BAC cloning (Shizuya *et al.*, 1992).

The BAC cloning system is based on the *Escherichia coli* F-factor, a low-copy plasmid that exists in a supercoiled circular form. The F-factor not only codes for genes that are essential to regulate its own replication but also controls its copy number (one or two copies pero cell) (Shizuya *et al.*, 1992). This fact allows a stable maintenance of large

DNA inserts reduces the potential for recombination between DNA fragments carried by the plasmid and allows for a viable large-scale DNA-purification (Figure 4). Additionally, the BAC vector incorporates a chloramphenicol resistence marker and a cloning system including: two cloning sites (*Hind*III and *Bam*HI) and several G+C rich restriction enzymes sites (e.g. *Not*I, *Eag*I, *Sma*I, *Bgl*I) for potential excision of the inserts. The cloning site is flanked by T7 and SP6 promoters for DNA sequencing of the insert segement (Shizuya and Kouros-Mehr, 2001). Several BACs cloned vectors have been developed and used in genome sequencing proyects such as pBeloBAC11 (Kim *et al.*, 1996), pBACindigo (Tomkins *et al.*, 2001) and pECBAC1 (Frijters *et al.*, 1997). The BAC vectors pBeloBAC11 and pBACindigo, contains and additionally component in the cloning system site-β-galactosidase (*lacZ*), which allows α complementation (Shizuya and Kouros-Mehr, 2001). This component allows clones with inserts to be readily identified as an X-gal cloro change.

BAC clones from the physical map can be readily used for functional analysis of the genome sequence. Furthermore, if the WGS approach is employed, selected BAC-clones can be used as substrates for gap closure in genome sequencing. This is because in the WGS approach, only the ends of the large-insert clones are sequenced and used to link the assembled sequence contigs and span putative long-range repetitive sequence regions.

3.2. Strategies for genome physical mapping with large-insert clones.

Several strategies have been developed and used to construct sequence-ready physical maps from large-insert clones for genomics studies. In the following subsections a short description of the main current strategies used are described. However, in most of the genome projects, a combined approach of them were required (Hong, 1997; Venter *et al.*, 1996). Examples of a minimal tiling path of BAC-clones, as well as a physical map, for *Bradyrhizobium leguminosarum* genome, are shown in Figure 5.

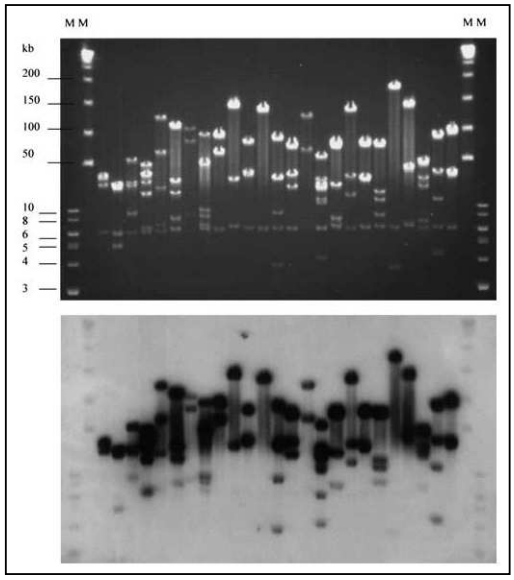


Figure 4: Twenty-five randomly selected *Bradyrhizobium japonicum* BAC clones digested with *Xba*I. (A) Ethidium bromide-stained pulse-field electrophoresis gel showing insert DNA above and below the 6-kb pBACindigo *Xba*I vector fragment band. **(B)** Autoradiograph of gel in *A*, after Southern transfer and probing with total *B. japonicum* genomic DNA(Tomkins *et al.*, 2001).

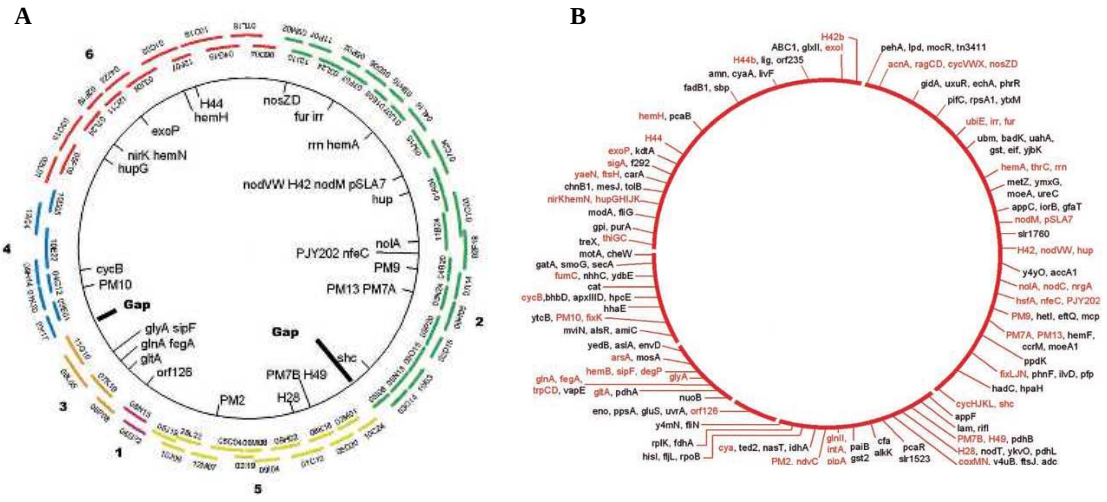


Figure 5: A: Six contigs illustrated by a selected set of BAC clones providing a complete tiling path of the *Bradyrhizobium japonicum* replicon. B: Complete physical map with integrated hybridization-based markers and BAC end sequences (Tomkins *et al.*, 2001).

3.2.1. Recurrent hybridization.

This method is a library colony hybridization-based strategy for global physical mapping of genomes. In this strategy, probes are first generated from the insert ends of a number of random large-insert clones and hybridized to the entire source library arranged on high-density colony filters. Then, a new set of clones are chosen from those that have not been hit in the first hybridization experiment and used to generate insert end probes to continue the hybridization of the entire source library spotted on filters. This process is repeated until most of the clones have been hit. Overlapping clones are organized into contigs according to the probes to which they hybridize. This approach has been used successfully to generate a BAC-based physical map of the *A. thaliana* genome (Arumuganathan and Earle, 1991).

3.2.2. Mapped DNA marker-based chromosome landing.

This is a hybridization- and/or PCR-based strategy for genome physical mapping. A prerequisite for using this approach is to have a high-density, regularly distributed, DNA marker genetic map in addition to the genomic DNA library of large inserts (Zhang and Wu, 2001). PCR or hybridization screening, depending on the marker in the genetic map, are used for clone fishing and anchoring in the large-insert library (Kurata et al., 1997). In the case of a BAC-library to be anchoring, a highly-saturated DNA marker genetic-map is required because of the small insert size of the BAC clones. In this strategy, DNA markers are selected from the genetic linkage map and used as probes or PCR primers to screen the BAC-library arranged on high-density colony or high concentration BAC-DNA filters. Overlapping clones are assembled into contigs according to the probes to which they hybridize or primer pairs with which they produce PCR products. The contigs are extended and anchored to the genetic linkage map according to the position of the DNA markers on the genetic map. This approach was applied to develop a BAC-based physical map of the euchromatic portion of the Drosophila genome (Hoskins et al., 2000). The advantage of this method is the development and integration of the physical map with linkage genetic map in the same process. However the prerequisition of a genetic map have restricted its wide applications.

3.2.3. Restriction fingerprints analysis.

This method is a restriction fragment fingerprint-based strategy for genome physical mapping. This strategy is based on the hypothesis that the clones derived from the same regions of a genome share similar restriction patterns (a large number of common bands). The restriction pattern of a clone is designated as the ‘fingerprint’ of the clone *(Marra et*

al., 1997). The fingerprints of the clones are visualized, entered and digitized in computers. Overlapping clones are assembled into contigs with computers according to the similarity of the restriction patterns of the clones (Fuhrmann *et al.*, 2003). This approach was used to develop a BAC-based physical map of the rice genome (Zhang and Wing, 1997).

4. Genome gene finding and annotation.

A crucial question for the whole-genome sequencing enterprise is: how informative are the sequences? In other words, when the complete genome sequence is available, for what fraction of the gene products will it be possible to reveal evolutionary relationships and predict functions?

The identification of genes in prokaryotic genomes has advanced to the stage at which nearly all protein-coding regions can be identified with confidence. Computational gene finders using Markov modeling techniques now routinely find more than 99% of protein-coding regions (Delcher *et al.*, 1999) and RNA genes (Lowe and Eddy, 1997). Once the protein-coding genes have been located, the most challenging problem is to determine their function (annotation). Today, about 40–60% of the genes in a new genome sequence can be classified almost automatically based on sequence similarity (Fraser *et al.*, 2000). This sequence similarity is the first step for assigning function to new proteins, but the transfer of functional assignments is difficult. On the other hand, to identify the function of the remaining 40–50% is still a laborious task. In particular, these unknown novel genes are often the most interesting from the scientific point of view, as most probably they encode proteins for some particular features of the organism of interest. As has been reflected from these facts, the annotation process is a key step for a successful genome project. During this process, best results are obtained when an approach, which combines the data automatically collected from databases with a human expert annotation, is used. In this sense, a couple of open source genome annotation software have been developed and used in several microbial genome projects (Meyer et al., 2003; Rutherford et al., 2000).

5. From computer analysis to experimentation: Functional genomics.

From computer based comparison of genomes, important conclusions can be drawn, including information about genome organization, conservation and probable evolutionary events (Van Sluys *et al.*, 2002). However, the most important outcome of these analyses is their utility for interpreting experimental results and directing new experimentation. Genome sequence analyses are used by experimenters in two main different ways. The first

involves the testing of computer predictions for specific gene products. Once a new complete genome sequences available, computer predictions are important for prioritizing experiments. The second way is to guide more global studies e.g. analysis of gene transcription or systematic expression analysis (Koonin *et al.*, 1996).

Functional genomics is a term applied for an overlapping set of experimental approaches that aims to use genome sequence data to facilitate the discovery of gene and proteins function, and to elucidate higher levels of interaction between genes and gene products (Pallen, 1999). These approaches include large-scale analysis of: gene transcription (transcriptome), protein expression (proteome), metabolites production (metabolomics), along with systematic mass mutagenesis and protein interaction maps (Colebatch *et al.*, 2002b). These experimental approaches are defining a new era in science, the "Post-genomic era".

5.1. Proteomics as a functional genomics tool.

Proteomics is defined as a large-scale study of proteins, usually by biochemical methods. Traditionally the word proteomics has been associated with displaying a large number of proteins from a particular cell line or organism on two dimensional polyacrylamide (2D-PAGE) gels (Anderson and Anderson, 1996; Wilkins *et al.*, 1996). In this sense proteomics studies started in the late 1970s when researchers started to build databases of proteins using the newly developed 2D-PAGE gel electrophoresis technique (O'Farrell, 1975). In the 1990s, biological mass spectrometry emerged as a powerful analytical method which removed most of the limitations of protein analysis (Mann *et al.*, 2001). Today, the term proteomics covers much of the functional analysis of gene products or 'functional genomics', including large-scale identification or localization studies of proteins, post-translation modification and interaction studies (Pandey and Mann, 2000).

How proteomic complement genomics? In the post-genomic era, large amounts of DNA sequences have been deposited in databases. However, as mentioned before, possession of only complete sequences of genomes is not sufficient to elucidate biological function. Proteomics complements genomics because it focuses on the gene products, which are the active elements of cells (Washburn and Yates, 2000). Information about modification of proteins is apparently not available from DNA sequences (e.g. isoforms and post-translational modifications) which can only be determined through proteomic methodologies. Additionally, protein–protein interactions can be studies only at the protein level (Pandey and Mann, 2000), through techniques like surface plasmone resonance.

5.1.1. Mass spectrometry: identification and analysis of proteins.

A standard proteomic analysis of an organism begins with the crucial step of isolation of its soluble proteins, followed by the separation and visualization of the protein mixture by one- or 2D-PAGE gel electrophoresis (Figure 6). Although essentially powerful, there are several limitations of 2D-PAGE gel electrophoresis: such as time consumption and laborious, difficulties in the visualization of low-abundant proteins, and problems to solubilize hydrophobic proteins (Eichacker *et al.*, 2004; Herbert *et al.*, 2001; Santoni *et al.*, 2000a). On the other hand, the advantages of 1D-PAGE electrophoresis include the easy solubilization of extremely acidic or basic proteins in SDS and the easy visualization. But one of the main disadvantages is the low resolution. After the invention of the 2D-PAGE gel electrophoresis, as mentioned the other significant advance in proteomics has been the development of biological mass spectrometers (MS), which allow the analysis and identification of gel-separated proteins(Mann *et al.*, 2001). This technique relies on the analysis of peptides obtained from *in gel* digested proteins by a sequence-specific protease such as trypsin. Peptides are eluted from gels and even a small set of peptides from a protein provides sufficient information for identification(Yates, 2000) (Figure 6).

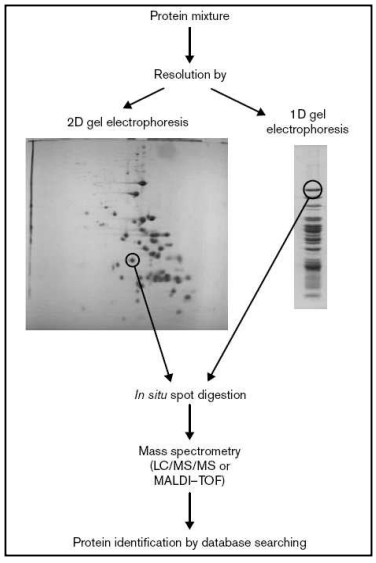


Figure 6: Schematic standard proteomic analysis of a protein mixture (Yates, 2000).

Mass spectrometric measurements are carried out in the gas phase on ionized analytes. By definition, a mass spectrometer consists of an ion source, a mass analyzer that measures the mass-to-charge ratio (*m*/*z*) of the ionized analytes and a detector that registers the number of ions at each *m*/*z* value(Aebersold and Mann, 2003) (Figure 7). Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), are the two techniques most commonly used to volatize and ionize the proteins or peptides for mass spectrometric

analysis(Aebersold and Mann, 2003; Pandey and Mann, 2000). ESI ionizes the analytes out of a solution and is therefore readily coupled to liquid-based (for example, chromatographic) separation tools (Figure 7a). MALDI sublimates and ionizes the samples out of a dry, crystalline matrix via laser pulses (Figure 7b). MALDI-TOF/MS is normally used to analyze relatively simple peptide mixtures, whereas integrated liquid-chromatography ESI-MS systems (LC-MS) are preferred for the analysis of complex samples.

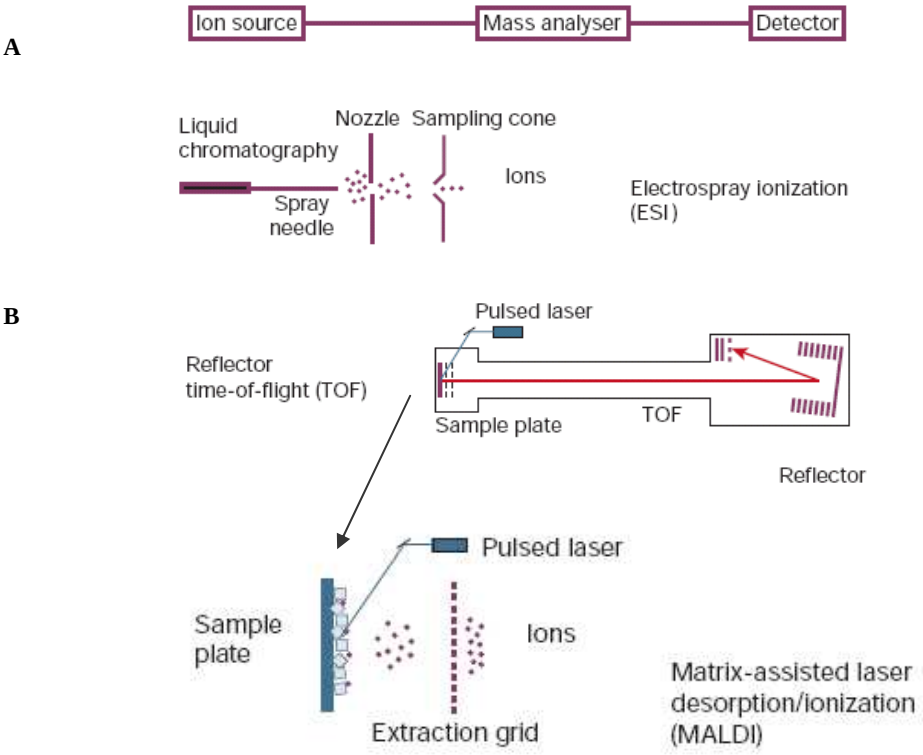


Figure 7: Common mass spectrometers used in proteome research. Ionization and sample introduction process in **A**: electrospray ionization (ESI), and **B**: matrix-assisted laser desorption/ionization (MALDI) (Aebersold and Mann, 2003)

Two main approaches for mass spectrometric protein identification are used (Figure 8). In the 'peptide-mass mapping' approach (Henzel *et al.*, 1993), the mass spectrum from the eluted peptide mixture is acquired, which results in a 'peptide-mass fingerprint'(PMF), of the protein being studied (Figure 8b). This mass spectrum could be obtained by a MALDI mass spectrometer, which results in a time-of-flight (TOF) distribution of the peptides comprising the mixture(Lay, 2001).

In a "two-step" procedure for unambiguous protein identification, MALDI fingerprinting is the first step. The second step relies on fragmentation of selected peptides in the mixture to get sequence information. For that, when a MADLI-TOF is used, a post-source decay

(PSD) analysis of selected peptides sequencing can be done (Gevaert *et al.*, 2001). Alternatively, when ESI mass spectrometer is used, selected peptides are ionized by 'electrospray ionization' directly from the liquid phase. The peptide ions are sprayed into a 'tandem mass spectrometer' which resolve peptides in a mixture and dissociate it into amino- or carboxy-terminal-containing fragments (Figure 8c). The main advantage of the second approach is that sequence information derived from several peptides is much more specific for protein identification than a list of peptide masses. Moreover, data obtained can be used to search both protein and nucleotide databases(Mann *et al.*, 2001).

5.1.2. Identification of post-translational modifications.

Post-translation modifications (PTMs) are covalent processing event that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more aminoacids (Mann and Jensen, 2003). PTMs events include, phosphorylation, glycosylation and sulphation, as well as many others, are difficult or impossible to be predicted from the sequences alone. Moreover, they are extremely important for protein structure and function as they can determine activity, stability, localization and turnover (Kuster and Mann, 1998; Wilkins *et al.*, 1999). One of the main and unique features of proteomics studies is the ability to systematically analyze post-translational processing and modifications of proteins. An example of this approach it is shown in Figure 8d.

5.1.3. Differential-display proteomics.

This approach is based on the identification of specific proteins, which are expressed in a particular condition, using a combination of 2D-gels electrophoresis, image analysis comparison software and mass spectrometry for protein identification (Figure 9). Currently, proteins from cells derived from two different conditions (Figure 9, A and B), are solubilized using appropriate methods. The protein mixture is then applied on a 'first dimension' run. After that, the strips or capillar gels, depending of the method employed, are subjected to reduction and alkylation and applied to a 'second dimension' SDS–PAGE gel run. The gels are then fixed and the proteins visualized by staining. The resulting protein patterns are recorded, quantified and compared using appropriate software, remaining one of the most labour-intensive parts of the two dimensional gel approach. Spots of interest are then excised and subjected to mass spectrometric analysis for its identification. Recently an improvement of this approach has been developed with the aim to reduce the number of gels and to facilitate the image analysis process (Tonge *et al.*, 2001). Two-dimensional difference gel electrophoresis (2D-DIGE) enables multiple protein extracts to be separated on the same 2D gel. This is made possible by labeling of

each extract using spectrally resolvable, size and charge-matched fluorescent dyes known as CyDye DIGE fluors (Marouga *et al.*, 2005).

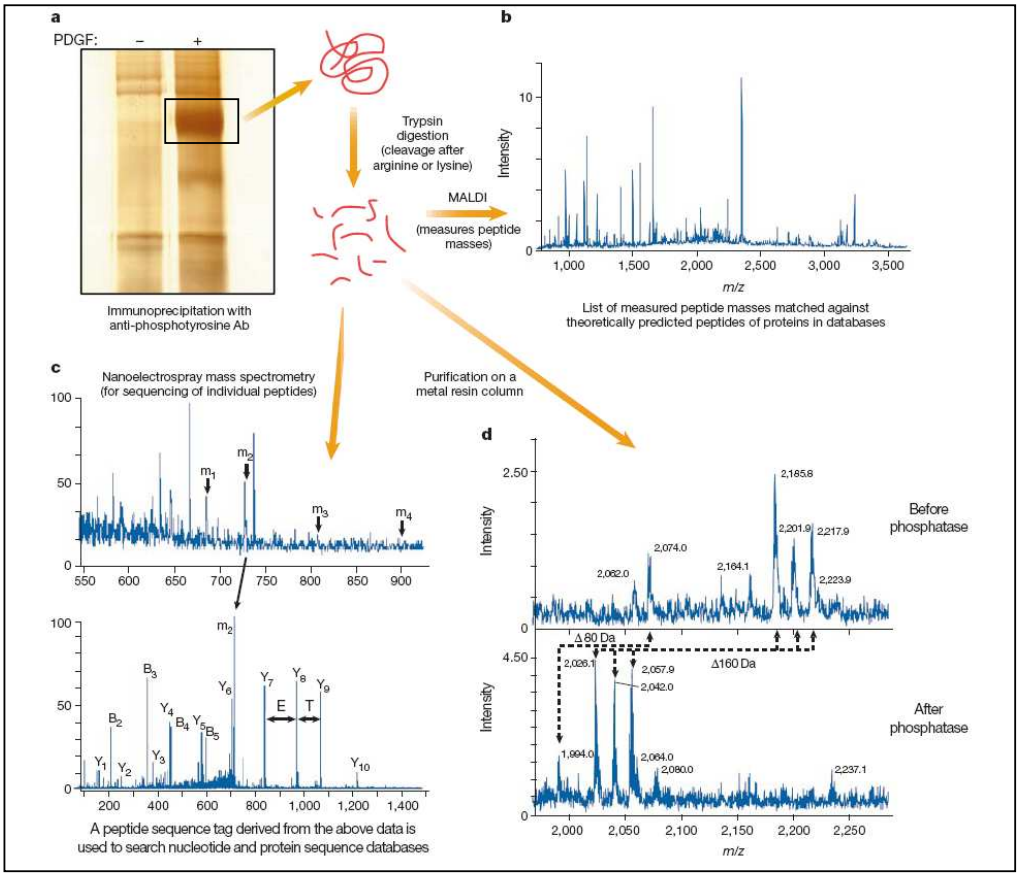


Figure 8: Schematic strategy for mass spectrometric identification of proteins, peptide sequencing and post-translational modifications studies. In this example a NIH 3T3 fibroblasts in response to a platelet growth factor (PDGF), was used as a sample study**A:** Responsive fibroblast cells are treated with PDGF, lysed and purified by immunoprecipitation. **B:** A band of interest is *in gel* digested and an aliquot of the peptide mixture, is analyzed by MALDI. **C:** An aliquot of the reminder peptide mixture is desalted and analyzed by ESI tandem mass spectrometry. The top panel shows the individual peptide peaks in the mass spectrum. The bottom panel shows how sequence can be derived by fragmentation of the chosen peptide (m2). **D**: Post-translational modification identification on a phosphopeptides enriched peptide mixture. Resulting peptides can be analyzed by MALDI before and after treatment with alkaline phosphatase. The panel shows a singly phosphorylated (showing a shift of 80 Da) and a doubly phosphorylated (showing a shift of 160 Da) peptide in the MALDI spectrum (Pandey and Mann, 2000)**.**

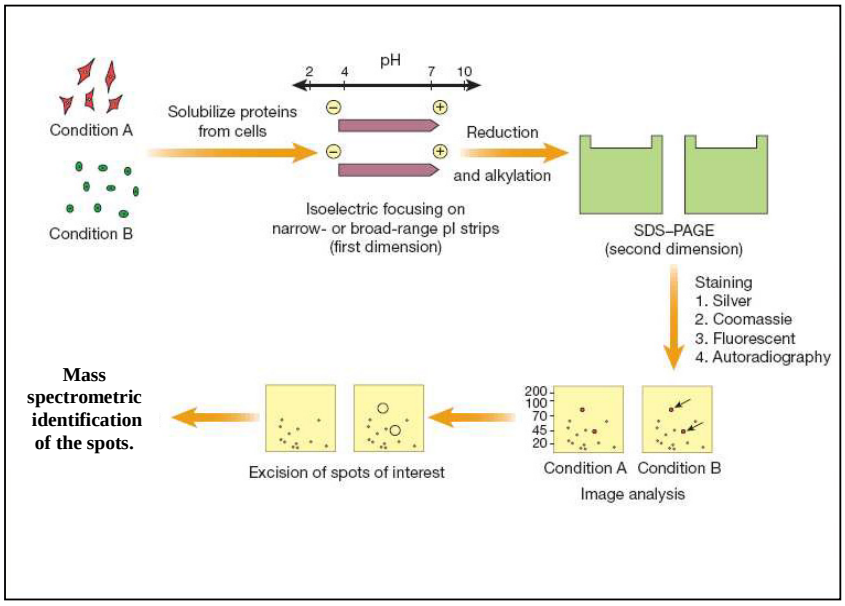


Figure 9: Schematic draw showing the 2D-gel differential display approach (Pandey and Mann, 2000).

5.2. Proteomics as a tool of plant-microbe interactions studies.

Proteomics serve as an ideal tool for the dissection of plant-microbe interactions. Firstly, it provides a wide overview of the proteins produced by both partners during their constant signal exchange. In particular, it allows the study of gene knockout effects within a protein network, as well as for the study of specific growth states. Secondly, it allows the detection of signal transduction pathways by following phosphorylation changes of proteins which are important for protein function (Rolfe *et al.*, 2003).

Although 2D-PAGE electrophoresis had been used quite early to study plant-microbe interactions e.g. mycorrhizal and rhizobial symbiosis, very few proteins have been identified, due to the limitation of both, electrophoretic and identification methods (Bestel-Corre *et al.*, 2004b). A pre-requisite for protein identification using a proteomic approach is the accessibility to genome sequence of the selected organism. Since the genome sequence of several symbiotic bacteria (Galibert *et al.*, 2001; Kaneko *et al.*, 2000; Kaneko *et al.*, 2002; Young *et al.*, 2006), as well as expressed sequence tag (EST) from their plants partners (Colebatch *et al.*, 2002a; Goff *et al.*, 2002; Yu *et al.*, 2002) have been published, new projects have been directed with the aim to establish models for studies and to conduct proteomics analysis. Recent publications have shown that major advances have been made in the microbial partner (Djordjevic *et al.*, 2003; Djordjevic, 2004; Guerreiro *et al.*, 1999; Natera *et al.*, 2000; Panter *et al.*, 2000), because of its ease of culturing and for the fact that it is a single cell, as compared to complex multicellular organism plant partner (Bestel-Corre *et al.*, 2002; Mathesius *et al.*, 2001; Mathesius *et al.*, 2002). These works mainly

focus so far, on the new discovery of novel proteins involved in symbiosis, some of their post-translational modifications, identification of specific isoforms of such proteins involved in specific pathways and finally to construct of biochemical pathways in which these novel proteins play their regulatory role (Bestel-Corre *et al.*, 2004b; Rolfe *et al.*, 2003).

6. *Azoarcus* sp. BH72, model bacterium for diazotrophic endophytes.

6.1. Taxonomy.

In the arid Punjab region of Pakistan, Kallar grass, a C4 plant (*Leptochloa fusca* (L.) Kunth), was introduced as a pioneer plant to improve the salt-affected soils of this region (Sandhu and Malik, 1975). Kallar grass has a high tolerance for waterlogged conditions, soil salinity and alkalinity (Khan, 1966). It's robust growth in those low-fertility soil conditions, together with the detection of acetylene reduction activity in its rhizosphere (Malik *et al.*, 1980), indicated a possible nitrogen fixing bacterium-plant association. An estimation of the diazotrophic population in the endorhizosphere and on the rhizoplane of Kallar Grass, as wells as in the nonrhizosphere soil, shows that the rhizoplane has 35 fold higher amount of bacteria than the nonrhizospheric soil (Reinhold *et al.*, 1986). *Azospirillum* species were dominant in the rhizoplane, while the endorhizosphere interestingly was dominated by a yellow-pigmented, motile, straight rod, unidentified diazotrophic bacterium (Reinhold *et al.*, 1986) which was later assigned to the genus *Azoarcus* spp. (Reinhold-Hurek *et al.*, 1993b) (Figure 10). Using rDNA-DNA homology and 16S rDNA sequences analysis, these isolates were grouped in a new separate rRNA branch in the β-subdivision of *Proteobacteria;* the genus *Azoarcus* (Hurek *et al.*, 1993; Reinhold-Hurek *et al.*, 1993b). From this group, two named species *A. indigens* and *A. communis* were proposed (Reinhold-Hurek *et al.*, 1993b). Strain BH72 was included into the genus *Azoarcus* but due to the lack of additional strains of this species it remained unnamed (Reinhold-Hurek *et al.*, 1993b). Three other related groups of diazotrophs have been recently placed under new genera, *Azovibrio restrictus*, *Azospira oryzae* and *Azonexus fungiphilus* (Reinhold-Hurek and Hurek, 2000). The plant-associated species of these genera posses several common physiological features, such as inability to use carbohydrate for growth, optimum temperature range from 37° to 42°C, generation time of approximately two hours, and growth on various organic acids or ethanol (Reinhold-Hurek *et al.*, 1993b). Additionally they share also their inability to be isolated from root-free soils, in contrast to *Azospirillum* spp. or rhizobia. Thus, they appear to be tightly associated with and ecologically dependent on plant (Reinhold-Hurek and Hurek, 1998a), fact observed in

other diazotrophic grass endophytes like *Herbaspirullum seropedicae* (James and Olivares, 1998). These species significativelly differ in this feature from additionally species of *Azoarcus*, which are mostly localized in a separated clade within the genus *Azoarcus* (Reinhold-Hurek and Hurek, 2000). These no plant-associated species include *Azoarcus toluliticus* (Zhou *et al.*, 1995) and *Azoarcus evansii* (Anders *et al.*, 1995) which were isolated from soils or sediments and are ables to degrade aromatic compounds. *Azoarcus anaerobicus* (Springer *et al.*, 1998) which is a respiratory but strictly anaerobic bacterium and strains degrading ethylenbenzene (Rabus and Widdel, 1995) or cyclohexane-1,2-diol (Harder, 1997). Their ocurred in anaerobic sediments and in sweage sludge indicates that their lifestyle *in situ* might be anaerobic (using nitrate as terminal electron acceptor) rather than microaerobic (fixing nitrogen and using O_2 as terminal electron acceptor) in contrast to the plant-associated strains. A genome sequence of *Azoarcus* sp. strain EbN1, a degrading ethylenbenzen strain within the soil-borne species, was recently published (Rabus *et al.*, 2005). To posses this information available will be of great advantage for genome comparison studies.

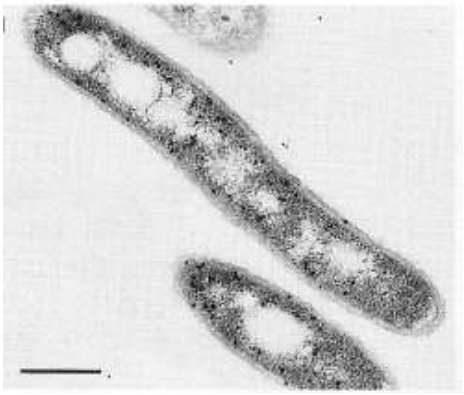


Figure 10: Transmission electron micrographs of *Azoarcus* sp. BH72. Bar 0.5 µm (Hurek *et al.*, 1995).

6.2. Plant infection and colonization.

In order to colonize the plant, some bacteria must found their way through cracks formed at the emergences of lateral roots or at the zone of elongation and differentiation of the roots (Rosenblueth, 2006). This is what was observed by microscopical studies, using immunological approaches and reported genes, in several grass endophytes including *Azoarcus* sp. BH72 (Hurek and Reinhold-Hurek, 2003) (Figure 11). Strain BH72 can colonize their original host (Kallar grass) and also rice seedlings, in a similar way, without causing symptoms of plant disease (Egener *et al.*, 1999; Hurek *et al.*, 1994b). Outer cell layers (exodermis, sclerenchyma) and the root cortex, are colonized inter- and intracellularly by *Azoarcus* sp. BH72 within 2-3 weeks. For this strain was found that the

aerenchyma, which forms in waterlogged plants being the main site for large micro colonies (Hurek *et al.*, 1991; Hurek *et al.*, 1994b) (Figure 11). Once inside the plants, strain BH72 can spread systematically and reach aerical tissues, most probably via xylema vessels (Hurek *et al.*, 1994b). Interestingly, even when root cells are colonized intracellularly, they are decaying or dead and the bacteria have not been observed to persist inside the living plant cytoplasm (Hurek *et al.*, 1994b). This feature of strain BH72 colonization is similar to plant pathogens, although the diazotrophic endophytes-grass interaction appears to be overall beneficial as in the rhizobial simbiosis. Additionally, no evidence for endosymbiosis in plant cells was detected for strain BH72 and in contrast to the rhizobia-legume nodule association (Reinhold-Hurek and Hurek, 1998a; Reinhold-Hurek and Hurek, 1998b). These atypically colonization of the model grass endophyte, suggests an intermediated types of interaction with plants, between plant-pathogens and plant-symbionts as rhizobia (Hurek and Reinhold-Hurek, 2003).

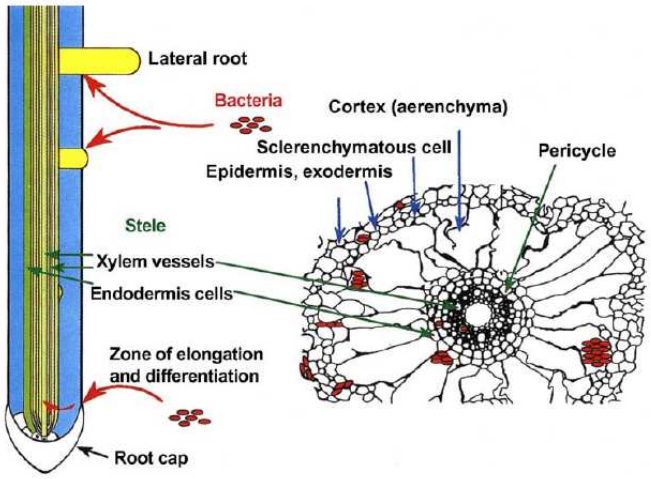


Figure 11: Possible sites of colonization and infection of diazotrophic endophytes in roots, shown in a sketch of a longitudinal (left) and a transversal (right) section of rice roots (Reinhold-Hurek and Hurek, 1998b).

Adherence of bacteria to the root host cells is one inicial step to entry into the plant in pathogens as wells as in symbiotic systems. Type IV pili are known to mediate attachment to host epithelia and to be a crucial virulence factor in human and animal pathogens (Bieber *et al.*, 1998) and may have a similar role in plant-microbe interactions. Strain BH72 possesses unusual short type IV pili, encoded by the *pilAB* locus which was demonstrated to be involved in the colonization of both, plant and fungal surfaces (Dörr *et al.*, 1998). This results indicates that strain BH72 may share some common traits in the initial steps of colonization with pathogenic bacteria.

The endophytes roots colonization could be an active process, which could be mediated by enzymes degrading plant cell wall polymers. In this sense two types of cellulolytic enzymes have been detected in *Azoarcus* sp. BH72 (Reinhold-Hurek *et al*., 1993a), from which an endoglucanase (EglA), was demonstrated to be involved in a rice root colonization (Reinhold-Hurek *et al.*, 2006). In contrast to most phytopathogens which possess cell wall-degrading enzymes, cellulose or its breakdown products do not support strain BH72 growth. Moreover, these enzymes -no common in pathogens- are not efficiently excreted into the culture supernatant, but remain bound to the cell surface (Reinhold-Hurek *et al.*, 1993a; Reinhold-Hurek *et al.*, 1993b). These features of the strain BH72 enzymes most probably play and important role in the host compatibility by not causing an aggressive attack to the plants cells.

On the other hand, recently it was demonstrated that the addition of Jasmonic acid (JA), a signal molecule important for stress responses and defense mechanisms in rice, decreased the physiologically successful colonization by the diazotroph *Azoarcus* sp., suggesting that plant defense responses might control endophytic ingression (Miche *et al.*, 2006). This was the first proteomic report on the response of roots of graminaceous plants towards colonization with endophytic bacteria.

6.3. Endophytic function.

Important points to understand the function of diazotrophic grass endophytes is the question if inside the plant there is a suitable microenvironment for nitrogen fixation and most crucial, whether the host plant profits from nitrogen fixation. As in other endophytic diazotrophs (Vande Broek *et al.*, 1996), in *Azoarcus* sp BH72 nitrogen fixation (Hurek *et al.*, 1987a) and expression of nitrogenase (*nif*) genes ocurred only in under microaerobic conditions and in absence of high concentration of of other nitrogen sources (Egener *et al.*, 1999). Additionally, using transcriptional fusions of *gusA* and *gfp* to the *nifH* gene, as well as labeled antibodies against the iron protein of nitrogenase (NifH), it was shown that in *Azoarcus* sp. BH72, *nif* genes were highly expressed in the root aerenchyma of Kallar grass (Hurek *et al.*, 1997b), as well as in the aerenchyma of young rice seedlings (Egener *et al.*, 1998). This results suggested that the interior of the rice roots provides a microenvironment suitable for N_2-fixation in gnotobiotic culture (Egener *et al.*, 1999; Reinhold-Hurek and Hurek, 1998b). They key question if the host plants profit from fixed nitrogen was also recently addressed for strain BH72 and its host Kallar grass. In N_2-deficient conditions, plant inoculated with the wild type strain BH72 (N_2-fixing test plants), grew better and accumulated more nitrogen than did plants inoculated with the

mutant strain (non-N_2-fixing control plants); total N-balance and natural ^{15}N abundance corroborated that fixed nitrogen was contributed. Moreover *nifH*-mRNA of strain BH72 was found to be predominant in plant roots (Hurek *et al.*, 2002). These results make the *Azoarcus* sp.-grass system a very interesting model system for a novel type of plant–microbe interaction.

6.4. Physiology and genetics of nitrogen fixation.

The understanding of regulatory cascades for N_2-fixation and assimilation is of importance due to the impact of N_2-fixation in the *Azoarcus* sp. BH72-grasses interaction. In strain BH72, as in most diazotrophs, the structural *nifHDK* genes of nitrogenase are localized in one operon (Hurek and Reinhold-Hurek, 1995). This operon is co-transcribed with a *nifY* and a ferredoxin gene which apparently is involved in the electron transport to the nitrogenase (Egener *et al.*, 2001). The essential transcription activator for the *nifHDK* operon NifA, as well as a NifL-like protein, which is usually present in *γ-Proteobacteria*, exist also in strain BH72 and are encoded in the *nifLA* operon (Egener *et al.*, 2002). Additionally, three different PII-like proteins (encoding by *glnB*, *glnK* and *glnY*), which are the central signal transmitter proteins in sensing the N-status of cells, are present in strain BH72 (Martin *et al.*, 2000). Interestingly, it has been reported that PII like proteins of strain BH72 have distinct roles for the physiological "switch off" and post-translational covalent modification of dinitrogenase reductase upon ammonium addition or anaerobiosis (Martin and Reinhold-Hurek, 2002).

In empirically optimized batch culture, at extremely low O_2 concentration (30nM), *Azoarcus* sp. BH72 cells can shift into a state of increased respiratory and N_2-fixation efficiency called "hyperinduction state" (Hurek *et al.*, 1994a; Hurek *et al.*, 1995). In the course of hyperinduction, strain BH72 cells form novel stack of intracytoplasmic membranes named "diazosomes" (Figure12). Diazosomes formation can be also induced reproducibly in the laboratory by co-cultivating strain BH72 with an ascomycete strain 2003 (related to *Acremonium alternatum*), which was isolated from the root interior of Kallar grass as well (Hurek and Reinhold-Hurek, 1999). This means that bacterial cells attached to the fungal mycelium and the fungal respiration may provide as well sufficient microerobic niches to shift to a hyperinduced state and form diazosomes. Interestingly, these structures are not present in cells fixing-nitrogen under standard conditions (e.g. around 2µM of dissolved oxygen) (Hurek *et al.*, 1995). Moreover, diazosome formation and N_2-fixation were abolished when 0.1% of NH_4Cl was added to the grothw medium or when mutants having a Nif^- phenotype were used in hyperinduced growths. These result demonstrated that the internal membrane stacks are functional linked to a effective nitrogen-fixation process

(Hurek *et al.*, 1995). Additionally, using labelled immunogold serum against the dinitrogenase reductase (NifH), was shown that the NifH protein was mainly located at the internal membrane stacks formed. Moreover, that the diazosomes formation coincides with the appearance of a NifH isoform of higher molecular mass which is membrane associated. These results added more evidence to the fact that most probably these specialized membranes are involved in efficient nitrogen fixation (Hurek *et al.*, 1995).

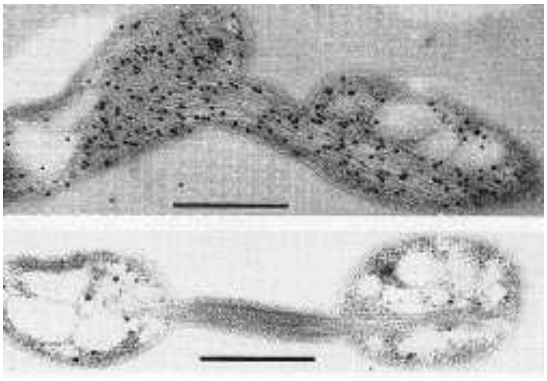


Figure 12: Transmission electron micrographs of the development of membrane stacks in nitrogen-fixing *Azoarcus* sp. BH72 in dual culture with *Acremonium alternatum*. In the upper picture, cells were labeled with an immunogold serum against dinitrogenase reductase. Bar 0.5 µm (Hurek *et al.*, 1995).

On the other hand, protein expression and uridylylation status of PII-like central regulatory proteins, suggest that this physiological state is indeed quite different from that of standard N_2-fixing conditions (Karg and Reinhold-Hurek, 1996; Martin *et al.*, 2000).

Complex internal membrane systems ocurr as specialized membranes in several, physiologically distint bacteria. These include phototropic bacteria, in where is located the photosynthetic apparatus, and chemolithtrotophics nitrifiying bacterias, as well as in methanotrophs, where appear to be involved in the methane oxidation (Hurek *et al.*, 1995). Since *Azoarcus* sp. is a chemoorganoheterotroph (Reinhold-Hurek *et al.*, 1993b), and do not share the same physiological requirements of those bacterias mentioned, it is not probable that the internal membranes in strain BH72 belong to any of the specialized types mentioned above (Hurek *et al.*, 1995). Intracytoplasmic membranes have also been described in *Azotobacter vinelandii* (Oppenheim and Marcus, 1970). They are vesicles originating from invaginations of the cytoplasmic membrane (Post *et al.*, 1982), nevertheless they do not exhibit the complexity of diazosomes in strain BH72 and moreover, their formation is regulated in a different way (Pate *et al.*, 1973; Post *et al.*, 1982). In this bacterium intracytoplasmic and cytoplasmic membranes seem to be largely identical (Post *et al.*, 1983), and nitrogenase was not found to be membrane associated (Robson, 1979). Also in *Franckia* sp., nitrogenase was located in the cytoplasm of vesicles and not on cytoplasmic membranes, by immunogold labeling (Meesters, 1987). This set of

evidence speaks in favor to a novel and unique type of specialized intracytoplasmic membrane in strain BH72.

7. Outline of this thesis

The aim of this thesis was to study an environmentally and biotechnologically relevant microorganism *Azoarcus* sp. strain BH72 from its complete genome sequence as a starting point, to its complementary proteome pattern under different growth conditions.
The two major goals of the first part of this work were the characterization of an *Azoarcus* sp. BH72 BAC library and the building of a physical map of strain BH72 chromosome (Chapter B). Insert size and restriction fragment length polymorphism (RFLP) determination, as well as end-sequencing of selected BAC clones were done with the aim to characterize a strain BH72 BAC library. A BAC minimal tiling path of strain BH72 chromosome was constructed using a combination of BAC-library screening by Southern hybridization and BAC clone end-sequencing. The BAC library was used for an independent analysis of the genome structure in comparison to a strain BH72 shotgun library with small insert sizes**.** Additionally, it was employed for contig assembly and gaps closure of the mentioned shotgun library. Functional classifications of BAC-end sequences which cover almost the entire genome were done. Assignment of these sequences to the genome assembly provided a high-density map of *Azoarcus* sp. BH72 genome. The physical map obtained was used for genome comparisons with the related *Azoarcus* sp. EbN1 genome sequence. The aim of this analysis was to found out if these two close related strains, which have different life-styles, share some genome similarities. Results obtained revealed a relatively low amount of syntheny between both mentioned strains, in agreement with their different life style. The second main objective of this thesis was the annotation and analysis of part of the genome sequence (Chapter C). In particular, genes which belong to the COG categories “Ion Transporters and Metabolism”, and “Carbohydrates Transport and Metabolism”, were studied (Chapter D). This analysis revealed several highlights in the genome sequence, which can be used as a starting point in future studies. In both COG categories analyzed, strain BH72 genome possesses the double amount of genes than the closest relative strain EbN1. Moreover, strain BH72 genome has one of the highest numbers of genes belonging to the “Ion transport and Metabolism” category in comparison to other genomes. Remarkable was the presence of a large number of genes coding for putative TonB-dependent outer membrane receptors, which at least doubled the number of genes described for *Azoarcus* sp. EbN1, and even more in comparison to other N_2-fixing symbionts studied. Strain BH72 has several putative receptors for the general classes of hydroxamates and catechol-type siderophores,

ferricitrate, vitamin B12 and colicins. However, no evidence for biosynthetic pathways for known hydroxamate or catecholate siderophores in its genome was found.

Finally, the main goal of the last part of this work was to carry out a proteomic analysis of *Azoarcus* sp. BH72 grown under N_2-fixing conditions, in pure- or in co-culture with *A. alternatum* strain 2003. The biological system used allows the identification of wilde-type membrane proteins differentially expressed in co-culture conditions probably related to diazosome structure or function. In this study a differential-display proteomic approach was used. With the aim to improve the solubilization and resolution of membrane proteins, a different technical approach, using capillar gels with isoelectric focused ampholites as a first dimension was done, instead of standard IEF strips. Image analysis shows strong differences in the membrane and cytoplasm proteomic patterns of strain BH72 N_2-fixing cells, grown in pure- or in co-culture conditions. In particulary, in both fractions most of the spots were down-regulated in co-culture conditions. These results reflect special adaptations in the bacterial metabolism to cope the metabolic requirements under the different conditions studied with a minor set of proteins. Proteins of interest were picked out from the gels and analyzed by MALDI-TOF mass spectrometry for its identification. MALDI-TOF protein identification and pathway reconstruction shows that *Azoarcus* sp. BH72 carbon metabolism in co-culture was adapted to utilize the exudates secreted by the fungus into the culture medium. This was reflected by the identification of up-regulated proteins involved in ethanol, acetate and lactate assimilation, compounds which were detected as fungus exudates. As expected nitrogen metabolism was very active in both conditions studied. Several key enzymes involved in the N_2-fixation process, as the nitrogenase complexes as well as enzymes related to its synthesis and activation were detected. Interestingly was the fact that proteins like NifH and NifM were identified down regulated in the cytoplasm and constitutively and up-regulated in the membrane fraction. These results give more evidence to the observation that the NifH is strongly associated to the membrane fraction in co-culture conditions, in particular to diazosomes, and moreover that these structures are involved in an active nitrogen fixation process.

An active protein synthesis, folding and processing, in both N_2-fixing conditions was reflected by the large number of proteins identified related to these process. The differential expression and localization of these proteins indicate a specialization of co-culture cells. Moreover, and in correlation with an intense N_2-fixation process, proteins involved in the respiratory chain and ATP synthesis were detected in the membrane fraction and up-regulated in co-culture conditions in concordance with the presence of diazosomes. Several proteins involved in the detoxification of reactive oxygen species were identified. This shows the important role of these types of proteins in actives N_2-

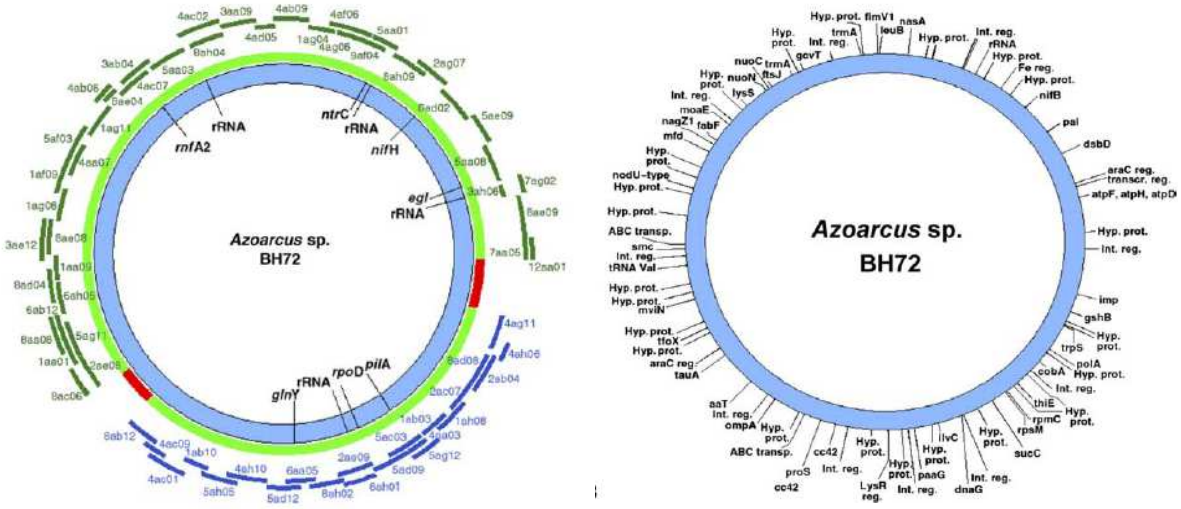
fixing cells, specially taking in mind that the nitrogenase complex is very sensitive to oxygen.

A large number of membrane proteins were identified using this proteomic approach. Three major constitutives spots were identified as porins, while isoformes from these proteins were detected up-regulated. This could indicate a probably role of these porins in bacteria-fungus interaction. A set of outer-membrane protein TonB-dependent were identified in both conditions. Although these proteins have high similarity with siderophores receptors, no evidence of siderophore production was detected in agreement with the lack of biosynthetic pathways for these compounds in the strain BH72 genome and with the iron-sufficient conditions used in both growth medium studied. Results obtained suggest that these receptors are involved in other process independently of iron aquisiton; most probably playing a role in the environmental signal perception, in bacteria-fungus interaction, or in biocontrol. The up-regulated identification of membrane protein involved in drug or antibiotic resistance is probably related to a bacterial response to fungal metabolites. Results suggest that this set of proteins play a crucial role in the establishment of a compatible interaction between *Azoarcus* sp. BH72 and the fungus partner.

The results obtained using this proteomic approach, contribute to a better understanding of the *Azoarcus* sp. BH72 physiology and ecology. Moreover the conditions-specific proteins identified here provide new targets for genome-wide mutagenesis to understand its role in *Azoarcus*-fungus and plant interaction.

CHAPTER B

Azoarcus sp. BH72 genome physical map



1. OBJECTIVES.

Within the *Azoarcus* sp. BH72 genome-sequencing project context, the objectives of this work were:

a) Characterization of an *Azoarcus* sp. BH72 bacterial artificial chromosome (BAC) library.

b) Construction of a physical map of *Azoarcus* sp. strain BH72 chromosome.

c) Contig assembly and validation of an *Azoarcus* sp.BH72 whole shot gun genome library.

d) Carry out genome comparison studies between *Azoarcus* spp. strains BH72 and EbN1.

2. CONTRIBUTION.

2.1. BAC library characterization: insert size and restriction fragment length polymorphism (RFLP) determination; end sequencing of selected BAC clones.

Randomly 82 BAC clones, from a BAC-library constructed previously (S. Reamon-Buettner), were *Dra*I digested and analyzed by pulse-field gel electrophoresis. In parallel, same clones were *Not*I digested, and products were loaded into agarose gels with the aim to obtain a RFLP. Additionally end-sequencing from those BAC clones was done. *In-gel* insert size and RFLP-patterns determined were compared to *in-silico* ones using the draft contig sequence of the *Azoarcus* sp. BH72 genome project and the end-sequences of BAC clones. This analysis was done using the pDRAW32 program. Additionally BAC-clone sequenceing was done by the cooperation patner in the University of Bielefeld (O. Kaiser).

2.2. BAC contig assembly of the *Azoarcus* sp. BH72 chromosome and validation of the chromosome assembly.

Genes known from previous molecular genetic functional studies on strain BH72 were used as “anchor probes” to screen the library by Southern hybridization for BAC clones carrying these genes. The ends of the inserts of positive clones were sequenced. These sequences were used to generate probes to screen for overlapping clones. Sequences were

aligned onto the draft genome sequence assembly by using Seqman in the DNAStar package. For gap closure, oligonucleotide and/or PCR probes were designed from the whole shotgun sequence assembly and used for hybridizing to high-density BAC-DNA filters. The arrangement of BAC-clones in a physical map was done in the University of Bielefeld with the programme BACCardi (D.Bartels), which is not freely available.

2.3 Functional characterization of BAC end-sequences and physical map construction.

The end-sequences of 109 BAC inserts that cover the entire genome were compared with the NCBI database using the BLASTX program. After removing low quality sequences, the hits that had a minimum of 100 bp sequence match (approx. 200) were manually evaluated and placed in different protein function categories. Assignment of these sequences to the genome assembly provided a high-density map of the *Azoarcus* genome.

2.4 Localization of *Azoarcus* sp. BH72 ribosomal RNA (*rrn*) operons.

To screen the BAC clones carrying 16S-rRNA genes and to assess their *Eco*RI restriction fragment pattern, different probes were used: the oligonucleotide probe TH15, specific for the genus *Azoarcus* (Hurek *et al.*, 1993), the general oligonucleotide probe 926f for *Bacteria* and *Archea* (Lane, 1991) and as well the gene probe TH3-5 (Hurek *et al.*, 1993).
To verify the genome assembly, *Eco*RI, *in-gel* size and RFLP of the 16S-rRNA clones were compared with *in-silico*16S-rRNA-probe hybridization and the size determination of each *Eco*RI fragment spanning the probe target region. In this analysis the contig draft sequence of the *Azoarcus* genome project and the pDRAW32 program were used.

2.5. Genome comparison.

The physical map obtained was compared with the gene arrangements in the genome of the soil strain *Azoarcus* sp. EbN1. For that, the relative position of known marker genes, including the 16S-rRNA genes in the strain BH72 genome were compared with the relative position in the genome of strain EbN1.
On the other hand, by using BLASTX analysis, the BAC-end sequence from the minimal tiling path obtained, were compared with the strain EbN1 genome. Additionally, the relative position of those BAC-end sequences as well as the insert sizes of the BAC clones, were compared with strain EbN1 genome.

3. RESULTS AND CONCLUSIONS.

3.1. A constructed *Azoarcus* sp. BH72 genome BAC-library was characterized. This library consist of 1176 clones which covers 23-fold the entire genome. The average inserts size was 90 kDa (ranging from 25 to 150 kDa), a value that correlated well with the *in-silico* average insert size analysis (81 kDa).

3.2. By assembly of BAC clones into contigs, using a combination of BAC-library screening and BAC clone end-sequencing, it was demonstrate that the draft assembly of the whole shotgun sequence into a unique contig was correct. Neverthelss, two gaps were not closed by screening the entire library. This indicate that certain regions appeared to be underrepresented, either due to problems in cloning of particular fragments, or to statistical reasons as would be expected for a restriction enzyme-based approach to DNA fragmentation.

3.3. Insert size and RFLP digestion comparison analysis of BAC clones, which cover almost the entire chromosome, additionally validated the genome assembly from the shotgun library. An example for eight randomly BAC clones, *in-silico* and *in-gel Not*I RFLP digestion comparison, is shown in Figure 1.

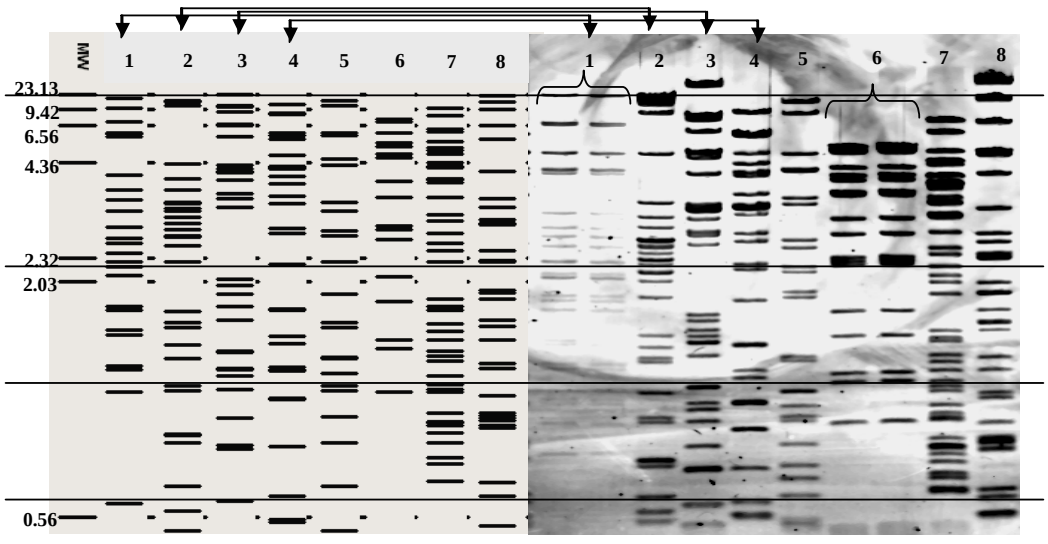


Figure 1: *In- silico* (left) and *in-gel* (right) *Not*I RFLP digestion of eight BAC clones constructed from *Azoarcus* sp. BH72 DNA. Molecular marker (lambda DNA digested with *Pst*I) is shows on the left.

3.4. Functional classification of BAC-end sequences from 109 BAC inserts which cover almost the entire genome were done. Assignment of these sequences to the genome assembly provided a high-density map of the *Azoarcus* sp. BH72 genome.

3.5. Using specific and general probes targeting different positions in the *Azoarcus* sp. strain BH72 16S-rRNA sequence, four clusters of BAC clones were identified that contain four different *rrn* operons according to RFLP. This analysis shows that *rrn* operons were not physically linked and well represented in the BAC library. Additionally, these results provided a correct assembly of the draft *Azoarcus* genome sequence in the region of *rrn* operons.

3.6. Comparison of the strain BH72 physical map with the gene arrangements in the genome of the soil strain *Azoarcus* sp. EbN1 revealed a relatively low amount of syntheny. This was reflected by a different relative position of the strain BH72 known marker genes in the strain EbN1 genome. Moreover, comparison of the BAC-end sequences with the genome of *Azoarcus* sp. EbN1 revealed considerable differences. Only 81% of the analyzed genes had homologues in the genome of strain EbN1. Out of those, only two third had highest hits in BLAST analyses with genes of strain EbN1, while one third had highest similarities to genes of other bacteria.

The entire analysis approach used in this study demonstrates the strength of the BAC technology to control the correct assembly in genome projects. Additionally, its utility in genome comparison analysis is demonstrated.

4. OUTLOOKS

4.1. The BAC library and the physical map obtained will provide the base for cloning and analysis of important genes. Operons of genetic pathways related to bacteria-plant interactions or of metabolic pathways are essential platforms for functional genomic analysis.

ELSEVIER

FEMS Microbiology Letters 249 (2005) 233–240

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Physical map of the *Azoarcus* sp. strain BH72 genome based on a bacterial artificial chromosome library as a platform for genome sequencing and functional analysis

Federico Battistoni [a], Daniela Bartels [b], Olaf Kaiser [c], Stella Marie Reamon-Buettner [a,1], Thomas Hurek [a], Barbara Reinhold-Hurek [a,*]

[a] *Laboratory of General Microbiology, University Bremen, P.O. Box 33 04 40, D-28334 Bremen, Germany*
[b] *Center for Biotechnology (CeBiTec), Bielefeld University, D-33594 Bielefeld, Germany*
[c] *Department of Genetics, Bielefeld University, D-33594 Bielefeld, Germany*

Received 1 May 2005; received in revised form 5 June 2005; accepted 9 June 2005

First published online 29 June 2005

Edited by M.Y. Galperin

Abstract

Azoarcus sp. strain BH72 is a Gram-negative proteobacterium of the β subclass; it is a diazotrophic endophyte of graminaceous plants and can provide significant amounts of fixed nitrogen to its host plant Kallar grass. We aimed to obtain a physical map of the *Azoarcus* sp. strain BH72 chromosome to be directly used in functional analysis and as a part of an *Azoarcus* sp. BH72 genome project. A bacterial artificial chromosome (BAC) library was constructed and analysed. A representative physical map with a high density of marker genes was developed in which 64 aligned BAC clones covered almost the entire genome.

Keywords: *Azoarcus*; BAC library; Physical map; Genome sequencing; Functional genomics

1. Introduction

Azoarcus sp. strain BH72 is a Gram-negative proteobacterium of the β subclass, isolated from surface-sterilised roots of Kallar grass (*Leptochloa fusca* L. Kunth) which is used as a pioneer plant on salt-affected, low fertility-soils in the Punjab of Pakistan [1,2]. It is regarded as a diazotrophic endophyte of graminaceous plants, a bacterial group sharing pathways of infection and colonisation patterns of grass and cereal roots [3–5]. Infection occurs at the emergence points of lateral roots and in the zone of elongation and differentiation above the root tip. The major colonisation sites are the outer root cell layers and the root cortex or aerenchyma in flood-tolerant plants; there is no evidence for an endosymbiosis in living plant cells as in the legume symbiosis [3,6]. Several other species of the genus *Azoarcus* have been isolated from soils that appear differ ecologically in being restricted to soil or sediments as habitats in contrast to plants, moreover they are localised on a phylogenetically distinct 16S rDNA branch [6–8]. The genome of one member of this clade, *Azoarcus* sp. EbN1, has recently been sequenced [9].

* Corresponding author. Tel.: +49 421 218 2370; fax: +49 421 218 9058.
E-mail address: breinhold@uni-bremen.de (B. Reinhold-Hurek).
[1] Present address: Drug Research and Medical Biotechnology, Fraunhofer Institute of Toxicology, Experimental Medicine, Nikolai Fuchs Strasse 1, 30625 Hannover, Germany.

0378-1097/$22.00
doi:10.1016/j.femsle.2005.06.015

the agarose by electroelution as was described by Strong et al. [26].

2.4. Cloning of DNA fragments into the BAC vector

The vector pBeloBAC vector was prepared as described by Shizuya et al. [20]. Ligations were performed in 150 μl reaction mixture using between 10–20 ng of vector and 250 ng of insert DNA for 16 h at 16 °C. After heat inactivation and dialysis against 10% PEG 600 for 90 min at room temperature, strain DH10B (ElectroMAX™ DH10B™ cells, Life Technology) was transformed by electroporation (3 μl of ligation mix). The BAC clone library was stored in microtiter plates at −70 °C.

2.5. Analysis of insert size

For BAC DNA preparation, cells from 30 ml LB medium were resuspended in 3 ml of Qiagen Buffer P1 (50 mM Tris–HCl, pH 8; 10 mM EDTA; 100 μg ml^{-1} RNase A), and plasmid DNA extracted by alkaline lysis. After two phenol–chloroform extractions, DNA was precipitated with isopropanol. Insert sizes were determined after restriction digestion of 1 μg of DNA with *Dra*I (NEB, New England Biolabs). Conditions for separation in a 1% agarose gel by PFGE were 6 V cm^{-1}, 5–15 s linear ramp, 12 °C, running time 14 h, included angle 120°. Low-range PFGE marker (NEB) was used as size standard. Restriction fragment analysis was carried out after restriction digestion with *Not*I by conventional agarose gel electrophoresis. In silico restriction analysis was carried out using the program pDRAW32 (http://www.acaclone.com/).

2.6. Screening of the BAC-library

High-density BAC DNA filters for hybridisation-based screening of the library were prepared using the Perfectprep® BAC-96 kit (Eppendorf). For southern blotting by standard procedures [27], digoxygenin-labelled gene probes were generated by PCR with primers designed for specific amplification of DNA fragments close to the ends of BAC insert sequences, and hybridised at conditions of high stringency. For oligonucleotide probes labelled with digoxygenin at the 5′ end, hybridisation was carried out in SET buffer [28].

Genes used for anchor probes were 16S ribosomal DNA (Acc. No. L15530) [28], *nifH* encoding the iron protein of nitrogenase (AA635586) [29], *glnY* encoding a PII-like protein (AF28105) [30], *pilA* encoding a type IV pilin (AF031954) [31], *rpoD* encoding a general sigma factor upstream of a type I intron [32], *eglA* encoding an endoglucanase (Reinhold-Hurek et al., submitted), and *ntrC* and *rnfA2* encoding a global regulator of N-assimilation and an electron transport complex protein (Sarkar et al., unpublished).

2.7. Sequence analysis of BAC inserts

Sequencing was carried out with 10 μg BAC DNA template, using the ALFExpress automated sequencer as outlined previously [33], with slight modifications of the thermocycler settings (initial denaturation at 97 °C for 50 s, denaturation step at 95 °C for 2 min, annealing/extension at 60 °C for 30 s, 30 cycles). Sequencing primers were were T7belocy5 (5′-GTAATACGACTCACTATAGGGCG-3′) and Sp6belocy5 (5′-GCTATTTAGGTGACACTATAGAATAC-3′). Additional sequencing of BAC-clones was carried out by Integrated Genomics GmbH (Jena, Germany) using ABI 3100 capillary sequencer systems (ABI, Weiterstadt, Germany).

High quality end-trimmed BAC end-sequences (defined as those having >100 bp insert sequences with at least phred20 quality) were used as queries in BLASTx searches against the non-redundant protein database of NCBI (National Centre of Biotechnology Information).

2.8. Assembly of BAC contigues

A minimal tiling path of BACs was constructed in a two-step approach by applying the software tool BACCardI [34]. In a first step all available BACs were automatically mapped onto the *Azoarcus* sp. BH72 genome by using their end sequences. In a second step, BACs of special interest were manually selected and marked. Subsequently, to establish a minimal tilling path these preselected BACs were joined by manually selecting additional BACs from the pool of mapped BACs.

3. Results and discussion

3.1. Library construction and characterisation

This work comprises at least two major goals, the establishment and characterisation of a BAC library and the building of a representative physical map for the genome of the diazotrophic endophyte *Azoarcus* sp. BH72, which is capable of colonising the interior of rice roots, one of the globally most important crops. A BAC library was constructed for *Azoarcus* strain BH72 that is suitable for physical mapping and subcloning of operons of interest. The library consists of 1176 clones. All of the 192 clones randomly sampled from the library contained inserts, thus the number of empty vectors was negligible. Size determination of 82 selected BAC clones from *Dra*I cleavage indicated an average insert size of 90 kb (ranging from 25 to 150 kb) by PFGE (Fig. 1A). This is comparable to other prokaryotic genomes, in particular those with similar GC content and similar genome size such as *Mycobacterium* [35,36].

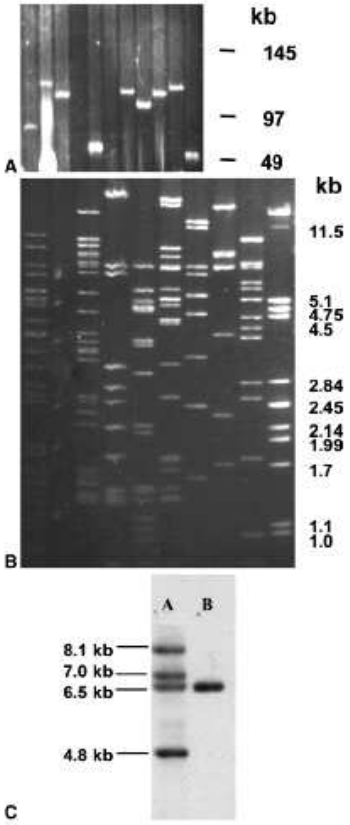


Fig. 1. Determination of insert size and restriction endonuclease patterns of BAC clones constructed from *Azoarcus* sp. strain BH72 DNA. (A) Nine randomly selected BAC clones digested with *Dra*I to assess the insert size. Position of molecular marker (NEB) shown on the right. (B) Eighteen randomly selected BAC clones digested with *Not*I to assess the restriction fragment length polymorphism. Molecular marker (lambda DNA digested with *Pst*I) is shows on the right. (C) Southern hybridisation of genomic DNA of strain BH72 (lane a) and BAC clone 2ae9 (lane b) with oligonucleotide probe TH15 which is specific for *Azoarcus* sp. 16S rRNA genes. Fragment sizes given in kb.

Restriction endonuclease patterns using *Not*I showed a high variation of patterns in randomly selected clones (Fig. 1B), suggesting that the library contained high numbers of independent clones. The genome of *Azoarcus* sp. BH72 was estimated to be 4.6 MB in size by the thermal renaturation method [2]. Therefore, a minimal set of 51 BAC clones would be necessary to cover the entire chromosome, suggesting that the library had 23-fold coverage of the entire genome.

3.2. BAC contig assembly of the Azoarcus *sp. BH72 chromosome and validation of the chromosome assembly*

Assembly of BAC clones into contigs was done by a combination of BAC-library screening, BAC clone end-sequencing and comparison with a draft assembly from shotgun sequence data of the *Azoarcus* sp. BH72 genome project [16]. After shotgun clone sequencing with 8-fold coverage of the *Azoarcus* genome, this draft assembly contained 68 contigs with at least 10 reads each. Genes known from previous studies on strain BH72 were used as "anchor probes" to screen the BAC library by Southern hybridisation for clones carrying these genes. Sequences obtained for the ends of these inserts were used to generate probes to screen for overlapping clones. In addition, randomly picked BAC clones were also subjected to end-sequencing. Sequences were aligned onto the draft genome sequence assembly by using BACCardI (Fig. 2) and Seqman in the DNA-Star package. For gap closure, oligonucleotide probes were designed from the sequence assembly and used for hybridising to high-density BAC-DNA filters. Positive signals were obtained for all the markers used in the hybridisations with an average of 2 hits per filter per probe, ranging from 1 to 5, indicating the presence of preferentially cloned regions in the chromosome. In total, the ends of 443 BAC clones were sequenced in both forward and reverse directions and were used to build two contigs covering almost the entire genome (Fig. 2). Two gaps were not closed by screening of the entire BAC library, which indicated that certain regions appeared to be underrepresented, either due to problems in cloning of particular fragments, or to statistical reasons as would be expected for a restriction enzyme-based approach to DNA fragmentation.

BAC clones were used for additional validation of the genome assembly from the shotgun library. In silico

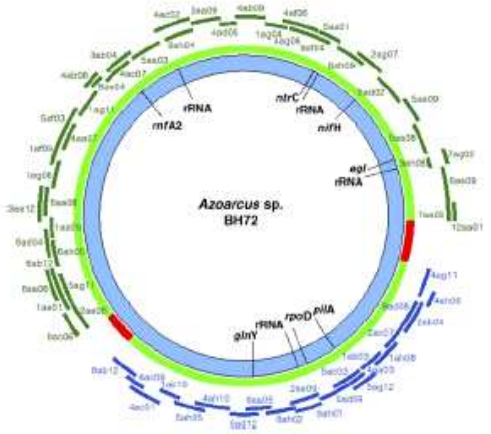


Fig. 2. Minimal set of BAC clones covering the *Azoarcus* sp. BH72 chromosome in two contigs. Initially known marker genes used to orient and anchor the contigs are indicated inside of the replicon. The position and identity numbers of BAC clones are given outside of the replicon. Inner blue circle represents the chromosome. Adjacent circle indicates regions covered by BAC clones (green) or gaps not covered by BACs in red.

insert size determination, based on the draft contig sequence of the *Azoarcus* sp. BH72 genome project and the end-sequences of BAC clones, showed an average insert size of 81 kb, a value that correlated well with the *in gel* analysis (90 kb). Selected BAC clones covering almost the entire chromosome (Fig. 2) were subjected to RFLP analysis using the restriction endonuclease *Not*I. The *in gel* size determination was compared to in silico size determination using the program pDRAW32. In visual inspections of the restriction patterns, they correlated very well. Additionally, restriction fragment sizes were determined from gels and in silico digests, and were considered to be in agreement when at least 60% of the fragments were identical. 100% of the analysed clones showed agreement with the draft genome assembly according to these criteria.

This provides strong evidence for a correct assembly of the genome sequence from the shotgun library and supports the observation that the genome of strain BH72 consists only of one replicon. It also demonstrates the strength of the BAC technology to control the correct assembly in genome projects. The possibility of subcloning of large DNA fragments from BAC clones will provide the base for cloning and analysis of important genes and whole operons or genetic pathway related to with bacteria-plant interactions or metabolic pathways, as well as an essential platform for functional genomic analysis.

3.3. Functional characterisation of BAC end-sequences

The end sequences of 109 BAC inserts that cover the entire genome were compared with the NCBI database using the BLASTX program. After removing low quality sequences the hits that had a minimum 100 bp sequence match at an *e*-value of at least 1×10^{-10} (approx. 200) were manually evaluated and placed in different protein function categories (Fig. 3A). The two largest groups in this analysis, 29% or 27%, were similar to hypothetical proteins or metabolism-related proteins, respectively. Most of the hypothetical proteins showed highest similarity to proteins of bacteria belonging to the β- or γ-subgroup of *Proteobacteria* (37% and 24%, respectively), as to be expected according to the taxonomic affiliation of *Azoarcus* sp. [2]; in particular similarities were found with the soil-borne species *Azoarcus* sp. EbN1 and plant-pathogenic bacteria, suggesting that *Azoarcus* sp. BH72 may share some unknown functions with bacteria of this life style [16]. A similar percentage (30%) was obtained in an analysis carried out for the nitrogen fixing symbiont *Bradyrhizobium japonicum* [19]. As in *Azoarcus* sp. BH72, also in *B. japonicum* a large group of proteins included hits to proteins related to bacterial metabolism (41%) [19]. Cell signalling and translation, or cell structure/membrane proteins comprised 7% or 6%, respectively. Another 5% of the sequences were

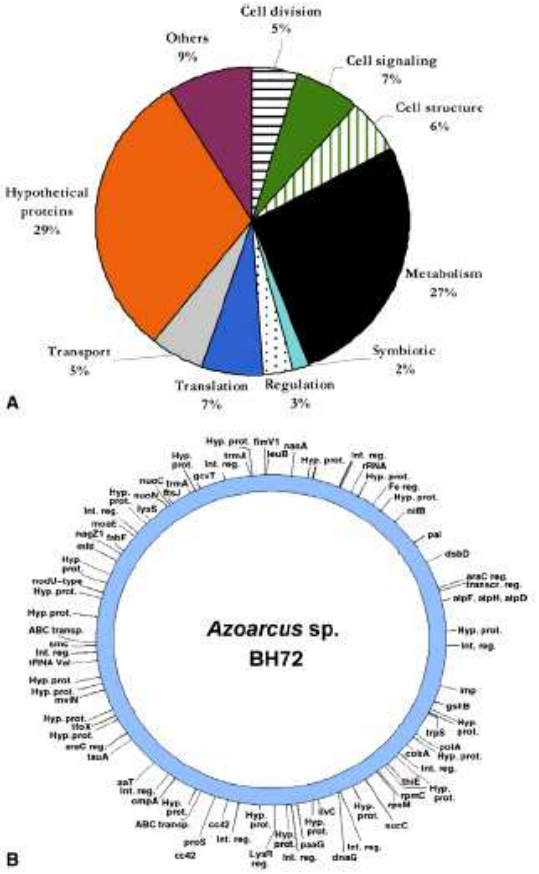


Fig. 3. (A) Functional characterisation of the best BLASTX hits obtained for the end-sequences of BAC clones, distribution into functional categories given in %. (B) Representative physical map of *Azoarcus* sp. BH72 developed from data obtained by pulse field electrophoresis gels, southern hybridisation and BAC-end sequencing. Positions of marker genes shown outside of the replicon.

associated with transport/binding proteins or cell division/DNA replication and repair. Finally, transcriptional and translational regulation, symbiotic and unclassified proteins were represented by 3%, 2% and 9% of the sequences. As can be seen from the localisation in the physical map (Fig. 2), marker genes related to nitrogen metabolism and nitrogen fixation such as *nifH*, *ntrC*, *glnY* or *rnfA2* are distributed in the chromosome and not necessarily physically linked.

Comparison of the BAC-end sequences with the recently published genome of a soil-borne species *Azoarcus* sp. EbN1 [9] revealed considerable differences. Only 81% of the analysed genes had homologues in the genome of EbN1. Out of those, only two third had highest hits in BLAST analyses with genes of EbN1, while one third had highest similarities to genes of other

Table 1
Localisation of *Azoarcus* sp. BH72 ribosomal RNA (*rrn*) operons

Operons	In silico/in gel *Eco*RI fragments for probe[a] (kb)			BAC clones[b]
	TH15	926-F	TH3-5	
1	6.6/6.5	21/20	21.4 + 6.7/20 + 6.5	2ae9, 6ah1
2	8.2/8.1	10/10	8.2 + 10/8.0 + 11	4ac2
3	7.0/7.0	9.0/9.0	7.0 + 9.0/7.0 + 9.0	2ab9, 1ae2
4	4.8/4.8	5.5/5.2	4.8 + 5.5/4.8 + 5.5	3ah6, 9ae9

[a] Probes for Southern blot hybridisation were oligonucleotide probes targeting *Azoarcus* 16S rDNA (TH15) or bacterial 16S rDNA (926-f), or a larger 16S rDNA gene probe derived by PCR from genomic DNA of strain BH72 as template (TH3-5). Expected results from in silico digestions given left to hybridisation results.
[b] Numbers of BAC clone numbers that carry the respective restriction fragments.

bacteria. Important marker genes such as nitrogenase genes (*nifHDK*) were not present in EbN1, which is not diazotrophic.

Assignment of these sequences to the genome assembly provided a high-density map with marker genes of the *Azoarcus* genome (Fig. 3B). Comparison of this map with the gene arrangements in the genome of the soil strain *Azoarcus* EbN1 revealed a relatively low amount of syntheny. Only five out of the BAC clones shown in Fig. 2 had insert sizes corresponding approximately to distances of genes found to be similar in strain EbN1 to their end-sequences, while most BAC-end sequences had counterparts more than 100 kb apart in the EbN1 genome. Also the relative position of the initially known marker genes (Fig. 2) in the genome was different in strain EbN1.

3.4. Localisation of Azoarcus sp. BH72 ribosomal RNA (rrn) operons

16S rRNA genes often occur in multiple copies in the chromosome of bacteria and may be difficult to assemble from shotgun libraries. To screen for BAC clones carrying 16S rRNA genes and to assess their restriction fragment pattern, several different probes were used. The end-labelled oligonucleotide probe TH15 is specific to 16S rDNA of the genus *Azoarcus* [28] (positions 240–222 according to the numbering of the *E. coli* 16S rRNA sequence). In Southern blot hybridisations, the probe hybridised to four different fragments of an *Eco*RI digest of chromosomal DNA of strain BH72 (Fig. 1C), indicating that four different copies of 16S rDNA were present. The general oligonucleotide probe 926f (corresponding to a conserved region of the 3′ end of bacterial 16S rDNA, sequence positions 901–926) detects 16S rDNA of *Bacteria* and *Archaea* [37]. Also this probe detected four fragments in Southern hybridisation (Table 1). However the larger gene probe TH3-5 (positions 168–240) [28] hybridised to eight different fragments due to an internal *Eco*RI restriction site (Table 1). These results showed clearly that four copies of the *rrn* operon are present in the *Azoarcus* sp. strain BH72 genome.

Screening of the BAC library yielded different BAC clones for each of the four 16S RNA copies found in the chromosome (at least two per copy) (Table 1), showing that the *rrn* operons were not physically linked and well represented in the BAC library. To verify the assembly of the whole genome shotgun approach, we used the contig draft sequence of the *Azoarcus* genome project for in silico 16S rRNA-probe hybridisation and thus size determination of each *Eco*RI fragment spanning the probe target region. As shown in Table 1, for each probe used we obtained a very good fragment size correlation between in silico analysis and Southern hybridisation results from BAC clones covering these regions according to their end sequences. This proved a correct assembly of the draft *Azoarcus* genome sequence in the region of *rrn* operons.

The number of *rrn* operons identified in *Azoarcus* (four) is similar to copy numbers in other relatively fast-growing, related Proteobacteria such as *Neisseria meningitidis*, *R. solanacearum*, *P. aeruginosa*, *Agrobacterium tumefaciens* and *Azoarcus* sp. EbN1 [38,39,9]. In contrast, the genome of *B. japonicum* that grows extremely slowly harbours only one copy despite a larger genome size [40]. This is in agreement with the supposition that 16S rDNA copy numbers may correlate with ecological strategies to respond more or less rapidly to changing environmental conditions [41].

Acknowledgements

This work was supported by grants in the GenoMik framework of the BMBF (Bundesministerium für Bildung und Forschung) to the competence center at Bielefeld University (Nos. 031U213D and 0313105) and to B.R.-H., University Bremen (No. 031U213D).

References

[1] Reinhold, B., Hurek, T., Niemann, E.-G. and Fendrik, I. (1986) Close association of *Azospirillum* and diazotrophic rods with different root zones of Kallar grass. Appl. Environ. Microbiol. 52, 520–526.

[2] Reinhold-Hurek, B., Hurek, T., Gillis, M., Hoste, B., Vancanneyt, M., Kersters, K. and De Ley, J. (1993) *Azoarcus* gen. nov., nitrogen-fixing proteobacteria associated with roots of Kallar grass (*Leptochloa fusca* (L.) Kunth) and description of two species *Azoarcus indigens* sp. nov. and *Azoarcus communis* sp. nov. Int. J. Syst. Bacteriol. 43, 574–584.

[3] James, E.K. and Olivares, F.L. (1998) Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. Crit. Rev. Plant Sci. 17, 77–119.

[4] Reinhold-Hurek, B. and Hurek, T. (1998) Interactions of gramineous plants with *Azoarcus* spp. and other diazotrophs: Identification, localization and perspectives to study their function. Crit. Rev. Plant Sci. 17, 29–54.

[5] Yanni, Y.G., Rizk, R.Y., Corich, V., Squartini, A., Ninke, K., Philip-Hollingsworth, S., Orgambide, G., De Bruijn, F., Stoltzfus, J., Buckley, D., Schmidt, T.M., Mateos, P.F., Ladha, J.K. and Dazzo, F.B. (1997) Natural endophytic association between *Rhizobium leguminosarum* bv. *trifolii* and rice roots and assessment of its potential to promote rice growth. Plant Soil 194, 99–114.

[6] Reinhold-Hurek, B. and Hurek, T. (1998) Life in grasses: diazotrophic endophytes. Trends Microbiol. 6, 139–144.

[7] Reinhold-Hurek, B., Tan Z. and Hurek, T. (2005) Genus *Azoarcus*. In: Bergey's Manual of Systematic Bacteriology (Garrity, G.M. Ed.), Vol. 2. Springer Verlag, New York (in press).

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

[9] Rabus, R., Kube, M., Heider, J., Beck, A., Heitmann, K., Widdel, F. and Reinhardt, R. (2005) The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. Arch. Microbiol. 183, 27–36.

[10] Hurek, T., Handley, L., Reinhold-Hurek, B. and Piché, Y. (2002) *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. Mol. Plant-Microbe In. 15, 233–242.

[11] Sevilla, M., Burris, R.H., Gunapala, N. and 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* mutant strains. Mol. Plant-Microbe In. 14, 358–366.

[12] Hurek, T., Egener, T. and Reinhold-Hurek, B. (1997) Divergence in nitrogenases of *Azoarcus* spp., *Proteobacteria* of the β-subclass. J. Bacteriol. 179, 4172–4178.

[13] Egener, T., Hurek, T. and Reinhold-Hurek, B. (1999) Endophytic expression of *nif* genes of *Azoarcus* sp. strain BH72 in rice roots. Mol. Plant-Microbe In. 12, 813–819.

[14] Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L. and Yang, H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296, 79–92.

[15] Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, B., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296, 92–100.

[16] Hurek, T. and Reinhold-Hurek, B. (2003) *Azoarcus* sp. strain BH72 as a model for nitrogen-fixing grass endophytes. J. Biotechnol. 106, 169–178.

[17] Tao, Q., Chang, Y.L., Wang, J., Chen, H., Islam-Faridi, M.N., Scheuring, C., Wang, B., Stelly, D.M. and Zhang, H.B. (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158, 1711–1724.

[18] Tauch, A., Homann, I., Mormann, S., Ruberg, S., Billault, A., Bathe, B., Brand, S., Brockmann-Gretza, O., Ruckert, C., Schischka, N., Wrenger, C., Hoheisel, J., Mockel, B., Huthmacher, K., Pfefferle, W., Puhler, A. and Kalinowski, J. (2002) Strategy to sequence the genome of *Corynebacterium glutamicum* ATCC 13032: use of a cosmid and a bacterial artificial chromosome library. J. Biotechnol. 95, 25–38.

[19] Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J. and Wing, R.A. (2001) A marker-dense physical map of the *Bradyrhizobium japonicum* genome. Genome Res. 11, 1434–1440.

[20] Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89, 8794–8797.

[21] Zhang, H.B. and Wing, R.A. (1997) Physical mapping of the rice genome with BACs. Plant Mol. Biol. 35, 115–127.

[22] Grant, S.G., Jessee, J., Bloom, F.R. and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. Proc. Natl. Acad. Sci. USA 87, 4645–4649.

[23] Shizuya, H. and Kouros-Mehr, H. (2001) The development and applications of the bacterial artificial chromosome cloning system. Keio J. Med. 50, 26–30.

[24] Reinhold-Hurek, B., Hurek, T., Claeyssens, M. and Van Montagu, M. (1993) Cloning, expression in *Escherichia coli*, and characterization of cellulolytic enzymes of *Azoarcus* sp., a root-invading diazotroph. J. Bacteriol. 175, 7056–7065.

[25] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, USA.

[26] Strong, S.J., Ohta, Y., Litman, G.W. and Amemiya, C.T. (1997) Marked improvement of PAC and BAC cloning is achieved using electroelution of pulsed-field gel-separated partial digests of genomic DNA. Nucl. Acids Res. 25, 3959–3961.

[27] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., Eds., (1987). Current Protocols in Molecular Biology. John Wiley & Sons, New York.

[28] Hurek, T., Burggraf, S., Woese, C.R. and 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.

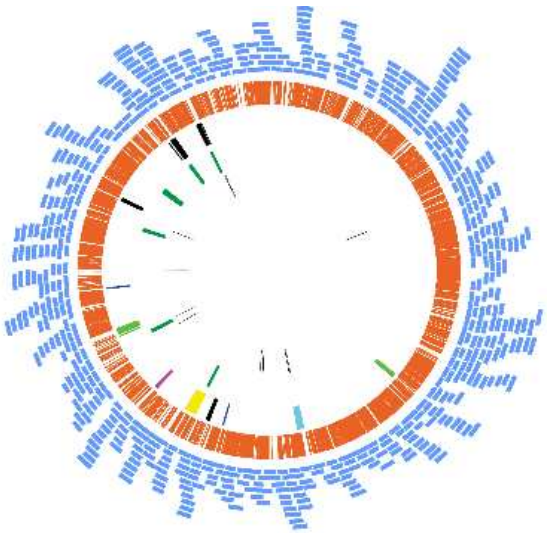
[29] Hurek, T., Van Montagu, M., Kellenberger, E. and Reinhold-Hurek, B. (1995) Induction of complex intracytoplasmic membranes related to nitrogen fixation in *Azoarcus* sp. BH72. Mol. Microbiol. 18, 225–236.

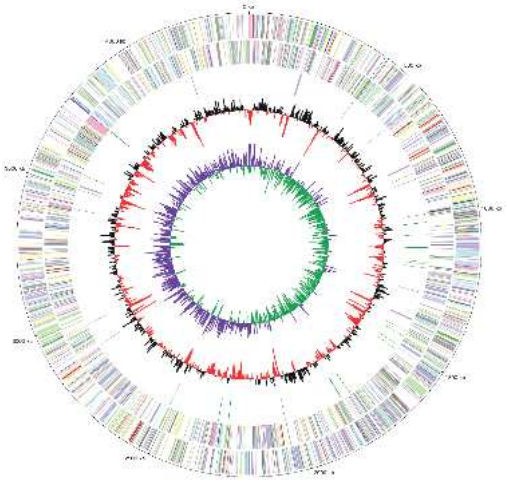
[30] Martin, D., Hurek, T. and Reinhold-Hurek, B. (2000) Occurrence of three P_{II}-like signal transmitter proteins in the diazotroph *Azoarcus* sp. BH72. Mol. Microbiol. 38, 276–288.

[31] Dörr, J., Hurek, T. and Reinhold-Hurek, B. (1998) Type IV pili are involved in plant-microbe and fungus–microbe interactions. Mol. Microbiol. 30, 7–17.
[32] Reinhold-Hurek, B. and Shub, D.A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. Nature 357, 173–176.
[33] Hurek, T., Wagner, B. and Reinhold-Hurek, B. (1997) Identification of N_2-fixing plant- and fungus-associated *Azoarcus* species by PCR-based genomic fingerprints. Appl. Environ. Microbiol. 63, 4331–4339.
[34] Bartels, D., Kespohl, S., Albaum, S., Drüke, T., Goesmann, A., Herold, J., Kaiser, O., Pfeiffer, F., Raddatz, G., Stoye, J., Meyer, F. and Schuster, S.C. (2005) BACCardI – a tool for the validation of genomic assemblies, assisting genome finishing and intergenome comparison. Bioinformatics 21, 853–859.
[35] Brosch, R., Gordon, S.V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B.G. and Cole, S.T. (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect. Immun. 66, 2221–2229.
[36] Brodin, P., Eiglmeier, K., Marmiesse, M., Billault, A., Garnier, T., Niemann, S., Cole, S.T. and Brosch, R. (2002) Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. Infect. Immun. 70, 5568–5578.
[37] Lane, D.J. (1991) 16S/23S rRNA sequencing In: Nucleic Acid Techniques in Bacterial Systematics (Stackebrandt, E. and Goodfellow, M., Eds.), Vols. 115–175. Wiley, Chichester.
[38] Gäher, M., Einsiedler, K., Crass, T. and Bautsch, W. (1996) A physical and genetic map of *Neisseria meningitidis* B1940. Mol. Microbiol. 19, 249–259.
[39] Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B.S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C. and Slater, S. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science 294, 2323–2328.
[40] Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S. (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. 9, 189–197.
[41] Klappenbach, J.A., Dunbar, J.M. and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66, 1328–1333.

CHAPTER C

Azoarcus sp. BH72 genome





1. OBJECTIVE

In the context of the *Azoarcus* sp. BH 72 genome project, the main objective of this work was the annotation and posterior analysis of a part of the genome sequence.

2. CONTRIBUTION *(*For details see Chapter D)

2.1. Annotation of the genes which belong to the COG categories "Ion Transporter and Metabolism" and "Carbohydrates Transport and Metabolism", was performed using the genome annotation system GenDB 2.2.

2.2. Analysis of the data obtained from the annotation of these genes was performed using different bioinformatics tools which are integrated in the GenDB 2.2 package (Chapter D, Table 3). Chemical detection of siderophores was performed as previously described (Arnow, 1937; Atkin *et al.*, 1970; Schwyn and Neilands, 1987).

3. RESULTS AND CONCLUSIONS.

Several conclusions from the analysis of both categories were obtained, which will be presented in detail in the next Chapter. Main results which contributed to the *Azoarcus* sp. BH72 genome publication (currently *in press*) are listed here:

3.1. The genome of strain BH72 contains a large number (22) of ORFs encoding genes for putative TonB-dependent outer membrane receptors. This is twice the number of genes described for *Azoarcus* sp. EbN1, a highly related soil bacterium, and even more in comparison to N_2-fixing symbionts, such as *B. japonicum* (13), *S. meliloti* (9) and *M. loti* (1). Moreover, two of these genes are not even present in the *P. fluorescens* Pf5 genome, a plant associated bacterium, known for its outstanding capacity in producing and uptaking a wide range of siderophores.

3.2. Strain BH72 has several putative receptors for the general classes of hydroxamates and catechol-type siderophores, ferricitrate, vitamin B12 and colicins and some receptors with unclear specificity. Surprisingly, the genome analysis of strain BH72 shows no evidence for the existence of the biosynthetic pathways for known hydroxamate or catecholate siderophores. Moreover, results from experimental studies, using the

Chromeazurol S (CAS) assay, which detects siderophores independent of their structure, also suggesting that strain BH72 lacks those pathways.

3.3. The *Azoarcus* sp. BH72 genome does not contain any of the genes required for a functional Entner-Doudoroff pathway or the oxidative branch of the pentose phosphate pathway, but the non-oxidative pentose phosphate pathway is complete. However, all the enzymes required for glycolysis via the Embden-Meyerhoff pathway, gluconeogenesis and the TCA cycle are present in the genome, including a full set of genes encoding a typical phosphoenolpyruvate: sugar phosphotransferase system (PTS). No specific outer membrane receptor or transporter for glucose or other carbohydrates was annotated.

3.4. Four complete and two incomplete copies of the TRAP-transport system (*dctPQM*), a family of high-affinity transporters for C4-dicarboxylates, were detected in the genome. C4-dicarboxylates are probably metabolized via the glyoxylate shunt pathway to the level of malate, which is then decarboxylated to pyruvate by the malic enzyme (MaeB). This underlines the fact, that the main carbons sources metabolized by strain BH72 are dicarboxylic acids and ethanol.

4. OUTLOOKS

Results presented above, represent just a small window out of what the *Azoarcus* sp. BH72 genome sequence is starting to show about diazotrophic, grass-endophyte lifestyle. The analysis of this part of strain BH72 genome sequence revealed several starting points for further researches which are presented below.

4.1. Functional analysis should focus on elucidation of the function of the large amount of outer membrane TonB-dependent receptors in *Azoarcus* sp. BH72 metabolism and ecology. In particular, reveal the function of those receptors which show no homology to genes present in strain EbN1 or *P. fluorescens* Pf5 genome will be of interest. Additionally, find out if these types of receptors play a role in signalling between strain BH72-plant and -ascomycete 2003 interaction.

A putative approach to achieve these objectives is making a mutagenesis with reporter genes of all OM TonB-dependent receptors founded. This approach will allow testing the specific expression of these receptors genes under different N_2-fixing growth conditions, e.g. pure culture, co-culture with *Acremonium alternatum,* as well as in the presence of rice

plants. Additionally, different iron-minimal medium could be also testing for gene expression with the aim to determine which of these receptors genes are involved in iron up take.

4.2. The lacking of siderophore production pathways in strain BH72, combined with the presence of high amount of receptors involved in the uptake of those molecules, leads to the hypothesis that maybe the plant, or the fungus partners, are able to produce those molecules. Experimental studies such as siderophore production detection using CAS indicator plates would be important to give some further hints. Moreover, it would be interesting to study if strain BH72 is able to use heterologues siderophores as iron sources.

4.3. Strain BH72 is not able to use carbohydrates as carbon sources (Reinhold-Hurek *et al.*, 2005). Functional studies with the aim to elucidate which role the annotated sugar PTS-systems of strain BH72 plays in its metabolism and ecology will be of interest. For this purpose, phenotype characterization of strain BH72 mutants in PTS- system, grown in different carbon sources, could give some ideas of function. Of particular interest will be to study if the lacking of this system has some effects during the plant-association and nitrogen-fixation processes.

4.4. Finally, since the *Oryza sativa* genome has been sequenced (Goff *et al.*, 2002; Yu *et al.*, 2002), the rice-*Azoarcus* system is an ideal model to study diazotrophic endophyte-grass interactions by functional genomic approaches.

ARTICLES

nature biotechnology

Complete genome of the mutualistic, N_2-fixing grass endophyte *Azoarcus* sp. strain BH72

Andrea Krause[1], Adarsh Ramakumar[1], Daniela Bartels[2], Federico Battistoni[1], Thomas Bekel[2], Jens Boch[3], Melanie Böhm[1], Frauke Friedrich[1], Thomas Hurek[1], Lutz Krause[2], Burkhard Linke[2], Alice C McHardy[2,6], Abhijit Sarkar[1], Susanne Schneiker[2,4], Arshad Ali Syed[1], Rudolf Thauer[5], Frank-Jörg Vorhölter[2,4], Stefan Weidner[2], Alfred Pühler[2,4], Barbara Reinhold-Hurek[1], Olaf Kaiser[2,4,7] & Alexander Goesmann[2,7]

Azoarcus **sp. strain BH72, a mutualistic endophyte of rice and other grasses, is of agrobiotechnological interest because it supplies biologically fixed nitrogen to its host and colonizes plants in remarkably high numbers without eliciting disease symptoms. The complete genome sequence is 4,376,040-bp long and contains 3,992 predicted protein-coding sequences. Genome comparison with the *Azoarcus*-related soil bacterium strain EbN1 revealed a surprisingly low degree of synteny. Coding sequences involved in the synthesis of surface components potentially important for plant-microbe interactions were more closely related to those of plant-associated bacteria. Strain BH72 appears to be 'disarmed' compared to plant pathogens, having only a few enzymes that degrade plant cell walls; it lacks type III and IV secretion systems, related toxins and an N-acyl homoserine lactones–based communication system. The genome contains remarkably few mobile elements, indicating a low rate of recent gene transfer that is presumably due to adaptation to a stable, low-stress microenvironment.**

Endophytic bacteria reside within the living tissue of plants without substantively harming them. They are of high interest for agrobiotechnological applications, such as the improvement of plant growth and health, phytoremediation[1] or even as biofertilizer[2]. Supply of nitrogen derived from fixation of atmospheric N_2 by grass endophytes, such as *Gluconacetobacter diazotrophicus* and *Azoarcus* sp. strain BH72, which has been shown to occur in sugarcane[3] and Kallar grass[2], is a process of potential agronomical and ecological importance.

Although the lifestyle of these endophytes is relatively well documented, the molecular mechanisms by which they interact beneficially with plants have only been poorly elucidated. A combination of features makes *Azoarcus* sp. strain BH72 an excellent model grass-endophyte[4]. (i) It supplies nitrogen derived from N_2 fixation to its host, Kallar grass (*Leptochloa fusca* (L.) Kunth); *in planta* it is usually not culturable, but can be detected by culture-independent methods based on *nifH*-encoding nitrogenase reductase, the key enzyme for N_2 fixation[2]. (ii) It colonizes nondiseased plants in remarkably high numbers: estimates range from 10^8 cells (culturable cells per gram root dry weight (RDW) of field-grown Kallar grass[5]) to 10^{10} cells (estimated on the basis of abundance of bacterial *nifH*-mRNA in roots)[2]. (iii) It is the only cultured grass endophyte shown by molecular methods to be the most actively N_2-fixing bacterium of the natural population in roots[2]. (iv) It also colonizes the roots of rice, a cereal of global importance, in high numbers (10^9 cells per g RDW) in the laboratory, and spreads systemically into shoots[6]. Plant stress response is only very limited in a compatible, that is, well-colonized rice cultivar[7]. Notably, *Azoarcus* sp. strain BH72 is capable of endophytic N_2-fixation inside the roots of rice[8].

For a wider application in agriculture, more knowledge is required on mechanisms of interaction and host specificities. Although the genome of a related species, strain EbN1, belonging to a branch of *Azoarcus* species that typically occurs in soils and sediments but not in association with plants[9], is available[10], phenotypic differences and phylogenetic distances of 5–6% suggest they might deserve the rank of a separate genus in future[9]. The plant-associated strain BH72—like many N_2-fixing endophytes grass endophytes—has not been detected in root-free soil[11]. In this study, we present the complete genome sequence of a diazotrophic grass endophyte, *Azoarcus* sp. strain BH72, and highlight features that may contribute to knowledge of the endophytic lifestyle of these plant-beneficial bacteria, which may be instrumental in developing biotechnological applications.

RESULTS

General features of the genome and mobile elements

The *Azoarcus* sp. strain BH72 genome sequence was obtained with a whole genome shotgun approach, the assembly being validated by a complete fosmid (**Fig. 1b**) and a bacterial artificial chromosome

[1]Laboratory of General Microbiology, University of Bremen, PO Box 330440, D-28334 Bremen, Germany. [2]Center for Biotechnology (CeBiTec), Bielefeld University, PO Box 100131, D-33501 Bielefeld, Germany. [3]Institut für Genetik, Martin-Luther-Universität, Weinbergweg 10, D-06120 Halle/Saale, Germany. [4]Lehrstuhl für Genetik, Bielefeld University, PO Box 100131, D-33501 Bielefeld, Germany. [5]Max-Planck-Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, D-35043 Marburg, Germany. [6]Present address: Bioinformatics & Pattern Discovery Group, IBM Thomas J Watson Research Center, Yorktown Heights, New York 10598, USA. [7]These authors contributed equally to the work. Correspondence should be addressed to Barbara Reinhold-Hurek (breinhold@uni-bremen.de).

Received 22 May; accepted 4 August; published online 22 October 2006; doi:10.1038/nbt1243

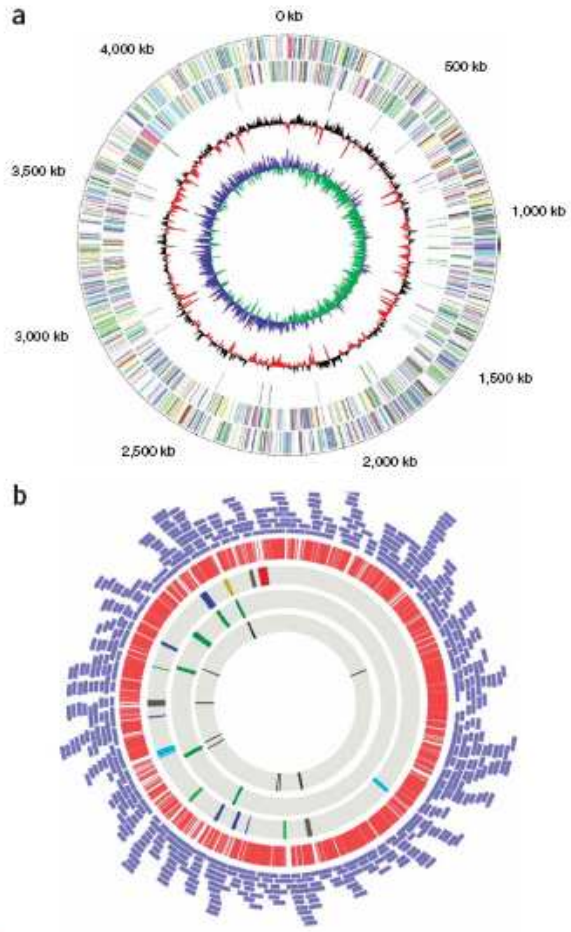


Figure 1 Circular representation of the *Azoarcus* sp. strain BH72 genome displaying relevant genome features and validation of the sequence assembly by a fosmid map. In the final consensus sequence each base matched at least phred40 quality. (a) From the outer to the inner concentric circle: circle 1, genomic position in kilobases; the origin of replication was clearly detectable by a bias of G toward the leading strand (GC skew); the start of the *dnaA* gene located in this region was defined as zero point of the chromosome; circles 2 and 3, predicted protein-coding sequences (CDS) on the forward (outer wheel) and the reverse (inner wheel) strands colored according to the assigned COG (clusters of orthologous groups) classes; leading strand, 2,074 CDS = 52.0%; lagging strand, 1,918 CDS = 48.0%; circle 4, tRNA genes (green) and the four rRNA operons (pink); circle 5, the G+C content showing deviations from the average (67.92%); circle 6, GC skew; a bi-directional replication mechanism suggested by a clear division into two equal replichores. (b) Fosmid map of the *Azoarcus* sp. strain BH72 chromosome. Each blue arc represents a single fosmid clone mapped onto the assembled sequence; circle 1, CDS with homologs in the chromosome of *Azoarcus* sp. *strain* EbN1 (e-value below e^{-30}); circle 2, gene clusters coding for surface-related proteins or other functions not related to proteins of *Azoarcus* sp. strain EbN1: exopolysaccharide/ lipopolysaccharide-related and pilus-related gene clusters (blue), flagella and chemotaxis related gene clusters (light blue), virulence-related gene clusters (red), proteins related to metabolism (gray), conserved hypothetical proteins or other proteins related to proteins of rhizobia or plant commensals, and various genes not present in *Azoarcus* sp. strain EbN1 (gold); circle 3, putative genomic islands predicted by the Pai-Ida program 1.1 (score > 3.8); circle 4, transposases and phage-related genes.

(BAC) map[12]. Characteristics of the single, circular chromosome and the predicted genes are shown in **Figure 1** and **Table 1**.

The genome contains remarkably few phage- or transposon-related genes, indicating a low degree of lateral transfer and genome rearrangements; just eight loci (**Fig. 1b**, circle 4) contain genes for integrases, recombinases, transposases or phage-related genes (**Supplementary Table 1** online). Only a few loci correspond to predicted anomalous gene clusters or putative pathogenicity islands (**Fig. 1b**, circle 3). In contrast, the genome of the *Azoarcus*-related soil isolate strain EbN1 contains 237 transposon-related genes[10]. Also rhizobial genomes harbor >100 transposases or phage-related genes (http://www.kazusa.or.jp/rhizobase/). Likewise, many plant-pathogenic proteobacteria contain high numbers of mobile elements[14]. High genomic plasticity might reflect the need for continuous adaptation to changing environments like soil or to host defense mechanisms. For nodule symbionts, soil is an alternative habitat in their life cycle; in contrast typical grass endophytes can not usually be isolated from root-free soil[11,13]. The comparatively low number of mobile elements in the endophyte BH72 might indicate a low rate of recent gene transfer and genome rearrangements, which is presumably due to adaptation to a stable, low-stress microenvironment inside plants.

Comparative genomics

Genome comparison revealed a surprisingly low degree of synteny between genomes of strain BH72 and the *Azoarcus*-related strain EbN1 (**Fig. 2**). At a low cutoff e-value of e^{-30}, the majority of predicted proteins (58%) in strain BH72 have some counterparts in strain EbN1 (**Fig. 1b**, circle 1). However, only 43% of these proteins were more closely related to those of EbN1 than to proteins of other strains. Other pathogenic or plant symbiotic proteobacteria have even less related genomes (**Supplementary Table 2** online). Because strains BH72 and EbN1 have a very different ecology, the differences may give important hints as to which genes are required specifically for the endophytic lifestyle. Several gene clusters of strain BH72 that are

Table 1 Genome features of the N_2-fixing endophyte *Azoarcus* sp. strain BH72 in comparison to the denitrifying soil bacterium *Azoarcus* sp. strain EbN1

Feature	*Azoarcus* sp. BH72	*Azoarcus* sp. EbN1
Size of chromosome (bp)	4,376,040	4,296,230
Plasmids	0	2[a]
G+C content, %	67.92	65.12
Coding sequences	3,992	4,133
Function assigned	3,418	2,560
Conserved hypothetical protein	517	628
Hypothetical protein	57	945
% of genome coding	91.2	90.9
Average length (bp)	999	945
Maximal length (bp)	6,330	6,132
% ATG initiation codons	86.57	76.46
% GTG initiation codons	10.40	16.01
% other initiation codons	n.d.[b]	n.d.
RNA elements		
rRNA operons	4	4
tRNAs	56[c]	58

[a]Plasmid 1 (207,355 bp), plasmid 2 (223,670 bp). [b]n.d., not determined. [c]One tRNA^{Ile} (azo_tRNA_0051) is disrupted by a self-splicing group I intron in the CAT anticodon loop[42].

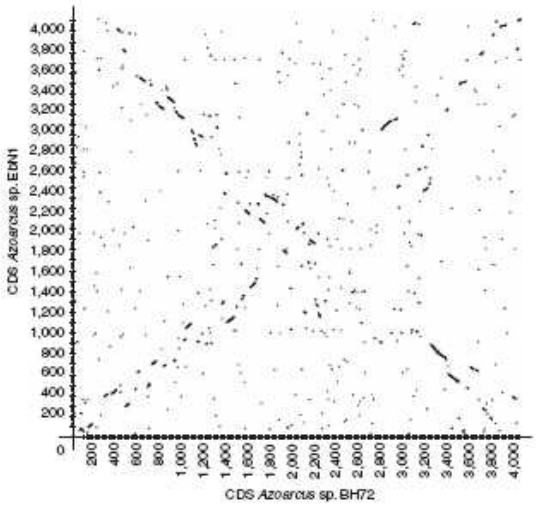


Figure 2 Synteny between the genomes of *Azoarcus* sp. strain BH72 and *Azoarcus* sp. strain EbN1. The genome of *Azoarcus* sp. strain EbN1 is adjusted to *dnaA* with its start codon as zero point of the chromosome. The diagram depicts *x-y* plots of dots forming syntenic regions between the two *Azoarcus* genomes. Each dot represents an *Azoarcus* sp. strain BH72 CDS having an ortholog in *Azoarcus* sp. strain EbN1, with coordinates corresponding to the CDS number in each genome. The orthologs were identified by best BLASTP matches of amino acid sequences (e-value $< e^{-30}$).

lacking in EbN1 or that are more similar to genes of other bacteria (**Fig. 1b**) harbor genes that encode proteins putatively involved in cell surface components or other features that may be required for the endophytic lifestyle (see below, **Fig. 3** and **Supplementary Table 1** online).

Carbon metabolism and signal transduction

Strain BH72 has a strictly respiratory type of metabolism and does not grow on any carbohydrate tested[9,15]. Aspects of putative carbon metabolism are shown in **Figure 4**. The inability to utilize common carbohydrates might contribute to a plant-compatible endophytic lifestyle because, in contrast to phytopathogens, the bacteria cannot grow and proliferate on the major cell wall constituents although a cellulase is present[16].

The major carbon sources for strain BH72 are dicarboxylic acids and ethanol[9]. Transport systems for C4-dicarboxylates (**Fig. 4**) might be of vital importance during the association with host plants, as in symbiotic rhizobia[17]. Ethanol might be important for association with flooded plants like rice, which accumulate ethanol under anoxic conditions, especially at root tips—one of the typical sites of colonization of strain BH72. Correspondingly, its genome harbors ten genes encoding putative alcohol dehydrogenases.

Strain BH72, despite being adapted to a relatively stable, low-stress microenvironment, shows a remarkable density of signal transduction systems (see details in **Supplementary Table 3** online). Thus it may be a good example for sophisticated signal transduction networks.

N_2 fixation and nitrogen metabolism

Azoarcus sp. strain BH72 appears to be highly adapted to environments poor in available nitrogen sources, which correlates with its role as an N_2-fixing endophyte (**Fig. 4**).

(i) A low-affinity glutamate dehydrogenase (GDH) for ammonium assimilation is lacking, a feature highly unusual in free-living bacteria, whereas it is present in strain EbN1. Only the high-affinity ATP-consuming assimilation system (GS[2x]-GOGAT) is present. (ii) Four genes encoding high-affinity ammonia transporters exist (*amtB/Y/D/E*), one of them with an additional regulatory domain. (iii) In contrast to the soil strain EbN1, structural genes for the molybdenum-dependent nitrogenase complex and all genes required for cofactor synthesis and maturation of the nitrogenase are present in strain BH72, one of them in two copies (*nifY*). Several putative low-potential electron donors for N_2 fixation were identified including two flavodoxin-encoding genes (*nifF1*, *nifF2*), 12 genes for ferredoxin-like proteins; two clusters encoding putative electron transport systems (*rnf1*, *rnf2*) might be instrumental for electron supply to ferredoxin during N_2 fixation. Several genes likely to be involved in the regulatory cascade are also listed in **Supplementary Table 1** online. Although in pure culture, N_2-fixing strain BH72 does not excrete substantial amounts of nitrogenous compounds[18], it supplies fixed nitrogen to its grass host[2]. The four ammonia transport proteins are putative candidates for export to the plant. Two transport systems for glutamate or glutamine as well as nine for branched-chain amino acids were also identified; however, the presence of periplasmic substrate-binding proteins suggests that these systems are used for import and not for export.

About 38 genes encoding enzymes and transporters involved in nitrate metabolism were identified (**Fig. 4** and **Supplementary Table 1** online). As in strain EbN1, genes required for assimilatory nitrate and

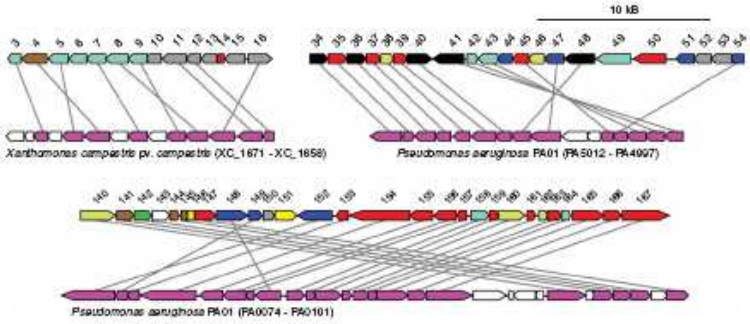


Figure 3 Gene clusters in *Azoarcus* sp. strain BH72 that are lacking in EbN1 or are more similar to genes of other bacteria. Synteny of selected clusters with gene clusters of other bacteria: *gum*-cluster, *Xanthomonas campestris* pv. *campestris* (locus_tag numbers XC_1671 - XC_1658); lipopolysaccharide-related cluster, *Pseudomonas aeruginosa* PA01 (locus_tag numbers PA5012 - PA4997); *sci*-cluster, *Pseudomonas aeruginosa* PA01 (locus_tag numbers PA0074 - PA0101); pink/white, genes present or not present in the gene cluster of strain BH72, respectively. Numbers refer to genes of *Azoarcus* sp. strain BH72 listed in **Supplementary Table 1** online. Highest similarities to proteins of other bacteria are depicted by the following colors: red, human or animal pathogens; green, root nodule symbionts (rhizobia); yellow-green, root-associated bacteria; turquoise, plant pathogens; other bacteria according to their phylogenetic affiliations: black, *Azoarcus* sp. strain EbN1; gray, beta-subgroup of *Proteobacteria*; blue, gamma-subgroup of *Proteobacteria*; yellow, alpha subgroup of *Proteobacteria*; brown, others.

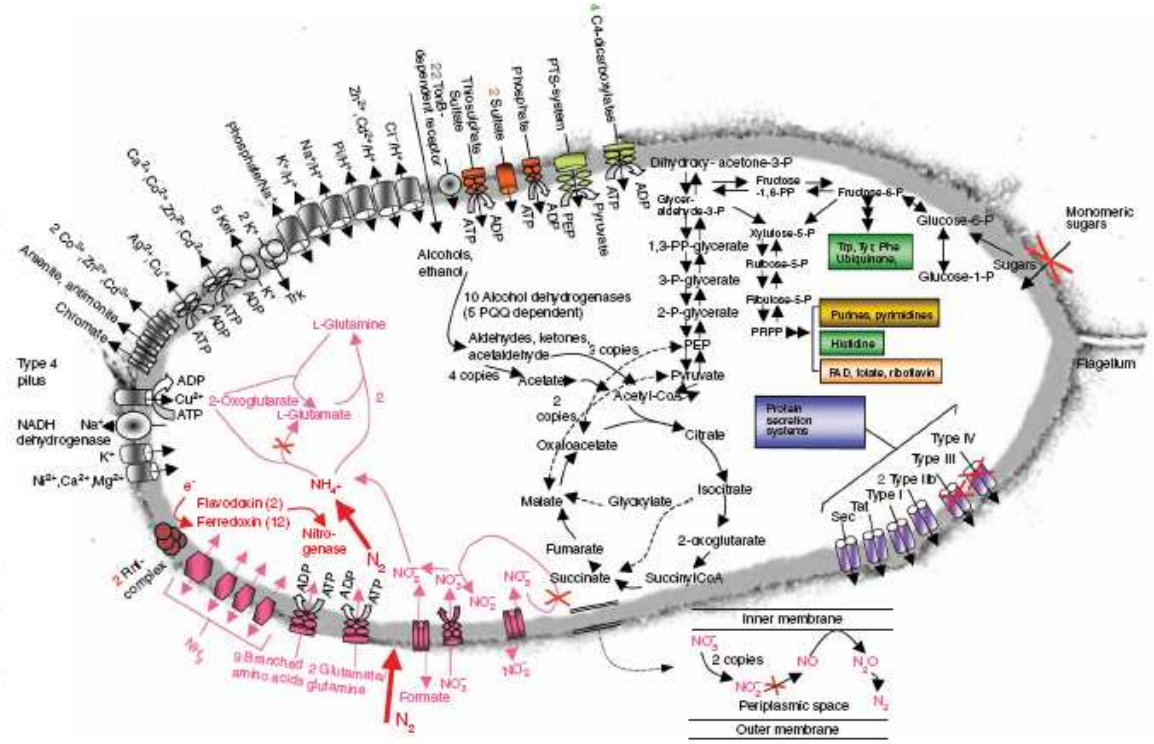


Figure 4 Overview of physiological features of *Azoarcus* sp. strain BH72. Depicted are carbohydrate metabolism, selected transporters and catabolic pathways for organic acids, nitrogen assimilatory and dissimilatory pathways, inorganic ion transport systems, and protein translocation systems. Crosses indicate pathways or reactions that are apparently not present in strain BH72. Details of metabolic features: the genome does not contain genes required for a functional Entner-Doudoroff pathway or the oxidative branch of the pentose phosphate pathway, however, the nonoxidative pentose phosphate pathway is complete. All enzymes required for glycolysis via the Embden-Meyerhoff pathway, gluconeogenesis and the TCA cycle are present, including genes encoding a typical phosphoenolpyruvate: sugar phosphotransferase system (PTS), but no specific outer membrane transporter for carbohydrates. Four complete and two incomplete copies of the TRAP-transport system (DctPQM) for C4-dicarboxylates are present. C4-dicarboxylates are probably metabolized via the glyoxylate shunt pathway, malate being decarboxylated by the malic enzyme (MaeB, two gene copies). Ten putative alcohol dehydrogenases are encoded in the genome (see also **Supplementary Table 1** online), five of which are PQQ-dependent enzymes not present in strain EbN1.

nitrite transport and reduction are present in strain BH72 in one copy (*nasFED*, *nirC*, *nasA*, *nasC*, *nirBD*). In contrast to strain EbN1, strain BH72 cannot conduct denitrification to N_2, as genes required for a nitrite reductase are missing in strain BH72. However, genes for subsequent denitrification reactions (*norCBQD* and *nosRZDYFLX* encoding reductases for nitric oxide and nitrous oxide, respectively) are present. The periplasmic localization of the nitrate reductase components (duplicated *nap* operon) and detoxification of NO_2^- via an NO_3^-/NO_2^- antiporter (NarK) might decrease toxicity of nitrite. In contrast, the denitrification pathway is complete in strain EbN1 for which denitrification appears to be a typical feature[10].

Iron-transport related proteins

In Gram-negative bacteria, TonB-dependent, outer-membrane proteins are responsible for the specific uptake of ferric-siderophore complexes, high-affinity iron chelators. They are also important for perception of environmental signals and are associated with pathogenicity of plant pathogens[19]. Strain BH72 possesses 22 genes encoding proteins related to iron transport (**Supplementary Table 1** online), twice as many as described for strain EbN1 and even more than other N_2-fixing endosymbionts (*Bradyrhizobium japonicum*, 13; *Sinorhizobium meliloti*, 9; and *Mesorhizobium loti*, 1). Two genes (*azo2156*, *azo3836*) are not even present in the *P. fluorescens* Pf5 genome, a plant-associated bacterium known for its capacity to produce and take up a wide range of siderophores[20], which contains 45 such genes. Although putative receptors for hydroxamate- and catechol-type siderophores, ferricitrate, vitamin B12, colicins and unknown substances are present, there was no evidence for biosynthetic pathways for known hydroxamate or catecholate siderophores[21]. Moreover, production of siderophores was not detected experimentally (**Supplementary Fig. 1a** online). Apparently, this strain is highly adapted to obtaining chelated iron from other sources, as fungi and monocotyledonous plants also produce siderophores[22]. The high number of putative receptors suggests a role not only for rhizosphere competence of strain BH72, but also for biocontrol.

Plant-associated lifestyle

Surface characteristics of bacteria are important factors for recognition by and interaction with the host. Several gene clusters putatively related to surface components of strain BH72 are lacking in strain EbN1, or are more highly related to genes of plant-associated or pathogenic bacteria.

Type IV pili are among the few factors known to affect endophytic colonization of grass diazotrophs[23]. Establishment of microcolonies on roots and fungal mycelium, and systemic spreading in rice are mediated by type IV pili[23]. The strain BH72 genome harbors 41 genes encoding proteins putatively involved in pilus assembly and regulation, whereas only 30 such genes were found in strain EbN1. Genes highly similar in both species encode proteins with conserved function such as assembly or regulation (for example, PilBCD-PilF-PilM-NOPQ-PilTU-PilZ or PilSR-PilGHIL). Other pilus proteins related mostly to phytopathogenic bacteria might be either pseudopilins involved in secretion (PilV/W/X) or putative tip adhesins (PilY1A/B, 31% and 39% sequence identity to *Ralstonia solanacearum* and *Xylella fastidosa*, respectively) that are lacking even in strain EbN1 and might be characteristic for interaction with plant surfaces (**Supplementary Table 1** online).

Other cell surface components that are often involved in recognition or virulence of pathogens are lipopolysaccharides, exopolysaccharides and capsular material. Intriguingly, many gene products putatively involved in their synthesis in strain BH72 are not highly related to those of the soil isolate EbN1, but to proteins of plant symbionts, pathogens or gamma- or alpha-proteobacteria (**Supplementary Table 1** online). Several genes of strain BH72 are similar to the *gum* operon for exopolysaccharide production in phytopathogenic *Xanthomonas campestris*[24]; genes encoding putative glycosyl transferases have considerable similarity to the rhizobial *pss* gene cluster (**Supplementary Table 1** online, 20–23), which is involved in exopolysaccharide polymerization, translocation and thus in plant-microbe interaction; a gene cluster related mainly to lipopolysaccharide synthesis is most similar to genes of gamma-proteobacteria including pathogens; these clusters did not show sufficient synteny to support the assumption of a very recent gene transfer (**Fig. 3**).

Motility. Flagella are pivotal for motility, adhesion, biofilm formation and colonization of the host. Strain BH72 is highly motile by means of a polar flagellum. At least 48 genes were identified that are generally required for biosynthesis and function of flagella and chemotaxis. They are located in three different noncontiguous clusters and are mostly related to genes of other beta-proteobacteria and a few pathogens of the gamma-subgroup (**Supplementary Table 1** online). There are three genes encoding flagellins (*fliC1, fliC2, fliC3*) and two encoding flagellar motor proteins, suggesting an important role for motility in the plant-associated lifestyle. In contrast, the nonmotile strain EbN1 does not possess a complete flagellar regulon[19].

Secretion and communication. Several genes encoding potential protein secretion systems were identified in the genome of strain BH72 (**Fig. 4**) for a sec-dependent pathway, a signal recognition particle (SRP)-mediated translocation and a twin arginine translocation (Tat) system, all of them targeting proteins through the inner membrane. Secretion of proteins through the entire cell envelope seems to be limited to only three varieties of pathways. Genes were identified that encode one type I secretion system and one autotransporter. Two gene clusters were detected that encode a type II secretion-related system (type IIb secretion system[25]), consisting only of GspDEFG.

Two other secretion systems are common to plant-associated bacteria, type III and IV secretion systems, which transport a wide variety of effector proteins into the extracellular medium or into the cytoplasm of eukaryotic host cells and affect interaction[26–29]. Intriguingly, neither system is present in strain BH72, probably preventing the export of toxic proteins to the host.

'Quorum sensing' is a common way of bacteria to communicate with each other or hosts by means of autoinducers that accumulate in the extracellular environment in a cell density-dependent manner[30]. Although there is evidence that autoinducer-dependent gene regulation occurs in *Azoarcus* sp. strain BH72 (Böhm, M. & Reinhold-Hurek, B., unpublished data), this strain appears to escape the usual communication systems. Widespread autoinducers of Gram-negative bacteria are N-acyl homoserine lactones (AHLs)[30]. There is no evidence for genes encoding an AHL-based quorum-sensing system in strain BH72; genes encoding the autoinducer synthetase (LuxI/LasI–type) or the responsible cytoplasmic autoinducer receptor (LuxR/LasR-type) are lacking. Furthermore, different bacterial sensor strains detecting presence of short- or long-chain AHLs did not yield a positive response toward strain BH72 (**Supplementary Fig. 1b** online). Also genes encoding the autoinducer-2 synthetase LuxS[30] are lacking. Gram-positive bacteria usually use peptides as autoinducers[30]. Genes expected for this system were not detected in strain BH72 either.

Virulence and interaction factors. The strain BH72 genome stands out by the lack of obvious genes involved in production of toxins. Moreover, common hydrolytic enzymes that macerate plant cell wall polymers and thus contribute to a phytopathogenic lifestyle and plant damage are rare: pectinase-encoding genes are absent; only a few genes encode putative glycosidases (*palZ, spr1, ndvC, eglA*), some of them with transmembrane helices. Detection of genes for membrane-bound enzymes is in agreement with the observation that strain BH72 does not secrete cellulases into the culture medium, but shows activities of a cell surface-bound endoglucanase (EglA) and exoglycanase that also hydrolyses xylosides[16], a major component of primary cell walls in grasses. A low production of macerating enzymes is likely to contribute to compatibility with the plant host, however these hydrolases might assist in endophytic colonization, as shown for the endoglucanase EglA[31].

There is no genomic evidence for the central process in the rhizobium-legume symbiosis, the induction of nodulation. Common *nodABC* genes required for the biosynthesis of the Nod-factor backbone are not present in strain BH72; only a few genes show some sequence similarity to other *nod* or *nol* genes (**Supplementary Table 1** online). However, like other grass-associated microbes such as *Azospirillum* a gene similar to *nodD* is present, in rhizobia encoding a central regulator for flavonoid-inducible gene expression.

Some other gene clusters in strain BH72 are also interesting targets for putative roles in plant-microbe interactions. One cluster shows similarity to genes that are localized in the *sci*-genomic island, affecting virulence of human pathogens. The genomic organization shows remarkable synteny with genes of *P. aeruginosa* (**Fig. 3**), arguing for a more recent gene transfer. Interestingly some homologs are also present in rhizobia. Other noticeable gene clusters code for mainly conserved hypothetical proteins or metabolism-related proteins that have orthologs in rhizobia, or of regulatory proteins (ColRS) important for root colonization of commensals (**Supplementary Table 1** online, no. 177–185, 188–196, 198–200).

DISCUSSION

The complete genome sequence of *Azoarcus* sp. strain BH72 offers insights into genomic strategies for an endophytic life style, and allows identification of various features that may contribute to their interaction with plants. The strain appears to be adapted to a relatively stable, low-stress microenvironment, since its genome contains remarkably few phage- or transposon-related genes in comparison to many soil bacteria or pathogens, indicating a low plasticity of the

genome. The lack of the typical communication system of Gram-negative bacteria based on AHLs also argues for a rather exclusive microhabitat.

Strain BH72 appears to be disarmed compared to plant pathogens by its in ability to metabolize carbohydrates coupled with the lack of a massive occurrence of cell-wall degrading enzymes. Moreover, this bacterium lacks known toxins and type III and IV secretion systems that are typically used by pathogens to transport effector molecules to their host. This might be instrumental for avoiding damage to the plant host despite a dense internal colonization, and only the small set of hydrolases identified may be required for penetrating into the plant tissue. Some specific features of nodule symbionts are also lacking, like most *nod/nol* genes required for nodule induction.

Genome comparison with the *Azoarcus*-related, nondiazotrophic, nonendophytic soil bacterium strain EbN1 revealed features likely to be important for plant-microbe interaction. Strain BH72 appears to be highly adapted to environments of low nitrogen availability, N_2-fixation playing a key role in its ecology. Several gene clusters that were lacking in strain EbN1 or were highly similar to genes of plant-associated or pathogenic bacteria were related to cell surface components that are often involved in recognition—gene products participating in the synthesis of exopolysaccharide, lipopolysaccharide, type IV pilus tips, the flagellar and chemotaxis apparatus. Further targets for studying interaction mechanisms were also identified by comparative genomics, such as virulence-related *sci* genes or genes encoding conserved hypothetical proteins shared with nodule symbionts. A large and diverse set of TonB-dependent receptors (22) might play a role in iron acquisition and biocontrol. In future functional genomic analyses, the role of these target genes for host compatibility will be elucidated, which is crucial for a wider agro-biotechnological application of N_2-fixing endophytes.

METHODS

Whole genome shotgun sequencing. DNA shotgun libraries with insert sizes of 1 kb and 2–3 kb in pGEM-T (Promega), and 8-kb fragments in pTrueBlue-rop (MoBiTec) vectors were constructed by MWG Biotech. Plasmid clones were end-sequenced on ABI 3700 sequencers (ABI) by MWG Biotech AG. Base-calling was carried out using PHRED[32,33]. High-quality reads were defined by a minimal length of 250 bp with an averaging quality value of ≥20. Finally, 60,715 high-quality reads, a total of 39,266 (5.26 ×), 18,070 (2.32 ×) and 3,379 (0.40 ×) end sequences (×'s indicate genome equivalents) from libraries with 1-kb, 2- to 3-kb and 8-kb inserts, respectively, were obtained.

Sequence assembly and assembly validation. Basecalling, quality control and elimination of vector DNA sequences of the shotgun-sequences were performed by using the software package BioMake (Bielefeld University) as previously described[34]. Sequence assembly was performed by using the PHRAP assembly tool (http://www.phrap.org). The CONSED/AUTOFINISH software package[35,36], supplemented by the in-house tool BACCardI[37], was used for the finishing of the genome sequence.

For gap closure, a BAC library with inserts of ~90 kb in pBeloBAC11 was constructed and BAC contigs were assembled[12]. Remaining gaps of the whole genome shotgun assembly were closed by sequencing on shotgun and BAC clones carried out by IIT GmbH on LI-COR 4200L and ABI 377 sequencing machines. So that we would obtain a high quality genome sequence, all regions of the consensus sequence were polished to at least phred40 quality by primer walking. Collectively, 1,374 sequencing reads were added to the shotgun assembly for finishing and polishing of the genomic sequence. Repetitive elements, that is, rRNA operons, were sequenced completely by primer walking on BAC clones. For assembly validation, a fosmid library with inserts of ~35–38 kb was constructed by IIT GmbH using the EpiFOS Fosmid Library Production Kit (Epicentre). End-sequencing of 672 fosmids was carried out on ABI 377 and MegaBACE 1000 (Amersham Biosciences) sequencing machines by IIT GmbH and on ABI 3730XL DNA analyzers by the sequencing group of the Max Planck Institut für Molekulare Genetik. For assembly validation, fosmid end sequences were mapped onto the genome sequence by employing the BACCardI tool.

Genome analysis and annotation. In a first step automatic gene prediction and annotation were performed using the genome annotation system GenDB 2.0 (ref. 38) as previously described[14]. In a second step the coding sequences (CDS) prediction was validated: a position weight matrix (PWM) was generated to score all translation starts, and visualization of CDS was done using GeneQuest (DNASTAR Inc.). Reinspection of starts was coupled to recomputing of homology (BlastP) and assessment of function. In this way 4.3% more ORFs were detected, and 15.5% of the start sites were changed in comparison to the prediction obtained by a combined GLIMMER and CRITICA approach[39]. Intergenic regions were checked again for CDS missed probably by the automatic annotation using the BLAST programs[40]. During manual annotation, the following criteria were applied: (i) for hypothetical proteins, amino acid identities to other proteins were less than 30% over the entire length of the protein; (ii) for conserved hypothetical proteins, amino acid identities to proteins of unknown function were more than 30%; (iii) for proteins with putative or probable functions, sequence identities to named gene products were >20% or 40%, respectively.

Genomic comparison. For comparative analyses, the annotated genome sequence of *Azoarcus* strain EbN1 (acc. nos. CR555306, CR555307, CR555308) was imported into GenDB. Comparisons of chromosomal sequences were carried out with GenDB[38].

Detection of regions with atypical G+C content. For detection of anomalous gene clusters or putative pathogenicity islands in bacterial genomes, the Pai-Ida program 1.1 (http://compbio.sibsnet.org/projects/pai-ida/) based on an iterative discriminant analysis[41] was used.

Database submission. The nucleotide sequence of the *Azoarcus* sp. strain BH72 chromosome was submitted to EBI under accession number AM406670–*Azoarcus* sp. BH72.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors grateful thank all the people involved in this project. This work was supported by grants of the German Federal Ministry of Education and Research (BMBF) (031U113D, 031U213D and 0313105) in the frame of the GenoMik network "Genome Research on Bacteria Relevant for Agriculture, Environment and Biotechnology" and the Fonds der Chemischen Industrie (to R.T.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Barac, T. *et al.* Engineered endophytic bacteria improve phytoremediation of water-soluble, volatile, organic pollutants. *Nat. Biotechnol.* 22, 583–588 (2004).
2. Hurek, T., Handley, L., Reinhold-Hurek, B. & Piché, Y. *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Mol. Plant Microbe Interact.* 15, 233–242 (2002).
3. Sevilla, M., Burris, R.H., Gunapala, N. & Kennedy, C. Comparison of benefit to sugarcane plant growth and 15N2 incorporation following inoculation of sterile plants with Acetobacter diazotrophicus wild-type and *nif*- mutant strains. *Mol. Plant-Microbe Interact.* 14, 358–366 (2001).
4. Hurek, T. & Reinhold-Hurek, B. *Azoarcus* sp. strain BH72 as a model for nitrogen-fixing grass endophytes. *J. Biotechnol.* 106, 169–178 (2003).
5. Reinhold, B., Hurek, T., Niemann, E-G. & Fendrik, I. Close association of *Azospirillum* and diazotrophic rods with different root zones of Kallar grass. *Appl. Environ. Microbiol.* 52, 520–526 (1986).
6. Hurek, T., Reinhold-Hurek, B., Van Montagu, M. & Kellenberger, E. Root colonization and systemic spreading of *Azoarcus* sp. strain BH72 in grasses. *J. Bacteriol.* 176, 1913–1923 (1994).
7. Miché, L., Battistoni, F., Gemmer, S. & Belghazi, M. Reinhold-Hurek, B. Differential colonization and up-regulation of jasmonate-inducible defence proteins in roots of *Oryza sativa* cultivars upon interaction with the endophyte *Azoarcus* sp. *Mol. Plant-Microbe Interact.* 19, 502–511 (2006).

8. Egener, T., Hurek, T. & Reinhold-Hurek, B. Endophytic expression of *nif* genes of *Azoarcus* sp. strain BH72 in rice roots. *Mol. Plant-Microbe Interact.* 12, 813–819 (1999).
9. Reinhold-Hurek, B., Tan, Z. & Hurek, T. in Bergey's Manual of Systematic Bacteriology. Vol. 2. (ed. G.M. Garrity) 890–901, (Springer Verlag, New York, 2005).
10. Rabus, R. *et al.* The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Arch. Microbiol.* 183, 27–36 (2005).
11. Reinhold-Hurek, B. & Hurek, T. Life in grasses: diazotrophic endophytes. *Trends Microbiol.* 6, 139–144 (1998).
12. Battistoni, F. *et al.* Physical map of the *Azoarcus* sp. strain BH72 genome based on a bacterial artificial chromosome (BAC) library as a platform for genome sequencing and functional analysis. *FEMS Microbiol. Lett.* 249, 233–240 (2005).
13. James, E.K. & Olivares, F.L. Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. *Crit. Rev. Plant Sci.* 17, 77–119 (1998).
14. Thieme, F. *et al.* Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris pv. vesicatoria* revealed by the complete genome sequence. *J. Bacteriol.* 187, 7254–7266 (2005).
15. Reinhold-Hurek, B. *et al. Azoarcus* gen. nov., nitrogen-fixing proteobacteria associated with roots of Kallar grass (*Leptochloa fusca* (L.) Kunth) and description of two species *Azoarcus indigens* sp. nov. and *Azoarcus communis* sp. nov. *Int. J. Syst. Bacteriol.* 43, 574–584 (1993).
16. Reinhold-Hurek, B., Hurek, T., Claeyssens, M. & Van Montagu, M. Cloning, expression in *Escherichia coli*, and characterization of cellulolytic enzymes of *Azoarcus* sp., a root-invading diazotroph. *J. Bacteriol.* 175, 7056–7065 (1993).
17. Yurgel, S.N. & Kahn, M.L. Dicarboxylate transport by rhizobia. *FEMS Microbiol. Rev.* 28, 489–501 (2004).
18. Hurek, T., Reinhold, B., Fendrik, I. & Niemann, E.G. Root-zone-specific oxygen tolerance of *Azospirillum* spp. and diazotrophic rods closely associated with Kallar grass. *Appl. Environ. Microbiol.* 53, 163–169 (1987).
19. Koebnik, R. TonB-dependent trans-envelope signalling: the exception or the rule? *Trends Microbiol* 13, 343–347 (2005).
20. Paulsen, I.T. *et al.* Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat. Biotechnol.* 23, 873–878 (2005).
21. Crosa, J.H. & Walsh, C.T. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol. Mol. Biol. Rev.* 66, 223–249 (2002).
22. Crowley, D.E., Wang, Y.C., Reid, C.P.P. & Szaniszlo, P.J. Mechanism of iron acquisition from siderophores by microorganism and plants. *Plant Soil* 130, 179–198 (1991).
23. Dörr, J., Hurek, T. & Reinhold-Hurek, B. Type IV pili are involved in plant-microbe and fungus-microbe interactions. *Mol. Microbiol.* 30, 7–17 (1998).
24. Katzen, F. *et al. Xanthomonas campestris* pv. campestris *gum* mutants: effects on xanthan biosynthesis and plant virulence. *J. Bacteriol.* 180, 1607–1617 (1998).
25. Filloux, A. The underlying mechanisms of type II protein secretion. *Biochim. Biophys. Acta* 1694, 163–179 (2004).
26. He, S.Y., Nomura, K. & Whittam, T.S. Type III protein secretion mechanism in mammalian and plant pathogens. *Biochim. Biophys. Acta* 1694, 181–206 (2004).
27. Krause, A., Doerfel, A. & Göttfert, M. Mutational and transcriptional analysis of the type III secretion system of *Bradyrhizobium japonicum*. *Mol. Plant-Microbe Interact.* 15, 1228–1235 (2002).
28. Christie, P.J. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. *Biochim. Biophys. Acta* 1694, 219–234 (2004).
29. Hubber, A., Vergunst, A.C., Sullivan, J.T., Hooykaas, P.J. & Ronson, C.W. Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion system. *Mol. Microbiol.* 54, 561–574 (2004).
30. Camilli, A. & Bassler, B.L. Bacterial small-molecule signaling pathways. *Science* 311, 1113–1116 (2006).
31. Reinhold-Hurek, B., Maes, T., Gemmer, S., Van Montagu, M. & Hurek, T. An endoglucanase is involved in infection of rice roots by the not cellulose-metabolizing endophyte *Azoarcus* sp. BH72. *Mol. Plant-Microbe Interact.* 19, 181–188 (2006).
32. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194 (1998).
33. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8, 175–185 (1998).
34. Kaiser, O. *et al.* Whole genome shotgun sequencing guided by bioinformatics pipelines–an optimized approach for an established technique. *J. Biotechnol.* 106, 121–133 (2003).
35. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202 (1998).
36. Gordon, D., Desmarais, C. & Green, P. Automated finishing with autofinish. *Genome Res.* 11, 614–625 (2001).
37. Bartels, D. *et al.* BACCardI–a tool for the validation of genomic assemblies, assisting genome finishing and intergenome comparison. *Bioinformatics* 21, 853–859 (2005).
38. Meyer, F. *et al.* GenDB–an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31, 2187–2195 (2003).
39. McHardy, A.C., Goesmann, A., Pühler, A. & Meyer, F. Development of joint application strategies for two microbial gene finders. *Bioinformatics* 20, 1622–1631 (2004).
40. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402 (1997).
41. Tu, Q. & Ding, D. Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. *FEMS Microbiol. Lett.* 221, 269–275 (2003).
42. Reinhold-Hurek, B. & Shub, D.A. Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* 357, 173–176 (1992).

CHAPTER D

Annotation and analysis of part of *Azoarcus* sp. BH72 genome sequence

"If you do not expect to, you will not discover the unexpected."

—Heraclitus

1. INTRODUCTION

1.1 Prokaryotic genomes sequencing.

Since the *Haemophilus influenzae* genome sequence was completed in 1995 (Fleischmann *et al.*, 1995), the number of publicly available prokaryotic genomes increased exponentially up today, when more than 300 microbial projects had been completed (http://www.ncbi.nlm.nih.gov/Genomes/). As a consequence of this revolution, "genomics" has emerged as the basis for the molecular analysis of complete organisms. Genomics comprises the sequencing of DNA templates, the combination or assembly of the resulting DNA sequence data, as well as the interpretation or annotation of the assembled DNA sequence. The improvements in sequencing during the last decade have been accompanied by free access to these sequences in public databases (table 1). The big amount of data produced and deposited in these databases can assist scientists in comparing genomes, in relating species evolutionarily, as well as, in performing functional genomics studies (proteomics, transcriptomics and metabolomics). The use of these approaches allow to investigate for example, the roles that microorganisms play in complex ecosystems and in global geochemical cycles, study their diversity, as well as predict their impact on the productivity and sustainability of agriculture and forestry (Celestino *et al.*, 2004; Koonin, 1997; Koonin and Galperin, 1997).

Table 1: Five commonly used public sequence databases.

Databases	URL
GOLD™ Genomes Online Database.	http://www.genomesonline.org
The Institute for Genomic Research-Microbial Database (TIGR).	http://www.tigr.org/tdb/mdb/mdbinprogress.html
The EMBL Nucleotide Sequence Database.	http://www.ebi.ac.uk/embl/index.html
National Center for Biotechnology Information- Microbial Genome (NCBI).	http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi
The Welcome Trust Sanger Institute- Microbial Genome.	http://www.sanger.ac.uk/Projects/Microbes/

To perform those studies high quality genomes sequences are necessary as a resource for the reason that a- functional genomic needs accurate datasets for probe design or protein identification; b- genome organization can only be studied with a single, complete sequence; and c- comparative genomics needs complete gene repertoires.

Two principal steps for obtain a high quality genome sequence are the sequencing and the annotation process. The process of genome annotation can be defined as assigning meaning to sequences data, which otherwise would be almost devoid of information (Meyer *et al.*, 2003). The genome can be understood and further research initiated, only after regions of interest were identified and putative functions were assigned to proteins that codify those areas. The best way to perform a high quality annotation is by human expert. They can incorporate the information from databases, and correlate it with their own experiences. Due to the big amount of data, which is necessary to be evaluate from one genome, partial automation procedures can be establish, therefore bioinformatics tools are also necessaries for the success of any genome project.

1.2. Genomes relevant for the environment: *Azoarcus* sp. BH72 genome project.

When the first microorganisms started to be sequenced, preferences were given to those microbes of medical or animal interest. Soon after, the scientific community started to focus on the sequencing of microbial genomes that were relevant for the agriculture, environmental and biotechnology. In this context the *Azoarcus* sp. BH72 genome sequencing project was initiated. This is one of the projects within the cluster I: "Plant growth promoting bacteria" projects coordinated by the Competence Center in Bielefeld University **(https://www.genetik.uni-bielefeld.de/GenoMik/agriculture.html)**.

1.3. Protein classification: Clusters of Orthologous Groups of proteins (COG) database.

In order to extract the maximum amount of information from the rapidly accumulating genome sequences, all conserved genes need to be classified according to their homologous relationships. The database of **C**lusters of **O**rthologous **G**roups of proteins (COG) is an attempt on a phylogenetic classification of proteins from completed genomes on the basis of the orthology concept (Tatusov *et al.*, 1997), **(http://www.ncbi.nlm.nih.gov/COG).** This database has also the purpose to serve as platform for functional annotation of newly sequenced genomes and for studies on genome evolution (Tatusov *et al.*, 2000). To facilitate functional studies, the COG have been classified into 18 broad functional categories including a class for only general functional prediction and a class of uncharacterized COG (Table 2)(Tatusov *et al.*, 2001; Tatusov *et al.*, 2003).

Table 2: COGs functional categories.

Code	COGs*	Domains#	Description
			Information storage and processing
J	217	6449	Translation, ribosomal structure and biogenesis
K	132	5438	Transcription
L	184	5337	DNA replication, recombination and repair
			Cellular processes
D	32	842	Cell division and chromosome partitioning
O	110	3165	Posttranslational modification, protein turnover, chaperones
M	155	4079	Cell envelope biogenesis, outer membrane
N	133	3110	Cell motility and secretion
P	160	5112	Inorganic ion transport and metabolism#
T	97	3627	Signal transduction mechanisms
			Metabolism
C	224	5594	Energy production and conversion
G	171	5262	Carbohydrate transport and metabolism#
E	233	8383	Amino acid transport and metabolism
F	85	2364	Nucleotide transport and metabolism
H	154	4057	Coenzyme metabolism
I	75	2609	Lipid metabolism
Q	62	2754	Secondary metabolites biosynthesis, transport and catabolism
			Poorly characterized
R	449	11948	General function prediction only
S	750	6416	Function unknown

*Number of clusters in each category. # Number of domains in each cluster. #Categories studied from *Azoarcus* sp. BH72 in this work.

The COG classification criterion was used in *Azoarcus* sp. BH72 genome project for annotation task distribution between annotators. Categories studied in this work are shown in grey in table 2.

2. OBJECTIVES

Within the *Azoarcus* sp. BH72 genome-sequencing project context, the particulars objectives of this work were:

A: Annotation of genes which belong to the COG categories “Ion Transporters and Metabolism”, and “Carbohydrates Transport and Metabolism” (approximately 400).

B: Analysis of the data obtained from the annotation of these two COG categories.

3. MATERIAL AND METHODS

3.1. Genome sequence analysis: annotation.

Annotation of *Azoarcus* sp. BH72 genome sequence was done using GenDB software package (Krause *et al.*, 2006). This programme was developed by the Department of Genetic of the University of Bielefeld (Meyer *et al.*, 2003). It is an academic project and its use is free for research applications. In each annotator station a web interface of version 2.0 of GenDB **(https://www.cebitec.uni-bielefeld.de/groups/brf/software/gendb-2.0/cgi-bin/login.cgi)** was used for manual gene annotation. Version 2.0 offers graphical user interface (GUI), which integrates the visualization of the genome context, bioinformatics results and annotation as wells as BLAST analysis tools (Figure 1). The basic databases are update in a regular basis (Table 3).

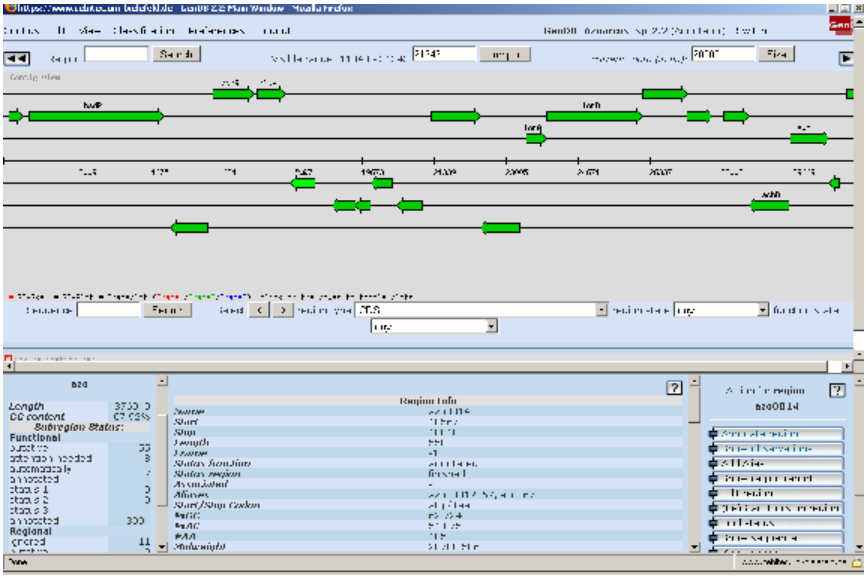


Figure 1: The GenDB 2.2 package for genome annotation. This screenshot shows the integrated genomic context visualization. Bioinformatics results and annotation windows are directly accessible for each predicted gene.

3.2. Comparison of genomes.

Numbers of coding genes from both categories were compared within three different groups of bacterial genomes. These bacterial genomes were grouped according to its predominant life style in: soil bacteria (*Azoarcus* sp. EbN1, *Geobacter sulfurreducens, Bacillus subtilis, Geobacillus kaustophilus, Nitrosomonas europea, Pseudomonas fluoresences* PF-1), plant pathogens bacteria (*Pseudomonas syringae, Ralstonia solanacearumn, Erwinia caratovora, Xylella fastidiosa, Xanthomonas axonopodis, Agrobacterium tumefaciens* C58), and beneficial plant associated bacteria (*Sinorhizobium meliloti, Mesorhizobium loti, Bradyrhizobium japonicum, Rhizobium etli, Rhizobium leguminosarum, Pseudomonas fluorescens* Pf-5). Information about each genome was collected from the NCBI database in the Genomic biology section, (http://www.ncbi.nlm.nih.gov/Genomes/).

Table 3: Main external Bioinformatics tools and databases integrated in the GenDB 2.0 package.

Tool/ database	Description	Reference/URL
BLAST	Pairwise similarity search for nucleotides and proteins sequences (**B**asic **L**ocal **A**lignment **S**earch **T**ool).	(Altschul *et al.*, 1990)
NCBI-nt	Non-redundant nucleotide database of the **N**ational **C**enter for **B**iotechnology **I**nformation	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=nucleotide
NCBI-nr	Non-redundant protein database of the **N**ational **C**enter for **B**iotechnology **I**nformation	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=protein
Blast/ KEGG	Pairwise similarity search against the **K**yoto **E**ncyclopedia of **G**enes and **G**enomes: integrated set of database on genes, proteins and metabolic pathways.	(Kanehisa *et al.*, 2004)
PSI-Blast/COG	**P**airwise **s**imilarity search against the **C**lusters of **O**rthologous **G**enes database, containing automatically classified gene products of full genomes.	(Schaffer *et al.*, 2001)
PSI-Blast SP	Pairwise similarity search against the Swiss-Prot protein sequence database.	(Schaffer *et al.*, 2001)
RPS Blast vs. CDD	Search against a Conserved domain database: includes protein domains from Pfam, SMART, and COG databases.	(Marchler-Bauer *et al.*, 2005)
TrEMBL	Non-redundant protein database of the European Molecular Biology Laboratory (**T**ranslated **EMBL** nucleotide database).	(Bairoch *et al.*, 2005)
Pfam	Search against a curated collection of protein families and domains using Hidden Markov Models (HMM) models.	(Bateman *et al.*, 2002)
Swiss-Prot	Manually curated protein database of the Expasy center (**Ex**pert **P**rotein **A**nalysis **S**ystem).	(Gasteiger *et al.*, 2003)
InterPro	Integrative resource including major protein signatures database (Prosite, Pfam, Prints, ProDOm, Smart, Tigrfams) and associated metatool (InterProScan).	(Mulder *et al.*, 2003)
COG	**C**lusters of **O**rthologous **G**enes database, containing automatically classified gene products of full genomes.	(Tatusov *et al.*, 2003)
SignalP	Signal peptide prediction using **N**eural **N**etworks (NN) and HMM.	(Nielsen *et al.*, 1997)
TMHMM	Transmembranes regions prediction based on HMM.	(Moller *et al.*, 2001)

3.3. Transporters classification.

Putative transporters were classified according to the transporters classification (TC) system (Busch and Saier, 2002) officially adopted by the International Union of Biochemistry and Molecular Biology (IUBMB). The system integrates structural, functional and evolutionary properties of proteins involved in transport and other transmembranes processes. The TC-DB (transport classification-database) is a web-based database of the TC-system (http://www.tcdb.org/). Numbers of transporters were compared with the closest related bacteria *Azoarcus* sp. EbN1 genome data (Rabus *et al.*, 2005).

3.4. ATP-binding cassette (ABC) transporters comparison.

To speculate how *Azoarcus* sp. BH72 copes with possible substrates from the environment, a comparison of number of predicted transporters was performed using different criteria.
Among: 1- all *Bacteria* genomes sequenced up to June 2005, 2- at least five genomes within different predominant life style groups (soil, plant pathogens, plant associated) and 3- at least five genomes within the mayor groups in *Bacteria* and *Archaea*.
For this purpose, the ABC-type transporters were chosen as a representative target gene set. The ABC-systems constitute one of the largest proteins superfamily and are ubiquitous in nature, being found in bacteria, eukarya and archaea (Higgins, 2001). Bacterial ABC-transporters are extremely versatile transport systems which vary greatly in size and function to either import or export a range of substances named as “allocrites” (Holland and Blight, 1999). The common feature of ABC transporters is the presence of highly conserved ATP-binding cassette which is the responsible for the generations of the energy for the transport of allocrites through out the ATP hydrolysis. Taking this in mind we assumed that one ABC-type transporter was present in the genome when an ATP-binding protein was present.
Numbers of predicted ABC-systems in all available bacterial genome sequences were compiled by searching in the genomes for the signature (Pfam=PF 00005) specifics for ABC-systems and compared with data obtained from the *Azoarcus* sp BH72 genome project. The Protein family database (Pfam), (http://www.sanger.ac.uk/Software/Pfam/) and the National Centre for Biotechnology Information database (NCBI), (http://www.ncbi.nlm.nih.gov/), were used to obtain data of all available sequenced bacterial genomes. Searching, blast and data analysis obtained were made using available tools in those databases. Transporters genomes comparison was done using the Pfam Genomic Distribution programme (http://www.sanger.ac.uk/cgi-bin/Pfam/genome_dist.pl), where a Pfam family for 2 or more species was compared using the 'Compare genomes' tool.

3.5. Central metabolism in *Azoarcus* sp. BH72: Sulfate, phosphate and iron transport and metabolism related genes.

Metabolic pathways were reconstructed automatically by GenDB programme (Meyer *et al.*, 2003), which integrate the information generated during the annotation process, with the databases listed in table 3. This programme gives an output graph of all potential pathways present in the *Azoarcus* sp. BH72 genome, with a corresponding GO, COG and E.C. number, function and name for each protein involved. When necessary, specific proteins sequences were check throught out the Comprehensive Enzyme Information System database (BRENDA) (http://www.brenda.uni-koeln.de/).

3.6. Siderophore biosynthesis pathways analysis in *Azoarcus* sp. BH72.

Putative biosynthesis pathways for catecholate siderophores were reconstructed automatically by the GenDB programme (Meyer *et al.*, 2003). Information was taken from the "Biosynthesis of siderophore group, non-ribosomal peptides" KEGG section. The output graph shows a KEGG pattern in which genes that belong to these pathways, with their corresponding E.C. numbers manually or automatically annotated, were highlighted by the programme. Additionally, all genes belonging to the enterobactin (a catecholate-type siderophore), biosynthesis pathway from *E.coli* (Crosa and Walsh, 2002), were check throught out *Azoarcus* sp. BH72 database using the GenDB Blast tool. On the other hand, all hydroxamate siderophore biosynthetic genes from *E.coli* (Martinez *et al.*, 1994), *S. meliloti* (Lynch *et al.*, 2001) and *R. leguminosarum* (Carter *et al.*, 2002) were check throught out *Azoarcus* sp. BH72 database using the GenDB Blast tool.

3.7. Siderophore detection in the growth medium.

Chemical detection of hydroxamates and catecholates siderophore-types, in bacterial supernatant of growth media was tested as previously described (Arnow, 1937; Atkin *et al.*, 1970). For that *Azoarcus* sp. BH72 was grown at 37°C on SM medium (Reinhold *et al.*, 1986), with and without the addition of 18mM Fe_3-EDTA, as well as in the same condition, but with the addition of 0,5 g/l yeast extract. From each culture medium 2 ml were taken and centrifuged 10 min. at 20000 xg. One ml of supernatant was used as a substrate in each chemical assay. All the glass material was treated overnight with 4N HCl and washed five times with high quality bidestilled water. Siderophore detection and growth as OD_{578nm} was measured during one week of growth.

For siderophore detection independent of their structure, a Chromeazurol-S (CAS) assay was used as was described before by Schwyn and Neilands (Schwyn and Neilands, 1987) with slightly modifications. The carbon source used was 0.1 M ethanol, instead of 9.9 mM glucose. Plates were incubated 4 days at 37°C.

4. RESULTS AND DISCUSSION.

4.1. Genome comparison.

General characteristics of bacterial genomes studied are shown in table 4. Within compared genomes, *Azoarcus* sp. BH72 contains the highest overall GC content (68%), follow by the plant pathogen bacterium *R. solanacearum* (67%).

Interestingly, bacteria which belong to the plant associated category, including N_2-fixing bacteria, possesses the biggest genome size among bacteria studied. Strain BH72 shared with this group a plant-associated lifestyle and, in the case of rhizobia, the feature to fix nitrogen. Results show that these features appear to be unrelated with the genome size and the GC content. Both parameters, size and GC content of *Azoarcus* sp. BH72, were slightly different to the closest related strain EbN1.

Table 4: General genome features of selected bacteria.

Genomes	Size (Mb)	GC content (%)	Predicted coding genes per genome size (%)
1) SOIL			
Azoarcus sp. EbN1	4.73	62.0	98.5
Bacillus subtilis	4.21	43.5	97.2
Geobacter sulfurreduncens	3.81	60.9	98.4
Geobacillus kaustophilus	3.59	52.1	96.9
Nitrosomonas europaea	2.81	50.7	92.3
Pseudomonas fluorescens Pf-1	6.44	60.5	98.4
2) PLANT PATHOGENS			
Agrobacterium tumefaciens C58	5.67	58.2	98.9
Erwinia carotovora pv. *atrospetica*	5.06	51.0	97.9
Pseudomona syringae pv. *tomato*	6.54	56.5	98.6
Ralstonia solanacearum	5.81	67.0	98.7
Xanthomonas axonopodis pv. *citri*	5.27	64.7	98.7
Xylella fastidiosa 9a5e	2.73	52.7	98.1
3) PLANT ASSOCIATED			
Bradyrhizobium japonicum	9.11	64.1	99.4
Mesorhizobium loti	7.6	61.0	99.2
Pseudomonas fluorescens Pf-5	7.07	63.3	87.0
Rhizobium leguminosarum	7.75	61.0	94.0
Rhizobium etli	6.5	58.0	92.4
Sinorhizobium meliloti	6.69	62.1	99.0
Azoarcus sp. BH72	4.38	68.0	91.2

Additionally, a comparison of the percentage of genes which belong to both studied COG category, in relation with the total number of gene-encoding proteins, from each genome studied, was done (Figure 2). Interestingly, strain BH72 has a higher amount of genes in both categories in comparison with the closest related strain EbN1 providing additional evidence that these two strains, which have different lifestyle, have a different genome composition. In both strains, a major number of genes were found in the ion transport and metabolism category. Moreover *Azoarcus* sp. BH72 contains one of the highest amounts of genes which belong to this category, among all genomes compared overcome only by *E. caratovora.* As in many plant-pathogens, the genome of *E. caratovora* posses a large amount of coding genes for extracellular plant cell-degrading enzymes (PCWDEs) (Bell *et al.*, 2004). Interestingly, in this bacterium, PCWDE production and full pathogenicity are linked to the production of the siderophores chrysobactin and achromobactin, which scavenge iron from the iron-poor environment of the host (Expert, 1999; Franza *et al.*, 2002). These observations on *E.caratovora* could explain the large amount of coding genes in both categories studied. Moreover, results obtained suggest that the large amount of coding genes in the Ion transport and metabolism COG category in *Azoarcus* sp. BH72 genome, a non-pathogenic bacterium, could provide some special mechanisms to adapt compete and survey in its specific niche.

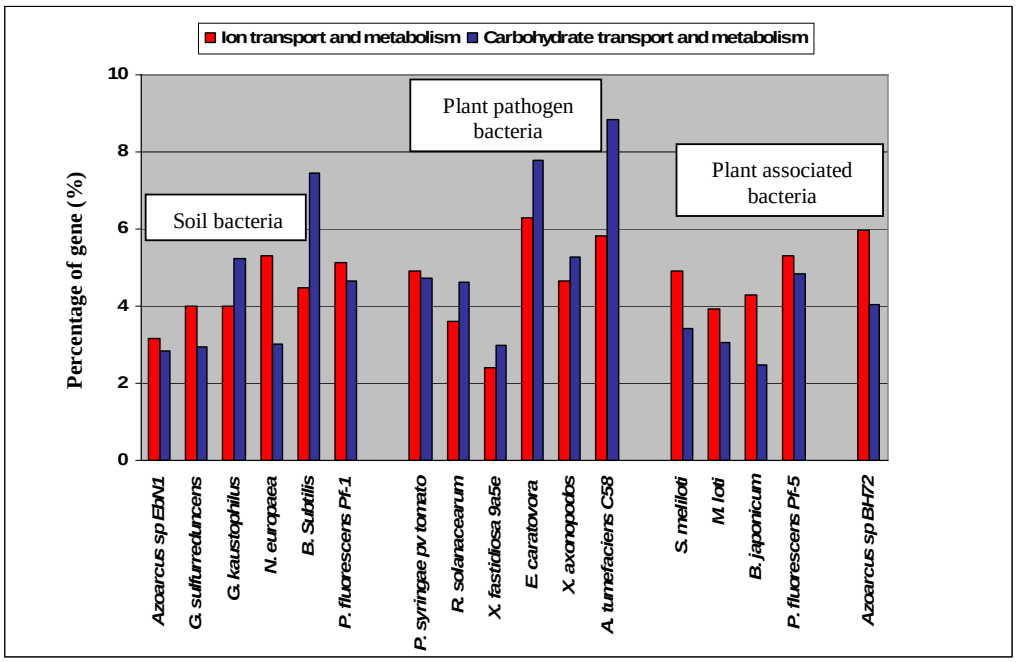


Figure 2: Percentage of encoding genes, among genomes compared, which belong to both COG categories studied.

4.2. Proteins of the "Ion Transport and Metabolism" C.O.G. category.

The high amount of genes in strain BH72 belonging to this category, compared with those of other genomes, brought up the need to be investigated by further analysis. Those genes were grouped according to their putative metabolic role assigned by similarity through the annotation process. The percentage of genes involved in each group, taken the total category gene number as the 100%, were graphed and shown in Figure 3. The highest percentage of genes in this category had homology to non-characterized genes in the databases (20%). On the other hand, most of the genes with putative known function are related to iron (16%) or sulphur (11%) transport and metabolism. Nitrogen uptake and metabolism represent 7% of the genes as we can expect from a genome of N_2-fixing bacteria. Additionally aromatic compound degradation consists also in 7% of the genes, a feature that shared with the specialized anaerobic aromatic-degrading *Azoarcus* strain EbN1 bacterium (Rabus *et al*., 2005). Genes involved in antibiotic resistance, which included several with high homology to permeases of the major facilitator superfamily (MFS), represent an additional 7%.

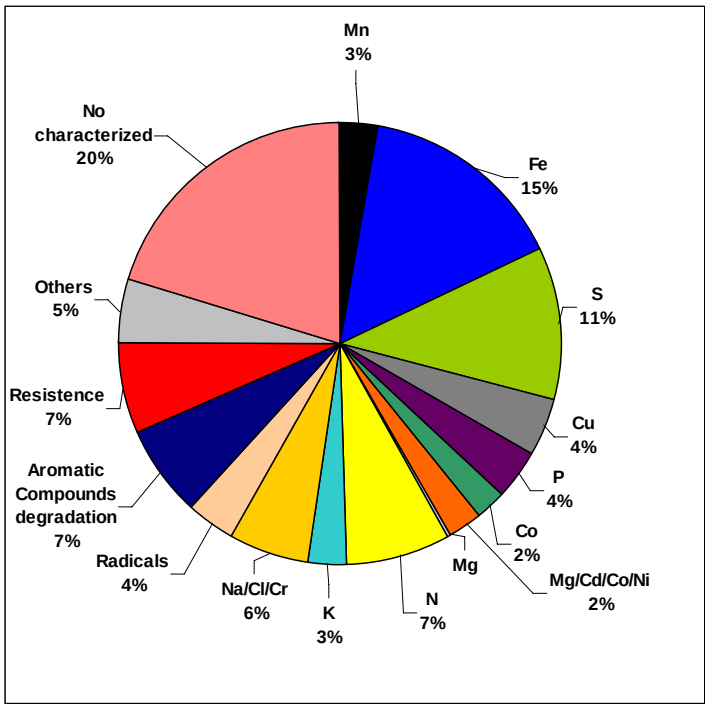


Figure 3: Functional classification of genes of *Azoarcus* sp. BH72 which belong to the category "Ion transport and metabolism".

4.2.1. Iron transport and metabolism related genes.

Taking in mind the high amount of genes probably involved in iron transport and metabolism, additional genome comparisons, as well as functional analysis were done.

Iron is an essential element for most organisms, including bacteria, but poses problems of toxicity and poor solubility. Faced with this problem imposed by their iron dependence, most bacteria have evolved several mechanisms to acquire iron from the various sources they may encounter in their diverse habitats, and to compete with other microbial organisms sharing the same space. When available, ferrous iron can be directly imported by the G protein-like transporter, FeoB (Andrews *et al.*, 2003). On the other hand, highly efficient ferric-iron acquisition systems are used to scavenge iron from the environment under iron-restricted conditions. These iron acquisition systems in bacteria have one of the two general mechanisms. The first, involves direct contact between the bacterium and the exogenous iron sources. The second mechanisms relies on molecules (siderophores and hemophores) synthesized and released by bacteria into the extracellular medium. These molecules can scavenge iron from various sources (Wandersman and Delepelaire, 2004).

In gram-negative bacteria, the iron sources recognition involves specific outer membrane receptors for a wide range of iron-containing molecules (Ratledge and Dover, 2000). A common feature of all these receptors is the involvement of the inner membrane TonB protein complex, (TonB-ExbBD). TonB-like proteins appear to exist in Gram-negative bacteria for the purpose of traducing the proton-motive force energy from the cytoplasmic membrane, where it is generated, to the outer membrane, where it is needed for active transport of iron siderophores, vitamin B12 and iron-binding proteins (Postle and Kadner, 2003). Subsequent transport of iron-compounds across the inner membrane involves specific, periplasmic binding protein-dependent ABC permeases (ABC-transport systems) (Figure 4).

4.2.1.1. Outer membrane TonB-dependent genes.

The genome of strain BH72 contains a large number (22) of encoding genes for putative TonB-dependent outer membrane receptors (Table 5). This is twice the number of genes described for *Azoarcus* sp. EbN1 and even more in comparison to N_2-fixing symbionts, such as *B. japonicum* (13), *S. meliloti* (9) and *M. loti* (1). Interestingly from this set of genes, four are not present in *Azoarcus* sp. EbN1 genome (azo0378, azo2156, azo2347 and azo3836). Additionally two of these genes (azo2156, azo3836) are not even present in the *Pseudomonas fluorescens* Pf5 genome, a plant associated bacterium known for its capacity to produce and uptake a wide range of siderophores (Loper and Henkels, 1999; Paulsen *et al.*, 2005). The genome of *P. fluorescens* Pf5 contains 45 genes for putative TonB-

dependent receptors, where sequences of 28 of these are similar to known receptors of ferric-siderophore complexes found in organisms (Paulsen *et al.*, 2005).

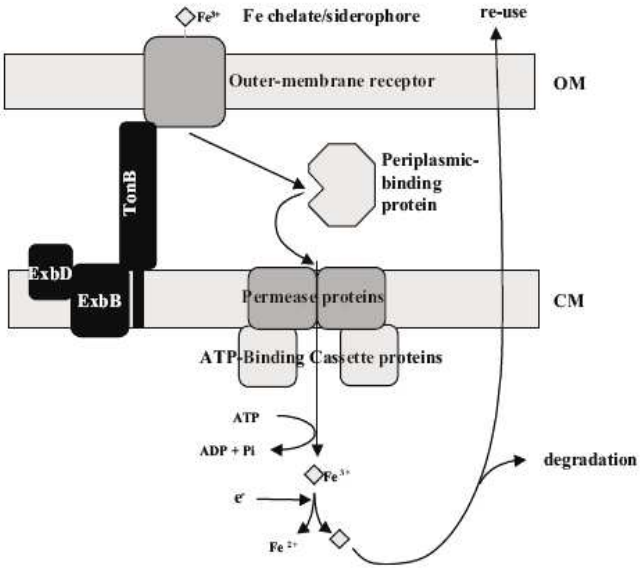


Figure 4: Schematic representation of siderophore-mediated iron uptake in Gram-negative bacteria (Andrews *et al*., 2003)**. OM:** outer membrane, **CM:** cytoplasmic membrane.

Siderophores are generally synthesized and secreted by bacteria, fungi and monocotyledonous plants, in response to iron restriction (Crowley *et al.*, 1991). These compounds usually form hexadentate octahedral complexes with ferric iron and typically employ hydroxamates, α-hydroxycarboxylates and cathecoles as extremely effective Fe^{3+} ligands (Winkelmann, 2002). Approximately 500 siderophores have been characterized which can be classify according to the functional groups they use as a ligands (Neilands, 1981, 1995). Putative ligands of each TonB-dependent outer membrane receptors annotated in the strain BH72 genome are shown in table 2. *Azoarcus* sp. BH72 has several receptors for hydroxamates and cathecoles types siderophores, ferricitrate, vitamin B12 and colicins as well as receptors whose precise specificity is not clear. Outer membrane TonB-dependent receptors were reported as important for perception of environmental signals and to be associated with pathogenicity of plant pathogens (Brito *et al.*, 2002; Koebnik, 2005; Marenda *et al*., 1998). The presence of this set of TonB-outer membrane genes in *Azoarcus* sp. BH72 genome suggests different capacities to acquire iron, as well as the putative specific signaling role of these proteins in *Azoarcus* sp. BH72 ecology.

Table 5: TonB-dependent outer membrane receptors annotated in *Azoarcus* sp. BH72 genome.

Gene number	Gene name	Gene product	Putative ligand	Best hit in *Azoarcus* EbN1 genome. (e-value)	Best hit in *P. fluorescens* Pf5 genome. (e-value)
azo0093	*irgA1*	**IrgA1**: putative TonB-dep. receptor.	Enterobactin (catecholate)	Hypothetical protein. eba3244, (7.8e-76)	TonB-dep. receptor. YP_258866, (2.0e-55).
azo0229	*fhuA1*	**FhuA1**: probable ferrichrome-iron TonB-dep. receptor.	Ferrichrome (hydroxamate)	TonB-dep. receptor. ebA3937, (3.1e-81).	TonB-dep. receptor.YP_258063, (4e-127).
azo0378[*]		Putative TonB-dep. receptor.	Alcalagine (hydroxamate), Vitamin B12.	No hits.	TonB-dep. receptor.YP_261992, (5.9e-98).
azo0607		**PfuA**: putative TonB-dep. receptor	Desferrioxamine B (hydroxamate)[&]	TonB-dep. receptor. ebA5333, (1.1e-75).	TonB-dep. receptor. YP_258546, (5.7e-83).
azo1830		Putative outer membrane receptor.	Enterobactin (catecholate), colicins.	Low similarity. eba2117, (3.0e-16).	Outer membrane ferric-enterobactin receptor. YP_259770, (2.4e-19).
azo2156[§]		Probable TonB-dep. receptor.	Ferri-citrate, colicins	No hits.	No hits.
azo2217[#]		Putative TonB-dep. receptor.	Ferri-citrate	TonB-dep. receptor. ebA6146, (2.1e-28).	No hits.
azo2347[*]	*fyuA*	**FyuA**: putative TonB-dep.receptor.	Bacteriocin, pesticin and yersinibactin (catecholate)	No hits.	TonB-dep. receptor. YP_257359, (1.7e-74).
azo2375	*fecA*	**FecA**: putative ferric-citrate TonB-dependent receptor.	Ferri-citrate	TonB-dep. receptor. ebA6149, (3.1e-26).	Ferric-vibriobactin receptor. YP_260722, (1 2.8e-67).
azo2396[#]		Putative TonB-dep. receptor.	Schizokinen (hydroxamate)	Low similarity. ebA6149, (3.5e-08).	No hits.
azo2400		Putative TonB-dep. receptor.	Vitamine B12,colicins catecholates	TonB-dep. heme receptor. ebA2117, (1.9e-14).	TonB- dep receptor, B12 family. YP_262579, (1 4.8e-21).
azo2511	*fpvA1*	**FpvA1**: putative TonB-dep.receptor.	Pyoverdine (hydroxamate)	TonB-dep. receptor. ebA5333, (4.3e-31).	Outer membrane ferri-pyoverdine receptor. YP_260420, (1 3e-133).
azo2610	*fhuA2*	**FhuA2**: putative ferrichrome-iron TonB-dep. receptor.	Ferrichrome (hydroxamate)	TonB-dep.receptor. ebA3937, (2e-124).	Outer membrane siderophore receptor. YP_258063, (1.4e-91)
azo2978[#]		Putative TonB-dep. receptor	Enterobactin (catecholate)	TonB-dep. receptor. ebA6096, (6.2e-22).	no hits
azo2994	*fpvA2*	**FpvA2**: probable TonB-dep.receptor.	Pyoverdine (hydroxamate)	TonB-dep. receptor. ebA5333, (1.3e-34).	Outer membrane ferri-pyoverdine receptor.YP_260588, (6e-157).
azo2997	*fhuF*	**FhuF**: putative TonB-dep.receptor.	Desferrioxamine B (hydroxamate)[&]	TonB-dep. receptor. ebA5333, (5.9e-61).	Outer membrane ferrichrome-iron receptor. YP_262764. (4e-129).
azo3023[#]		Probable TonB-dependent receptor.	Ferri-citrate	TonB-dep. receptor. ebA6096, (0.0).	no hits
azo3555		Putative TonB-dep. receptor.	Vitamin B12, catecholates	TonB-dep. receptor. ebA4013, (2.9e-24).	TonB-dep receptor, B12 family. YP_262579. (7.8e-78).
azo3556		Putative TonB-dep. receptor.	Vitamin B12, catecholates	TonB-dep.receptor. ebA4013, (0.0).	TonB-dep. receptor, B12 family. YP_262579. (5.4e-21).
azo3836[§]		Probable TonB-dependent receptor.	-	no hits	no hits
azo3844		Putative TonB-dep. receptor	-	Low similarity. ebA3937, (1.5e-09).	TonB-dep. receptor. YP_258866. (5.5e-12).
azo3852	*irgA2*	**IrgA2**: Putative TonB-dep. receptor	Enterobactin (catecholate)	TonB-dep. receptor. ebA5333, (1.1e-75).	TonB-dep. receptor. YP_262764. (2.4e-54).

[*]No similarity in strain EbN1 genome. [#]No similarity in *P. fluorescens* genome. [§]Only presents in strain BH72 genome.
[&]Fungic siderophore.

4.2.1.2. Absence of siderophore biosynthetic pathways.

Surprisingly, blast analysis results showed no evidence for known hydroxamate or catecholate siderophore biosynthetic pathways in *Azoarcus* sp. BH72 genome. These observations are in agreement with experimental results obtained in this work. No positive reaction was detected when specific test for, hydroxamate or catecholate, were applied to the supernatant of strain BH72 grown in iron limited SM medium. Moreover, no positive reaction was observed when Chromeazurol S (CAS) assay was used to detect siderophores independent of their structure as visualized by a yellow halo around the colonies (Figure 5). In these experiments *Azoarcus tolulyticus* Td1 and *Azospirillum brasilense* Sp7 strains were also tested for siderophore production, while *Pseudomonas fluoresences* and *Sinorhizobium* sp. NGR1 strains were used as a positive control (Carson *et al.*, 2000; Paulsen *et al.*, 2005).

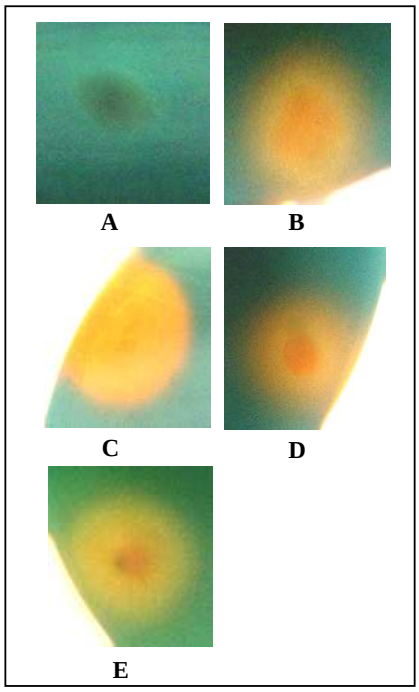


Figure 5: Siderophore production on Chromeazurol S (CAS) indicator plates. A: *Azoarcus sp.* BH72, **B:** *Pseudomonas fluorescens*, **C:** *Azoarcus tolulyticus* Td1, **D:** *Azospirillum brasilense* Sp7, **E:** *Sinorhizobium sp.* NGR1.

As it is shown in figure 5, positive reactions were obtained in the close related soil-borne bacterium *A. toluliticus* Td1 (C) and in the diazotrophic bacterium *A. brasilense* Sp7 (E), but not in *Azoarcus* sp. BH72. Apparently, strain BH72 is highly adapted to obtaining chelated iron from other sources, since fungi and monocotyledonous plants also produce siderophores (Crowley *et al.*, 1991). To confirm this hypothesis additional CAS assays were carry out with the aim to test if the fungus partner *Acremonium alternatum* strain 2003, isolated also from

Kallar grass (Hurek *et al*., 1995), has the capacity to produce these compounds. Positive siderophore production on CAS plates was only detected, by fungus 2003, when *Azoarcus* sp. BH72 was streaked in contact with the fungal mycelium (Figure 6).

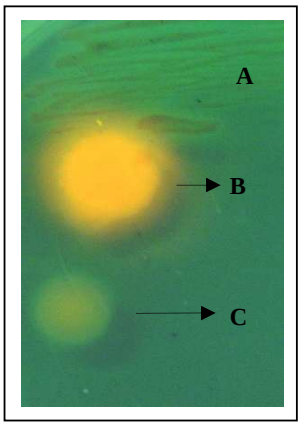


Figure 6: Siderophore production on Chromeazurol S (CAS) indicator plates. A: *Azoarcus* sp. BH72, **B:** *Azoarcus* sp. BH72 in contact with *Acremonium alternatum*, **C:** *Acremonium alternatum.*

Moreover, when autoclaved cells of strain BH72, were placed onto CAS plate in contact with *A. alternatum* mycelium, no positive siderophore production was detected (data not shown). These results suggest that is not the contact between the two partners, which activates the siderophore biosynthesis in *A. alternatum*. A hypothesis, in which a putative diffusible signal compound produced by the bacterium is involved, will be interesting to be tested. Taken in mind all results obtained, we can speculate that most probably in their natural environment strain BH72 use as iron source the heterologous ferric-siderophores synthesized and excreted by the fungal partner, as was reported for other bacteria (Jurkevitch *et al.*, 1993; Llamas *et al.*, 2006; Loper and Henkels, 1999). However, additionally a putative siderophore production by the plant partner and posterior internalization by strain BH72 might also be possible.

4.2.1.3. tonB-exbBD complex genes.

The energy required to allow the outer membrane receptors to deliver the ferric-siderophore complex into the periplasmic spaces, derives from a complex of proteins (TonB, ExbB, ExbD) in the cytoplasmic membrane driven directly or indirectly by the protonmotive force (Postle and Kadner, 2003). Two complete operons which code for a TonB-ExbBD complex are present in *Azoarcus* sp. BH72 genome, and have a high degree of similarity in sequences with the same *tonB-exbBD1* operon in *Azoarcus* sp. EbN1 (ebA1839, 1841 and 1842). Only one of those operons in strain BH72 has the same gene arrangement as in strain EbN1, except for the opposite orientation (Figure 4).

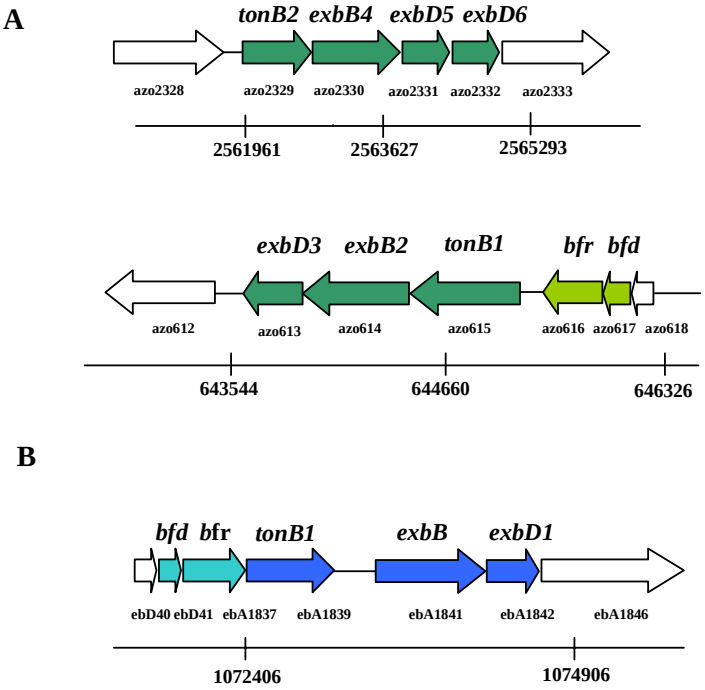


Figure 7: Schematic representation of TonB-complex genes arrangement. A: Strain BH72 genome, **B:** Strain EbN1 genome.

Additionally, strain BH72 genome presents two extra *exbBD* operons. One of them *exbB1D1D2*, has also a high degree of similarity with the mentioned *exbBD1* operon from strain EbN1. The third one *exbB3D4*, has high degree of similarity and the same gene arrangement to a different *exbBD* operon in strain EbN1 (ebA1468 and ebB176). In these two extra *exbBD* operons, no *tonB* gene was annotated close to those operons in both strains. This indicates that most probably these extra operons are involved in another type of transport process.

4.2.1.4. ABC-iron related transporters.

In bacteria, iron compounds are usually transported across the cytoplasm membrane by an ABC-transport system (Koster, 2001). Two complete and one incomplete, ABC-transport systems operons, were annotated in strain BH72 genome as probably involved in iron transport across the cytoplasm membrane (Table 6). The incomplete operon, in which the substrate-binding protein is missing, and the *fbpA1B1C1* operon, share high degree of similarity and the same gene arrangement, with ABC-transporters operons in strain EbN1 genome. Interestingly, in strain BH72 the incomplete operon is close to two outer membrane TonB-dependent receptors (azo3555 and azo3556), which most probably are part of the same

operon. Those receptors share high degree of homology with vitamin B12 TonB-dependent receptors (Table 5). Moreover, they are located close to genes related with vitamin B12 metabolism, which could mean that most probably are involved in vitamin B12 uptake. On the other hand, the additional complete operon *fbpA2C2B2*, has significant sequence similarity with ABC-iron related transporters in the strain EbN1, only in the ATPase-coding gene (Table 6).

Table 6: ABC-iron transport system (TC 3.A.1.10.1), in *Azoarcus* sp. BH72 genome.

Gene number.	Gene name.	Gene product.	Best hits gene product and organism (Blast P-value).	Best hits in *Azoarcus* sp. EbN1. (Blast P-value).
azo2205	*fbpA1*	Iron (III) transport system, probable substrate-binding protein.	Putative iron binding protein. *Azoarcus* sp. EbN1, (6e-159).	ebA4918
azo2206	*fbpB1*	Iron (III) transport system, probable permease protein.	Binding-protein-dependent transport systems, inner membrane component. *Dechloromonas aromatica* RCB. (0.0)	Putative iron utilization protein.ebA4893, (0.0).
azo2207	*fbpC1*	Iron (III) transport system, probable ATP-binding protein.	Probable transport system ATP-binding protein. *Azoarcus* sp. EbN1, (1e-166).	ebA4892
azo3825	*fbpA2*	Ferric transport system, putative substrate-binding protein	Extracellular solute-binding protein, family 1. *Syntrophobacter fumaroxidans* MPOB. (3.3e-68).	-
azo3826	*fbpC2*	Ferric transport system, probable ATP-binding protein.	Putative ABC transport ATP-binding subunit. *Bordetella parapertussis* 12822. (8.3e-77).	Putative iron transport system ATP-binding protein. ebA2111, (4.8e-57).
azo3827	*fbpB2*	Ferric transport system, putative permease protein.	Probable ABC-transporter permease protein. *Chromobacterium violaceum* ATCC 12472. (1.6e-95).	Putative iron utilization protein. ebA4893, (2.8e-15).
azo3553	-	ABC-type transporter, putative ATPase component.	ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components. *Dechloromonas aromatica* RCB. (2.4e-59)	ATP-binding ABC transporter protein. ebA4020, (1e-35).
azo3554	-	ABC transporter system, probable permease protein.	Probable transmembrane ABC transporter protein, permease. *Azoarcus* sp. EbN1, (4e-132)	ebA4019

4.2.1.5. Iron regulators related genes.

Usually bacteria regulate their iron metabolism in response to iron availability. In *E.*coli and many other bacteria, this regulation is mediated by the ferric-uptake regulator protein (Fur), which acts as positive repressor (Escolar *et al.*, 1999). Depending on the bacteria, 50 to 100 genes are negatively regulated by Fur in iron-rich media. These genes are involved in diverse functions including iron and heme acquisition, virulence factors production, and acid resistance (McHugh *et al.*, 2003; Vasil and Ochsner, 1999). An exception was reported in *S.meliloti* and *R. leguminosarum,* in which Fur is involved in the manganese metabolism. In those bacteria, an additional Fur-like protein was reported to be involved in the iron

metabolism regulation (Chao *et al.*, 2004; Diaz-Mireles *et al.*, 2004; Platero *et al.*, 2004; Viguier *et al.*, 2005).

Two genes, azo2578 and azo0644 which codified for a putative Fur protein, were annotated in *Azoarcus* sp. BH72 genome. Both have high degree of similarity with genes present in the genome of strain EbN1, ebB167 and ebA1809, which were also annotated as *fur*-like genes. Nevertheless, additional functional genomics approaches are necessary to elucidate the exact function of those regulators genes in *Azoarcus* sp. BH72 metabolism.

An alternative iron-uptake system regulation was described in *E.coli*, which involve the ferric-citrate transport system is code by the *fecIR fecABCDE* operon. The transcription of the *fec* transport genes is subjected to double control. Cells first detect iron deficiency and respond by synthesis of the regulatory proteins FecI and FecR, which initiate transcription of the *fec* transport genes. Additionally, transcription of the *fecIR* regulatory genes and the *fec* transport genes is repressed by the Fe^{2+}- Fur complex (Braun *et al.*, 2003). Two copies of this type of regulatory systems are present in *Azoarcus* sp. BH72 genome. One *fecIRA* (azo2374, azo2375, azo2376), has a high degree of similarity to the *E.coli* ferric-dicitrate transport system, while the other (azo2511, azo2512 and azo2513), has similarity to the *P.aeruginosa* pyoverdine transport system. Interestingly none of them are located close to an ABC-transport system. Form these operons mentioned; only the outer membrane receptor component genes have similarity with strain EbN1 genes.

4.2.1.6. Intracellular iron homeostasis.

As mentioned, iron is an essential element but has problems of solubility and toxicity. That means that bacteria should be able to control the gene expression of these systems, as was described just above and moreover, the exactly cytoplasmic amount of this metal. *Azoarcus* sp. BH72 genome encode for three *bfr* copies genes (azo247, azo616, and azo1784), which codify for iron storage bacterioferritin proteins, as well as a *bfd* gene (azo617), which codify for a bacterioferritin-associated ferredoxin protein. Bacterioferritins and associated proteins were reported to be involved in the intracellular iron-homeostasis control, as well as in oxygen radical defense (Andrews et al., 2003; Wandersman and Delepelaire, 2004). *Azoarcus* sp. BH72 *bfr* and *bfd* genes, have they paralogous in strain EbN1 genome, with the same gene arrange only in two bacterioferritin copies (azo616 and azo1784). Interestingly, one of this set of gene (azo616, azo617), are located in the same operon with *tonB1exbB2D3* (Figure 7). This gene arrange, support more the idea that the mentioned TonB-complex, is involved in iron uptake.

4.2.1.7. Ferrous-iron transport related gene.

Additionally to the high-affinity iron transport systems described, a *feoB* (azo0606) gene, which codified for a ferrous iron transport protein B, is present in strain BH72 genome. The FeoB is involved in the iron II uptake under anaerobic conditions, in which this ion is available (Kammler et al., 1993). Interestingly the *feoB* gene has not similarity in the strain EbN1 genome, showing different abilities to uptake iron in both related strains.

4.2.2. Sulfur and phosphorus transport and metabolism related genes.

4.2.2.1. Sulfur assimilation

Taken in mind that the second high amount of genes (11%), within the "Ion transport and metabolism" COG category, were classified as probably involved in sulfur transport and metabolism (Figure 3), further genome comparisons were done. As iron, sulfur plays an important role in bacterial metabolism. Sulfur is cofactor of several enzymes (e.g. aconitase, respiratory chain complexes, sulfite and nitrite reductases), as well as a component of the amino acids methionine and cysteine. Most of genes related with sulfur metabolism annotated in this category code for transporters or transferases. Putatives pathways, enzymes and genes are shown in figure 8. Strain BH72 genome has two *sulP* gene copies, which code for an H^+-driven sulfate permease (SulP). Additionally, an ABC-transporter operon *cysPTWA* for sulfate and thiosulfate is present in strain BH72 genome. Both *sulP* copies and *cysPTWA* operon sequences have high degree of similarity and the same gene arrangement, with genes that codified for the same type of transporters in strain EbN1.

A thiosulfate sulfurtransferase encoded by *tst*, could convert thiosulfate to sulfide, which can be directly used for cysteine synthesis via the CysM and CysK isoenzyme (azo1077 and azo2399). On the other hand, assimilatory sulfate reduction probably occurs via APS (adenylylsuphate) and PAPS (3´-phosphoadenylyl-sulfate), in contrast to *Azoarcus* sp. EbN1 and *Allochromatium vinosum* (Neumann *et al.*, 2000; Rabus *et al.*, 2005). A sulfate adenylyltransferase (CysD) and a bifunctional enzyme sulfate adenylyltransferase/APS kinase (CysNC), activates sulfate to APS and PAPS respectively, which is subsequently reduced to sulfite by PAPS-reductase (CysH). The latter is reduced to sulfide by sulfite-reductase (CysI) (Figure 8). Most probably and as in strain EbN1 the structural *cys* genes are transcriptionaly controlled by the LysR-type regulator CysB present in *Azoarcus* sp. BH72 genome (azo434).

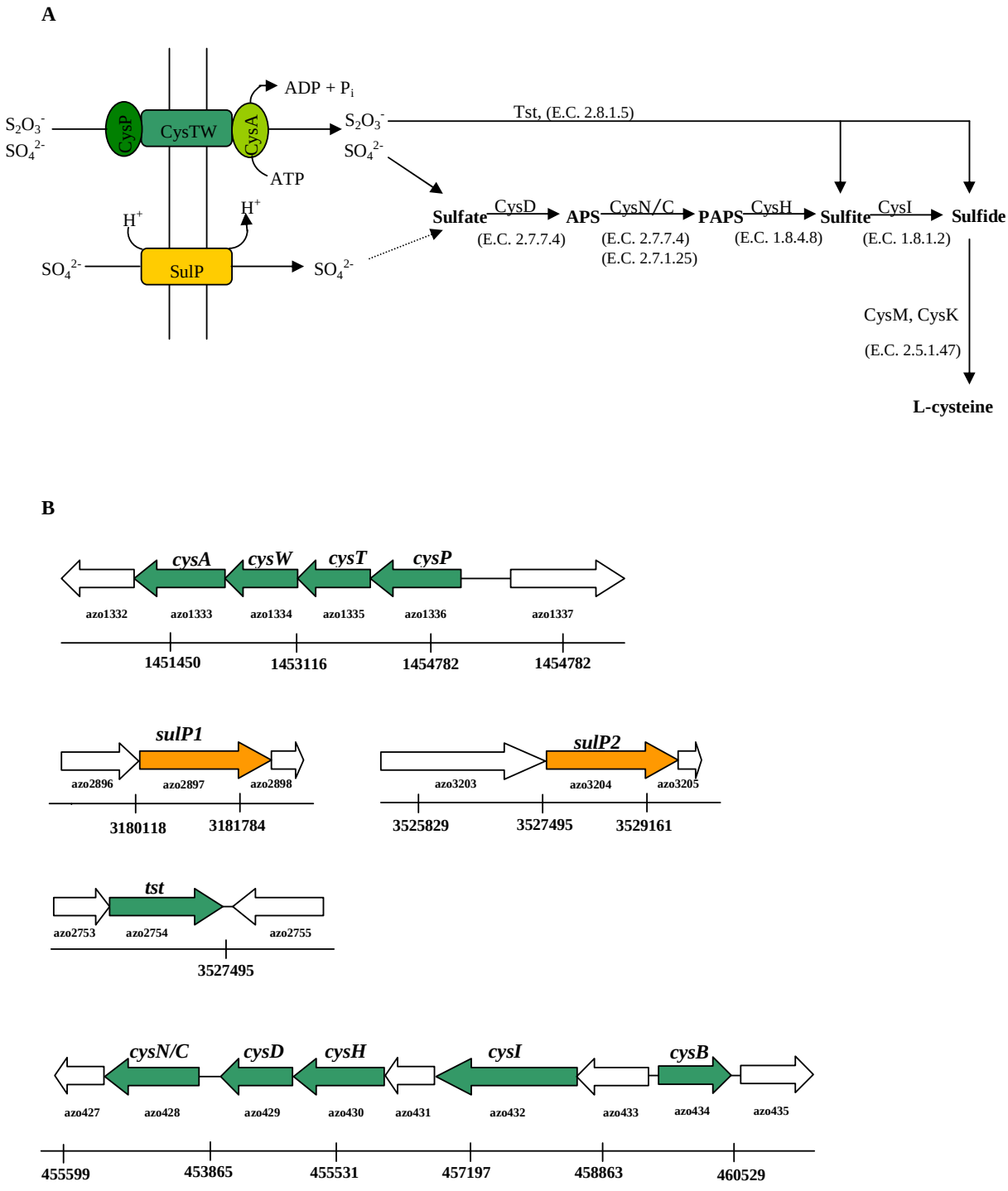


Figure 8: Transport and assimilation of sulphate. A: pathways. **B:** genes.

4.2.2.2. Phosphate assimilation

Phosphate uptake is probably mediated by an ABC-transporter (*pstSCAB*), from which two copies of the *pstS* gene were present in the genome sequence (Figure 9). The gene order within the *pst* operon is similar to strain EbN1 and *E.coli*, except for the *phoU* gene (azo2827), which is absent in operons of both *Azoarcus* strains. This *phoU* encodes a putative regulator of the operon and is present in other genome position as in strain EbN1. In *E.coli* the *pst* operon was reported to be induced under phosphate starvation (Aguena *et al.*, 2002). Additionally, strain BH72 genome presents two probable low–affinity phosphate transporters proteins (azo2653 and azo2135), which belong to the Inorganic Phosphate Transporter (PiT) family. These types of transporter in *E.coli* were found constitutively expressed, and dependent on the proton motive force for energy (Harris *et al*., 2001).

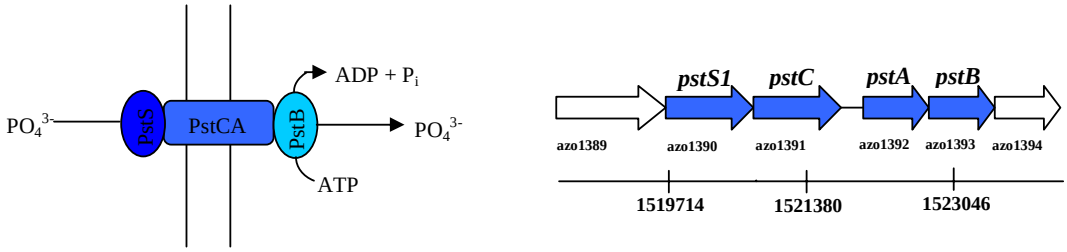


Figure 9: Transport and assimilation of phosphate. Left: transporters. **Right:** genes.

4.3. Protein of the "Carbohydrate Transport and Metabolism" C.O.G. category: Central metabolism in *Azoarcus* sp. BH72.

4.3.1. C4-dicarboxylates transport and glyoxylate shunt

C4-dicarboxylates like succinate, fumarate and malate are transported and metabolized by bacteria under aerobic or anaerobic conditions. In aerobic growth, the substrates serve as carbon sources and become oxidized to CO_2 in the citric acid cycle (Janausch *et al*., 2002). C4-dicarboxylate carriers have to fulfill a large number of different functions in bacteria, and catalyse symport (uptake or efflux) of substrate with H^+ or Na^+ (Janausch *et al.*, 2002). Major carbon sources for strain BH72 are dicarboxylic acids and ethanol (Reinhold-Hurek *et al.*, 2005). In agreement with this, four complete and two incomplete copies of the TRAP-transport system (*dctPQM*) were annotated in the genome, which are a family of high-affinity transporters for C4-dicarboxylates. This number of transporters is surprisingly low, in comparison with the 15 copies reported in strain EbN1 genome (Table 7). The TRAP-transporter family represents a novel family of secondary carriers composed of two membrane integral proteins and a periplasmic solute binding protein (Kelly and Thomas,

2001). This transport system is unique in combining a high-affinity periplasmic solute-binding protein (DctP) with an H^+ -driven permease (DctM) (Figure 10A) (Forward *et al.*, 1997; Rabus *et al.*, 1999). In rhizobia as well as in *Rhodobacter capsulatus*, C4-dicarboxylates transport systems and the corresponding pathways are induced only in the presence of external C4-dicarboxylates, where a two component regulatory system is involved (Forward *et al.*, 1997; Janausch *et al.*, 2002; Shaw *et al.*, 1991). The genes *dctRS* coding for sensor and regulator, are also present in strain genome, having the same order as in strain EbN1 (Rabus et al., 2005). These genes are located downstream from one set of C4-dicarboxylate transporter (*dctP2Q2M2*) (Figure 10B), which means that most probably this operon is under the control of this two component regulatory system.

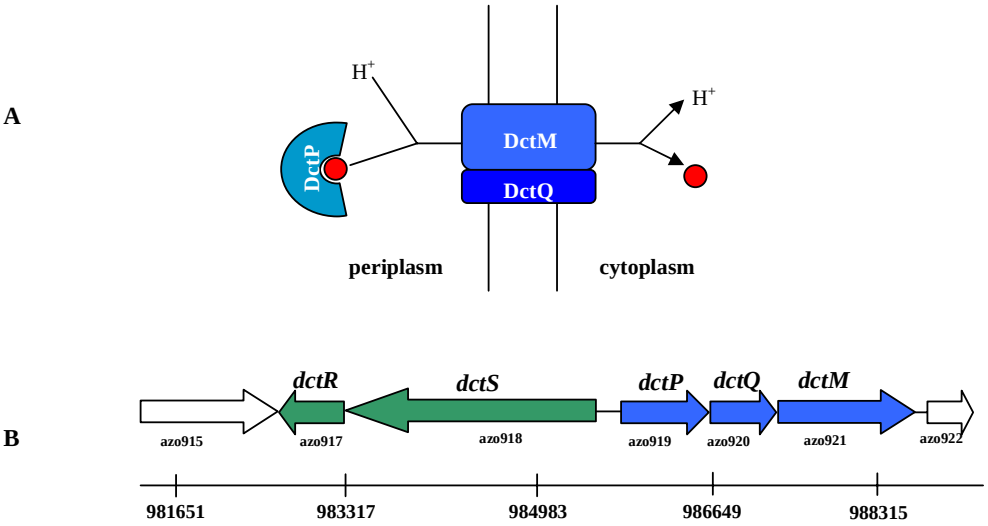
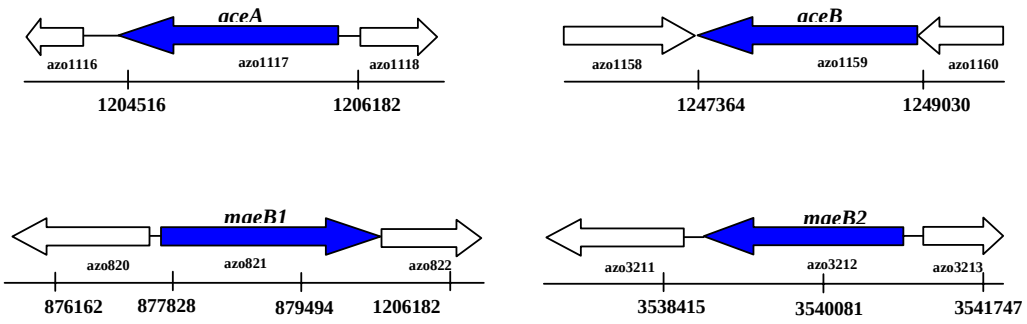


Figure 10: Uptake of C4-dicarboxylates by TRAP transporter. A: transporter; **B:** gene arrangement of one of the operons. **DctP:** periplasmic solute binding protein; **DctM**: H^+-driven permease; **DctQ**: transmembranes of unknown function.

As in symbiotic rhizobia this C4-dicarboxylate transporters might be of vital importance during the association with the host plant (Yurgel and Kahn, 2004). C4-dicarboxylates are probably metabolized via the glyoxylate shunt pathway to the level of malate, which is then decarboxylated to pyruvate by the malic enzyme (MaeB) (Figure 11). Terminal oxidation of pyruvate is then performed by pyruvate dehydrogenase complex and the TCA-cycle. As in strain EbN1 two paralogs of malic enzyme encoding genes (*maeb1* and *maB2*), are present in the genome of strain BH72.

A



B

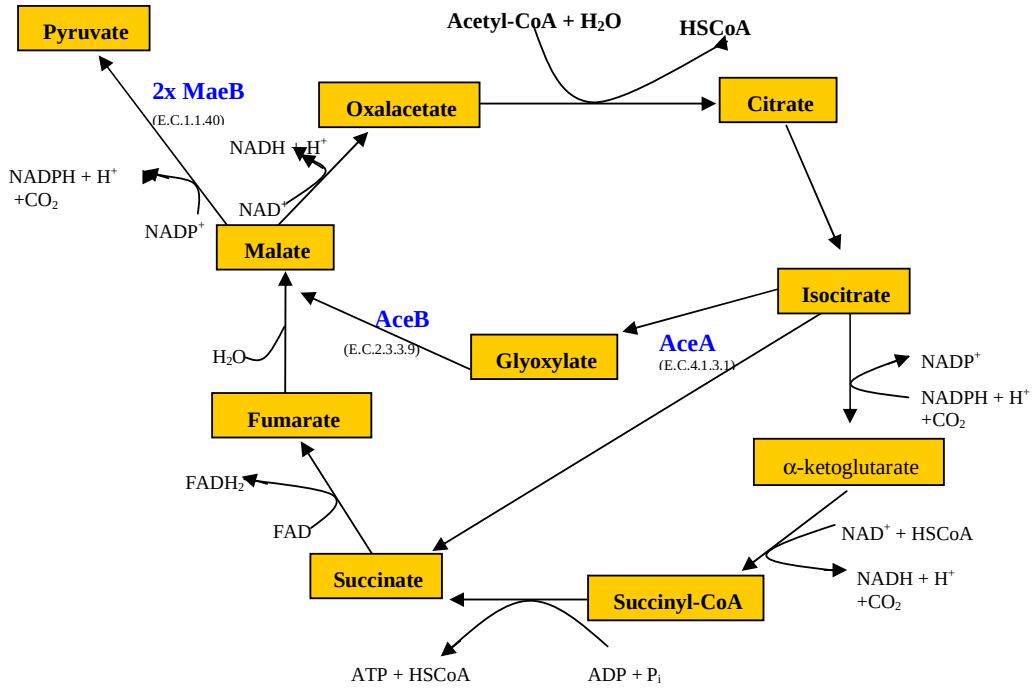


Figure 11: Glyoxylate shunt. A: genes and **B:** pathway**.** Specific enzymes involved in the glyoxylate shunt are present in blue. **AceA**: isocitrate synthase; **AceB**: malate synthase; **MaeB**: malic enzyme; 2x, two copies.

4.3.2. Tricarboxilic acid (TCA) cycle, glycolysis, gluconeogenesis and phosphoenolpyruvate: sugar phosphotransferase system (PTS).

Like the other plant-associated species of *Azoarcus* spp., strain BH72 has a strictly respiratory type of metabolism and does not grow on carbohydrates (Reinhold-Hurek and Hurek, 2000). The genome does not contain any of the genes required for a functional Entner-Doudoroff pathway or the oxidative branch of the pentose phosphate pathway, but the non-oxidative pentose phosphate pathway is complete (Figure 12). However, all the enzymes required for glycolysis via the Embden-Meyerhoff pathway, gluconeogenesis (Figure 13), and the TCA cycle (Figure 14) are present in *Azoarcus* sp. BH72 genome.

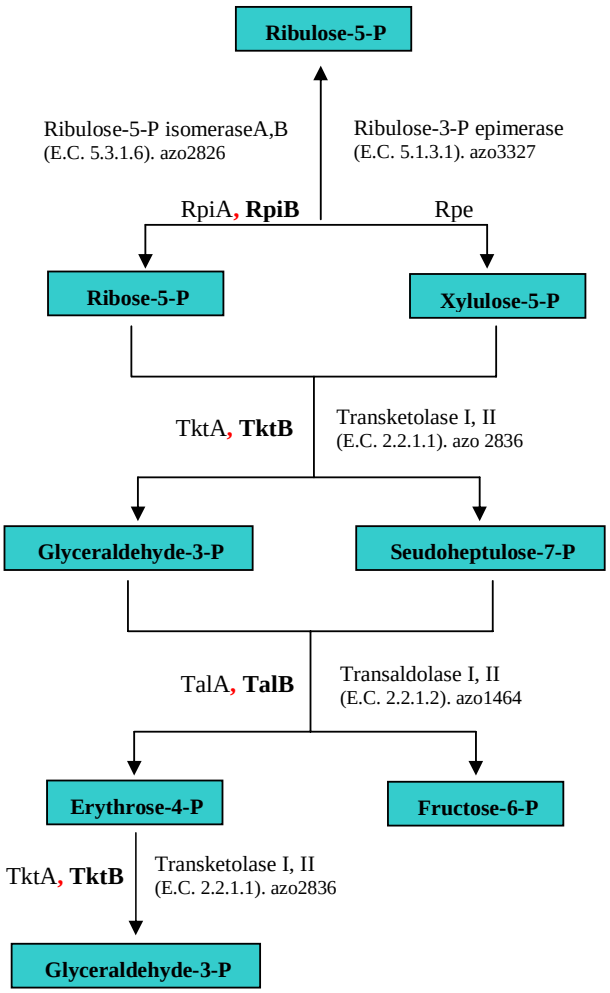


Figure 12: Non-oxidative branch of Pentose pathway. Enzyme named labeled in black bold are present in the genome**.**

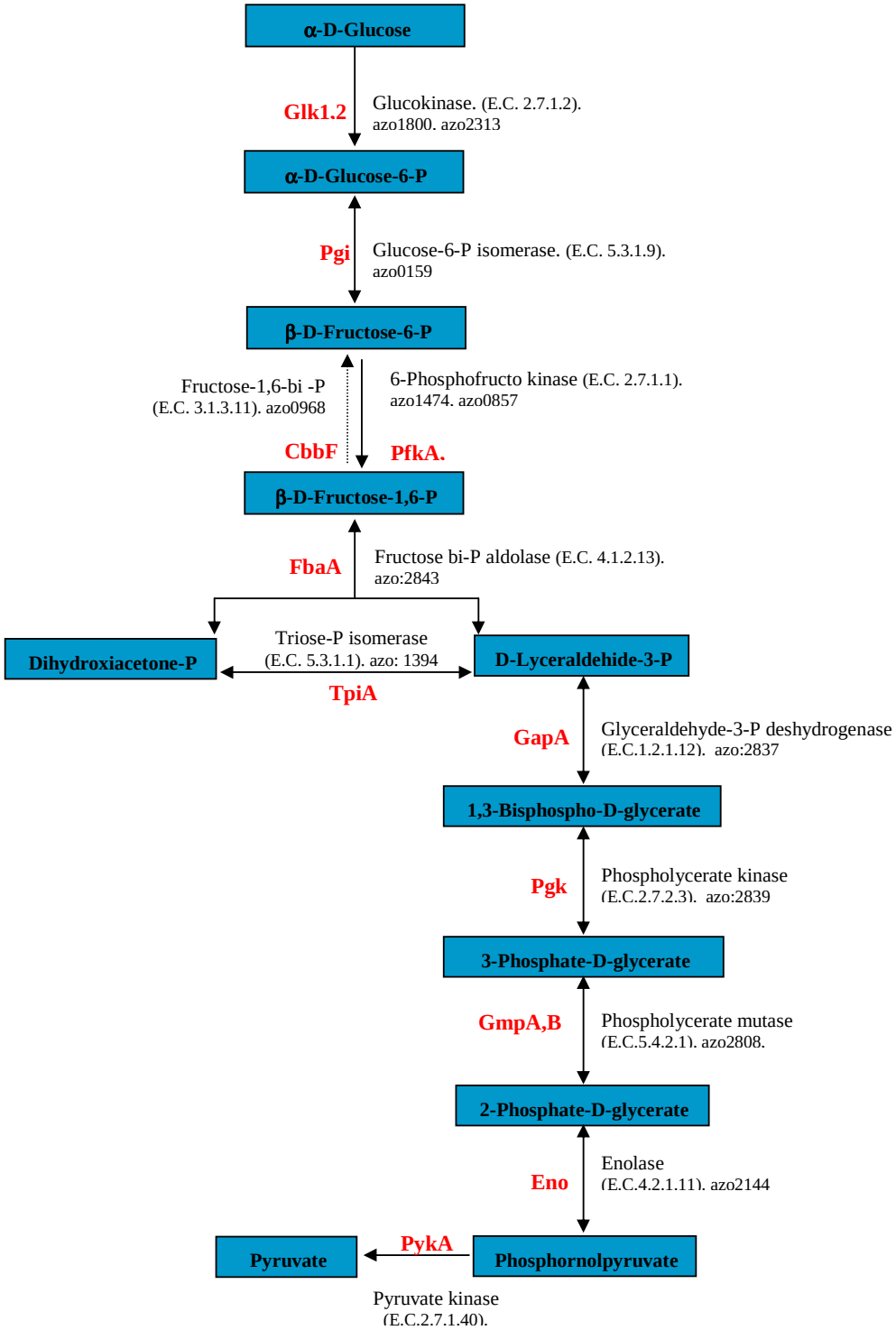


Figure 13: Glycolysis pathway in *Azoarcus* sp. BH72.

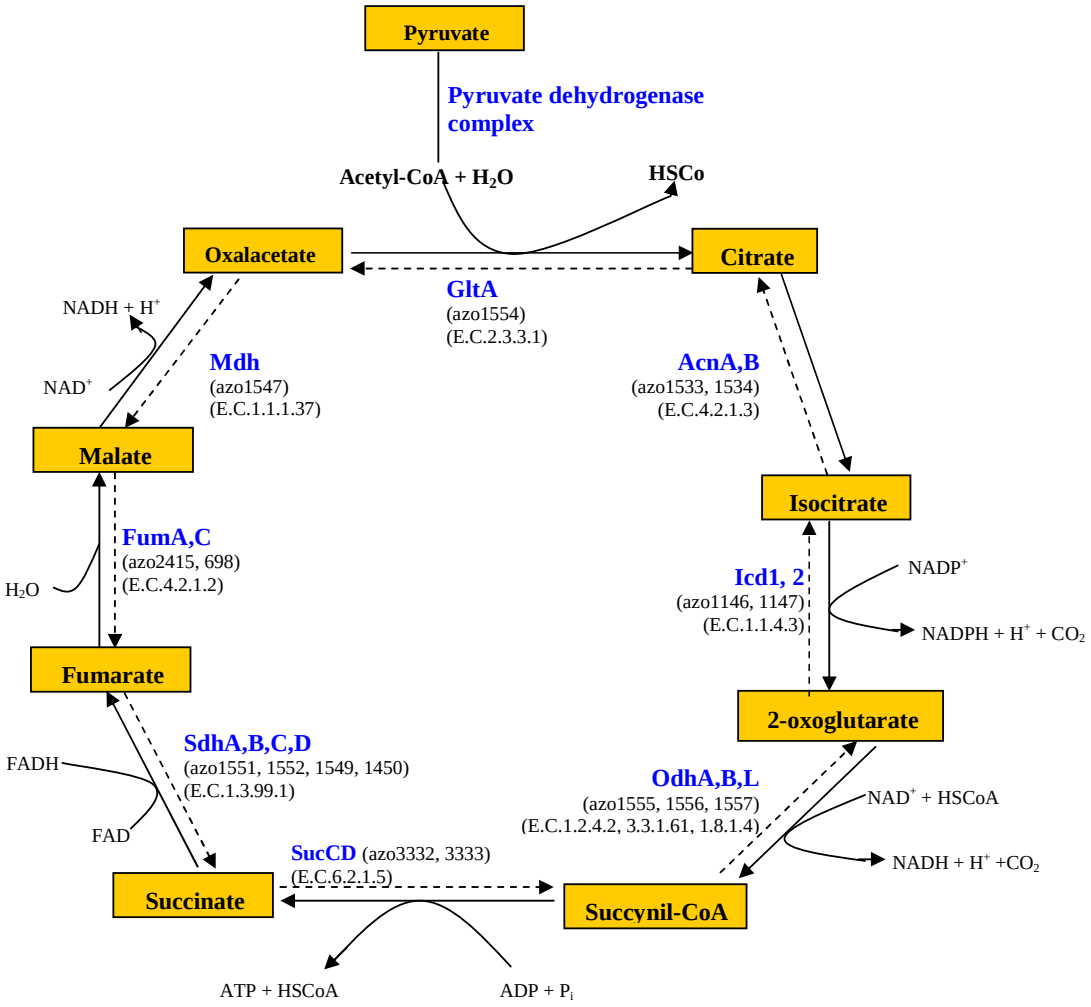


Figure 14: Tricarboxylic acid cycle pathway in *Azoarcus* sp. BH72. GltA: citrate synthase; **AcnAB**: aconitase isoenzyme; **Icd**: isocitrate dehydrogenase; **OdhA**,**B,L**: 2-oxoglutarate dehydrogenase complex; **SucCD**: succinyl-CoA synthase; **SdhABCD**: succinate dehydrogenase; **FumAC**: fumarase isoenzyme; **Mdh**: malate dehydrogenase.

A full set of genes encoding a typical phosphoenolpyruvate: sugar phosphotransferase system (PTS), is present in strain BH72 genome (Figure 15A). This system catalysis the concomitant transport and phosphorylation of its sugar substrates in a process termed "group translocation" (Postma *et al.*, 1993) (Figure 15B). The PTS-system transports mainly glucose, but also mannose, and mannitol into the cell (Gosset, 2005; Postma *et al.*, 1993). In the case of strain BH72, the PTS organization is similar to *E.coli* and to one of the four present in strain EbN1 genome, which was annotated as a mannose transporter (Table 7). No specific outer membrane receptor or transporter for glucose or other carbohydrates was annotated in the *Azoarcus* sp. BH72 genome, which might explain why

carbohydrates do not serve as carbon source (Reinhold-Hurek and Hurek, 2000). The function of these specialized transporters is not clear, and further functional analyses are necessary to clarify its function.

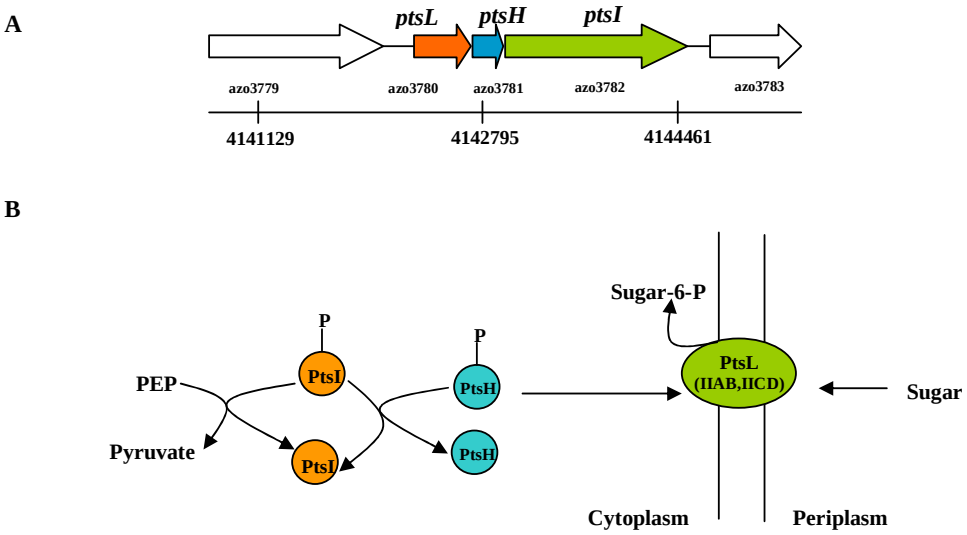


Figure 15: Uptake of sugar by PTS transports system in *Azoarcus* sp. BH72. A: transporter**, B:** genes. **PtsI:** phosphoenolpyruvate protein kinase, **PtsH:** phosphocarrier protein, **PtsL**: permease with two specific domains, **IICD:** contains the sugar binding site, **IIA:** contains the primary phosphorylation site, **IIB:** contains the phosphoryl transfer site.

4.4. Transporters.

4.4.1. Classification of general transporters.

Close to 200 genes, probably involved in transport and other membrane processes were classified as described in material and methods. Results are summarizing in table 7, where also the numbers of the same type of transporters annotated in the genome of the closest related strain EbN1 is shown. As mentioned, strain BH72 posses the double amount of coding genes for TonB-dependent outer membrane receptors (see section 2.1). Additionally, large differences between both strains were found in the “Porters: uniporters, symporters and antiporters family” (T.C.2.A). In particular strain EbN1 has significantly more of coding genes which belong to the MFS (2.A.1), and RND (2.A.6) superfamilies, as well as the TRAP (2.A.56) family, as was discussed (see section 3.1).

Table 7: Overview of proteins potentially involved in transport and other transmembranes processes in strain BH72. The numbers of gene affiliating with each TC-DB category are listed. A certain degree of over estimation has to be considered at this point.

TC-DB category	Description	N° hits in strain	
		BH72	EbN1
1.A.α-Channels			
1.A.11.	The Chloride Channel (ClC) Family	1	7
1.A.20.	The gp91phox Phagocyte NADPH Oxidase-associated Cytochrome b558 (CytB) H+-channel Family	1	5
1.A.22.	The Large Conductance Mechanosensitive Ion Channel (MscL) Family	1	1
1.A.35.	The CorA Metal Ion Transporter (MIT) Family	1	1
1.B.β-Channels			
1.B.1	The General Bacterial Porin (GBP) Family	3	6
1.B.6.	The OmpA-OmpF Porin (OOP) Family	5	5
1.B.9.	The FadL Outer Membrane Protein (FadL) Family	1	1
1.B.14.	Outer Membrane receptors (OMR) Family	22*	10
1.B.17.	The Outer Membrane Factor (OMF) Family	4	5
1.B.19.	The Glucose-selective OprB Porin (OprB) Family	-	1
1.B.31	The General Bacterial Porin (GBP) Family	3	6
1.B.33	The Outer Membrane Protein Insertion Porin (OmpIP) Family	1	1
1.B.39	The Bacterial Porin, OmpW (OmpW) Family	1	1
1.B.42.	The Outer Membrane Lipopolysaccharide Export Porin (LPS-EP) Family	1	1
2.A. Porters: uniporters, symporters and antiporters			
2.A.1.	Major Facilitator Superfamily (MFS)	25	39
2.A.2	The Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family	1	-
2.A.4.	The Cation Diffusion Facilitator (CDF) Family	4	3
2.A.6.	The Resistance-Nodulation-Cell Division (RND) Superfamily	11	27
2.A.7.	Drug/Metabolite Transporter Superfamily (DMT)	11	7
2.A.20.	The Inorganic Phosphate Transporter (PiT) Family	2	2
2.A.36.	The Monovalent Cation:Proton Antiporter-1 (CPA1) Family	1	2
2.A.37.	The Monovalent Cation:Proton Antiporter-2 (CPA2) Family	6	2
2.A.38.	The K+ Transporter (Trk) Family	4	6
2.A.47.	The Divalent Anion:Na+ Symporter (DASS) Family	1	1
2.A.49.	The Ammonium or Ammonia Transporter (Amt) Family	4	2
2.A.51.	The Chromate Ion Transporter (CHR) Family	1	3
2.A.56.	Type C4 Dicarboxilate Transporters (TRAP)	6	15
2.A.58.	The Phosphate:Na+ Symporter (PNaS) Family	1	-
2.A.59.	The Arsenical Resistance-3 (ACR3) Family	1	1
2.A.72.	The K+ Uptake Permease (KUP) Family	1	2
2.C. Ion gradient-driven energizers			
2.C.1.	The TonB-ExbB-ExbD/TolA-TolQ-TolR (TonB) Family of Auxiliary proteins for Energization of Outer Membrane Receptors (OMR)-mediated Active Transport.	13	9

3.A. Disphosphate bond hydrolysis-driven transporters			
3.A.1.	The ATP-binding Cassette (ABC) Superfamily	62	61
3.A.3.	The P-type ATPase (P-ATPase) Superfamily	4	13
3.D. Oxidoreduction-driven transporters			
3.D.1.	The Proton or Sodium Ion-translocating NADH Dehydrogenase (NDH) Family	15	27
4.A. Phosphotransfer-driven group translocators			
4.A.6	The PTS Mannose-Fructose-Sorbose (Man) Family	1	1
8.A. Auxiliary transport proteins			
8.A.5.	The K+ Transport/Nucleotide-binding Regulatory Domain/Protein (KTN) Family	2	1
8.A.7	The Phosphotransferase System Enzyme I (EI) Family	1	4
9.A. Recognized transporters of unknown biochemical mechanism			
9.A.19	The Mg2+ Transporter-E (MgtE) Family	1	19
9.B. Putative uncharacterized transport proteins			
9.B.21	The Frataxin (Frataxin) Family	1	-
9.B.37	The HlyC/CorC (HCC) Family of Putative Transporters	1	4
Others		4	-

***Transporters which have high differences in number between both strains are shown in grey.**

The MFS-transporters are single-polypeptide secondary carriers capable of transporting small solutes in response to chemiosmotic ion gradients. This large and diverse superfamily includes over a thousand sequenced members, which specificity covers a wide range of compounds (Pao *et al.*, 1998). On the other hand, characterized members of the RND superfamily catalyze also a broad range of substrate efflux via an H^+ antiport mechanism (Tseng *et al.*, 1999). Interestingly, strain BH72 posses a small amount of genes, in comparison with strain EbN1, which code for transporters belonging to P-ATPase (3.A.3) superfamily, proton or sodium ion-translocating dehydrogenase (3.D.1), and Mg^{2+}-transporter-E (9.A.19) families.

This analysis clearly shows different ability for substrates transport across the membranes on both strains. As mentioned, these two related strains possess different type of lifestyle: soil-borne vs plant-association. Most probably the different number of transporters in both strains might be as a consequence of an evolutive adaptation to their natural environments where different types of nutrients are available.

4.4.2. Comparison of number of predicted ATP-binding cassette transporters among bacterial genomes.

ATP-binding cassettes (ABC) transporters are transport systems, which are widespread in *Bacteria*, but also in *Archaea* and *Eukarya* (Higgins, 2001). ABC-transporters have been experimentally characterized with specificities which range from small to large molecules, as well from highly charged to highly hydrophobic substrates (e.g. inorganic ions, sugars,

aminoacids, proteins or complex polysaccharides) (Schmitt and Tampe, 2002). These transporters show a common global organization which includes three types of molecular components. Typically, it consists of two integral membrane proteins (permeases), each having six transmembranes segments, two peripheral membrane proteins, which bind and hydrolyze ATP, and a periplasmic (or lipoprotein) substrate-binding protein (Figure 16). The ATP-binding protein component is the most conserved, the membrane protein component is somewhat less conserved, and the substrate-binding protein component is most divergent in terms of sequence similarity (Tomii and Kanehisa, 1998).

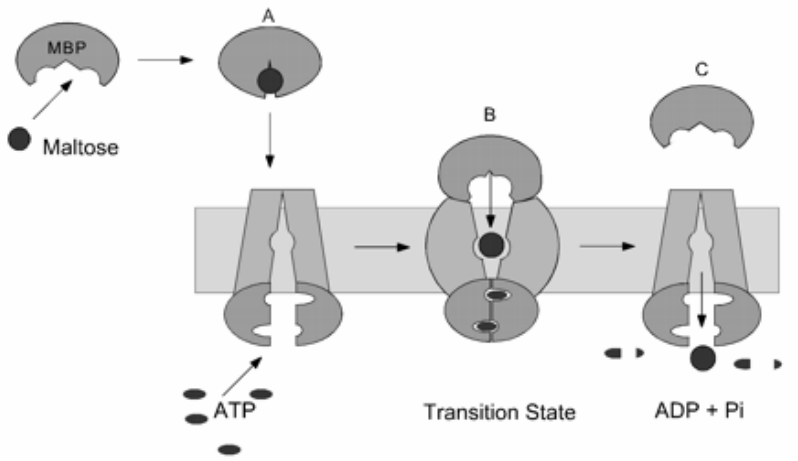


Figure 16: Model for maltose transport in *E.coli*. A: Maltose binding protein (MBP), in a closed conformation with maltose bound interacts with the transporter to initiate transport and hydrolysis. **B:** In the presumed transition state for ATP hydrolysis, MBP protein is tightly bound to the transporter in an open conformation that has a lower affinity for maltose, and the transmembranes helices have reoriented to expose an integral sugar-binding site to the periplasm. **C:** Following ATP hydrolysis, maltose is transported, and MBP is released as the transporter returns to its original conformation (Davidson and Chen, 2004).

Numbers of predicted ABC-systems in all available bacterial sequences genomes were compiled by searching for the signature (Pfam=PF 00005), specifics for ABC-systems, and compared with data obtained from *Azoarcus* sp BH72 genome project.

Strain BH72 genome with 62 copies, does not encode an exceptional number of such transporters (Figure 17). Largest copies numbers were found in symbionts rhizobia e.g. *M. loti* (225), *S. meliloti* (218) and *B. japonicum* (210); as well as in the plant-pathogenic bacterium *A. tumefaciens* (217). These organisms share with strain BH72 a common plant-interaction life style, and with rhizobia strains, the ability to fix nitrogen. Interestingly, these bacteria also possess large genome size among all compared bacteria. This result indicates no direct correlation between the ability to fix nitrogen, and a high number of encoding ABC-transporters genes. Moreover, both *Azoarcus* strains show almost the same genome size, as well as ABC-transporter gene copy number (Table 7 and Figure 17).

Taking in mind that larges genomes are expected to contain the genetic potential to degrade a broader spectrum of substrates, and to encode a larger number of corresponding transporters (Harland et al., 2005), a number of ABC-transporters per genome size (in Mbp) normalization was done (Figure 18).
This analysis shows no constant ratio of ABC-transporters/Mbp among bacteria, with genomes which encode a minimum of 7.2 (*Bacteroides thetaiotaomicron*), to a maximum of 46 (*Treptonema denticola*: 45.8); *Azoarcus* sp. BH72 with 14.2, is located in the middle range. Results obtained are in agreement with Lombardot, (2004) (Lombardot, 2004) and in disagreement with Harland *et al.*, (2005) (Harland *et al.*, 2005). Harland and co-workers showed that bacteria generally encode the same number of ABC-systems per genome unit, with the exception of nitrogen-fixing bacteria, which encoded significantly more. The discrepancy could be by the fact that Harland and co-workers made a comparison of only 72 genomes, while here all genomes sequenced up to June 2005 were compared (approx. 200).

4.4.3. Comparison of number of predicted ABC-transporters systems between different niches and between Bacteria and Archaea.

The niches: soil, plant-pathogen and beneficial plant-associated, in which each bacterium predominantly exists was taken into account for comparison of numbers of ABC-transporters (Figure 19). Results show that beneficial plant-associated bacteria had significantly more (186.8±44.3) number of predicted ABC-systems, than the others two categories (64.3±23.5 and 77.60±42.0 in soil and pathogen bacteria respectively). Interesting, no significant differences were found between soil and plant-pathogen bacteria. Additionally, numbers of ABC-transporters per Mbp, between these categories, showed no significant differences (data not shown). Taking in mind these results we can speculate that plant-symbionts bacteria have been develop special capacities to transport allocrites, which most probably are available in the specific environment, with the aim to adapt to these particular conditions. *Azoarcus* sp. BH72 with 62 ABC-transport systems can be grouped within the soil-bacteria category. This result is interestingly taking in mind that strain BH72 is a grass-associated N_2-fixing bacterium. Results obtained support more the hypothesis that the interactions of N_2-fixing grass endophytes with *Gramineae* appear to differ from typical symbionts or pathogenic interactions (Hurek and Reinhold-Hurek, 2003).

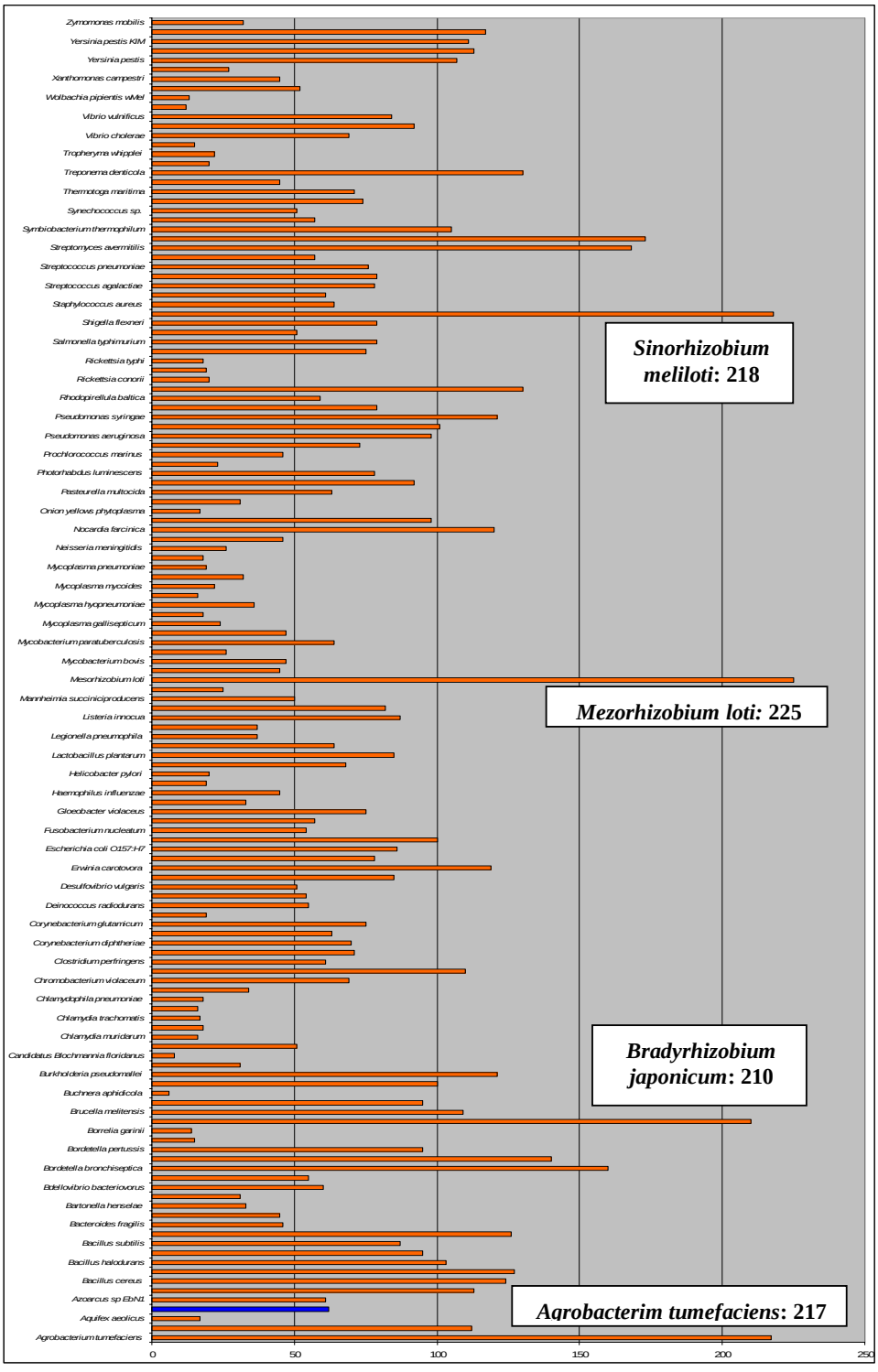


Figure 17: Number of predicted ABC transporters in *Azoarcus* sp. BH72 (blue) and other members of Bacteria (red). Species names are present in alphabetic order. Data was collected from all bacterial genomes available (June 2005) (ABC transporters, ATP-binding component-Pfam: PF00005, e-value<10^{-10}). Note: Due to space limitations some species names are missing.

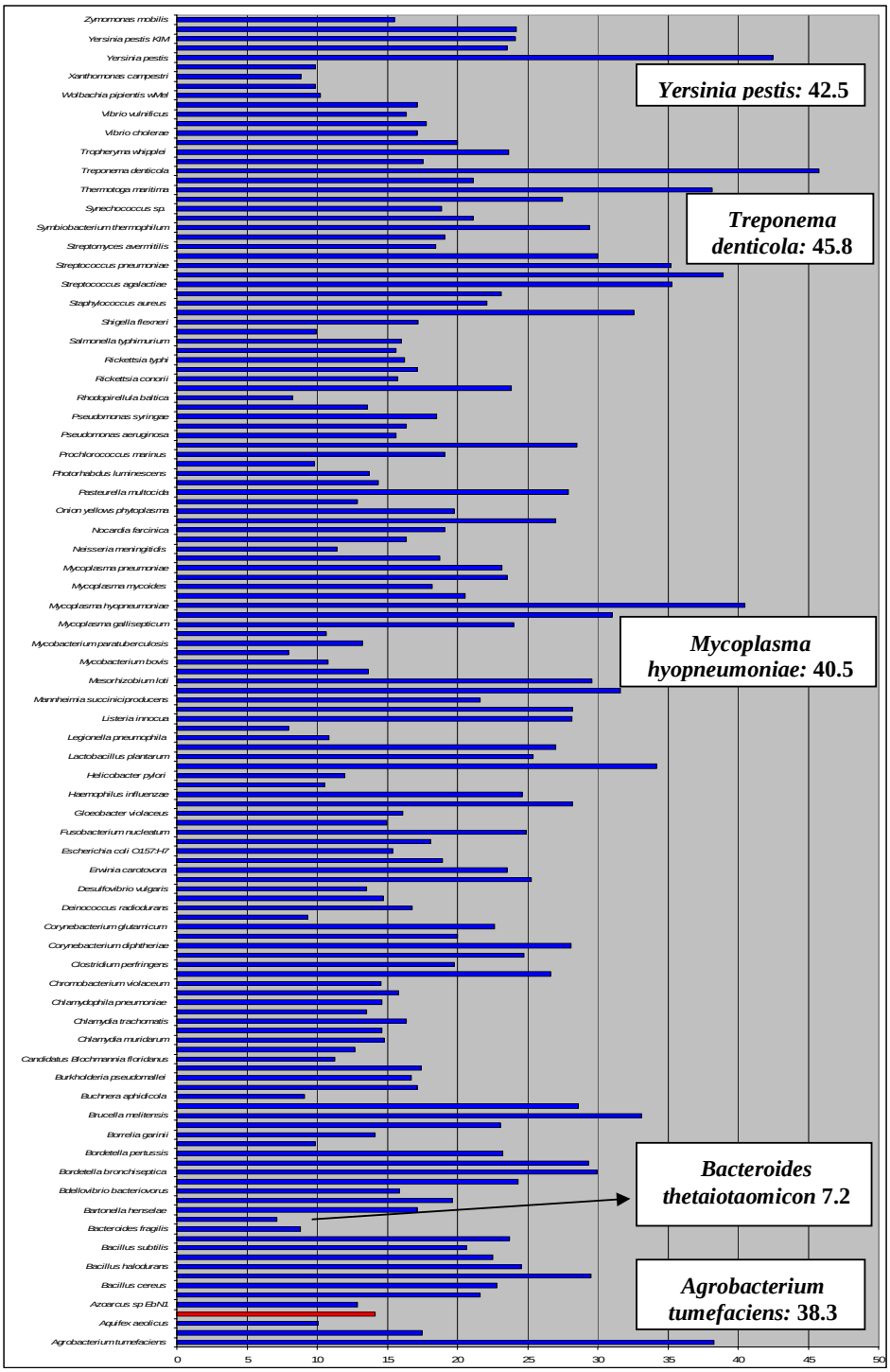


Figure 18: Number of predicted ABC transporters (normalized to genome size) in *Azoarcus* sp. BH72 (red) and other members of Bacteria (blue). Species names are present in alphabetic order. Data was collected from all bacterial genomes available (June 2005) (Pfam: PF00005, e-value<10^{-10}). Note: Due to space limitations some species names are missing.

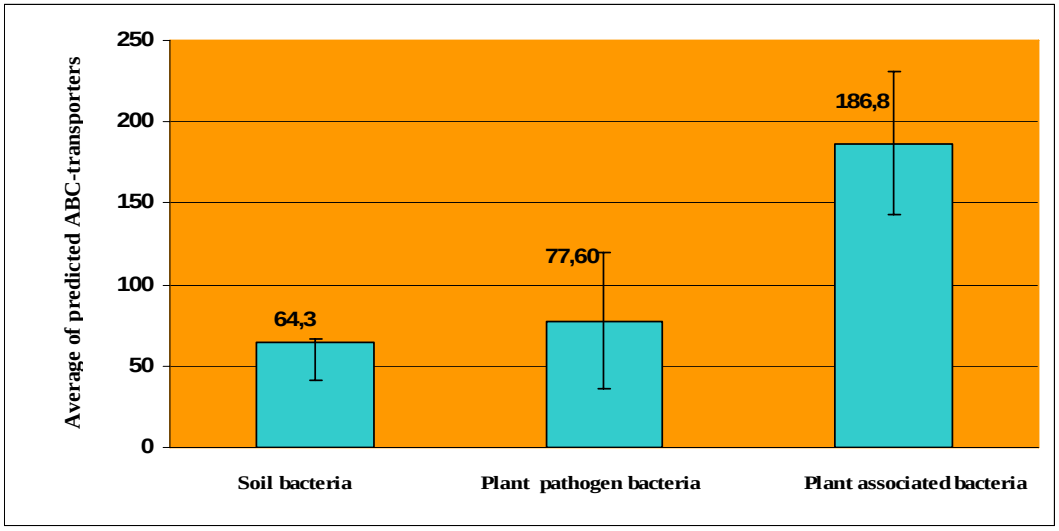
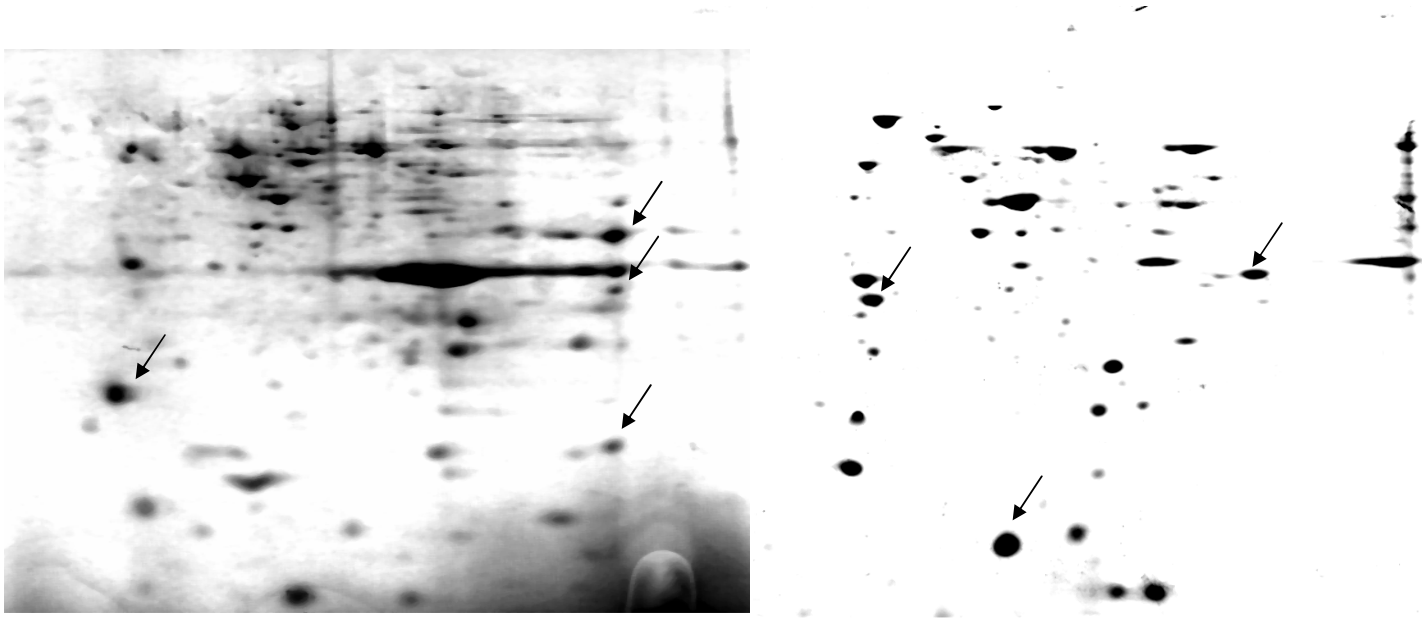


Figure 19: Average of predicted ABC systems (PF0005, e-value < 1e-10) in soil, plant-pathogen and plant- symbionts bacteria.

On the other hand, average number of predicted ABC-system was compared within at least five genomes of each mayor groups in *Bacteria* and *Archaea*. Analysis reveals no correlation in ABC-systems number and domains or bacteria species within selected genomes (data not shown). These results are in agreement with previous work which did not obtain any correlation between ABC-system numbers and phyla or Gram-stain (Harland *et al.*, 2005). Same results were obtained when genome size normalization was done (data not shown). Taking these results into account, we can conclude that the ABC-systems are not distributed in a particular number between species analyzed.

CHAPTER E

Functional genomics in *Azoarcus* sp. BH72, a proteomic approach



1. OBJECTIVES

The aim of this study was to perform a functional genomic analysis of *Azoarcus* sp. BH72 grown under N_2-fixing conditions, in pure or in co-culture with *Acremonium alternatum*, focused on a differential-display proteomic approach.
Co-culture N_2-fixing growth conditions were studied with the aim to detect and identified putatives proteins involved in diazosomes structure or function.

2. CONTRIBUION

2.1. Pure and co-culture N_2-fixing growth conditions and protein sample preparation.

Pure- and co-culture growths cells were obtaines in SM and FU medium respectively (Hurek *et al.*, 1995; Reinhold *et al.*, 1985). During approximately 2 weeks of co-culture growth, O_2 consumption and acetylene reduction (as an indicator of N_2-fixation), were determined using a gas chromatograph. Co-culture cells were harvest when the pO_2 was 4.0% in the culture gas phase, while pure culture cells, grown in a fermenter, were harvest when the OD_{578} reached a value of 0.70.

2.2. Protein extraction, solubilization and 2D-gels electrophoresis.

Dual culture bacterial cells were separated from fungal spores and hyphae by treatment with detergents and differential filtration as previously described (Hurek *et al.*, 1995). Membrane and cytoplasmic proteins, from cells which derived from both conditions, were isolated by pressure disruption and ultracentrifugation as described before with slightly modifications (Martin and Reinhold-Hurek, 2002). The membrane pellet and the cytoplasmic fraction were dissolved in extraction buffer and extracted with Tris/HCl-buffered phenol (pH 8.0). Proteins were precipitated overnight at –20°C with 0.1 M ammonium acetate in methanol. Final pellets were resuspended in sample buffer (5M urea, 2M thiourea, 2% (w/v) CHAPS (SIGMA), 2% (w/v) ASB-14 (SIGMA), 10mM DTT and 2% v/v ampholines pH 3.5-10 (BioRad)). Protein concentration was determined using a RCDC Protein assay kit (BioRad).
Isoelectric focusing in glass capillar tubes as a first dimension and SDS-PAGE as a second dimension, were performed for protein separation according to their isoelectric point and molecular weight respectively, as described before with slightly modifications (Hurek *et al.*, 1995)

2.3. Differential-display proteomics.

This approach was used with the aim to study the proteomic pattern of strain BH72 cells grown under different N_2-fixing conditions by 2D-PAGE and the posterior identification of interesting spots. Proteomic patterns were analyzed by using the Image Master software which allows for the quantitative comparison of protein spots in the gels. In the comparison analyses, the proteome patterns from N_2-fixing cells grown under pure culture condition were taken as references, while the mutant strain BH1599 was used as a negative control. This pilus mutant strain, is unable to establish a tight association with *A. alternatum* and to form diazosomes (Dörr *et al.*, 1998). Spots of interest were quantified, excised from the gels and freezed at -20°C. Trypsin digestion and mass spectrometry (MALDI-TOF) analysis was done for protein identification in the University of Greiswald (C. Sharf) (Chapter E, manuscript Materials and Methods).

2.4. Ethanol and siderophore detections.

An alcohol dehydrogenase enzymatic activity test (Kagi and Vallee, 1960) was used with the aim to determine if *A. alternatum* was able to exudate ethanol, in pure or in co-culture.
On the other hand, specific and general siderophore-type production were assayed chemically in the supernatant of SM and FU media, of strain BH72, strain BH1599 or *A. alternatum*, in pure and co-culture (Arnow, 1937; Atkin *et al.*, 1970; Schwyn and Neilands, 1987).

3. RESULTS AND CONCLUSIONS

3.1. Image analysis shows that in both, membrane and cytoplasm fractions, most of the spot were down regulated in co-culture cells (53% and 46% respectively). These results suggest a differentiation process in co-culture cells, as they have to cover their metabolic requirements with a smaller set of enzymes.

3.2. As expected proteins involved in the N_2-fixing process and subsequent assimilation of the ammonia, were detected in both conditions studied. These included the nitrogenase complex and proteins which are related to its synthesis and activation. Additionally, a putative nitrogenase mobile electron donor, a ferredoxin, and a component of the hydrogenase complex, which recycle the protons produced by the nitrogenase activity, were identified.
On the other hand, one of the key regulators in the N_2-metabolism NtrC, as well as several proteins involved in N_2- assimilation, were also identified.

3.3. Almost all the enzymes involved in the TCA and glyoxylate cycle were detected. In particular, identification of a malate transporter in pure culture and a malic enzyme in both conditions shows that this C4-dicarboxylate compound is actively transported and metabolized in agreement with the growth conditions used. Additionally, several enzymes involved in the glycolysis-gluconeogenesis pathways were also identified, most probably to be used in the gluconeogenesis as a synthesis pathway for glucose and other compounds.

3.4. Ethanol was detected as a fungal exudate and as a carbon source for strain BH72 in co-culture growths. These results are in agreement with the presence of several enzymes up-regulated from the pyruvate metabolism, which are involved in the ethanol assimilation. Moreover, this is in agreement with the capacity of strain BH72 to grow on ethanol as sole carbon source (Reinhold-Hurek *et al.*, 2005). Additionally, enzymes involved in the assimilation of acetate and lactate (also detected as a fungal exudate), were identified as up-regulated in co-culture conditions. In context these results show that the *Azoarcus* sp. BH72 C-metabolism, in co-culture conditions, is adapted to the C-sources exudate by the fungal partner.

3.5. By using the proteome pattern of the mutant strain BH1599 as a negative control, specific proteins expressed by the wild-type in co-culture and probably related with diazosomes structure or function were able to identified. Results demonstrate that this unusualls structures are involved in an efficient and coordinated respiration, ATP synthesis and N_2-fixation process as was suggested in a previous work (Hurek *et al.*, 1995). This hypothesis is supported by the detection of several up-regulated and membrane associated proteins involved in the mentioned processes. Moreover, membrane proteins probably involved in diazosomes structure were also identified using this proteomic approach.

3.6. Identification of several proteins related with detoxification of reactive oxygen species, chaperonig process under stress conditions and PHB synthesis, indicated that N_2-fixing conditions are stressful for *Azoarcus* sp. BH72.

3.9. Outer membrane (OM) TonB-dependent receptors for iron compounds and vitamin B12 were identified in both conditions. Taking in mind that both media used were iron-sufficient, a putative role of these OM identified in iron metabolism was discarded. Moreover, no siderophore production was detected in the supernatant of all studied conditions, indicating that most probably these OM receptors are involved in other processes such as external signal sensing, bacteria-host interaction or biocontrol.

4.0. OM proteins involved in the drug or antibiotic resistance were identified identified. Most probably, this set of proteins is related to the resistance against antibiotics produced by the fungus and moreover, are involved in a compatible interaction between both partners.

4. OUTLOOKS

Results obtained here contribute to a better understanding of the *Azoarcus* sp. BH72 ecology as well as the role of the diazosomes in the metabolism of strain BH72.

This analysis represents the basis for functional proteomic analysis of *Azoarcus* sp. BH72. The condition-specific proteins identified here provide new targets for genome-wide mutagenesis e.g. the OM TonB-dependent receptors identified, as well as the membrane proteins probably involved in diazosomes structure.

This manuscript will be send for publication to Journal of Bacteriology

PROTEOMICS ANALYSIS OF *Azoarcus* sp. BH72 GROWN UNDER DIFFERENT N_2-FIXING CONDITIONS

Running title: Proteomic analysis of the N_2-fixing *Azoarcus* sp. BH72.

Federico Battistoni[a], Janina Oetjen[a], Christian Scharf[b], Tim Karg[c] and Barbara Reinhold-Hurek[a]*.

[a] Laboratory of General Microbiology, University Bremen, PO. Box 33 04 40, D-28334 Bremen, Germany;
[b] Functional genomics laboratory. Department of Otorhinolaryngology Head and Neck Surgery. Medical School, Greifswald University. Walther-Rathenau-Str. 43-45 D-17475 Greifswald.
[c]

keywords: *Azoarcus* sp. BH72, nitrogen fixation, functional analysis, proteomics, dual culture.

* Corresponding author. Tel. +49 421 218 2370; fax: +49 421 218 9058. *e-mail address:* breinhold@uni-bremen.de

ABSTRACT

Azoarcus sp. BH72, a Gram-negative, N_2-fixing bacterium that belongs to the β-proteobacteria, was isolated from the endorhizosphere of Kallar grass, a pioneer plant on low fertility soils in the Punjab of Pakistan. In laboratory conditions strain BH72 can supply nitrogen to the natural host Kallar grass (Hurek *et al.* 2002), moreover it shows also a similar colonization pattern of rice roots in comparison to its original host (Hurek *et al*., 1994). In co-culture with the fungal strain 2003, an ascomycete isolated from the rhizosphere of Kallar grass, *Azoarcus* sp. BH72 can develop into a hyperinduced state characterized by high specific rates of respiration and efficient N_2-fixation. Ultrastructural analysis of cells in the course of hyperinduction revealed that complex intracytoplasmic membrane systems (diazosomes) are formed, which were absent under standard N_2-fixing conditions (Hurek *et al*., 1995). With the aim to identify the proteins expressed differentially in this hyperinduced state and in particular the ones involved in diazosome formation or function, a differential- display proteomic approach was conducted on strain BH72 grown under both N_2-fixation conditions. Membrane and cytoplasmic proteins were isolated from *Azoarcus* sp. BH72 as well from a non-diazosome forming mutant strain BH1599.

Different 2D-PAGE gels protein profiles were compared using ImageMaster software and spots of interest were analyzed by mass spectrometry (MALDI-TOF) for its identification. Results suggests that a differentiation process ocurred in strain BH72 cells during co-culture which is characterized for metabolics and structurals protein changes to scope with the new microenvironmet. Moreover, in co-culture conditions an efficient and coordinated respiration and N_2-fixation process ocurred in the diazosomes. According to that, several proteins probably involved in the diazosomes structure or function were identified in this work.

INTRODUCTION

Oxygen plays an important role in biological nitrogen fixation. In their natural environment diazotrophs often have to deal with very low O_2 concentration which allows energy generation and N_2-fixation simultaneously, as nitrogenase the key enzyme for nitrogen fixation is oxygen-sensitive. N_2-fixation takes place in grass root-associated bacteria such as *Azospirillum* spp. and *Azoarcus* spp. at dissolved oxygen concentration (DOC) in the range of a few micromolars (Hurek *et al.,* 1987a; Reinhold-Hurek *et al.,* 1993b). *Azoarcus* sp. BH72, a Gram-negative, N_2-fixing bacterium which belongs to the β-proteobacteria, was isolated from the endorhizosphere of Kallar grass, a C4 plant grown in the Punjab of Pakistan, on low-fertility soils that is highly tolerant of soil salinity, alkalinity and waterlogged conditions (Reinhold *et al.,* 1986; Reinhold-Hurek *et al.,* 1993b). Strain BH72 supplied nitrogen to its host Kallar grass plant in an "unculturable" state (Hurek *et al.,* 2002); moreover *in situ* experiments demonstrated that *Azoarcus* sp. nitrogenase genes (*nif*) are expressed in the aerenchyma of uninoculated field-grown Kallar grass plants (Hurek *et al.,* 1997a). In *gnotobiotic* culture strain BH72 shows also a similar colonization pattern of rice roots in comparison to its original host (Hurek *et al.,* 1994b). Under these conditions the *nifH* genes are also strongly expressed in the appoplast of roots cortex of rice (Egener *et al.,* 1999). An unusual feature of *Azoarcus* sp. BH72 is the formation of intracytoplasmic membrane stacks, diazosomes (Hurek *et al.,* 1995). Cells fixing nitrogen under standard conditions (e.g. around 2µM of dissolved oxygen) do not harbour these membranes. They become apparent in the course of hyperinduction, a physiological state of augmented activity and efficiency of nitrogen fixation in optimized batch cultures at extremely low O_2-concentrations (Hurek *et al.,* 1995). Mutational analysis of structural genes for nitrogenase (*nifHDK*) and immunohistochemical localization of the dinitrogenase reductase demonstrated that these highly organized structures are involved in an efficient nitrogen fixation process (Hurek *et al.,* 1995). A reproducible induction of these intracytoplasmic membrane stacks occurs in co-cultures of strain BH72 with an ascomycete strain 2003 (related to *Acremonium alternatum*), which was isolated from surface-sterilized roots of Kallar grass (Hurek and Reinhold-Hurek, 1999). Bacteria that are adhered to the actively respiring mycelium may experience extremely low DOCs which are maintained over a log period of time. Most probably

these conditions trigger induction of diazosomes however; biotic inducer might also be involved (Hurek *et al.,* 1995). Global changes in protein pattern of total cellular and membrane proteins were detected in diazosome-containing bacteria in comparison with cell fixing N_2 under standard conditions. In diazosomes containing cells, a protein homologous to the *gln*B gene product PII, an intracellular signal transmitter involved in regulation of the nitrogen metabolism, was identified by its-terminal aminoacid sequence, and may have a different state of uridylation (Karg and Reinhold-Hurek, 1996; Martin *et al.*, 2000). These observations suggest that the hyperinduced physiological status is indeed quite different from that of standard N_2-fixing cells.

In post-genomic era, proteomics is one of the best strategies used to reveal the dynamic expression of whole proteins in cells and their interactions. High resolution two-dimensional gel electrophoresis (2D-PAGE) combined with matrix-assisted laser desorption ionization time of flight (MALDI-TOF) and bioinformatics analysis is widely used for a large-scale study of protein identification based on peptide mass fingerprint (PMF) (Pandey and Mann, 2000). A proteome is the protein complement expressed by a genome. A proteomic analysis of an organism under different conditions, can contribute to understand of how a microbe adapts its metabolism to different environments. Identification of differential displayed proteins could be used to determine the genes responding to relative physiological actions and clarify the function of these genes.

In this work for the first time a high-throughput functional analysis using a proteomic approach was used in *Azoarcus* sp. BH72 with the aim to study and identify proteins differentially expressed under standard or in co-culture with *A. alternatum,* nitrogen fixing grow conditions. *Azoarcus* sp. BH72 membrane and cytoplasmic proteins were isolated from cells grown on pure or co-culture. Different 2D-PAGE gels protein profiles were compared and spots of interest were analyzed by MALDI-TOF for its identification using an unreleased *Azoarcus* sp. BH72 genome data. Result show that membrane and cytoplasmic 2D-PAGE proteomic patterns of strain BH72 N_2-fixing cells in pure- and co-culture have strongly significant differences which reflected adaptations in bacteria metabolism to cope with different environmental conditions studied. Metabolic pathways were reconstructed in both conditions and new proteins were identified and correlated with previous and current data obtained by using additional functional approaches. Results obtained here contribute to a better understanding of the *Azoarcus* sp. BH72 ecology as well as the role of the diazosomes in the strain BH72 metabolism. This analysis represents a landmark to functional proteomic analysis of *Azoarcus* sp BH72 and provides a complementary strategy for other high-throughput global analysis approaches. Moreover the conditions-specific proteins identified here will provide new target for genome-wide mutagenesis.

MATERIALS AND METHODS

Bacterial strains, media and growth conditions.

Azoarcus sp BH72 was isolated from the root interior of Kallar grass, *L. fusca* (L.) Kunth, grown in the Punjab of Pakistan (Reinhold *et al.*, 1986). Strain BH1599 is an *Azoarcus* pilus mutant unable to form diazosomes (Dörr *et al.*, 1998). The *A. alternatum*, ascomycete strain 2003 has been isolated from the rhizosphere of Kallar grass (Hurek *et al.*, 1995). *Azoarcus* sp. BH72 and the fungal strain 2003 were grown at 37° and 30°C, respectively. Both were cultured on VM-ethanol agar plates (Reinhold *et al.*, 1985). For standard nitrogen fixation in pure culture, *Azoarcus* sp. BH72 was grown microaerobically in SM medium (Reinhold *et al.*, 1985) using a Biostat®B fermenter (B. Braun Melsungen Biotech International). Precultures were grown in 20 ml of SM medium with overnight shaking, and cells were washed two times in nitrogen-free SM medium. Two and a half liters of nitrogen-free SM medium were inoculated with washed cells to an optical density at 578 nm (OD_{578}) of 0.005. The gas phase of the culture was adjusted to 99.4% nitrogen and 0.6% O_2. Temperature and stirring were setted to 37°C and 600 rpm respectively. Cultured were harvests when the OD_{578} reached the 0.65 value. For dual cultures of *Azoarcus* strains and the fungal isolate 2003, bacterial and fungal cells were precultured separately on VM-ethanol agar and then cultivated together in sealed 1-liter serum flasks containing 200 ml of "Fungi-Medium" (FU) and sealed with rubber stoppers as previously described (Hurek *et al.*, 1995). Cells were harvested when the gas phase pO_2 had decreased to 4.0% measured in a gas chromatograph. For this propose, from the headspace of cultures sealed with rubber stoppers, samples were taken with a syringe and directly applied to the gas chromatograph for measurements. The chromatograph features, oxygen and ethylene determination proceeding were done as previously described (Karg and Reinhold-Hurek, 1996).

Protein extraction and solubilization

Prior to protein extraction, dual culture bacterial cells were separated from the fungal spores and hyphae by treatment with detergents and differential filtration as previously described (Hurek *et al.*, 1995). Cells were resuspended at 4°C in 50mM buffer phosphate (pH 7), 600mM NaCl and disrupted by four passages through a French press cell at 1500 lb/in^2. Cells debris were removed by centrifugation (20 min at 20000 X g) and the remaining supernatant was used for membrane sedimentation by ultracentrifugation (2 hrs. 200,000 X g). The remaining supernatant was referred to as the soluble cytoplasmic fraction, whereas the isolated membrane fractions were washed twice with 50 mM sodium phosphate buffer (pH 7.0) - 600 mM NaCl to remove nonspecifically or loosely bound proteins. The insoluble pellet after ultracentrifugation was referred to as the membrane fraction. The membrane pellet and the cytoplasmic fraction were dissolved in extraction buffer and extracted with Tris/HCl-buffered phenol (pH 8.0). Proteins were precipitated overnight at –20°C with 0.1 M ammonium acetate in methanol. Final pellets were resuspended in 50-80 µl of sample buffer (5M urea, 2M thiourea, 2% (w/v) CHAPS

(SIGMA), 2% (w/v) ASB-14 (SIGMA), 10mM DTT and 2% v/v ampholines pH 3.5-10 (BioRad)). Protein concentration was determined using a RCDC Protein assay kit (BioRad).

Two dimensional gels electrophoresis

Isoelectric focusing in glass tubes as a first dimension, SDS-PAGE as a second dimension were carried out as previously described (Hurek *et al.*, 1995) with slightly modifications. Approximately 200µg of protein in a sample volume of 50µl to 80µl was loaded on the top of the tube gels. The following voltage programme was applied to the Isoelectric focusing cell (BioRad): 10 min at 100 V, 10 min at 200 V, 2 hrs at 300 V and 18 hrs at 400 V. The extruded gels were incubated 20 min in equilibration buffer. For second dimension, vertical gel electrophoresis was used with a 12.5% (wt/vol) polyacrylamide gel as described by Laemmli (Laemmli, 1970), with a gel size of 13 by 16 cm.

Staining procedures and image analysis.

Coomassie staining of the gels was done according to standard procedures, using a staining solution (35% (v/v) methanol, 10% (v/v) acetic acid and 0.05% (w/v) Coomassie brilliant blue R-250 (SERVA, Germany)). Gels were destained with 30% (v/v) methanol, 5% (v/v) acetic acid solutions, and stored at 4°C in storage buffer (18% (v/v) ethanol, 3% (v/v) glycerol) until analysis. Stained gels were scanned at 600 dots per inch on a UMAX Power Look III Scanner (UMAX, Data System Inc. Taiwan), and image analysis was carried out with the ImageMaster® 2-D Elite version 4.01 software (Amersham Biosciences). Spot detection, quantification, background subtraction, gel-to-gel matching and differential display analysis were performed on three independent biological repetitions. Intensities of individual spots were normalized based upon the total density of all spots detected in each gel. Apparent masses of each protein spot were calculated by the use of the software, according molecular weight ladder used (Fermentas Life Sciences).

Preparation of peptide mixtures for MALDI-MS

Proteins were manually excised from Colloidal Coomassie Brilliant Blue stained 2-D gels and transferred into 96 well microplates loaded with 100 µL Lichrosolv water per well. Digestion with trypsin and subsequent spotting of peptide solutions onto the MALDI-targets were performed automatically in an Ettan Spot Handling Workstation (Amersham-Biosciences, Uppsala, Sweden) using a modified standard protocol. Briefly, gel pieces were washed twice with 100 µL 50 mM ammoniumbicarbonate/ 50% (v/v) methanol for 30 min and once with 100 µL 75% (v/v) ACN for 10 min. After 17 min drying 10 µL trypsin solution containing 20 ng/µL trypsin (Promega, Madison, WI, USA) in 20 mM ammoniumbicarbonate was added and incubated at 37 °C for 120 min. For peptide extraction gel pieces were covered with 60 µL

50% (v/v) ACN/ 0.1% (w/v) TFA and incubated for 30 min at 37 °C. The peptide containing supernatant was transferred into a new microplate and the extraction was repeated with 40 µL of the same solution. The supernatants were dried at 40 °C for 220 min completely. Peptides were dissolved in 2.2 µL of 0.5% (w/v) TFA/ 50% (v/v) ACN and 0.7 µL of this solution were directly spotted on a MALDI-target. Then, 0.4 µL of matrix solution (50% (v/v) ACN/ 0.5% (w/v) TFA) saturated with CHCA was added and mixed with the sample solution by aspirating the mixture five times. Prior to the measurement in the MALDI-TOF instrument the samples were allowed to dry on the target for 10 to 15 min.

MALDI-TOF MS analysis

The MALDI-TOF measurement of spotted peptide solutions was carried out on a Proteome-Analyzer 4700 (Applied Biosystems, Foster City, CA, USA). The spectra were recorded in reflector mode in a mass range from 800 to 3700 Da with a focus mass of 2000 Da. For one main spectrum 25 sub-spectra with 100 shots per sub-spectrum were accumulated using a random search pattern. If the autolytic fragments of trypsin with the mono-isotopic $(M+H)^+$ m/z at 1045.5 and/or at 2211.1 reached a signal to noise ratio (S/N) of at least 10, an internal calibration was automatically performed using at least one peak for one- or two peaks for two-point-calibration. Calibration was performed manually for the less than 1% samples for which the automatic calibration failed.

Additionally, MALDI-TOF-TOF analyses were performed for the five strongest peaks of the TOF-spectrum after subtraction of peaks corresponding to background, keratin and trypsin fragments. For one main spectrum 20 sub-spectra with 125 shots per sub-spectrum were accumulated using a random search pattern. The internal calibration was automatically performed as one-point-calibration if the mono-isotopic arginine (M+H)+ m/z at 175.119 or lysine (M+H)+ m/z at 147.107 reached a signal to noise ratio (S/N) of at least 5.

After calibration a combined database search of MS and MS/MS measurements was performed using the GPS Explorer software (Applied Biosystems, Foster City, CA, USA) with the following settings: (i) MS peak filtering: mass range from 800 to 3700 Da; minimum S/N filter of 10; peak density of 50 peaks per range of 200 Da and maximal 200 peaks per protein spot; mass exclusion list contained background peaks and trypsin fragments with an exclusion tolerance of 100 ppm (ii) MS/MS peak filtering: mass range from 60 Da to a mass that was 20 Da lower than the precursor mass; peak density of 50 peaks per 200 Da and maximal 65 peaks per MS/MS; minimum S/N filter of 10 (iii) database search: precursor tolerance 35 ppm and MS/MS fragment tolerance 0.65 Da.

The peptide search tolerance was 35 ppm but the actual RMS value was between 5 and 15 ppm. Peak lists were compared with an *Azoarcus* sp BH72 specific database using the Mascot search

engine (Matrix Science Ltd, London, UK). Peptide mixtures that yielded a mowse score of at least 49 were regarded as positive identifications ($P < 0.05$).

Analysis of fungal exudates by HPLC.

To analyse the exudates produced by *A. alternaturm* a HPLC system (Sykam, Gilching, Germany) with a S1000 pump and a RI-detector ERC 7512 (Erma, Tokyo, Japan) was used. Peaks were evaluated with the help of an integration programme (Sykam). 100 µl of sterile filtered supernatant from co-cultured cells as well as from *A. alternatum* grown in FU-medium was loaded onto a stainless steel column with a DVB-styrene-copolymer as stationary phase. As negative control 100 µl of FU-medium was applied. Material was eluted from the column with 1 mM sulphuric acid at 0.8 ml/min. A standard including lactate, formiate, acetate, propionate, buryrate, isopropanol, valerate, n-butanole and capronoate was used, where each compound was adjusted to 100 µM. As additional standards 1 mM solutions of glucose, formiate, acetaldehyde, formaldehyde, succinate, methanol and ethanol were applied.

Ethanol production by the fungus in the growth medium.

Ethanol production by *A. alternaturm,* grown in pure- and co-culture with *Azoarcus* sp. BH72, was indirect measured by using an alcohol dehydrogenase (EC: 1.1.1.1) enzymatic activity assay (Kagi and Vallee, 1960). For that, 500 µl of culture supernatant were taken with a sterile syringe and centrifuged at 4°C (10 min 14000 X g). The supernatant kept on ice, was used as a substrate for the enzymatic assay. The ethanol production was measured during one week of pure and co-culture growth. Two independent biological experiments were done with four different cultures per condition.

Siderophore detection in the SM and FU culture media.

Detection of hydroxamates and catecholes siderophore-types on the FU and SM media supernatant was done as previously described (Arnow, 1937; Atkin *et al.*, 1970). For siderophore detection independent of their structure a Chromeazurol-S (CAS) assay was used as was described (Schwyn and Neilands, 1987).

RESULTS AND DISCUSION

Comparison of two-dimensional membrane and cytoplasmic proteome pattern of *Azoarcus* sp. strain BH72 grown under pure and co-culture N_2-fixing conditions.

Proteome patterns of membrane and cytoplasmic fraction from *Azoarcus* sp. strain BH72 grown under different N_2-fixing condition were compared with the aim to detect specific protein expressed in co-culture with *A. alternatum*. ImageMaster® software was used to process three

different gels that came from independent experiments out of twelve biological repetitions per condition. This analysis programme allows for the quantitative comparison of protein spots while adjusting for experimental variations caused by problems such as irregular protein loading or staining variations. For gels alignments, the spot matching alignment algorithm was used. In the comparison analyses, the proteome patterns from N_2-fixing cells grown under pure culture condition were taken as references, while the mutant strain BH1599 was used as a negative control. This pilus mutant strain, is unable to establish a normal association with *A. alternatum* and to form diazosomes (Dörr *et al.*, 1998). We assume that no contamination of bacterial cellular proteins with fungal hyphae or spore proteins took place using this procedure, indicating that additional protein spots of dual cultures were not caused by contaminating fungal proteins (Hurek *et al*., 1995; Karg and Reinhold-Hurek, 1996).

Pure-culture membrane and cytoplasmic proteome patterns of strain BH72 revealed significant differences from dual-culture ones. Approximately 150 spots per condition were separated on the Coomassie-stained 2D gels within the pH 3-10 and size range 10-122 kDa (Figure 1). Gels image analysis reveals that in membranes fractions 53% of the spots were absent or down-regulated while 18% were novel or up-regulated in co-culture conditions (Figure 1E and F). From the last group, 9 novel spots were detected in the wild type proteome pattern (Figure 1F, black circles). These proteins most probably are related to the presence of specific internal membrane formed under co-culture conditions. Within the cytoplasmic fraction, 46% and 10% of the spots were down- and up-regulated respectively compared to the reference gel pattern (Figure 1A and B). In this case 13 novel induced spots were found in the wild type proteome pattern (Figure 1B, black circles). A cut-off of 2 folds was used in the ImageMaster® software to report that a specific spot was up-regulated or down-regulated. The observed down regulation of co-culture proteins in both fractions indicates that under these conditions, cells must have a most specialized metabolism to cover their metabolic requirements with fewer enzymes.

All of the most abundant protein present on Coomassie-stained 2D gels were picked and analyzed by mass spectrometry (MALDI-TOF). In total approximately 800 proteins spots were evaluated from several gels in the pH range of 3-10. Peptide mass fingerprint (PMF) data from several spots with the same coordinates on gels containing proteins were shown to be the same. This redundant protein identification validates gel image analysis and confirmed results. The protein products of 180 unique genes were identified using this approach, which represent 4.5% of the genome coding genes. From these group 140 genes products were assigned EC number in the *Azoarcus* sp. BH72 database. These numbers were used to search the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for possible pathways.

Several proteins in *Azoarcus* sp. BH72 produced multiple spots probably due to post-translational modifications. In most cases the shift in position is horizontal, suggesting that the

modification influences only the pI and leaves the molecular mass substantially unchanged (Figure 1 and table 1). Exceptions to that were the cases of the NifH, NifD and NuoG proteins from which two isoforms with different molecular weight were found (Figures 1A, E, F, spots n°1, 2, 97). The case of NifH is in agreement with previous reports which described that under nitrogen growth conditions strain BH72 expresses two isoforms of the NifH protein (Hurek *et al.*, 1995). Likely the modifying group of NifH represented an ADP-ribosylation (Martin and Reinhold-Hurek, 2002).

It is well known that there are limitations in the present technology for resolution and identification of integral membrane proteins by 2D-gel electrophoresis and MALDI-TOF analysis. Hydrophobic proteins are not easily solubilized in the non-ionic detergents used for isoelectric focusing, and the hydrophobic fragments obtained after trypsin treatment are difficult to ionize for mass spectrometric analysis (Aebersold and Mann, 2003; Santoni *et al.*, 2000b). As described in materials and methods, as a first dimension a isoelectric focus in capillar tube was used instead of IPG (immobilized pH gradients) strips. Additionally, with the aim to improve the solubilization of membrane proteins, a detergent combination of CHAPS and ASB-14, both to a final concentration of 2%, was used in the sample buffer. Moreover, in the capillar gels 2% of the ASB-14 detergent was employed. Results obtained by using this methodology show a high number of recovered membrane proteins. As it is shown in table 1, 19 membrane proteins, including outer membrane ones, were identified. By SMART (Schultz *et al.*, 1998) and TMMH (Meyer *et al.*, 2003) blast analysis, transmembrane domains were detected in 3 of the aminoacid sequences. Moreover, 5 hypothetical secreted or conserved hypothetical secreted proteins were also identified, in this case 2 of the protein sequence posses a transmembrane domain (table 1, n° 159 and 161). This means that the methodology for membrane protein extraction and separation was successfully applied. Nevertheless a degree of under-representation of membrane proteins is expected. This is reflected in the fact that from the 5 ABC-transporter systems identified, only the ATP-binding protein or the periplasmic binding proteins subunits were detected (table1). This was also observed in proteomics studies of membrane fractions in *S. meliloti* and *A. borkumensis* (Djordjevic *et al.*, 2003; Sabirova *et al.*, 2006).

Metabolic activity profiles in *Azoarcus* sp. BH72

Nitrogen metabolism

As expected the pathways involved in the N_2-fixing process and subsequent assimilation of the ammonia formed were active in all the conditions tested. The three subunits of the nitrogenase complex NifHDK were identified under the conditions studied, as well as enzymes related to its synthesis (NifS, NifU, NifN and NifE) and activation (NifM). Previous works in strain BH72 demonstrated that in pure-culture the NifH subunit protein is uniformly disperse in the

cytoplasm, while in co-culture conditions was mainly located on internal stacks membranes formed (diazosomes)(Hurek *et al.*, 1995). In agreement to that, in this work the NifH protein was identified to be down-regulated in co-culture cytoplasmic fractions, while it was always present in membrane fractions. Moreover, the NifM protein, which is required for activation and stabilization of NifH, was identified to be down regulated in the cytoplasm and up-regulated in the membrane fraction. This data give more evidence to the observation that the NifH protein is strongly associated to the membrane fraction in co-culture, in particulary to diazosomes as was described (Hurek *et al.*, 1995).

A mobile electron donor ferredoxin V FdxD protein, was identified to be constitutively expressed in the cytoplasmic fractions. In *Azoarcus* sp. BH72 genome the *fdxD* gene is strategically located downstream of the *nifHDK* genes flanked by another ferredoxin gene (*fdxG*), and the *nifQ* gene. Previous work in *Azoarcus* sp. BH72 has demonstrated that the *nifHDK* genes are co-transcribed in one operon together with a ferredoxin gene *fdxN*. Mutational analysis showed that FdxN play an important but not essential role in nitrogen fixation. A *fdxN* mutant (strain BHΔfdxN$^-$), had a reduced nitrogenase activity in pure and co-culture, nevertheless comparable amounts of nitrogenase Fe protein were present in both conditions (Egener *et al.*, 2001). This reduced activity was explained to be cause of a less efficient electron transport to dinitrogenase reductase in absence of FdxN and not as a destabilization of *nifH* mRNA. Most probably the residual nitrogenase activity found in the FdxN$^-$ mutant could be explained by the presence of the FdxD protein identified in this work.

The large subunit of the hydrogenase complex (HupL), was detected to be constitutively expressed in strain BH72 N_2-fixing cells and membrane associated. In endosymbiotic bacteria the Hup complex allows the oxidation of the hydrogen produced by nitrogenase. This symbiotic hydrogen oxidation has been shown to reduce the energy losses associated with nitrogen fixation and to enhance productivity in certain legume hosts (Ruiz-Argüeso *et al.*, 2001). In *Rhizobium leguminosarum* bv. viciae UPM791 was shown that the structural and most of the accessory *hup* genes are only expressed in the nodules, and that they are under the control of NifA (Ruiz-Argüeso *et al.*, 2001). Since to be that in *Azoarcus* sp. BH72 the Hup complex also plays an important role in the hydrogen oxidation under N_2-fixing conditions, a fact which is particularly important in co-culture conditions where larges amounts of hydrogen are formed as a product an intensive N_2-fixation process.

One of the key regulators of *Azoarcus* sp. BH72, involved in the nitrogen metabolism, was also identified under the conditions studied. The NtrC protein, a member of a two component regulatory system (NtrBC), was constitutively expressed in the cytoplasm and interestingly down-regulated in the membrane fraction. The NtrBC two-component system plays an important role in sensing and inducing N-regulatory cascades involved in N_2-fixation and N-assimilation (Ninfa *et al.*, 1995). In *Azoarcus* sp. BH72 it was demonstrate that NtrBC activates

nifLA gene expression, moreover that NifA acts as an essential activator for the *nifHDK* genes (Egener *et al.*, 2002). These previous observations are in agreement with results obtained here in which the constitutive expressions of NtrC as well as of the NifHDK proteins were shown during N_2-fixation growths.

In the context of N-assimilation, the strain BH72 genome shows the presence of the high-affinity ATP-consuming assimilation system (GS-GOGAT). From this system, two copies of the glutamine synthetase gene (*glnA* and *glnT*) were annotated, which encode for a glutamine synthetase type I and III protein respectively (Krause *et al.*, 2006). In *Herbaspirillum seropedicae* the *glnA-ntrBC* operon is subjected to NtrBC dependent transcription and activity control (Persuhn *et al.*, 2000). The situation is more complex in *Rhizobium*, where three GS encoding genes exist. In this case, the *glnII* promoter is fully dependent on the positive control by the NtrC protein (Patriarca *et al.*, 1992). In the present work, the glutamine synthetase type I protein (GlnA), was identified constitutively expressed in both N_2-fixing conditions studied.

Interestingly, the two subunits CarA and CarB of a carbamoyl phosphate synthase (glutamine-hydrolyzing) protein were detected to be constitutively expressed in the cytoplasm and up-regulated in the membrane respectively. The Car protein is involved in the glutamine assimilation by forming glutamate and carbamoyl-phosphate as a part of the glutamate and pyrimidine metabolism. Since no GOGAT protein was detected under the conditions studied, most probably the glutamine assimilation to glutamate could occur in *Azoarcus* sp. BH72 by using this pathway. In particular, the CarB subunit was detected as a novel protein in the wild-type strain, most probably as a consequence of an intense N_2-fixation process which occurred in the diazosomes forming cells.

An ATP-binding protein GlnQ1, which is a subunit of an ABC-glutamine transporter (GlnQ1MPH), was detected associated to the membrane fraction. The regulation of the *glnHPQ* operon is diverse in bacteria. In *Escherichia coli* and *Salmonella typhimurium* this operon is regulated negatively by nitrogen availability in particular by the intracellular glutamine pool (Kustu *et al.*, 1979). Particulary in *E. coli*, an NtrC binding site was found upstream of *glnH* indicating that most probably this operon is under the control of the NtrC regulator (Nohno and Saito, 1987). In *Azoarcus* sp. BH72 no putative NtrC binding site was detected upstream of the *glnQ1MPH* operon which means that most probably this operon is not NtrC-regulated. Moreover, taking in mind that this protein was detected down regulated in the wild type strain, and constitutive in the BH1599 (table 1, n°14), a putative N-regulation through the glutamine pool most probably is taking place. A suitable explanation is that in the wild type co-cultures cells, high amounts of glutamine are produced as a consequence of an intense N_2-fixation process. On the other hand, in strain BH1599 the N_2-fixation process is no so efficient with the concomitant less production of glutamine. The high amount of glutamine produced in the wild

type most probably repress the expression of the *glnQ1* gene, while in the mutant strain this amount is not enough.

An additional set of proteins related to the N-metabolism were identified in the cytoplasm fraction. These include the two subunits constitutively expressed of the glycine cleavage multienzyme complex (GCV), the up-regulated PurD protein, as well as the alpha subunit of the urease protein, (UreC). This last protein was detected novel in the wild type strain. Interestingly the GCV complex as well as the PurD protein use glycine as a substrate. The PurD protein, form N1-(5-phospho-D-ribosyl) glycinamide, an intermediate of the purines metabolism. More probably this protein is regulated by glycine and its up-expression could be as a consequence of a high amount of glycine produced by an active N_2-fixing process in co-culture conditions. On the other hand, the GCV complex catalyses the oxidative cleavage of glycine into: CO_2, NH_3 and a C1 unit. In *E.coli* the GCV encoding genes (*gcvTHP*), are positively regulated by glycine (Heil *et al.*, 2002). *Azoarcus* sp. BH72 genome encodes for a *gcvTHP* operon which has the same genetic arrangement as in *E. coli*, which means that most probably the expression of this complex is due to a glycine induction. Moreover, the C1 units generated as a product of the reaction are not end products. Like in *E.coli,* these compounds could be used as substrates in other metabolic pathways like the biosynthesis of purines, thymidine, histidine, coenzyme A and methionine (Heil *et al.*, 2002), from where proteins were also identified in strain BH72 (table 1).

The UreC protein, which is part of the urease complex, catalyzes the hydrolysis of urea to form ammonia and CO_2 as a part of the urea cycle. In *Sinorhizobium meliloti* the gamma subunits of urease, as well as the urea-amide binding protein were identified to be down regulated in protein extract of bacteroids. Results obtained in this bacterium are consistent with the shift from ammonium assimilation to ammonium export and the idea that bacteroids are nitrogen proficient (Natera *et al.*, 2000). On the other hand in *R. leguminosarum* bv. viciae moderate levels of urease activity were detected in vegetative cells as well as in pea bacteroids. This activity did no require urea for induction and was partially repressed by the addition of ammonium into the medium (Toffanin *et al.*, 2002). These observations show no clear regulation of the urease expression and activity in rhizobia under N_2-fixing conditions where most probably an ammonium repression is involved. Results obtained in *Azoarcus* sp. BH72 suggests a urease ammonium-independent regulation since both conditions studied were no rich in ammonium.

Carbon metabolism

Almost all the enzymes involved in the tricarboxilic acid (TCA) cycle were identified in the different conditions studied (table 1 and Figure 2), showing the key role of this pathway in the *Azoarcus* sp. BH72 metabolism for energy production, as well as for source of different metabolic intermediates. This is reflect with the fact that numerous enzymes involved in the

amino acid, fatty acid or cofactor biosynthetic pathways as propanoate or porphyrins, which use TCA intermediate compounds as precursors, were also detected (table 1).

The glyoxylate cycle was very active under the conditions studied. Constitutive expressed and membrane associated a protein with homology to a NAD^+-dependent malic enzyme (MaeB2), which catalyzes the formation of pyruvate from malate, was detected. Pyruvate formed can be used as a precursor of the valine, leucine and isoleucine biosynthesis pathways, from where enzymes were also identified (table1). Alternatively, pyruvate could be converted to acetyl-CoA by the pyruvate dehydrogenase complex, from which a dihydrolipoamide dehydrogenase LpdA enzyme was found. This enzyme was detected constitutively expressed in the wild type strain and down regulated in the BH1599 strain (table 1, n° 41), which reflect and alternative way to form acetyl-CoA in co-culture mutant strain. C4-dicarboxylate acids like succinate, fumarate and malate are internalized and metabolized by bacteria under aerobic or anaerobic conditions (44). In pure culture *Azoarcus* sp. BH72 was grown in SM medium with L-malate as carbon source (Reinhold *et al.*, 1985), a fact which explain the MaeB2 expression. The added concentration of L-malate in the FU medium was chosen to be 1% from those taken under pure culture conditions with the aim to promote the growth of strain BH72 in the beginning and to bypass the lag-phase (Hurek *et al.*, 1995). By HPLC-analysis it was ensured that the L-malate was completely metabolized, reflecting that the expression in co-culture of MaeB2 is due to its role in the pyruvate metabolism. In rhizobia, C4-dicarboxylates supplied by the host plant are major carbon and energy sources during symbiosis (Yurgel and Kahn, 2004). In particular, *S. meliloti* genome encode two distinct malic enzymes, Tme (Smc1126) and Dme (Smc00169) which are $NADP^+$- and NAD^+-dependent, respectively. Dme is required for symbiotic N_2-fixation by bacteroids and is a part of the pathway for the conversion of C4-dicarboxylic acids to acetyl-CoA (Djordjevic, 2004). In *Azoarcus* sp. BH72 genome, also two types of malic enzymes, MaeB1 (azo0821) and MaeB2 (azo3211) are present (Krause *et al.*, 2006). In this work we were able to identify MaeB2 in pure and co-culture, which was annotated as a NAD^+-dependent enzyme. These results indicate that, like in *S. meliloti*, the NAD^+-dependent copy of the malic enzyme is used in both N_2-fixing conditions studied. The role of the second malic enzyme copy present in strain BH72 genome remains unknown. Additionally, one component of a C4-dicarboxylate transporter system DctP2, was detected down regulated in co-culture cells. In rhizobia as well as in *R. capsulatus*, C4-dicarboxylate transport systems, and the corresponding pathways, are induced only in the presence of external C4-dicarboxylates, where a two component regulatory system is involved (Janausch *et al.*, 2002; Yurgel and Kahn, 2004). The coding genes for a sensor and regulator proteins (DctRS), are present in strain BH72 genome (azo0917 and azo0918). These genes are located upstream of one set of the C4-dicarboxylate transporter operon (*dctP2Q2M2*) from where the DctP2 subunit was identified. Results obtained show that in *Azoarcus* sp. BH72, the C4-dicarboxylate system DctP2Q2M2 is a

malate transporter which is regulated by the external amount of malate via the two component system *dctRS*.

An additional key enzyme from the glyoxylate cycle, the isocitrate lyase AceA protein, which is involved in the conversion of isocitrate into succinate and glyoxylate, was detected up-regulated in co-culture in agreement with the growth conditions used. The succinate formed could maintain the levels needed for a continue flow of the TCA cycle, as well undergo to other key biosynthetic pathways, e.g. porphyrins and heme synthesis. On the other hand, the glyoxylate also formed in this reaction could be transformed to malate, which can be used in the TCA cycle or alternatively to form pyruvate as was described before.

The glycolysis-gluconeogenesis pathways were also active in all the conditions studied. Several proteins involved in this pathway were identified (table 1), most probably with the aim to be used as a synthesis pathway for glucose, and other key structural compounds. This hypothesis is supported by the fact that strain BH72 can not use glucose and other carbohydrates as carbon sources (Reinhold-Hurek *et al.*, 2005) and additionally, by the presence of two up-regulated proteins involved in the synthesis and transport of exopolysacharides (WbnF and Wza2). Moreover, a probable CDP-4-dehydro-6-deoxyglucose reductase AscD protein, involved in the starch and sucrose metabolism was detected in the cytoplasm of pure culture cells (table 1, n°168).

In *Azoarcus* sp. BH72, the TCA and glyoxylate cycles are linked to the gluconeogenesis pathways by the phosphoenolpyruvate carboxykinase Pck enzyme, which was up-regulated in co-culture. Pck mutants in *S. meliloti* showed a reduce level in nitrogen fixation, while in *R. leguminosarum* no apparent symbiotic phenotype was detected (Osteras *et al.*, 1997). However, this enzyme was found only in proteins extracted from *S. meliloti* cultured cells (Djordjevic, 2004). In contrast to rhizobia, under N_2-fixing conditions the role of this protein as a link between these two key pathways in the *Azoarcus* sp. BH72 metabolism especially in co-culture growth seem to be evident.

The pyruvate metabolism appears to plays a key role in *Azoarcus* sp. BH72, especially in co-culture, where a shift in carbon sources was detected. In co-culture growths, strain BH72 was totally dependent on fungus exudates since no other carbon source as glucose was included in the FU medium (Hurek *et al.*, 1995). HPLC analysis showed that under pure culture growth conditions *A. alternatum* excretes acetate, lactate, formiate and unidentified compound (final concentration in mM: 1.13, 0.89 and 1.0 respectively; 100% for the unidentified). Using the same approach it was also demonstrated that under co-culture conditions, the amount of acetate and lactate but not formiate decreased drastically (final concentration in mM: 0, 0, and 0.83 respectively; 60% for the unidentified). These results are in correlation with experiments which showed that *Azoarcus* sp. BH72 can grow on acetate and lactate but not on formiate as a sole carbon source. Moreover, these physiological results are in agreement with the fact that several

enzymes, involved in the metabolism of these compounds, were detected up-regulated in co-culture growth conditions. It seems that the acetate assimilation started with its conversion into acetyl-CoA by the AscB enzyme. This enzyme was detected constitutively expressed in the membrane and novel expressed in wild type co-culture cytoplasm cells. This is most probably because a close contact between strain BH72 and the fungus exudates than in strain BH1599 which can not properly attach to the fungus mycelium. Acetyl-CoA formed can be directly used in several metabolic pathways including TCA, propanoate metabolism and fatty acids biosynthesis from where enzymes were identified (table1). Alternatively, acetyl-CoA can be converted by the ThlA enzyme, which was detected constitutively expressed, into acetoacetyl-CoA. This last compound could be directed into the poly-β-hydroxibutyrate (PHB) biosynthetic pathway, from where two enzymes involved in their synthesis and regulation were identified to be up-regulated (table1). PHB is an intracellular reserve of carbon and reducing energy that bacteria accumulate when a nutrient, other than carbon is limiting for growth (Dawes and Senior, 1973). In N_2-fixing bacteria different factors are involved in the regulation of PHB synthesis including nitrogen limitation (Hurek *et al.*, 1987b; Hurek *et al.*, 1995), ammonia (Sun *et al.*, 2000), high or low O_2 concentrations (Hurek *et al.*, 1987b). Results obtained in this work support the idea that most probably the PHB synthesis is regulated by an O_2-limitation in the conditions studied rather than other factors mentioned.

Only one probable enzyme involved in lactate assimilation, the lactoylglutathione lyase GloA protein, was detected as up-regulated in BH1599 co-culture cells. GloA is involved in the formation of methylglyoxal and glutathione from lactoylglutathione, a compound derivate from lactate. Both compounds methylglyoxal and glutathione can undergo into the glycine and threonine metabolism from where several enzyme were also detected (table1).

Strongly up-regulated in the cytoplasm, two copies of the alcohol dehydrogenase ExaA2 and ExaA3 enzymes, as well as the quinoprotein ExaB1, were detected (table1). The ExaB1 protein was only present in the BH1599 strain, while four isoforms of the ExaA2 protein were present in co-culture cells (Figure 1B, n°48). By their relative positions in the 2D-gels the alcohol dehydrogenase spots identified here were the same spots up-regulated previously observed in 2D-gels of total proteins extracted from diazosome-containing cells (Karg and Reinhold-Hurek, 1996). This set of enzymes is involved in the ethanol oxidation to acetaldehyde. The acetaldehyde formed most probably is converted to acetate by the NAD^+-aldehyde dehydrogenase AldA protein, which was also detected up-regulated in co-culture cells, and then to acetyl-coA by the mentioned AscB enzyme. These results clearly show that the enzymes necessary for ethanol assimilation were very active in strain BH72 co-culture cells in agreement with the fact that this compound can be used as sole carbon source (Reinhold-Hurek *et al.*, 2005). HPLC analysis was also used with the aim to detect the presence of ethanol in the FU medium. By comparing the retention times from the unidentified exudate, with those from the

alcohol standards, was excluded that is an alcohol. Additionally, an indirect ethanol measurement using an alcohol dehydrogenase enzymatic approach was applied. This assay was used for ethanol detection in FU medium of *A. alternatum* grown, in pure- or co-culture with strain BH72 or strain BH1599, during one week. Results show that the activity of the enzyme, which reflects the amount of ethanol in the medium, was high during the first 3 days and decrease to no activity after 5 days. The same result was observed in both pure- and co-culture growth conditions, with the only difference that in co-culture the activity decreased faster (data not shown). These results show that *A. alternatum* produced and alternatively also metabolized ethanol, while strain BH72 is able to use also this compound as a carbon source in agreement with the up-regulation of enzyme involved in the ethanol assimilation in co-culture conditions. On the other hand, these results explain the impossibility to detected ethanol using the HPLC approach, since no activity was detected after 5 days.

Energy production

The process of nitrogen fixation is highly energy demanding as nitrogenase requires 16 molecules of ATP for the reduction of 1 molecule of N_2 to ammonium. During this process, which occurrs under extremely low oxygen concentrations, a respiratory chain with high affinity to oxygen closely coupled to ATP production is required (Saeki, 2004). This hypothesis is supported in *Azoarcus* sp. BH 72 by results obtained in which several subunits of the NADH-ubiquinone oxidoreductase complex (Nuo), involved in the respiratory chain, were detected membrane associated (table 1). Moreover, a PetA1 and CycH proteins, which are part of a cytochrome and involved in its synthesis respectively, were identified also membrane-associated (table 1). The PetA1 protein was found constitutively expressed, while the CycH was detected novel in co-culture wild type cells. This specific up-regulation is correlated with the presence of specialized internal membranes in strain BH72 co-culture cells where most probably this protein is located. In the cytoplasm fraction three electron transfer flavoproteins EtfA1, Etf1 and IsiB were identified (table 1). In particular, the Etf1 protein was detected up-regulated in correlation with an intensive, high energy demanding, N_2-fixation process in co-culture conditions. Finally, several subunits of the ATP synthase were detected membrane-associated and in some cases up-regulated in co-culture grown conditions (table1). In a context these results are in agreement with the hypothesis in which diazosomes formed in strain BH72 under co-culture conditions are not only involved in the N_2-fixation, but also in the respiration process (Hurek *et al.*, 1995).

Protein synthesis and folding

An active protein synthesis in both N_2-fixing conditions studied is reflected by the large number of proteins identified related to this process particulary in co-culture (Table 1). This suggests a

differentiation process during co-cultures growths and is in concordance with the detection of smaller amount of spot proteins in co-culture 2D-proteome patterns.

A large number of proteins involved in chaperoning were identified including the constitutive expressed GroES1, DnaK, Tig, PpiB, Hsp18 and ClpX proteins (table 1). The first three appeared to be the major spot in the gels and their role in protein folding, degradation and in the assembly of large protein complexes are well known in bacteria (Bukau and Horwich, 1998). Particularly DnaK was identified also as down regulated in membrane fraction. Different DnaK isoforms were reported in *Agrabcterium tumefaciens* and *Desulfovibrio vulgaris* probably as a product of posttranslational modifications (Chhabra *et al.*, 2006; Rosen and Ron, 2002). This fact could be the explanation for the two DnaK representing spots detected in the strain BH72 membrane and cytoplasm proteome. Additionally, PpiB and ClpX are chaperones which accelerates the folding of proteins (Herrler *et al.*, 1994), while the Hsp18 chaperonine belongs to the small heat shock protein family, which are involved in the prevention of the accumulation of unfolded intermediates during stress periods (Veinger *et al.*, 1998). Members of these chaperone families mentioned were reported to be important during the establishment of the rhizobia-host symbiotic association and expressed in bacteroids cells, as well as induced under nutrient stress conditions (Djordjevic *et al.*, 2003; Djordjevic, 2004; Munchbach *et al.*, 1999). Results obtained from *Azoarcus* sp. BH72 suggest that the N_2-fixation process could be a stressful event for the bacterium, as many of the chaperones identified play a key role during stress response. This hypothesis is supported also by the presence of additional up-regulated proteins in strain BH72 with chaperoning function like GroeL1, and a conserved hypothetical peptidyl-prolyl cis-trans isomerase protein (table 1). Differentially expression of chaperone sets in pure- or co-culture cells, as well as in the membrane and the cytoplasmic fraction indicates a shift in the cellular status and moreover, where an active specific protein synthesis is taking place. This observation is correlated with the presence of internal cytoplasmic membranes in strain BH72 co-culture cells, which most probably require for an extra set of protein for its building and maintenance. An additional evidence for that is the identification of enzymes related to the processing of proteins. These include the constitutive expressed Pcm protein, as well as the up-regulated PrlC, Prc2 and a probable Zn dependent peptidase (table1). The Pcm and the PrlC proteins were detected in the membrane fraction, while interestingly the last one was only present as novel protein in wild-type cells.

Iron uptake and metabolism related proteins.

Five outer membrane (OM) TonB-dependent receptors for iron compounds, a substrate-binding component (FbpA1) of an ABC iron-transporter system and a copy of the bacterioferritin protein Bfr3, were identified (table 1). In Gram-negative bacteria, these type of receptors, transporters and storage proteins are responsible for the specific uptake of ferric-siderophore

complexes under iron-limited conditions, as well as in the maintenance of the intracellular iron homeostasis (Andrews *et al.*, 2003). Interestingly, the OM receptor azo3556 as well as the FbpA1 and Bfr3 proteins were found to be up-regulated in co-culture. In particular, the FbpA1 protein was detected up-regulated only in wild-type co-culture cells, suggesting a specific role in this strain. The mentioned receptor, azo3556, has high homology with OM receptors TonB-dependent for vitamin B12. In strain BH72, this gene is located within a genomic region which contain gene related with the vitamin B12 uptake and synthesis. Moreover, vitamin B12 was supplied in the FU medium used for co-culture growths. Taking in mind these observations, we can assume that this receptor is involved in the vitamin B12 uptake and most probably, is regulated by the vitamin B12 present in the growth medium.

Blast analysis of the additional OM receptors sequences identified showed a high degree of similarity with putative OM TonB-dependent receptors for ferric-citrate, hydroxamate and catecholate siderophore-types. Siderophore compounds are synthesized and secreted by bacteria under iron limitation (Andrews *et al.*, 2003). With the aim to find a putative correlation between these OM receptors and transports systems, with the synthesis and excretion of siderophores in the conditions studied, different tests were applied to the supernatant of the fungus medium. Chemical assays for specific or general siderophore detection shows no positive results in both growths conditions used. Moreover, no positive results were obtained during one week of pure or co-cultures growths of strain BH72 or strain BH1599 and *A. alternatum*. These results are in agreement with the absence of siderophores biosynthetic pathways in strain BH72 genome (Krause *et al.*, 2006). Furthermore, in this study both growths media conditions used were iron rich, as this element was supplied in the form of Fe^{3+}-EDTA. Strain BH72 genome encode for a FeoB protein (azo0606), in *E.coli*, FeoB is involved in iron II uptake under anaerobic conditions in which this ions is available (Kammler *et al.*, 1993). Taking this in mind, most probably under the anaerobic iron-sufficient condition used in this work, a FeoB protein is involved in the iron uptake.

In *S. meliloti* and *R. leguminosarum* siderophore OM TonB-dependent receptors were found iron-regulated, while mutants in these receptors were not affected in its capacity to fix nitrogen (Lynch *et al.*, 2001; Yeoman *et al.*, 2000). That means that the siderophore OM TonB-dependent receptors are not necessary during symbiosis. Results obtained in strain BH72 suggest an iron-independent regulation of these OM receptors proteins under both N_2-fixing conditions studied and moreover, that these receptors are involved in other process independently of iron acquisition from siderophores.

In *Ralstonia solanacearum* it was described that the binding of the PrhA protein, an OM protein TonB-dependent siderophore receptor, to a non-diffusible structure of the cell wall, initiated a regulatory cascade that finally induced the *hrp* (hypersensitive response and pathogenicity) genes. These genes are crucial for the interaction with host and non-host plants (Marenda *et al.*,

1998). On the other hand, was reported that OM TonB-dependent receptors are also important for the perception of environmental signals and are associated with pathogenicity of plant pathogens (Koebnik, 2005). Taking in mind these observations we can speculate that in strain BH72 these OM TonB-dependent receptors could be involved in the perception of environment signals or in bacteria-fungus interaction. In this sense interestingly is the case of the OM receptor azo0378 which was detected constitutively expressed in the wild-type strain and down regulated in the strain BH1599 (table 1, n°135). This kind of expression could be due to an inefficient association between strain BH1599 and the fungus, as this strain can not attach to the fungus.

Cellular protection and detoxification

Production of reactive oxygen species (ROS), are an intrinsic property of aerobic metabolism. Particularly in N_2-fixing bacteria, ROSs are additionally generated by the high requirement for iron as a key component in the nitrogenase and several electron-transport proteins (Santos *et al.*, 2001b). Additionally, nitrogenase is extremely oxygen sensitive which means that the inactivation of ROSs is very important for a correct N_2-fixation process. According to that several proteins involved in the detoxification of ROSs were identified in all the conditions. Moreover, some of them were up-regulated in co-culture growth conditions, such as AhpC, SodC, Ohr, Tpx (only in the strain BH1599), and a conserved hypothetical peroxiredoxin (table1). The antioxidant AhpC appeared in the wild-type co-culture membrane fractions, as a multiple spots with slightly different molecular weights and pIs (table 1 and Figure 1F, n° 112). The same observation was made in *Desulfovibrio vulgaris,* suggesting stress-dependent posttranslational modifications (Chhabra *et al.*, 2006). Several proteins included the two major spots SodB and KatA, involved in the detoxification of ROSs, were found to be up-regulated in *S. meliloti* bacteroids. Mutants from both proteins were drastically affected in their symbiotic capacity reflecting a key protective role in their symbiotic process (Djordjevic *et al.*, 2003; Djordjevic, 2004). These observations show that the main importance of the detoxification proteins appeared to be in highly N_2-fixing conditions, a feature also shown by *Azoarcus* sp. BH72 in co-culture growths.

Motility and adhesion

Flagellins are part of the flagella structure, and its role in motility, adhesion, biofilm formation and in colonization of the host organism is well known (Moens and Vanderleyden, 1996). Constitutively expressed and associated to the membrane fraction two copies of the flagellin protein FliC2 and FliC3 were identified, suggesting a quite unusual role for both proteins.

On the other hand, a putative PilP protein which is essential for the biogenesis of type IV pili system (Filloux, 2004), was identified as novel in the mutant strain. As strain BH1599 is a pilus mutant, maybe the overexpression of this protein is because no PilA is present to be assembly.

Structural elements: membrane proteins

Membrane proteins are of particular interest in proteomic studies. In bacteria they play important functions in signal transduction pathways, transport, cell interactions and other processes. A set of three major constitutive spots in the membrane fraction were identified as outer membrane porins (table 1). They were annotated in *Azoarcus* sp. BH72 as belonging to the general bacterial porin (GBP) family (T.C.:1.B.1). The function of this category is diverse, some porins are cation-selective, others are anion-selective and still others are selective for specific compounds (e.g., sugars, nucleotides, phosphate, pyrophosphate) (Nikaido, 2003). In *S. meliloti* a set of outer membrane proteins with unknown function were also identified in cultured and nodule bacteria, suggesting an important role of these proteins in the bacterial survival under several conditions (Djordjevic, 2004). Interestingly isoforms of the porins mentioned in strain BH72 were up-regulated in co-culture (Figure 1F and 1G, n° 137, 138 and 139). In many host-microbe interactions, porin proteins are virulence factors (Massari *et al.*, 2003). The presence of new isoforms in co-culture possibly indicates that these porins could play a role in strain BH72-fungus interaction.

Proteins related to the cell envelope biogenesis, integrity and shape were detected in both conditions studied. Constitutively expressed a protein with similarity to OstA, an organic outer membrane tolerance protein was identified. In *E.coli*, OstA is involved in the *n*-hexane tolerance level as well as in the cell envelope biogenesis and integrity (Braun and Silhavy, 2002). The *E.coli ostA* gene is co-transcribed with *surA* located upstream, which encodes a periplasmic peptidyl-prolyl cis-trans isomerase with additional chaperone activity (Behrens *et al.*, 2001). *Azoarcus* sp. BH72 has the same gene arrangement as *E.coli*, indicating that these two proteins probably have the same function. Interestingly, a putative peptidoglycan-associated lipoprotein Pal, member of the OmpA family, was detected down regulated in the wild type strain and constitutive in the strain BH1599 (table 1). The mayor role of Pal in the outer membrane integrity in association with peptidoglycan was reported in *E.coli* as well as in *Haemophilus influenzae* (Cascales *et al.*, 2002; Parsons *et al.*, 2006). Additionally, a rod shape-determining protein MreB, was identified down-regulated in co-culture cells. Previous microscopical observations in strain BH72 co-culture cells detected frequently dumbbell-shaped bacteria (Hurek *et al.*, 1995). These observations are in correlation with the absence in co-culture of the MreB protein and with the down-regulation in the wild-type of the Pal protein. Moreover, theses results show that during co-culture growths, strong changes are taking place in the wild-type outer membrane.

Under the growth conditions studied two proteins related to the transport and/or efflux of compounds were identified. An AggA protein, member of the outer membrane efflux protein (OEP) family, was detected constitutively expressed. Additionally, a membrane fusion protein (azo0244) member of the HlyD family was found down regulated in the wild-type strain and constitutive expressed in the strain BH1599. Moreover, several isoforms of this protein were also detected up-regulated in the mutant strain (Figure 1G, n°141). Both proteins are part of a type I secretion system involved in the secretion of a number of proteins and molecules, and require the help of members belonging to the ABC transporter family (Nikaido, 2003). The presence of these two proteins suggests different efflux process in both strains during co-culture conditions.

Finally, two different outer membrane proteins (OprM3 and OmpW), involved in drug or antibiotic resistance were up-regulated (table 1). The OprM3 protein, which is a probable outer membrane efflux protein, member of the resistance-nodulation-cell division (RND) superfamily (T.C.: 2.A.6.-), was detected as novel in co-culture wild-type cells. In *Pseudomonas aeruginosa* was reported that OprM protein is the major outer membrane efflux porin involved in intrinsic multiple antibiotic resistance (Hancock and Brinkman, 2002). Additionally, the outer membrane protein precursor, member of the bacterial porin OmpW family (T.C.:1.B.39), was detected up-regulated in both strains. OmpW was reported to be related to the ampicillin and tetracycline resistance in *E.coli* (Xu *et al.*, 2006), while in *Vibrio cholerae* the expression was found to be dependent on *in vitro* culture conditions such as temperature, salinity, and availability of nutrients or oxygen (Nandi *et al.*, 2005). The induction of these proteins in co-culture are most probably related to the resistance against antibiotics synthesized by *A. alternatum* in co-culture, while in the case of OmpW, an addition oxygen regulation could be involved in co-culture growth conditions. To possess this set of antibiotic resistance proteins might allow strain BH72 to establish a compatible interaction with *A. alternatum.*

Diazosomes related proteins

As mentioned, internal membrane stacks (diazosomes), are formed in *Azoarcus* sp. BH72 during co-culture with *A. alternatum* and at low O_2 concentration (Hurek *et al.*, 1995). The approach used in this work allows to study of specific proteins which are expressed only in diazosome forming cells which could most probably be involved in the diazosome structure or function. In this sense, several membrane and membrane associated proteins were up-regulated in the wild-type strain (table1). From this set two of them (table 1, n°148 and 149) were annotated as conserved hypothetical membrane proteins. In both protein sequences, a transmembrane motif was found. In particular, the protein azo3667 was also present in the mutant BH1599 strain, but significantly less (wt/mut: 0.59± 0.29/0.2±0.17). These results suggest that these proteins could be involved in diazosome formation or function. Additionally

proteins probably located in these specialized structures are the mentioned NifH and CycH protein involved in the N_2-fixation and respiration process. Moreover, proteins like Pnp and PrlC were detected as well up-regulated in the membrane fraction of the wild-type cells (table1). These proteins are involved in RNA processing and messenger RNA degradation, as well as in oligopeptide cleavage respectively. Result obtained are in correlation with the presence of diazosomes in strain BH72 co-culture cells, where process such as N_2-fixation, respiration and ATP-synthesis are very active.

Miscellaneous

Strongly up-regulated in the cytoplasmic fraction was detected a protein (azo1228) with 77% similarity to a choline dehydrogenase (EC 1.1.99.1) and related flavoproteins from *Rubrivivax gelatinosus* PM1. Choline dehydrogenase catalyzes the four-electron oxidation of choline to glycine-betaine a potent osmoprotectant described in *Haemophilus elongata* and *S. meliloti* (Gadda and McAllister-Wilkins., 2003; Osteras *et al.*, 1998). The expression of this protein in co-culture is correlated to the presence of cholinechloride as well as betainhydrochloride in the FU medium used as an osmoprotectant.

On the other hand, interestingly a probably stringent starvation protein, SspA2 was identified constitutively expressed in the cytoplasmic fraction. This protein is synthesized when cells are exposed to protein starvation, and they are involved in the regulation of several proteins during exponential and stationary-phase growth (Hansen *et al.*, 2005). The presence of this protein is an indication that the cells in both conditions are stressed, most probably because in pure culture they have been harvest in a late exponential phase, while in co-culture after a week when the nutrients starting to be scared. Nevertheless, an additional functions than the protein starvation should not be discarded.

In the membrane fraction and up-regulated in co-culture a probable competence precursor ComL lipoprotein was identified. This type of protein is required for an efficient transformation by species-related DNA fact that shows that during co-culture growth a DNA exchange between *Azoarcus* strains is taking place. Also in the membrane fraction, a protein with homology to a sporulation initiation inhibitor ParA3 was detected constitutively expressed. ParA3 proteins are involved in the sporulation inhibition in bacteria.

REFERENCES

1. **Aebersold, R., and M. Mann.** 2003. Mass spectrometry-based proteomics. Nature **422:**198-207.
2. **Andrews, S. C., A. K. Robinson, and F. Rodriguez-Quinones.** 2003. Bacterial iron homeostasis. FEMS Microbiol Rev **27:**215-237.
3. **Arnow, and L. E.** 1937. Colorimetric determination of the components of 3,4-dihydroxyphenylalanine-tyrosine mixtures. Journal of Biological Chemistry **118:**531-537.

4. **Atkin, C. L., J. B. Neilands, and H. J. Phaff.** 1970. Rhodotorulic acid from species of Leucosporidium, Rhodosporidium, Rhodotorula, Sporidiobolus, and Sporobolomyces, and a new alanine-containing ferrichrome from Cryptococcus melibiosum. J Bacteriol **103:**722-733.
5. **Behrens, S., R. Maier, H. de Cock, F. X. Schmid, and C. A. Gross.** 2001. The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. Embo J **20:**285-294.
6. **Braun, M., and T. J. Silhavy.** 2002. Imp/OstA is required for cell envelope biogenesis in Escherichia coli. Mol Microbiol **45:**1289-1302.
7. **Bukau, B., and A. L. Horwich.** 1998. The Hsp70 and Hsp60 chaperone machines. Cell **92:**351-366.
8. **Cascales, E., A. Bernadac, M. Gavioli, J. C. Lazzaroni, and R. Lloubes.** 2002. Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. J Bacteriol **184:**754-759.
9. **Chhabra, S. R., Q. He, K. H. Huang, S. P. Gaucher, E. J. Alm, Z. He, M. Z. Hadi, T. C. Hazen, J. D. Wall, J. Zhou, A. P. Arkin, and A. K. Singh.** 2006. Global analysis of heat shock response in Desulfovibrio vulgaris Hildenborough. J Bacteriol **188:**1817-1828.
10. **Dartigalongue, C., and S. Raina.** 1998. A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in Escherichia coli. Embo J **17:**3968-3980.
11. **Dawes, E. A., and P. J. Senior.** 1973. The role and regulation of energy reserve polymers in micro-organisms. Adv Microb Physiol **10:**135-266.
12. **Djordjevic, M. A.** 2004. Sinorhizobium meliloti metabolism in the root nodule: a proteomic perspective. Proteomics **4:**1859-1872.
13. **Djordjevic, M. A., H. C. Chen, S. Natera, G. Van Noorden, C. Menzel, S. Taylor, C. Renard, O. Geiger, and G. F. Weiller.** 2003. A global analysis of protein expression profiles in *Sinorhizobium meliloti:* discovery of new genes for nodule occupancy and stress adaptation. Molecular Plant Microbe Interactions **16:**508-524.
14. **Dörr, J., T. a. Hurek, and B. Reinhold-Hurek.** 1998. Type IV pili are involved in plant-microbe and fungus-microbe interactions. Mol. Microbiol. **30:**7-17.
15. **Egener, T., T. Hurek, and B. Reinhold-Hurek.** 1999. Endophytic expression of *nif* genes of *Azoarcus* sp. strain BH72 in rice roots. Molecular Plant Microbe Interactions **12:**813-819.
16. **Egener, T., D. E. Martin, A. Sarkar, and B. Reinhold-Hurek.** 2001. Role of a ferrodoxine gene cotranscribed with the *nifHDK* operon in N_2 fixation and nitrogenase "switch off" of *Azoarcus* sp. strain BH72. Journal of Bacteriology **183:**3752-3760.
17. **Egener, T., A. Sarkar, D. E. Martin, and B. Reinhold-Hurek.** 2002. Identification of a NifL-like protein in a diazotroph of the beta-subgroup of the *Proteobacteria*, *Azoarcus* sp. strain BH72. Microbiology **148:**3203-3212.
18. **Filloux, A.** 2004. The underlying mechanisms of type II protein secretion. Biochim Biophys Acta **1694:**163-179.
19. **Gadda, G. a., and E. E. McAllister-Wilkins.** 2003. Cloning, expression and purification of choline deydrogenase from the moderate halophile *Halomonas elongata*. Appl. Environ. Microbiol. **69:**2126-2132.
20. **Hancock, R. E., and F. S. Brinkman.** 2002. Function of pseudomonas porins in uptake and efflux. Annu Rev Microbiol **56:**17-38.
21. **Hansen, A. M., Y. Qiu, N. Yeh, F. R. Blattner, T. Durfee, and D. J. Jin.** 2005. SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli*. Mol Microbiol **56:**719-734.
22. **Heil, G., L. T. Stauffer, and G. V. Stauffer.** 2002. Glycine binds the transcriptional accessory protein GcvR to disrupt a GcvA/GcvR interaction and allow GcvA-mediated activation of the Escherichia coli gcvTHP operon. Microbiology **148:**2203-2214.

23. **Herrler, M., H. Bang, and M. A. Marahiel.** 1994. Cloning and characterization of ppiB, a Bacillus subtilis gene which encodes a cyclosporin A-sensitive peptidyl-prolyl cis-trans isomerase. Mol Microbiol **11:**1073-1083.
24. **Hurek, T., T. Egener, and B. Reinhold-Hurek.** 1997. Divergence in nitrogenases of *Azoarcus* spp., *Proteobacteria* of the ß-subclass. Journal of Bacteriology **179:**4172-4178.
25. **Hurek, T., L. Handley, B. Reinhold-Hurek , and Y. Piché.** 2002. *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. Molecular Plant Microbe Interactions **15:**233-242.
26. **Hurek, T., B. Reinhold, I. Fendrik, and E. G. Niemann.** 1987. Root-zone-specific oxygen tolerance of *Azospirillum* spp. and diazotrophic rods closely associated with Kallar grass. Applied & Environmental Microbiology **53:**163-169.
27. **Hurek, T., B. Reinhold, and E. G. Niemann.** 1987. Effect of oxygen on ammonium-grown continuous cultures of *Azospirillum*-spp. and diazotrophic rods closely associated with Kallar grass. Canadian Journal of Microbiology **33:**919-922.
28. **Hurek, T., and B. Reinhold-Hurek.** 1999. Interactions of *Azoarcus* sp. with rhizosphere fungi, p. 595-614. *In* A. Varma and B. Hock (ed.), Mycorrhiza, 2nd ed. Springer Verlag, Berlin.
29. **Hurek, T., B. Reinhold-Hurek, M. Van Montagu, and E. Kellenberger.** 1994. Root colonization and systemic spreading of *Azoarcus* sp. strain BH72 in grasses. Journal of Bacteriology **176:**1913-1923.
30. **Hurek, T., M. Van Montagu, E. Kellenberger, and B. Reinhold-Hurek.** 1995. Induction of complex intracytoplasmic membranes related to nitrogen fixation in *Azoarcus* sp. BH72. Molecular Microbiology **18:**225-236.
31. **Hurkman, W. J., and a. C.K.Tanaka.** 1986. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. Plant Physiol **81:**802-806.
32. **Janausch, I. G., E. Zientz, Q. H. Tran, A. Kroger, and G. Unden.** 2002. C4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta **1553:**39-56.
33. **Kagi, J. H., and B. L. Vallee.** 1960. The role of zinc in alcohol dehydrogenase. V. The effect of metal-binding agents on thestructure of the yeast alcohol dehydrogenase molecule. J. Biol. Chem. **235:**3188-3192.
34. **Kammler, M., C. Schon, and K. Hantke.** 1993. Characterization of the ferrous iron uptake system of *Escherichia coli*. J. Bacteriol. **175:**6212-6219.
35. **Karg, T., and B. Reinhold-Hurek.** 1996. Global changes in protein composition of N_2-fixing *Azoarcus* sp. strain BH72 upon diazosome formation. Journal of Bacteriology **178:**5748-5754.
36. **Koebnik, R.** 2005. TonB-dependent trans-envelope signalling: the exception or the rule? Trends Microbiol **13:**343-347.
37. **Krause, A., A. Ramakumar, D. Bartels, F. Battistoni, T. Bekel, J. Boch, M. Böhm, F. Friedrich, T. Hurek, L. Krause, B. Linke, A. McHardy, A. Sarkar, S. Schneiker, A. Syed-Ali, R. Thauer, F. Vorhölter, S. Weidner, A. Pühler, B. Reinhold-Hurek, O. K. and, and A. Goesmann.** 2006. Life inside cereals: Genomic insights into the lifestyle of the mutualistic, N2-fixing endophyte Azoarcus sp. strain BH72. Nat. Biotech. submitted.
38. **Kustu, S. G., N. C. McFarland, S. P. Hui, B. Esmon, and G. F. Ames.** 1979. Nitrogen control of Salmonella typhimurium: co-regulation of synthesis of glutamine synthetase and amino acid transport systems. J Bacteriol **138:**218-234.
39. **Laemmli, U. K.** 1970. Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature **227:**680-685.

40. **Lynch, D., J. O'Brien, T. Welch, P. Clarke, P. O. Cuiv, J. H. Crosa, and M. O'Connell.** 2001. Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. J. Bacteriol. **183:**2576-2585.
41. **Marenda, M., B. Brito, D. Callard, S. Genin, P. Barberis, C. Boucher, and M. Arlat.** 1998. PrhA controls a novel regulatory pathway required for the specific induction of *Ralstonia solanacearum hrp* genes in the presence of plant cells. Molecular Microbiology **27:**437-453.
42. **Martin, D., T. Hurek, and B. Reinhold-Hurek.** 2000. Occurrence of three P_{II}-like signal transmitter proteins in the diazotroph *Azoarcus* sp. BH72. Molecular Microbiology **38:**276-288.
43. **Martin, D. E., and B. Reinhold-Hurek.** 2002. Distinct roles of P_{II}-like signal transmitter protein and *amtB* in regulation of *nif* gene expression, nitrogenase activity, and posttranslational modification of NifH in *Azoarcus* sp. strain BH72. Journal of Bacteriology **184:**2251-2259.
44. **Massari, P., S. Ram, H. Macleod, and L. M. Wetzler.** 2003. The role of porins in neisserial pathogenesis and immunity. Trends Microbiol **11:**87-93.
45. **Meyer, F., A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, and A. Puhler.** 2003. GenDB-an open source genome annotation system for prokaryote genomes. Nucleic. Acids. Res. **31:**2187-2195.
46. **Moens, S., and J. Vanderleyden.** 1996. Functions of bacterial flagella. Crit. Rev. Microbiol. **22:**67-100.
47. **Munchbach, M., P. Dainese, W. Staudenmann, F. Narberhaus, and P. James.** 1999. Proteome analysis of heat shock protein expression in Bradyrhizobium japonicum. Eur J Biochem **264:**39-48.
48. **Nandi, B., R. K. Nandy, A. Sarkar, and A. C. Ghose.** 2005. Structural features, properties and regulation of the outer-membrane protein W (OmpW) of Vibrio cholerae. Microbiology **151:**2975-2986.
49. **Natera, S. H., N. Guerreiro, and M. A. Djordjevic.** 2000. Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. Mol Plant Microbe Interact **13:**995-1009.
50. **Nikaido, H.** 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev **67:**593-656.
51. **Ninfa, A. J., M. R. Atkinson, E. S. Kamberov, J. Feng, and E. G. Ninfa.** 1995. Control of bacterial nitrogen assimilation by the NRI-NRII two component system of enteric bacteria., p. 67-88. *In* T. J. Silhavy and J. A. Hoch (ed.), Two Component Systems of Bacteria. American Society for Microbiology, Washington D.C.
52. **Nohno, T., and T. Saito.** 1987. Two transcriptional start sites found in the promoter region of Escherichia coli glutamine permease operon, glnHPQ. Nucleic Acids Res **15:**2777.
53. **Osteras, M., E. Boncompagni, N. Vincent, a. M.-C. Poggi, and D. L. Rudulier.** 1998. Presence of a gene encoding choline sulfatase in *Sinorhizobium meliloti bet* operon: choline-*O*-sulfate is metabolized into glycine betaine. Proc. Natl. Acad. Sci. U S A **95:**11394-11399.
54. **Osteras, M., S. A. O'Brien, and T. M. Finan.** 1997. Genetic analysis of mutations affecting pckA regulation in Rhizobium (Sinorhizobium) meliloti. Genetics **147:**1521-1531.
55. **Pandey, A., and M. Mann.** 2000. Proteomics to study genes and genomes. Nature **405:**837-846.
56. **Parsons, L. M., F. Lin, and J. Orban.** 2006. Peptidoglycan recognition by Pal, an outer membrane lipoprotein. Biochemistry **45:**2122-2128.
57. **Patriarca, E. J., M. Chiurazzi, G. Manco, A. Riccio, A. Lamberti, A. De Paolis, M. Rossi, R. Defez, and M. Iaccarino.** 1992. Activation of the Rhizobium leguminosarum glnII gene by NtrC is dependent on upstream DNA sequences. Mol Gen Genet **234:**337-345.
58. **Persuhn, D. C., E. M. Souza, M. B. Steffens, F. O. Pedrosa, M. G. Yates, and L. U. Rigo.** 2000. The transcriptional activator NtrC controls the expression and activity of glutamine synthetase in Herbaspirillum seropedicae. FEMS Microbiol Lett **192:**217-221.

59. **Reinhold, B., T. Hurek, and I. Fendrik.** 1985. Strain-specific chemotaxis of *Azospirillum* spp. Journal of Bacteriology **162:**190-195.
60. **Reinhold, B., T. Hurek, E.-G. Niemann, and I. Fendrik.** 1986. Close association of *Azospirillum* and diazotrophic rods with different root zones of Kallar grass. Applied & Environmental Microbiology **52:**520-526.
61. **Reinhold-Hurek, B., T. Hurek, M. Gillis, B. Hoste, M. Vancanneyt, K. a. Kersters, and J. De Ley.** 1993. *Azoarcus* gen. nov., nitrogen-fixing proteobacteria associated with roots of Kallar grass (*Leptochloa fusca* (L.) Kunth) and description of two species *Azoarcus indigens* sp. nov. and *Azoarcus communis* sp. nov. Int. J. Syst. Bacteriol. **43:**574-584.
62. **Reinhold-Hurek, B., Z. Tan, and T. Hurek.** 2005. p. 890-901. *In* G. M. Garrity (ed.), Bergey's Manual of Systematic Bacteriology., vol. 2. Springer Verlag, New York.
63. **Rosen, R., and E. Z. Ron.** 2002. Proteome analysis in the study of the bacterial heat-shock response. Mass Spectrom Rev **21:**244-265.
64. **Ruiz-Argüeso, T., J. M. Palacios, and a. J. Imperial.** 2001. Regulation of the hydrogenase system in Rhizobium leguminosarum. Plant and Soil **230:**49–57.
65. **Saeki, K.** 2004. Electron transport to nitrogenase: diverse routes for a common destination.
66. **Santoni, V., M. Molloy, and T. Rabilloud.** 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis **21:**1054-1070.
67. **Santos, R., D. Herouart, S. Sigaud, D. Touati, and A. Puppo.** 2001. Oxidative burst in alfalfa-Sinorhizobium meliloti symbiotic interaction. Mol Plant Microbe Interact **14:**86-89.
68. **Schultz, J., F. Milpetz, P. a. Bork, and C. P. Ponting.** 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA **95:**5857-5864.
69. **Schwyn, B., and J. B. Neilands.** 1987. Universal chemical assay for the detection and determination of siderophores. Anal Biochem **160:**47-56.
70. **Sun, J., X. Peng, J. Van Impe, and J. Vanderleyden.** 2000. The ntrB and ntrC genes are involved in the regulation of poly-3-hydroxybutyrate biosynthesis by ammonia in Azospirillum brasilense Sp7. Appl Environ Microbiol **66:**113-117.
71. **Toffanin, A., E. Cadahia, J. Imperial, T. Ruiz-Argueso, and M. Palacios.** 2002. Characterization of the urease gene cluster from Rhizobium leguminosarum bv. viciae. Arch Microbiol **177:**290-298.
72. **Veinger, L., S. Diamant, J. Buchner, and P. Goloubinoff.** 1998. The small heat-shock protein IbpB from Escherichia coli stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. J Biol Chem **273:**11032-11037.
73. **Xu, C., X. Lin, H. Ren, Y. Zhang, S. Wang, and X. Peng.** 2006. Analysis of outer membrane proteome of Escherichia coli related to resistance to ampicillin and tetracycline. Proteomics **6:**462-473.
74. **Yeoman, K. H., F. Wisniewski-Dye, C. Timony, J. B. Stevens, N. G. deLuca, J. A. Downie, and A. W. Johnston.** 2000. Analysis of the *Rhizobium leguminosarum* siderophore-uptake gene *fhuA*: differential expression in free-living bacteria and nitrogen-fixing bacteroids and distribution of an *fhuA* pseudogene in different strains. Microbiology **146 (Pt 4):**829-837.
75. **Yurgel, S. N., and M. L. Kahn.** 2004. Dicarboxylate transport by rhizobia. FEMS Microbiol Rev **28:**489-501.

TABLE 1: Differentially expressed proteins derived from both membrane and cytoplasmic protein fractions of N_2-fixing *Azoarcus* sp. BH72 cells grown on pure- or co-culture with *Acremonium alternatum*.

FIGURE 1: 2D-proteome pattern of cytoplasm (A, B and C) and membrane (D, E and F) fractions of N_2-fixing *Azoarcus* sp. BH72 cells. In panels A and D strain BH72 was grown in pure culture, while in panels B, C, E and F was grown in co-culture with *Acremonium alternatum*. C and G: proteome pattern of mutant *Azoarcus* sp. BH1599. Isoelectric focus was performed using a capillar tube gel of pH 3-10. The focused proteins were separated on 12% SDS-PAGE gel and were stained with Coomassie blue colloidal stain. Spots shown in grey were detected down-regulated in co-cultures cells (panels A and E). In slim black circles are shown spots which were detected up-regulated in co-culture cells (panels B, C, D and G). Spots shown in thick black circles are unique in wild-type strain cells (panels B and F) or in mutant strain cells (panel C and G). Numbers indicate isoforms proteins listed in table 1.

FIGURE 2: Schematic representation of major carbon metabolic pathways identified in N_2-fixing *Azoarcus* sp. BH72 cells proteome. The names refer to the following proteins identified and listed in table 1: DctP2: C4-dicarboxylate-binding periplasmic protein, ExaA1,A2: quinoprotein ethanol dehydrogenase, AldA: aldehyde dehydrogenase (NAD+), AcsB: acetyl-coenzyme A synthetase, ThlA: acetoacetyl-Coa thiolase, LpdA: dihydrolipoamide dehydrogenase, GloA: lactoylglutathione lyase, AceA: isocitrate lyase, MaeB2: Probable malic enzyme ($NADP^+$) and PckG: phosphoenolpyruvate carboxykinase. Arrows indicates pathways expressed: constitutive

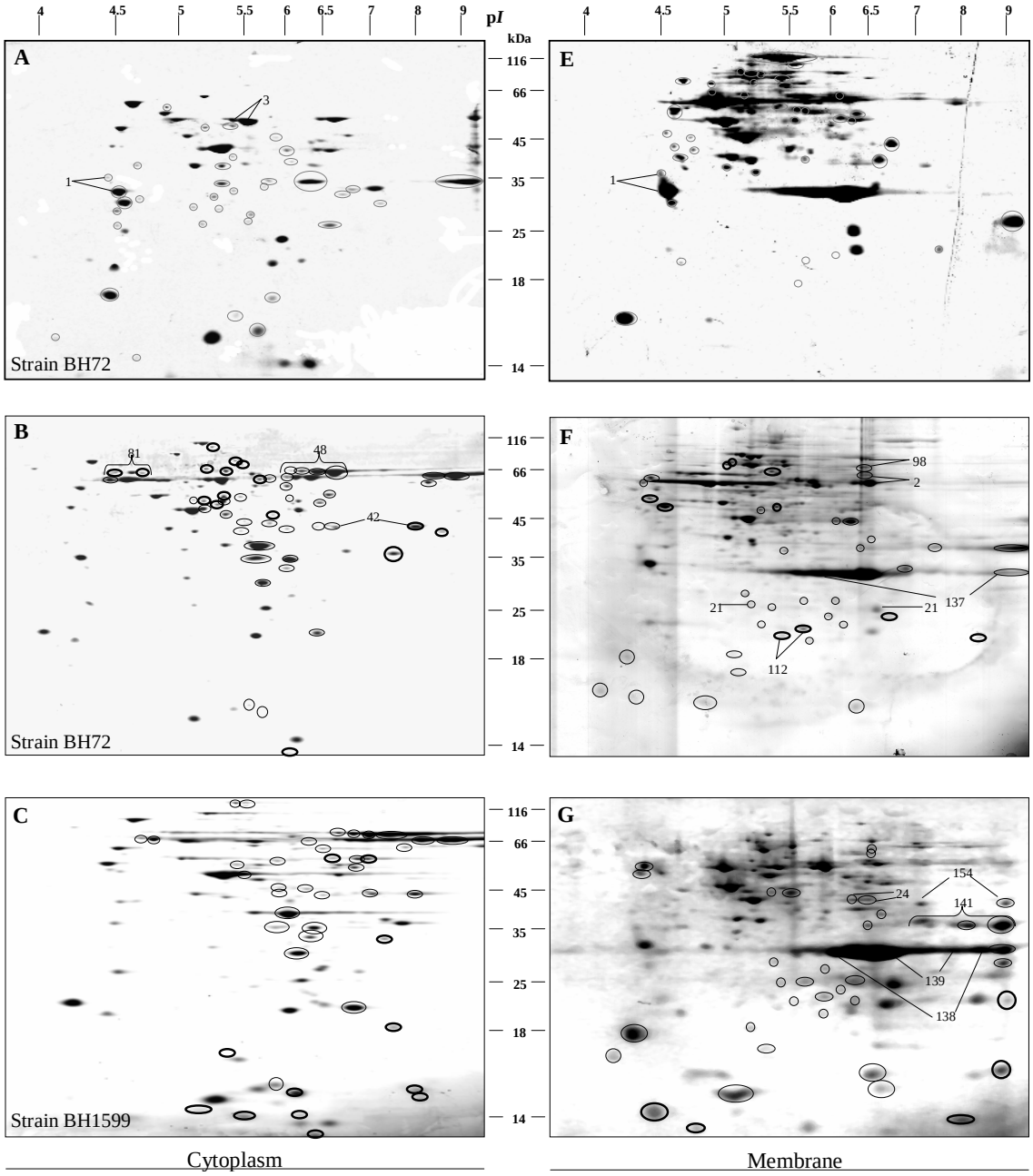

4
4.5
5
5.5
6
6.5
7
8
9
pI
kDa
116
66
45
35
25
18
14
A
Strain BH72
3
1
E
1
B
Strain BH72
81
48
42
F
98
2
21
21
137
112
C
Strain BH1599
G
154
24
141
139
138
Cytoplasm
Membrane

Figure 1

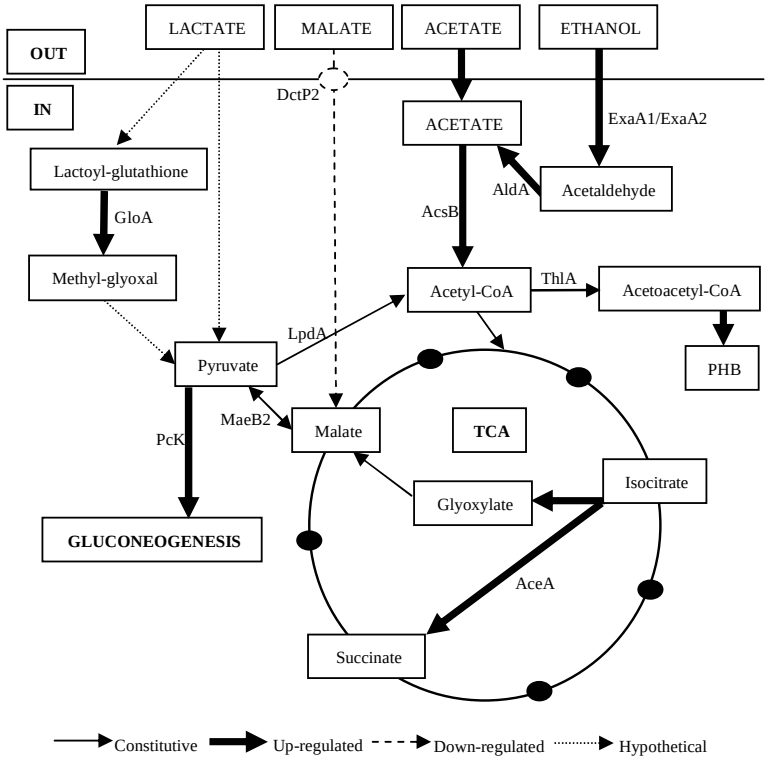

LACTATE
MALATE
ACETATE
ETHANOL
OUT
IN
DctP2
ACETATE
ExaA1/ExaA2
Lactoyl-glutathione
AldA
Acetaldehyde
GloA
AcsB
Methyl-glyoxal
Acetyl-CoA
ThlA
Acetoacetyl-CoA
LpdA
Pyruvate
PHB
MaeB2
Malate
TCA
PckA
Isocitrate
Glyoxylate
GLUCONEOGENESIS
AceA
Succinate
Constitutive
Up-regulated
Down-regulated
Hypothetical

Figure 2

Spot n°.[a] Function	Mol wt (kDa)	pI	Acces. n°[b]	Protein name[b]	Cov. (%)	Differential abundance[c]
Nitrogen metabolism						
1:C, M Iso.	31,2	4,7	azo0538	NifH: nitrogenase iron protein probable	30	↓(2,62± 0,51); CON
2:C, M Iso.	55,3	6,3	azo0539	NifD: nitrogenase molybdenum-iron protein alpha chain	59	CON
3: C Iso., M	59,1	5,68	azo0540	NifK: nitrogenase molybdenum-iron protein beta chain	45	CON
4:C	25,2	4,78	azo0543	NifY1: NifY protein	56	↓
5:C, M	32,0	5,1	azo0546	NifM: probable peptidylprolyl isomerase	49	↓; ↑
6:C	43,7	6,43	azo0552	NifS**:** cysteine desulphurase	34	CON
7:C	31,6	4,95	azo0553	NifU:Probable nitrogen fixation protein NifU.	42	↓(2,70± 0,69)
8:M	54,3	5,6	azo0562	NifE: nitrogenase iron-molybdenum cofactor biosynthesis protein	34	CON
9:M	48,8	5,83	azo0561	NifN: probable nitrogenase iron-molybdenum cofactor biosynthesis protein.	42	↓
10:C	13,0	6,7	azo0528	FdxD:Ferredoxin V (FdV) (Ferredoxin plant-type)	43	CON
11:M	66,3	6,19	azo3787	HupL: probable ferredoxin hydrogenase, large chain	27	CON
12:C,M	53,1	5,87	azo0735	NtrC: nitrogen regulation protein NR(I)	52	CON; ↓
13:C	51,9	5,29	azo0738	GlnA: Glutamine synthetase I	39	CON
14:M	26,7	7,08	azo0439	GlnQ1: probable glutamine transport ATP-binding protein.	34	wt/mut.: ↓/CON
15:C	60,3	5,3	azo3504	UreC : urease alpha subunit	33	↑ novel wt
16:C	13,0	4,0	azo1286	GcvH: glycine cleavage system H protein	65	CON
17:C	36,2	6,5	azo1287	GcvT:glycine cleavage system T protein	45	CON
18:C	40,4	6,4	azo1381	CarA: Carbamoyl-phosphate synthase small chain	43	CON
19:M	116,8	5,2	azo1382	CarB: carbamoyl-phosphate synthase	31	↑ novel wt
TCA, glyoxylate shunt, gluconeogenesis and pyruvate metabolism						
20:M	64,8	6,05	azo1551	SdhA: succinate dehydrogenase, flavoprotein subunit	55	CON
21:M, M Iso.	27,1	6,19	azo1552	SdhB: succinate dehydrogenase, iron-sulfur subunit.	45	CON; iso. novel co-culture
22:C, M	41,2	5,24	azo3332	SucC: probable succinyl-CoA synthetase, beta chain	56	CON
23:C, M	31,0	6,54	azo3333	SucD: Probable succinyl-CoA synthetase, alpha chain	44	CON; ↑ novel wt
24:C, M Iso.	48,3	6,15	azo1554	GltA: probable citrate synthase	37	CON; Iso.1: wt/mut.: CON/↓(4, 76± 1,355); Iso.2: CON
25: M	41,7	5,74	azo1556	OdhB: dihydrolipoamide S-succinyltransferase	47	↑
26:C	49,7	6,83	azo1557	OdhL: Probable dihydrolipoamide dehydrogenase	32	CON
27:M	82,0	5,58	azo3211	Maeb2:Probable malic enzyme	23	CON
28:C	98,3	5,7	azo1533	AcnA: aconitate hydratase	27	↑ novel wt
29:C, M	92,7	5,49	azo1534	AcnB: aconitate hydratase	38	CON; wt/mut: CON/↓
30:C	35,5	5,78	azo1547	Mdh: malate dehydrogenase	52	CON
31:C, M	47,9	5,6	azo1117	AceA: isocitrate lyase.	46	↑
32: C	40,0	5,63	azo3860	YeaU: tartrate dehydrogenase	34	CON
33:M	68,3	5,5	azo0820	PckG: probable phosphoenolpyruvate carboxykinase	21	↑
34:C	45,9	4,79	azo2144	Eno: enolase probable	16	CON
35:C	30,9	5,63	azo1096	DapA:dihydrodipicolinate synthase	42	↓
36:C	28,2	5,7	azo2576	DapB: dihydrodipicolinate reductase	28	↓
37:C	27,9	5,89	azo2808	GpmA: phosphoglycerate mutase probable	46	CON
38:M,C	41,5	5,75	azo2839	PgK: phosphoglycerate kinase probable	41	CON ; ↑
39:C	38,2	5,7	azo2843	FbaA : fructose-bisphosphate aldolase	38	↓
40:M, C	71,5	5,77	azo2414	AcsB: probable acetyl-coenzyme A synthetase	37	CON ; ↑ novel wt
41:M	62,7	5,8	azo1371	LpdA: dihydrolipoamide dehydrogenase	38	wt/mut.: CON/ ↓ (3,2 ± 1)
42:C Iso.	40,6	6,55	azo2172	ThlA: probable acetoacetyl-Coa thiolase	28	CON
43:M	50,4	5,79	azo3638	Pta: conserved hypothetical phosphate acetyltransferase	30	↓
44:C	13,6	5,0	azo3228	GloA: lactoylglutathione lyase	57	↑ novel mut.
45:C	63,7	5,72	azo1702	Acyl-CoA dehydrogenase	34	↑ novel co-culture
46:C	27,5	5,1	azo0790	PaaF1: putative enoyl-CoA hydratase	37	↓
47:C	55,1	6,4	azo2939	Alda: aldehyde dehydrogenase (NAD+)	33	↑
48:C Iso.	77,3	7,72	azo2972	ExaA2 :quinoprotein ethanol dehydrogenase probable	24	↑
49:C	41,7	8,23	azo2975	ExaA3: quinoprotein ethanol dehydrogenase probable	25	↑
50:C	15,6	6,9	azo2933	ExaB1: probable cytochrome c550	53	↑ novel mut.
51:M	22,4	5,28	azo3283	Wza2: conserved hypothetical polysaccharide export protein.TM.	56	↑

52:M	37,3	6,1	azo0852	WbnF: Nucleoside-diphosphate-sugar epimerase	28	↑ novel wt
Propanoate metabolism						
53:M	73,3	5,76	azo0688	Pcca: probable propionyl-CoA carboxylase alpha chain	39	CON
Amino acids biosynthesis						
54: M	36,9	6,42	azo3156	IlvC: ketol-acid reducto isomerase	54	↓
55:C	65,7	5,7	azo0632	IlvD: dihydroxy-acid dehydratase	33	↑ novel wt
56:C	26,3	4,86	azo3345	HisA: phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	47	CON
57:M	47,0	5,51	azo0818	HisD: Histidinol dehydrogenase	33	↑
58:C	45,2	6,6	azo2801	GlyA: serine hydroxymethyltransferase	31	CON
59:C	39,2	5,7	azo1042	Asd: aspartate-semialdehyde dehydrogenase	39	CON
60:C	34,3	6,26	azo2399	CysK: cysteine synthase A	61	↓
61:C	55,4	5,3	azo3162	LeuA1: 2-isopropylmalate synthase	51	CON
Fatty acids biosynthesis						
62:C	27,2	5,9	azo2130	FabI: probable enoyl-[acyl-carrier-protein] reductase	55	CON
Butanoate metabolism						
63:C, M	26,1	6,32	azo1023	PhbB2: acetoacetyl-CoA reductases	17	CON, ↑
64:M	23,6	4,11	azo1024	PhbF: polyhydroxyalkanoate synthesis repressor	45	↑
65:C	35,9	6,1	azo3867	AcoA2: probable acetoin dehydrogenase, alpha subunit	32	↑ novel wt
Cofactors biosynthesis						
66:C	39,1	5,37	azo0319	RibAB: GTP cyclohydrolase II	46	↑
67:C	89,3	4,14	azo1626	AcpP: Acyl carrier protein	49	↓
68:M	42,5	5,25	azo0995	HemX: putative uroporphyrin-III C-methyltransferase.	44	CON
69:M	49,6	6,2	azo0862	AccC1: probable biotin carboxylase	45	↓ (6,6 ± 4,3)
70:M	34,5	5,42	azo3537	CbiH :Precorrin-3B C17- methyltransferase		↓
Pyrimidine and purine biosynthesis						
71:C	13,7	4,0	azo3700	TrxC2: probable thioredoxin-disulfide reductase	70	↑ novel mut.
72:C	44,7	5,8	azo2895	PurD:phosphoribosylamine--glycine ligase	24	↑
73:C	60,1	5,76	azo2146	PyrG: CTP synthetase	36	CON
74:C	11,9	6,52	azo2658	TrxA2: putative thioredoxin-disulfide reductase	58	CON
75:C	11,9	4,8	azo1030	TrxA1: thioredoxin-disulfide reductase	58	↓ wt
76:C	33,9	5,44	azo1363	TrxB2: thioredoxin-disulfide reductase	25	↓
77:C	23,6	5,98	azo1473	Adk: adenylate kinase	49	CON
78:C	15,4	5,6	azo0923	Ndk: nucleosidediphosphate kinase	39	CON
Protein synthesis, folding and reparation						
79:M, C	35,7	5,41	azo3391	RpoA: DNA-directed RNA polymerase	63	CON
80:C	20,3	6,34	azo3429	NusG: transcription antitermination protein	42	↑ novel mut.
81:C, C Iso., M	62,9	5,19	azo1072	RpsA: 30S ribosomal protein S1.	25	wt/mut.: CON/ ↓ (4,85± 1,98); iso. novel co-culture; wt/mut.: ↓ (2,23 ± 0,89)/ ↓
82:M	28,1	7,0	azo1909	RpsB: 30S ribosomal protein S2.	59	↑ novel mut.
83:M	14,5	7,07	azo0718	RpsF: 30S ribosomal protein S6.	40	↑ novel mut.
84:C	16,3	5,51	azo0721	Rpl1: 50S ribosomal protein L9	63	CON
85:C	21,5	5,97	azo0754	RplY: 50S ribosomal protein L25	67	CON
86:C	20,7	4,84	azo0086	Efp: putative elongation factor P	56	CON
87:C, M	43,1	5,42	azo3419	TufA: elongation factor Tu (EF-TU)	33	CON
88: C, M	43,1	5,42	azo3431	TufB: elongation factor Tu (EF-TU)	33	CON
89:C, M	5,09	77,4	azo3420	FusA2 : Elongation factor EF-G.	47	CON ; ↓
90:C	51,9	5,29	azo1908	Tsf: elongation factor EF-Ts	23	↓
91:C	41,7	4,94	azo1497	PrfB: peptide chain release factor.	24	CON
92:M	75.3	5.24	azo2103	Pnp: polyribonucleotide nucleotidyltransferase	29	↑ novel wt
93:M	23,9	5,0	azo0848	Pcm: putative protein-L-isoaspartate O-methyltransferase.	32	CON
94:C	49,3	7,2	azo2806	Prc2: carboxy-terminal processing protease precursor	41	↑
95:M	77.2	5.25	azo2877	PrlC: oligopeptidase A	70	↑ novel wt
96:C	54,1	8,0	azo0765	Probable Zn dependent peptidase	39	↑
Respiratory chain and energy production						
97:M	22,9	4,73	azo1398	NuoC:respiratory-chain NADH dehydrogenase, chain C.	41	↓ (2,63 ± 0,54)
98:M, M Iso.	81,9	6,13	azo1402	NuoG: NADH-ubiquinone oxidoreductase chain G	35	CON, iso. novel wt
99:M	18,5	6,89	azo1404	NuoL: NADH-ubiquinone oxidoreductase, chain I.	70	CON
100: M	21,0	6,43	azo0960	PetA1: probable ubiquinol-cytochrome c reductase iron-sulfur protein. TM	54	CON

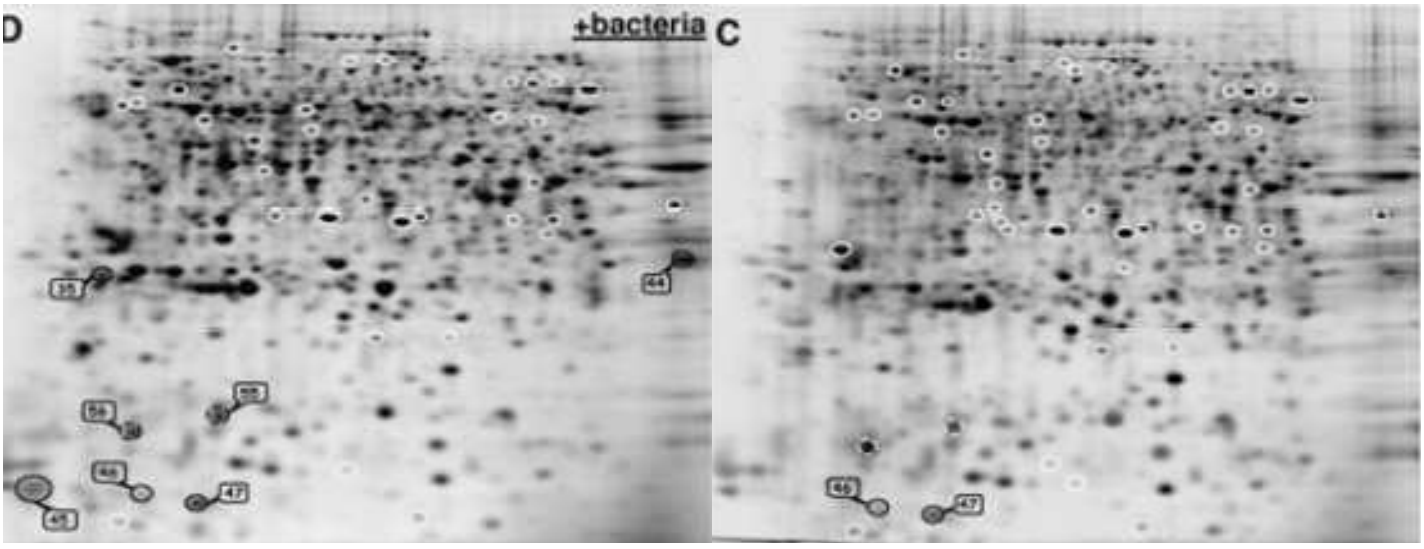
101:M	43,2	4,96	azo3927	CycH: conserved hypothetical cytochrome c-type biogenesis protein. TM	39	↑ novel in wt
102:C	31,0	5,36	azo1700	EtfA1: probable electron transfer flavoprotein, alpha subunit.	70	CON
103:C	59,7	6,42	azo2151	Etf1: probable electron transfer flavoprotein-ubiquinone oxidoreductase	29	↑
104:C	19,7	4,13	azo0954	IsiB: probable flavodoxin	51	↓
105:M	55,3	5,61	azo0157	AtpA: ATP synthase alpha chain.	38	CON
106: C, M	50,6	5,23	azo0159	AtpD: ATP synthase beta chain	65	CON
107:M	18,8	4,5	azo0156	AtpH: putative ATP synthase delta chain.	26	↑
108:C	21,5	5,97	azo1358	Ppa: probable inorganic pyrophosphatase	37	CON
Detoxification related proteins						
109:C	21,5	5,84	azo1466	SodB: superoxide dismutase probable	62	CON
110:C	17,5	7,4	azo0522	SodC: superoxide dismutase [Cu-Zn] precursor	34	↑
111:C, M Iso.	20,6	5,74	azo0769	AhpC: alkyl hydroperoxide reductase subunit C	68	CON, ↑ novel wt
112:C	55,6	5,4	azo0770	AhpF: alkyl hydroperoxide reductase subunit F	37	↓
113:C	15,0	6,2	azo2405	Ohr: probable organic hydroperoxide resistance protein	28	↑
114:M	24,5	6,4	azo2663	Conserved hypothetical peroxiredoxin	31	↑
115:C	17.2	5.2	azo0105	Tpx: probable thiol peroxidase.	42	↑ novel mut.
116:C	31,4	4,8	azo2548	Dyp: Dye-decolorizing peroxidase	37	↓
117:M	18,3	4,62	azo0616	Bfr3: Bacterioferritin	25	↑
Chaperones						
118:C, M	68,7	4,69	azo1063	DnaK: chaperone protein. (Heat shock protein 70)	64	CON ; ↓ (3,4 ± 1,2)
119:C	48,0	4,93	azo2072	Tig: trigger factor	35	CON
120:C	10,4	5,43	azo0973	GroES1: chaperonin	80	CON
121:C	57,9	4,9	azo0974	GroeL1: 60 kDa chaperonin	31	↑ novel co-culture
122:C	18,2	5,49	azo1056	PpiB: peptidyl-prolyl cis-trans isomerase B	70	CON
123:C,M	15,9	5,67	azo3243	Hsp18: putative small heat shock protein (chaperonine).	71	CON
124:M	67,9	5,52	azo1568	PpiD: putative peptidyl-prolyl cis-trans isomerase.TM	46	wt/mut. : CON/ ↓ (3,2 ± 1,0)
125:M	95,8	5,64	azo1566	ClpB: probable ATP-dependent Clp protease, ATP-binding subunit.	47	↓
126:C	71,8	5,17	azo1425	HtpG: probable chaperon protein HtpG.	46	↓
127:C	46,5	5,52	azo2070	ClpX:ATP-dependent Clp protease ATP-binding subunit	39	CON
128:C, M	17,5	4,54	azo2062	Conserved hypothetical peptidyl-prolyl cis-trans isomerase.	24	CON ; ↑
Membrane proteins						
129:M	76,4	5,9	azo3023	Probable TonB-dependent outer membrane receptor	41	CON
130:M	78,7	5,81	azo2396	Probable TonB-dependent outer membrane receptor	31	CON
131:M	50,0	4,76	azo2978	Probable TonB-dependent outer membrane receptor	15	CON
132:M	71,6	5,93	azo3556	Probable TonB-dependent outer membrane receptor	40	↑
133:M	74,8	5,57	azo0378	Probable TonB-dependent outer membrane receptor	56	wt/mut. : CON/ ↓ (3,9 ± 1,2)
134:M	79,0	5,05	azo2887	Osta: organic solvent tolerance protein precursor.	18	CON
135:M, M Iso.	39,4	8,43	azo3017	Outer membrane porin protein precursor. TM	45	CON, iso. novel co-culture
136:M, M Iso.	38,0	6,97	azo3290	Outer membrane porin protein precursor	70	CON, iso. novel co-culture
137:M, M Iso.	37,8	9,11	azo3291	Outer membrane porin protein precursor	16	CON, iso. novel co-culture
138:M	53,3	6,65	azo0654	AggA:putative outer membrane efflux protein	43	CON
139:M, M Iso.	41,4	7,71	azo0244	Membrane fusion protein	65	wt/mut.: ↓/CON; iso. novel mut.
140:M	36,7	4,98	azo0174	MreB : rod shape-determining protein	35	↓
141:M	49,9	5,8	azo0246	OprM3: probable outer membrane efflux protein	22	↑ novel wt
142:M	26,3	6,9	azo3741	OmpW: outer membrane protein W precursor	23	↑
143:M	18,1	6,59	azo0416	Pal: putative peptidoglycan-associated lipoprotein.	69	wt/mut.: ↓ (3,6 ± 1,5) / CON
144:M	29,1	5,63	azo3212	ParA3: sporulation initiation inhibitor protein Soj	23	CON
145:M	30,2	7,1	azo1017	ComL: probable competence lipoprotein precursor	46	↑
146:M	11,0	6,28	azo3667	Conserved hypothetical membrane protein. TM		↑
147:M	31,1	5,3	azo2269	Conserved hypothetical membrane protein. TM	33	↑ novel wt
Motility and adhesion						
148:M	50,0	4,76	azo2693	FliC2: flagellin.	47	CON
149:M	49,8	5,78	azo2704	FliC3: flagellin	40	CON
150:M	19,0	5,6	azo3647	PilP: putative type 4 fimbrial biogenesis protein.	32	↑ novel mut.
Transporters						
151:C	37,2	8,79	azo0919	DctP2: probable C4-dicarboxylate-binding periplasmic protein	55	↓
152:M, M Iso.,C	41,7	8,23	azo2795	Putative periplasmic substrate binding protein	51	wt/ mut.: ↓/CON; iso. novel co-culture; ↑ novel wt

153:M	31,4	5,01	azo0677	UrtD: Conserved hypothetical ABC transporter, ATP binding protein.	30	↑
154:C	37,0	8,33	azo2205	FbpA1: Iron transport system substrate-binding protein probable.	45	↑ novel wt
155:M	61,5	5,74	azo1275	Yjjk: probable ABC transporter, ATP-binding protein.	39	↓
156:C	22,0	7,4	azo3936	CcmA: probable heme exporter protein A	11	↑
Hypothetical secreted proteins						
157:M	51,4	4,86	azo3760	Hypothetical secreted protein. TM	50	↓
158:M	22,3	8,69	azo3886	Hypothetical secreted protein	58	↑ novel mut.
159:M	23,3	9,65	azo0869	Conserved hypothetical secreted protein. TM	40	↑ novel mut.
160:M	26,9	5,25	azo3738	Conserved hypothetical secreted protein.	40	↑
161:C	13,7	7,77	azo2563	Conserved hypothetical secreted protein	56	↑
162:M	63,4	5,58	azo0758	Conserved hypothetical secreted protein	53	CON
Miscellaneous						
163:C	56,9	8,01	azo1228	Choline dehydrogenase	29	↑
164:C	23,2	6,31	azo0963	SspA2: probable stringent starvation protein A	42	CON
165:C	35,2	7,23	azo0296	probable metallo-beta-lactamase superfamily protein	60	wt/ mut.: CON/↓
166:C	36,2	6,0	azo3496	AscD: probable CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydrase reductase.	31	↓
167:C	12,1	5,7	azo3319	Fkbp: peptidyl-prolyl cis-trans isomerase	41	↑ novel mut.
168:M	48,5	6,19	azo3637	GlmU: DP-N-acetylglucosamine pyrophosphorylase / diamine N-acetyltransferase	39	wt/ mut.: CON/ ↓
Conserved hypothetical						
169:C	26,0	4,85	azo0574	YebC: Conserved hypothetical protein	29	wt/ mut.: ↓/CON
170:C	22,31	4,81	azo1131	Conserved hypothetical protein	59	↓
171:C	15,6	4,3	azo0475	Conserved hypothetical protein	30	↑ novel mut.
172:M	24,0	6,6	azo1168	Conserved hypothetical protein	29	↑
173:C	40,2	5,6	azo2298	Conserved hypothetical protein	35	↑
174:M	30,0	5,7	azo2696	Conserved hypothetical protein	17	↑ novel wt
175:C	30,8	5,36	azo2934	Conserved hypothetical amino acid-binding protein.	73	↑
176:C	33,2	6,25	azo2935	Conserved hypothetical protein	49	↑
177:C	22,3	7,1	azo3874	Conserved hypothetical secreted protein	25	↑
178:M	75,7	7,8	azo3950	Conserved hypothetical protein	49	CON

[a] C: cytoplasmic fraction, M: membrane fraction, Iso: isoforme. TM: transmembrane domains (based on the THMM2, a transmembrane helix prediction method based on a hidden Markov model). [b] Accession numbers and protein name are presented according to the annotated genome (Krause *et al.*, 2006). [c] For differential abundance calculations pure culture N_2-fixing gels were used as reference. A cutoff of 2 folds was used to define a spot as up or down regulated. For these, relative spot intensities calculated from three experiments are given with standard deviation. ↑: up-regulated or novel in co-culture, ↓: down regulated or absent in co-culture, CON: constitutive in both growth conditions. Wt: wild type strain BH72, Mut: mutant strain BH1599.

CHAPTER F

Proteomics studies of the *Azoarcus* sp. BH72-rice interaction



1. OBJECTIVES.

The main objective of this part was to establish a protein identification procedure by using a matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometer.

2. CONTRIBUTION.

2.1. Protein sample preparation for MALDI-TOF analysis.

For peptide mass fingerprints (PMF), proteins of interest were manually excised from the gel with a scalpel on a clean bench and placed into a 1.5-ml protein low-binding Eppendorf tube. Destaining, washing and proteolytic digestion were done by using previously described protocols (Hellman, 2000) with slightly modifications. Overnight digestion at 37°C with modified sequencing-grade trypsin (Roche, Mannheim, Germany) was performed in 0.1 M NH_4HCO_3. A treated piece of gel containing the protein band was transferred to a microcentrifuge tube, and 0.63 µg of trypsin per µl was added. Digestion was allowed to proceed overnight at 37°C. The digested peptides were extracted with acetonitrile (MeCN)/trifluoro acetic acid (TFA) as described (Shevchenko *et al.*, 1996), with modifications. Samples were evaporated to dryness in a vacuum centrifuge and were redissolved in 0.1% TFA. Tryptic peptides were then desalted using C18 ZipTip (Millipore Corp. Beverly, MA, U.S.A.) and were eluted in 1 µl MeCN/water 1:1, 0.1% TFA.

2.2. MALDI-TOF analysis for protein identification.

PMF was performed using a Voyager DE-PRO mass spectrometer (Perspective Biosystems, Farmingham, MA, U.S.A.). Peptide extract and a matrix-saturated solution of cyano-4-hydroxycinnamic acid, (Sigma) in 50% MeCN and 0.1% TFA were mixed (0.5 µl each), were loaded on the MALDI laser target, and were allowed to crystallize at room temperature for a few minutes before analysis.The mass spectrometer was operated in the reflector mode with an accelerating voltage of 20 kV, and the average spectrum was determined for 500 laser shots per spectrum. Each spectrum was externally mass calibrated by using a standard peptide mixture (Sigma) of angiotensin I, human adrenocorticotropic hormone (ACTH) 1-17 clip, ACTH 18-39 clip, and ACTH 7-38 clip. Protein identification was carried out with the ProFound Advanced software from ProteoMetrics, using the nonredundant NCBI (National Center for Biotechnology Information) database for the search, with parameters as follows: one trypsin missed cleavage allowed, a fixed carbamidomethyl Cys modification, a variable Met oxidation, and variable pI

and molecular mass ranges (according to the values expected), until significant score was obtained for the best matched protein and at least three matched peptides.

MPMI Vol. 19, No. 5, 2006, pp. 502–511. DOI: 10.1094/MPMI-19-0502. © 2006 The American Phytopathological Society

Upregulation of Jasmonate-Inducible Defense Proteins and Differential Colonization of Roots of *Oryza sativa* Cultivars with the Endophyte *Azoarcus* sp.

Lucie Miché,[1] Federico Battistoni,[1] Sabrina Gemmer,[1] Maya Belghazi,[2] and Barbara Reinhold-Hurek[1]

[1]Laboratory of General Microbiology, Faculty of Biology and Chemistry, University of Bremen, PO. Box 33 40 40, D-28334 Bremen, Germany; [2]INRA-UMR 6175 PRC, Service de spectrométrie de masse pour la protéomique, 37380 Nouzilly, France

Submitted 13 July 2005. Accepted 27 December 2005.

The endophyte *Azoarcus* sp. strain BH72 expresses nitrogenase (*nif*) genes inside rice roots. We applied a proteomic approach to dissect responses of rice roots toward bacterial colonization and jasmonic acid (JA) treatment. Two sister lineages of *Oryza sativa* were analyzed with cv. IR42 showing a less compatible interaction with the *Azoarcus* sp. resulting in slight root browning whereas cv. IR36 was successfully colonized as determined by *nifH::gusA* activity. External addition of JA inhibited colonization of roots and caused browning in contrast to the addition of ethylene, applied as ethephon (up to 5 mM). Only two of the proteins induced in cv. IR36 by JA were also induced by the endophyte (SalT, two isoforms). In contrast, seven JA-induced proteins were also induced by bacteria in cv. IR42, indicating that IR42 showed a stronger defense response. Mass spectrometry analysis identified these proteins as pathogenesis-related (PR) proteins (Prb1, RSOsPR10) or proteins sharing domains with receptorlike kinases induced by pathogens. Proteins strongly induced in roots in both varieties by JA were identified as Bowman-Birk trypsin inhibitors, germinlike protein, putative endo-1,3-beta-D-glucosidase, glutathion-*S*-transferase, and 1-propane-1-carboxylate oxidase synthase, peroxidase precursor, PR10-a, and a RAN protein previously not found to be JA-induced. Data suggest that plant defense responses involving JA may contribute to restricting endophytic colonization in grasses. Remarkably, in a compatible interaction with endophytes, JA-inducible stress or defense responses are apparently not important.

Additional keywords: proteome, β-glucuronidase.

Several nitrogen-fixing, non–nodule inducing bacteria have been recognized as endophytes of suborder *Gramineae*, such as *Azoarcus* spp. in Kallar grass and rice (Hurek et al. 1994; Reinhold et al. 1986), *Herbaspirillum seropedicae* in sugar cane (James and Olivares 1998) and sorghum (James et al. 1997), or *Gluconacetobacter diazotrophicus* in sugar cane (James et al. 1994). These species differ markedly from other symbiotic or plant-colonizing bacteria such as *Rhizobium* or *Azospirillum* spp. in that they are tightly associated with plants: they do not survive well in, and often can not be isolated from, root-free soil (James and Olivares 1998; Reinhold-Hurek and Hurek 1998b).

These nitrogen-fixing endophytes share pathways of infection and colonization of grass roots (James and Olivares 1998; Reinhold-Hurek and Hurek 1998b), which were also detected in rhizobia colonizing rice (Yanni et al. 1997). Infection occurs at the emergence points of lateral roots and in the zone of elongation and differentiation above the root tip. The major colonization sites are the outer root cell layers and the root cortex or aerenchyma in flood-tolerant plants; as for plant some pathogens, ingress into the stele and specifically into xylem cells is also observed for endophytes, albeit rarely. Although intracellular colonization occurs in addition to intercellular localization, there is no evidence for an endosymbiosis in living plant cells (James and Olivares 1998; Reinhold-Hurek and Hurek 1998a) as in the legume symbiosis. High numbers of culturable bacterial cells in roots have been reported (up to 10^8 per gram of root dry weight); however, no symptoms of disease are found under natural conditions (Barraquio et al. 1997; Reinhold et al. 1986). Thus, the host interactions of these endophytes lie between a pathogenic and endosymbiotic lifestyle.

With respect to possible roles of these endophytes, there are numerous reports of plant-growth promotion (Hurek et al. 1994; James and Olivares 1998; Yanni et al. 1997). In addition, the contribution of fixed nitrogen to a grass by a specific endophyte has recently been proven in a few cases, such as for *Gluconacetobacter diazotrophicus* and sugar cane (Sevilla et al. 2001), *Azoarcus* sp. strain BH72 and Kallar grass (Hurek et al. 2002), or *Klebsiella* sp. and wheat (Iniguez et al. 2004). In gnotobiotic cultures, also, rice is internally colonized by *Azoarcus* sp. strain BH72 (Hurek et al. 1994) and nitrogenase genes are expressed in the aerenchyma of rice roots (Egener et al. 1999).

Since the rice genome sequence is known (Goff et al. 2002; Yu et al. 2002) and the genome of strain BH72 is currently sequenced (Hurek and Reinhold-Hurek 2003), mechanisms of initial interactions can be studied by functional genomic and proteomic approaches in this system. Information on responses of any tissues of graminaceous plants to endophytic colonization is scarce. Except for expressed sequence tags reported for the root to stem transition zone of the sucrose-accumulating plant sugar cane infected by *Gluconacetobacter* or *Herbaspirillum* spp., there are no data on responses of roots, the main colonization sites in rice and other grasses. Moreover, informa-

Corresponding author: Barbara Reinhold-Hurek; E-mail: breinhold@uni-bremen.de; Telephone: +(49) 421-218 2370; Fax: +(49) 421-218 9058.

Current address for L. Miché: CIRAD-UMR BGPI, Campus International de Baillarguet, TA 41/K, 34398 Montpellier Cedex 05, France.

tion about the response of rice roots to microbial colonization in general is scarce, since the important pathogens in paddy rice culture mostly attack the aerial part of the plant.

In general, jasmonic acid (JA) and salicylic acid (SA) are key signaling phytohormones in numerous plant responses to stresses such as pathogen attack and exposure to fungal elicitors (Berger 2002; Kunkel and Brooks 2002); however, recent studies suggest that, in rice especially, JA plays an important role in defense mechanisms (Kim et al. 2003; Rakwal and Komatsu 2000). Also, in symbiotic interactions such as arbuscular mycorrhiza, plant defense responses are triggered during early stages (Liu et al. 2003). In the graminaceous plant barley, colonization by an arbuscular fungus leads to elevated levels of JA (Hause et al. 2002). For dicotyledonous plants (*Rosidae* spp.), it has been suggested that plant defense responses mediated by ethylene and SA are involved in the regulation of enteric endophytic colonization (Iniguez et al. 2005). Here, we applied proteome analysis of *Oryza sativa* roots to study the extent to which a plant defense response might be involved in the interaction of an *Azoarcus* sp. with different cultivars.

RESULTS AND DISCUSSION

Rice varieties are differentially colonized by an *Azoarcus* sp.

Physiologically successful root colonization of the endophyte *Azoarcus* sp. strain BH72 can be monitored by visualization of transcriptional activation of bacterial nitrogenase genes in or on roots, using an isogenic reporter strain (BHGN3.1) carrying a transcriptional *nifH::gusA* fusion (Egener et al. 1999). While intense β-glucuronidase (GUS) staining had previously been observed for *Oryza sativa* cv. *nipponnbare* (Egener et al. 1999), a screen of different *indica*-type rice varieties revealed differential colonization of roots by *Azoarcus* strains; cultivar IR36 showed *nifH::gusA* expression throughout the root system (Fig. 1A, panel 1), in concordance with the previous observation of endophytic colonization of this variety (Hurek et al. 1994), whereas GUS staining was much less intense for IR42 (Fig. 1A, panel 2). Moreover, IR42 developed a slight brownish root color upon inoculation with *Azoarcus* sp. strain BHGN3.1, indicative of a less compatible interaction with the bacterium. This root browning morphologically resembled a defense response of roots to pathogens and is putatively due to the accumulation of phenolic compounds. Varieties *O. sativa* cvs. IR36 and IR42 are sister lineages from the same cross that have similar disease-resistance profiles; however, in contrast to IR42, IR36 is moderately resistant to the brown grasshopper biovar 3 and is more tolerant to flooding. Thus there are no obvious explanations for different colonization responses to bacterial endophytes. Future applications of endophytes will, however, require a better understanding of host-range restrictions to these bacteria.

Elevated levels of JA prevent physiologically successful colonization.

To gain insight into the signaling pathways that might be involved, we tested how rice roots respond to known signaling molecules, such as jasmonate, salicylate, or ethylene. JA (100 μM) or SA (1 mM) were added to the plant medium of cv. IR42 without bacterial inoculation. SA did not induce visible changes of root color (not shown), whereas addition of JA resulted in brownish roots (Fig. 1C). Cross-sections showed that, in comparison with that in untreated roots (Fig. 1B), the brownish compounds were accumulated in the same region in

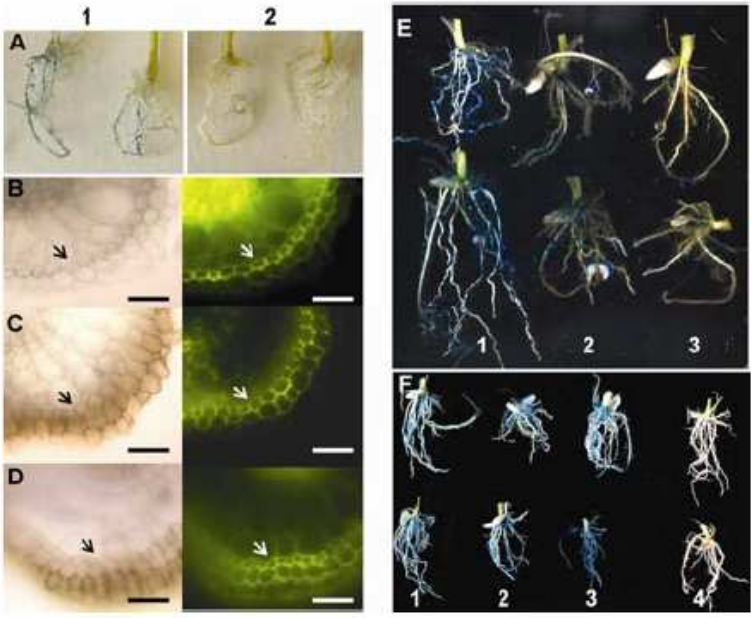


Fig. 1. Response of rice roots to bacterial inoculation and external application of jasmonic acid (JA). **A,** Roots of *Oryza sativa* cv. IR36 (panel 1) or cv. IR 42 (panel 2) inoculated with strain BHGN3.1 at 10^7 cells per gram of quartz sand and, at day 15 after inoculation, stained for *nifH::gusA* activity with X-Gluc. **B** through **D,** Light microscopic (panel 1) or fluorescence microscopic images (panel 2) (excitation 450 to 490 nm) of hand sections of roots. Arrows point to sclerenchyma. Bars = 50 μm. Transversal sections of roots from seedlings of *O. sativa* cv. IR42 **B,** not inoculated, **C,** not inoculated but treated with 100 μM JA at day 9, or **D,** inoculated with *Azoarcus* sp. strain BHGN3.1 at 10^8 cells per gram of quartz sand at day 1 and harvested at day 12. **E,** Inhibition of physiologically successful colonization by external application of JA, two plants per treatment shown. Seedlings of *O. sativa* cv. IR36 inoculated and stained for β-glucuronidase (GUS) activity as in A and with no JA (panel 1), with 15 μM JA (panel 2), or with 40 μM JA (panel 3) added at day 1. Neutralized D,L-malic acid (200 mg/liter) was added as a carbon source to enhance infection. **F,** Physiologically successful colonization in the presence of ethylene applied as ethephon, two plants per treatment shown. Experiment as in E, with Panel 1, no ethephon, Panel 2, 1 mM ethephon, and Panel 3, 5 mM ethephon added. Panel 4, unlike panels 1 through 3 was not stained for GUS activity. Upper plant, no ethephon added; lower plant, 4 mM ethephon added.

JA-treated roots (Fig. 1C) and roots inoculated with high numbers of bacteria (Fig. 1D); the outer cell layers adjacent to sclerenchyma showed increased cell-wall browning (Fig.1, panel 1) and increased fluorescence (Fig. 1, panel 2). This suggested that plant defense responses might be involved in differential bacterial host specificity and that JA might participate in the signaling pathway leading to rice root changes in less-compatible interactions with the endophyte.

Since external application of JA mimicked the morphological changes of roots of IR42 to the *Azoarcus* sp., it could be anticipated that JA triggers a root response preventing endophytic colonization. Therefore, we investigated whether elevated levels of JA affected colonization rates in a compatible endophyte-rice interaction. External JA was added to roots in concentrations that still allowed growth of rice seedlings (15 or 40 μM, respectively). The more-compatible variety IR36 was used, and root colonization was enhanced by addition of an external carbon source (Hurek et al. 1994). Intense GUS staining (*nifH::gusA* fusion) of roots was detected in inoculated, untreated plants throughout the root system (Fig. 1E, panel 1). GUS staining decreased with the application of increasing JA concentrations (Fig. 1E, panels 2 and 3) and was almost diminished at 40 μM JA. This suggested that JA triggered a defense response preventing physiologically successful colonization by this endophyte.

In a similar experiment, JA was replaced by ethylene. It was supplied as ethephon to the root medium in a concentration range used by others (0.1 to 1 mM ethephon), and no effect on root color or GUS staining was observed (Fig. 1F). Even the addition at 4 to 5 mM ethephon, which led to partial senescence of shoots (not shown), did not cause root browning or inhibition of endophyte colonization (Fig. 1F). Thus, the observed response in the incompatible cultivar was not likely to be provoked by ethylene.

JA and SA are well known key signaling phytohormones in numerous plant responses to stresses, such as pathogen attack in dicots (Glazebrook 1999), in which SA is necessary for the full expression of both local resistance and SAR (systemic acquired resistance), including pathogenesis-related (PR) proteins. The JA pathway typically regulates response to abiotic stress, defenses against insect herbivores, necrotrophic fungal pathogens, and even against some biotrophic pathogens (Turner et al. 2002). However, recent studies suggest that, in rice shoots especially, JA plays an important role in defense mechanisms. SA levels are comparatively high in rice and did not significantly increase after inoculation with *Pseudomonas syringae* or the biotrophic fungal pathogen *Magnaporthe grisea* (Silverman et al. 1995). In contrast, JA possesses the ability to induce a variety of defense-related proteins that are not induced by SA but, e.g., by *M. grisea* (Kim et al. 2003; Rakwal and Komatsu 2000; Rakwal et al. 1999; Silverman et al. 1995). In rice roots, consistent with our observations, application of JA concentrations of <10 μM does not provoke chlorosis or other symptoms of senescence that occur at concentrations of 20 to 40 μM (Moons et al. 1997).

The root proteomes of *O. sativa* cvs. IR36 and IR42 are almost identical.

In order to gain insight into the differential response of the two rice varieties, we applied a proteomic approach. First, the root proteomes of plants grown under aseptic conditions without inoculation were compared for *O. sativa* cv. IR36 and IR42 by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). With approximately 1,000 spots detectable per gel, a better resolution with fewer distortions of the protein patterns than in several other studies on rice roots (Konishi et al. 2005; Moons et al. 1997; Salekdeh et al. 2002) was achieved. The protein pattern of both varieties without treatment was very similar (Fig. 2A and B). Only four abundant proteins were found to be reproducibly different (201 to 204). Their identification by mass spectrometry (MS) (Table 1) suggests that they are isoforms of the same proteins, with slightly different electrophoretic mobility in both varieties.

Only a few JA-inducible proteins are upregulated in a compatible interaction of an *Azoarcus* sp. with IR36.

In order to investigate the response of rice roots in more detail, we compared root proteomes of both *O. sativa* cultivars IR36 and IR42 in response to application of JA and to inoculation with *Azoarcus* sp. strain BHNG3.1. Among about 1,000 spots detected on the different 2D-PAGE gels, 47 showed reproducible induction patterns after bacterial inoculation (Fig. 2C and D). Out of those, we focused for further analysis on the spots that were also induced by JA treatment. The other ones, which are labeled in Figure 2 without numbering, appeared to be largely of bacterial origin and will be identified upon completion of the *Azoarcus* sp. strain BH72 genome project (Hurek and Reinhold-Hurek 2003).

In comparison with proteins induced upon JA treatment (Fig. 2E and F), only two spots were detected that were induced both by *Azoarcus* inoculation and by JA in the more compatible cultivar IR36 (Fig. 2C and E). However, induction in the presence of JA was stronger (Fig. 3). These proteins (spot 46 and 47) were identified by MS (Table 2). They both showed homology to SalT, a protein found to be induced by salt and drought stress in rice roots (Claes et al. 1990). SalT induction by salt and JA application has also been detected in rice seedlings (Moons et al. 1997), which suggests an overlap between salt- and JA induction of proteins. Recent studies determined that SalT is a cytoplasmic mannose-binding rice lectin that responds to a wide range of stresses, suggesting its implication in a global mechanism of response to environmental stresses (de Souza Filho et al. 2003) and involvement in intracellular regulation and signaling (Van Damme et al. 2004). In our gels, two isoforms could be detected. They may result from posttranslational modifications such as phosphorylation, as four putative phosphorylation sites could be predicted by NetPhos 2_0 on the SalT amino-acid sequence. However, a clear answer about the actual site of phosphorylation could not be obtained from the peptides identified by MS.

For protein spots 55 and 56, a weak induction (1.5 to 1.7 times) could be observed in IR36 following bacterial inocula-

Table 1. Protein spots differing in roots of uninoculated *Oryza sativa* cultivars IR36 and IR42

Spot		Best matched protein			
No.	pI/MM (kDa)	Protein name	Accession no.[a]	pI/MM (kDa)	% Coverage[b]
201	6.09/57.62	Dihydrolipoamide dehydrogenase precursor, putative	3489480	7.21/52.61	44
202	4.91/49.76	Peroxidase	46981333	5.51/37.78	10
203	6.36/56.47	Dihydrolipoamide dehydrogenase precursor, putative	34894800	7.21/52.61	38
204	5.03/48.78	Peroxidase	46981333	5.51/37.78	11

[a] National Center for Biotechnology Information accession numbers.
[b] Proteins identified by nano-liquid chromatography-electron spray ionization-tandem mass spectrometry.

tion (Fig. 3), however no JA effect could be observed in cv. IR36. MS identified both proteins as members of the PR-10 class of PR proteins, PR10-b and RSOs-PR10, respectively. They are induced by pathogen attack in a wide variety of plants (Hashimoto et al. 2004). RSOsPR10 was rapidly induced in roots of rice seedlings upon salt and drought stress as well as blast-fungus infection and JA application (Hashimoto et al. 2004). This indicates that a weak defense response towards *Azoarcus* spp. might also be induced in this cultivar IR36. However, since a wide range of proteins was upregulated by JA treatment in rice roots (Fig. 2E and D, and discussed below) that were not induced by *Azoarcus* inoculation in IR36, these endophytes appear to elicit only a mild stress or defense response in a compatible interaction.

JA-inducible PR proteins upregulated in a less compatible interaction of *Azoarcus* sp. with IR42.

In comparison with that in IR36, the root proteome of cultivar IR42 showed a lower number overall of upregulated proteins upon inoculation with the endophyte (Fig. 2D), as can be expected after a less successful colonization. However, several proteins that were typically induced by JA were also upregulated in IR42 upon inoculation with bacteria (Fig. 2D and F). In common were spots 46 and 47, which showed similar induc-

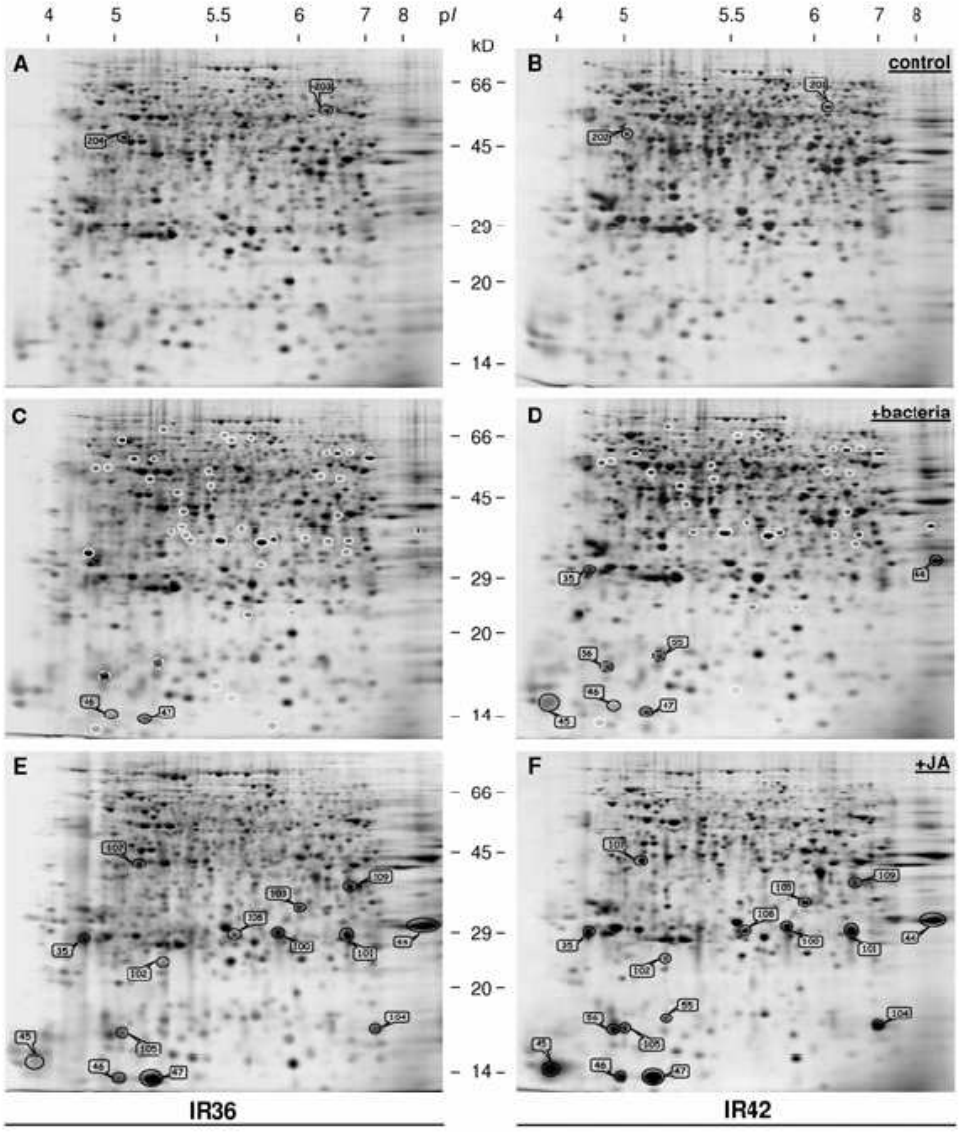


Fig. 2. Coomassie-stained two-dimensional gels of protein extracts from 20-day-old rice roots. Comparison of root proteins induced by bacterial inoculation and jasmonic acid (JA) treatment in *Oryza sativa* cv. IR36 (A, C, E) and cv. IR42 (B, D, F). **A** and **B,** Control roots of rice, harvested 15 days after transfer to plant medium only. **C** and **D,** Roots harvested 15 days after inoculation with *Azoarcus* sp. strain BHGN3.1. White circles indicate proteins induced by bacterial inoculation only. Numbered spots indicate proteins that are induced by both bacterial inoculation and JA treatment. **E** and **F,** Roots harvested 15 days after transfer to plant medium and 3 days after JA treatment (100 µM). Labeled spots indicate some of the proteins that are induced following JA treatment.

tion levels as in cv. IR36 (Figs. 2D, F and 3). Two PR proteins (spots 55 and 56) were weakly upregulated in response to bacteria, as in cv. IR36; however, in cv. IR42, they were also induced by JA application (Fig. 3).

Interestingly, the additional spots upregulated in IR42 only were PR proteins. Protein spot 45 was present at a low basal level in cultivar IR42 and was induced (twofold) or strongly induced (sixfold) by bacterial inoculation or JA treatment, respectively (Fig. 3), while it was hardly detectable in cv. IR36. It was identified as protein Prb1 (Table 2), which belongs to a family of PR proteins, PR-1. This is a dominant protein group induced by pathogens and is commonly used as a marker for—normally SA-related—SAR (van Loon et al. 1998). It was induced in roots of rice seedlings after salt stress or JA treatment (Moons et al. 1997), as well as in JA-treated stems (Rakwal and Komatsu 2000). Thus Prb1 may be a marker for the induction of some defense reaction in cv. IR42 towards *Azoarcus* spp.

Proteins 35 and 44 were inducible by JA in both varieties; however, only in cv. IR42 were they upregulated in inoculated roots (Table 2). They are very similar proteins (55% amino-acid identity) despite a different isoelectric point (pI) value (Table 2). They show similarities to a new family of receptor-like protein kinases (RLK) that were induced in potato infected by *Erwinia carotovora* (Montesano et al. 2001). The first member of this family, PvRK20-1, was identified in the roots of common bean (Lange et al. 1999). Interestingly, the transcript level of *PVRK20-1* was induced during pathogen attack by *Fusarium* spp. but remained at the level of the control or was even suppressed upon infection with the nodule-forming symbiont *Rhizobium tropici*. Relatives of this RLK family were

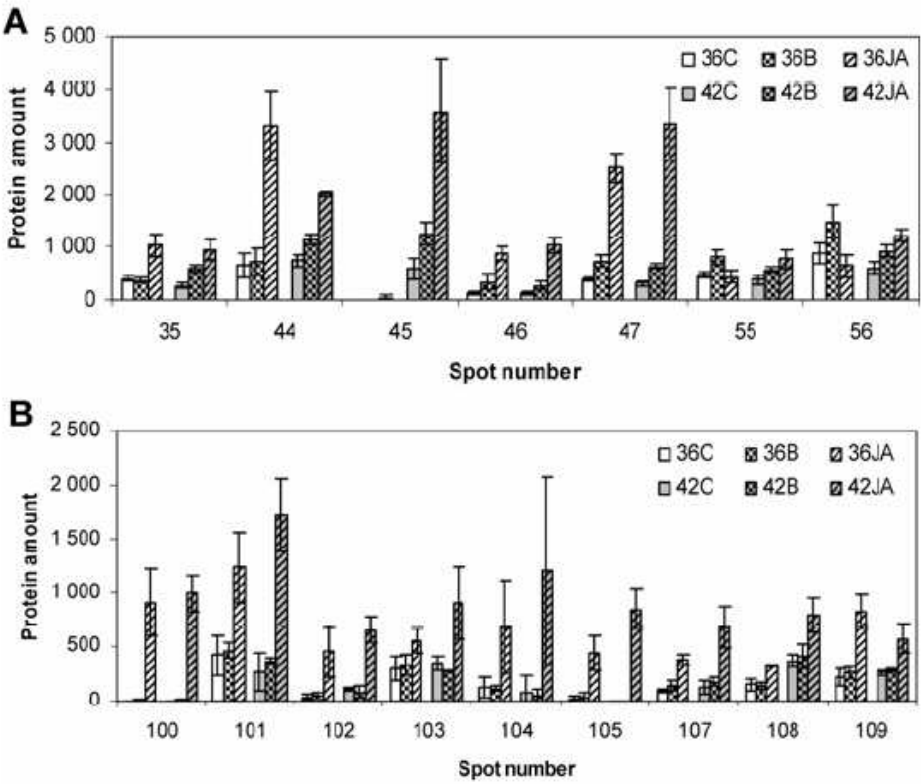


Fig. 3. Quantification of proteins in rice roots induced **A**, by inoculation with *Azoarcus* sp. and jasmonic acid (JA) treatment or **B**, by JA treatment only. Relative spot intensities calculated from three experiments are given with standard deviations (error bars). Two *Oryza sativa* cultivars IR 36 (36) or IR42 (42) were compared. C = Uninoculated controls; B = inoculated with bacteria *Azoarcus* sp.; JA = treated with JA.

Table 2. Rice root proteins induced both by *Azoarcus* inoculation and by jasmonic acid (JA) treatment

Spot		Best matched protein			
No.	pI/MM (kDa)	Protein name	Accession no.[a]	pI/MM (kDa)	% Coverage
35	4.48/30.26	Putative receptorlike protein kinase	19387274	5.01/27.92	38[b,c]
44	8.18/32.09	Membrane binding protein (putative receptor-like protein kinase)	46399155	7.99/29.24	11[c]
45	3.79/14.67	Pathogenesis-related protein (Prb1)	33440014	4.32/17.11	42[c]
46	4.84/14.31	Salt-stress induced protein (SalT)	134190	5.20/15.18	59[c]
47	5.22/13.98	Salt-stress induced protein (SalT)	134190	5.20/15.18	45[b]
55	5.32/18.64	Pathogenesis-related protein (PR-10b)	9230757	4.88/17.43	41[b]
56	4.75/17.74	Pathogenesis-related protein (RSOsPR10)	38678114	4.79/16.90	84[b,c]

[a] National Center for Biotechnology Information accession numbers.
[b] Protein identified by matrix-assisted laser desorption-ionization time-of-flight spectrometry.
[c] Protein identified by nano-liquid chromatography-electron spray ionization-tandem mass spectrometry.

also identified in *Arabidopsis thaliana*. The *At-RLK3* gene is expressed at basal level in the entire plant but is activated upon oxidative stress and preferentially during incompatible interaction with the pathogen *Ralstonia solanacearum* (Czernic et al. 1999). This would indicate a role for this new family of RLK in the early stages of recognition during plant-microbe interactions, with the activation of the receptor-encoding gene leading to the amplification of the signal through increased RLK protein level. It would also correlate with our observation that these proteins are not induced in the compatible rice variety upon infection with *Azoarcus* strains.

It is also worthwhile noting that spots 35 and 44 encode only the receptor part of those RLK, which possess a duplicated domain containing four conserved cysteines (DUF26) whose function is unknown. Induction of such a truncated version was also observed after inoculation of suspension-cultured rice cells by rice blast fungus as well as by JA treatment (Kim et al. 2003) and in wounded rice leaf sheat (Shen et al. 2003). The fact that this domain is both induced by microorganisms recognized as pathogenic by the plant and also directly by JA treatment, raises a question about the effectors that actually activate this kind of receptor.

Proteins induced by JA treatment only.

JA application to roots of rice seedlings led to a general decrease in the amount of most proteins, especially in cv. IR42, and to a marked accumulation of a set of defense-related proteins (Fig. 2E and F; Table 3). This is similar to previous observations of others (Moons et al. 1997). A wide range of proteins was induced by JA treatment in both cultivars that were not induced by *Azoarcus* inoculation, which indicates that there are only partial overlaps of both responses.

By MS analysis, we identified several additional proteins that were relatively strongly induced (Table 3; Fig. 3). Identification of some of them adds further insight into the root proteins regulated by JA in rice, since they had not yet been found as induced in roots. Most proteins were PR proteins. Protein spots 100 and 104 were identified as BBIs (Bowman-Birk inhibitors), which encode serine protease inhibitors. Rice BBI genes are both developmentally regulated and rapidly induced by JA and ethylene as well as in response to pathogen attacks (Qu et al. 2003; Rakwal et al. 2001). Moreover, overexpression of rice RBBI2-3 resulted in strong plant resistance against a fungal pathogen, supporting the view that proteinase inhibitors play an important role in the rice defense system (Qu et al. 2003). Interestingly, two forms of BBI could be detected in our 2D-PAGE gels. Apart from the classical 16-kDa BBI (spot 104) found in monocotyledonous plants (*Liliopsida* spp.), another BBI with an unusually greater molecular mass was also induced by JA (spot 100). This protein was also found in stems of rice seedlings treated with JA (Rakwal and Komatsu 2000) and was identified as having an additional third cys-rich domain in the N-terminus (Qu et al. 2003).

As a putative ACC (1-aminocyclopropane-1-carboxylate) oxidase (Table 3), protein spot 107 may catalyze ethylene production. The phytohormone ethylene regulates many different processes in plants and has been implicated in defense responses

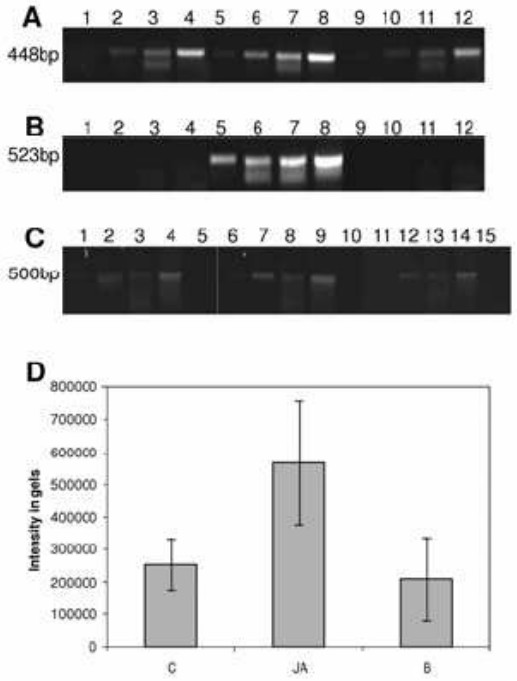


Fig. 4. Analysis of expression of *SalT* and Bowman-Birk trypsin inhibitor in root RNA of *Oryza sativa* cv. IR36 by reverse transcriptase-polymerase chain reaction (RT-PCR) analysis. **A** through **C**, Amplification products of RT-PCR for **A**, Bowmann-Birk trypsin inhibitor, **B**, *SalT*, and **C**, *actin1*. In A and B lanes 1 through 4, uninoculated roots; lanes 5 through 8, jasmonic acid (JA)-treated roots; lanes 9 through 12, roots inoculated with *Azoarcus* sp. Lanes 1 through 4, 5 through 8, and 9 through 12 each show products of increasing cycle number (24, 28, 34, or 40 cycles, respectively). **C**, Lanes 1 through 4, 6 through 9, and lanes 11 through 14 show products of increasing cycle number (24, 28, 34, or 40 cycles, respectively) from uninoculated, JA-treated, or inoculated roots, respectively. Lanes 5, 10, and 15 show the DNA controls of these three PCR reactions with inactivated reverse transcriptase. **D**, Quantification of Bowman-Birk trypsin inhibitor–specific RT-PCR reactions from fluorescence intensities of products obtained after 40 cycles; columns show mean values and standard deviations of fluorescence intensities from three independent repetitions of uninoculated control roots (C), JA-treated roots (JA), and roots inoculated with bacteria (B).

Table 3. Rice root proteins induced only by jasmonic acid (JA) treatment

Spot		Best matched protein			
No.	pI/MM (kDa)	Protein name	Accession no.[a]	pI/MM (kDa)	% Coverage
100	5.83/30.55	Bowman Birk trypsin inhibitor, putative	7619799	5.40/29.25	10[b]
101	6.46/29.46	GTP-binding nuclear protein Ran	5360230	6.66/25.02	27[c]
102	5.32/26.39	zwh0010.1 (oxalate oxidase/germinlike protein)	5852087	6.30/25.84	15[b,c]
103	5.96/33.98	Glucan endo-1,3-beta-D-glucosidase, putative	22830909	6.96/36.91	11[c]
104	6.95/17.99	Trypsin inhibitor (Bowman-Birk)	476550	7.90/16.08	23[b]
105	4.88/17.55	Probenazole-induced protein PBZ1 (PR-10a)	7442204	4.88/16.68	25[c]
107	5.13/43.20	1-Aminocyclopropane-1-carboxylate oxidase, putative (ACC oxidase)	51536083	5.35/35.54	11[c]
108	6.68/30.18	Glutathione *S*-transferase, putative	34914740	5.67/25.35	49[c]
109	6.51/37.84	Class III peroxidase 59 precursor	5777629	6.08/36.77	16[c]

[a] National Center for Biotechnology Information accession numbers.
[b] Protein identified by matrix-assisted laser desorption-ionization time-of-flight spectrometry.
[c] Protein identified by nano-liquid chromatography-electron spray ionization-tandem mass spectrometry

(Iniguez et al. 2005; Wang et al. 2002). JA and ethylene have also been shown to cooperate in regulating the expression of many genes, and at least some JA-inducible genes are not inducible in plants unable to produce or sense ethylene (Wang et al. 2002). Two defense genes have been shown to be responsive to ethylene in rice, *OsPR1a* (Agrawal et al. 2000) and *OsBBIP1* (Rakwal et al. 2001), a member of the Bowman Birk trypsin inhibitor proteins. Concomitant induction of BBIs and the putative ACC oxidase upon JA treatment in our rice roots points to the importance of ethylene as a potential modulator of stress responses in rice.

Spot 102 was identified as a germinlike oxalate oxidase (Table 3). These enzymes are only found in the "true cereals," in which they are involved in defense responses to invasion by fungal pathogens. Their possible modes of action include generation of microcidal concentrations of hydrogen peroxide, elicitation of hypersensitive cell death, and others, as reviewed by Lane (2002).

Protein spot 103 belongs to family 17 of the glycosyl hydrolases, a β-1,3-glucanase (a PR2). Similar proteins were induced in rice leaves treated with JA (Rakwal et al. 1999). Class PR2 of the PR proteins can inhibit fungal cell growth by hydrolytic degradation of fungal cell walls (van Loon and van Strien 1999).

Protein 105 was identified as PBZ1 (probenazole-induced protein). PBZ1 belongs to the PR-10 class of PR proteins (Hashimoto et al. 2004; McGee et al. 2001). Their biological functions are still unknown, although it has been suggested that some possess a ribonuclease activity (Bantignies et al. 2000). PBZ1 has been reported to accumulate in rice roots treated with JA (Moons et al. 1997), in leaves from whole plant treated with blast fungus (Kim et al. 2003), and in suspension-cultured rice cells in which six isoforms of PBZ1 were induced in response to *M. grisea* as well as JA treatment (Kim et al. 2003).

Protein spots 108 and 109, identified as putative glutathione *S*-transferase (GST) and peroxidase precursor, were only slightly induced by JA treatment (Table 3). Proteins of this type are known to be involved in cellular protection through reactive oxygen species detoxification. The induction of GST by JA in rice was previously reported (Rakwal et al. 1999), supporting the hypothesis that JA is probably involved in mediating hypersensitive response in rice. Consistent with our results, the induction of a peroxidase precursor matching spot 109 was also reported in rice roots treated with JA (Moons et al. 1997). Apart from peroxide detoxification, this secreted peroxidase may also be involved in lignin biosynthesis (Passardi et al. 2004).

Spot 101 was identified as a RAN protein, a member of GTPases that are involved in the active transport of proteins through nuclear pores and that have been discovered in the *Arabidopsis* genome in four copies (Vernoud et al. 2003). To our knowledge, this is the first time that this protein has been found to be JA-inducible.

Gene expression analysis by reverse transcriptase-polymerase chain reaction.

In order to verify whether differential protein accumulation was accompanied by elevated transcriptional activity, two proteins were selected that either accumulated upon inoculation and JA treatment (SalT) or only upon JA treatment (BBI). Plants were treated as for proteome studies, and RNA extracts from roots were subjected to reverse transcriptase-polymerase chain reaction (RT-PCR) with specific primers (Fig. 4). To obtain results previous to saturation of the PCR reaction, samples were taken at several different cycle numbers and were subjected to gel electrophoresis. As a control for RNA quality, rice *actin 1* (Yamaguchi et al. 2002) was used as a constitutively expressed gene (Fig. 4C). The gene encoding SalT was strongly upregulated in JA-treated roots, as expected; however, an induction upon inoculation was not evident (Fig. 4B). The gene encoding the Bowman-Birk trypsin inhibitor was expressed in control roots and inoculated roots and appeared to be upregulated in JA-treated roots (Fig. 4A). Band intensities were quantified for three independent repetitions (Fig. 4D), and a statistical analysis was carried out. According to unpaired *t*-test, the ratios of JA-treated to control roots were signifycantly different ($P < 0.0035$) from the ratio of inoculated to control roots for the *BBI* RT-PCR, while it was not significant for the *actin 1* gene, suggesting that *BBI* was upregulated only upon JA treatment. Similar results were obtained in a second plant experiment (not shown). Despite an overall concurrence of protein and RNA expression analyses, discrepancies in gene expression ratios and protein induction rates may be affected by protein or mRNA stability, depending on timing of the sampling. For example, the *OsBBPI* induction after JA application is maximal at 6 h and then gradually declines (Rakwal et al. 2001), which might explain the low gene-induction values we detected three days after application of JA.

Concluding remarks.

This is the first proteomic report on the response of roots of graminaceous plants towards colonization with endophytic bacteria. Addition of JA, a signal molecule important for stress responses and defense mechanisms in rice, decreased the physiologically successful colonization by the diazotroph *Azoarcus* sp., suggesting that plant defense responses might control endophytic ingress. This response was not observed for ethylene. Concomitantly, it has been proposed for endophytes of dicotyledonous plants (*Rosidae*), that plant defense naturally limits infection, albeit based on SA and ethylene as signaling molecules (Iniguez et al. 2005). A proteome approach allowed us to detect only a small overlap in the responses of roots of rice seedlings towards colonization by N_2-fixing endophytes and towards JA treatment. In rice variety *O. sativa* cv. IR36 being physiologically successfully colonized by the bacterium, remarkably, only two of the proteins induced by JA were also induced by the endophyte. This indicates that JA-induced stress responses or PR proteins are not important in a compatible interaction with rice. In contrast, the proteome of a sister lineage, cv. IR42, showing a less-compatible interaction with an *Azoarcus* sp., displayed more overlap (though limited) with JA-induced proteins. They notably included PR-proteins or proteins sharing domains with receptorlike kinases induced by pathogens. This suggests that plant defense responses involving JA do play a role in restricting endophytic colonization in grasses when the host-bacterium interaction is less compatible. It is remarkable, though, that in well-colonized plants, these responses appear to be very limited. In contrast, for other root-colonizing bacteria, such as plant-growth-promoting *Pseudomonas fluorescens*, a wide range of defense-related genes were induced in the dicotyledonous *Arabidopsis* (Wang et al. 2005). For future research aiming for applications of endophytic bacteria in a wide range of hosts, it will be important to elucidate which factors are suppressing the defense response in a compatible variety to allow an efficient endophytic colonization.

MATERIALS AND METHODS

Plant growth and treatments.

Rice inoculation was done as previously described (Egener et al. 1999). Briefly, *Oryza sativa* subsp. *indica* (cv. IR36 or IR42) seeds were surface-sterilized and placed to germinate on agar plates (Hurek et al. 1994). The sterile three- to five-day-old seedlings were transferred to glass tubes containing 5 g of sterilized quartz sand (Merck, Darmstadt, Germany) saturated

with nitrogen-free plant medium containing a small amount of neutralized malic acid (5 mg/l), and were inoculated with a midlogarithmic preculture of *Azoarcus* sp. strain BHGN3.1 (wild-type BH72 containing *nifH::gusA* [Egener et al. 1999]) to reach a final concentration of 10^7 bacteria per gram of quartz. Plants were grown for two weeks in a growth chamber at 80% relative humidity, 30°C, and 15 kLux light intensity, with a day and night cycle of 14 and 10 h, respectively. Controls were done with noninoculated rice seedlings. For JA treatment, 100 μM (±) JA (racemic mixture; Sigma, St. Louis) was added to the plant medium of noninoculated seedlings three days before harvesting for proteome analysis. For other experiments, JA was added during transplanting at the concentrations mentioned in the text. Exposure to ethylene was achieved by addition of ethephon to the root medium during transplanting, at concentrations from 0.1 to 5 mM (Camposan Extra, 660 g/liter; Bayer, Leverkusen, Germany). Roots were stained for activity of GUS with X-glu, and light and fluorescent microscopic images were taken with an RGB video camera, as previously described (Egener et al. 1999).

Preparation of total protein extracts.

Proteins from control and treated plant roots were prepared essentially as by Majoul and associates (2000), with modifications. For roots inoculated with bacteria, successful colonization as *nifH::gusA* activity was controlled for few plants by staining with X-glu (Egener et al. 1999). Roots from about 30 independently treated seedlings (1 g) were first ground with quartz powder in liquid nitrogen, then at 4°C in 5 ml of extraction buffer (0.7 M sucrose, 0.5 M Tris, 30 mM HCl, 0.1 M KCl, 50 mM EDTA; 2% [vol/vol] 2-mercaptoethanol and 2 mM phenyl methylsulfonyl fluoride [PMSF] were added just before use). Samples were then extracted with Tris/HCl-buffered phenol (pH 8.0). Proteins were precipitated overnight at –20°C with 0.1 M ammonium acetate in methanol and were rinsed and dried under a stream of nitrogen gas. The recovered pellet was solubilized in 200 μl of protein buffer (5 M urea, 2 M thiourea, 2% [wt/vol] CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), 2% [wt/vol] SB 3-10 [*N*-decyl-*N,N*-dimethyl-3-amino-1-propanesulfonate], 1% [wt/vol] DTT [dithiothreitol], 2% [vol/vol], pH 3 to 10, carrier ampholytes [Bio-Rad, Hercules, CA, U.S.A.]). Protein concentrations in the various extracts were determined using the RC DC protein assay (Bio-Rad) with bovine serum albumin as the standard, and samples were stored at –80°C.

2D-PAGE.

For the first dimension, immobilized nonlinear pH gradient strips were used (Immobiline DryStrips, pH 3 to 10, NL, 18 cm; Amersham Bioscience, Uppsala, Sweden). Strips were rehydrated overnight with protein buffer containing 1 mg of protein extracts. Isoelectrofocusing was performed for a total of 25 kV/h at 20°C, using the MultiphorII system (Amersham Biosciences) using a three-phase program. The first phase was set at 500 V, the second was a linear gradient from 500 to 3,500 V, and the final phase was set at 3,500 V. Prior to the second dimension, the gel strips were equilibrated for 2 × 15 min in equilibration buffer containing 30% (wt/vol) glycerol, 6 M urea, 2% (wt/vol) sodium dodecyl sulfate (SDS), 50 mM Tris-HCl (pH 6.8). DTT (2% wt/vol) was added to the first equilibration solution and iodoacetamide (2.5% wt/vol) to the second. Equilibrated gel strips were run overnight on 12.5% acrylamide gels (30:1). Wide-range molecular markers (Sigma) were loaded beside the strip of some repetition gels to determine apparent molecular masses of the proteins.

Protein staining and analysis of 2D-PAGE gels.

Coomassie staining of the gels was done according to standard procedures, using a staining solution containing 35% (vol/vol) methanol, 10% (vol/vol) acetic acid, and 0.05% (wt/vol) Coomassie brilliant blue R-250 (Serva, Heidelberg, Germany). Gels were destained with 30% (vol/vol) methanol, 5% (vol/vol) acetic acid solutions and were stored at 4°C in storage buffer (18% [vol/vol] ethanol, 3% [vol/vol] glycerol) until analysis.

Stained gels were scanned at 600 dots per inch on a UMAX Power Look III scanner (UMAX, Data Systems, Inc., Taipei, Taiwan), and image analysis was carried out with the ImageMaster 2-D Elite version 4.01 software (Amersham Biosciences). Spot detection, quantification, background subtraction, gel-to-gel matching, and differential display analysis were performed on three independent biological repetitions. Variations are shown in Figure 3. Intensities of individual spots were normalized based upon the total density of all spots detected in each gel. Apparent masses and pIs of each protein spot were calculated by the use of the software, according to the strip gradient and molecular weight ladder used.

Protein identification by MS.

Differentially expressed proteins reported in this study were either identified by matrix-assisted laser desorption-ionization time-of-flight (MALDI-TOF) or by peptide sequencing using liquid chromatography/tandem MS (LC-MS/MS).

For peptide mass fingerprinting (PMF), spots of interest were manually excised from Coomassie-stained gels and were digested overnight at 37°C with modified sequencing-grade trypsin (Roche, Mannheim, Germany). The digested peptides were extracted with acetonitrile (MeCN)/trifluoro acetic acid (TFA) as described by Shevchenko and associates (1996), with modifications. Samples were evaporated to dryness in a vacuum centrifuge and were redissolved in 0.1% TFA. Tryptic peptides were then desalted using C18 ZipTip (Millipore Corp. Beverly, MA, U.S.A.) and were eluted in 1 μl MeCN/water 1:1, 0.1% TFA. PMF was performed using a Voyager DE-PRO mass spectrometer (Perspective Biosystems, Farmingham, MA, U.S.A.). Peptide extract and a matrix-saturated solution of cyano-4-hydroxycinnamic acid, (Sigma) in 50% MeCN and 0.1% TFA were mixed (0.5 μl each), were loaded on the MALDI laser target, and were allowed to crystallize at room temperature for a few minutes before analysis.

The mass spectrometer was operated in the reflector mode with an accelerating voltage of 20 kV, and the average spectrum was determined for 500 laser shots per spectrum. Each spectrum was externally mass calibrated by using a standard peptide mixture (Sigma) of angiotensin I, human adrenocorticotropic hormone (ACTH) 1-17 clip, ACTH 18-39 clip, and ACTH 7-38 clip. Protein identification was carried out with the ProFound Advanced software from ProteoMetrics, using the nonredundant NCBI (National Center for Biotechnology Information) database for the search, with parameters as follows: one trypsin missed cleavage allowed, a fixed carbamidomethyl Cys modification, a variable Met oxidation, and variable pI and molecular mass ranges (according to the values expected), until significant score was obtained for the best matched protein and at least three matched peptides.

For MS/MS analyses, digestion of protein spots was modified. Excised spots were reduced with 10 mM DTT for 45 min at 56°C and then were alkylated with 55 mM iodoacetamide in the dark for 30 min. After enzymatic digestion peptides were sequentially extracted with 1% formic acid for 10 min and 0.1% formic acid (vol/vol) in 50% (vol/vol) MeCN for 15 min. Supernatants were pooled, were evaporated in a vacuum centrifuge, and were redissolved in 0.1% formic acid

(vol/vol) in 2% (vol/vol) MeCN. LC-MS/MS experiments were performed on an electron spray ionization-Q-TOF (Q-TOF Ultima Global, Waters/Micromass, Manchester, U.K.) equipped with a nano–high-pressure liquid chromatography (Cap LC, Waters/Micromass), using a ten-port zero dead volume valve (Vici-Valco Instruments, Schenhon, Switzerland), enabling fast sample loading on a precolumn (LC-Packings PepMap C18, 5 μm, 100 Å, 300 μm × 5 mm) at a flow rate of 20 μl/min isocratically with solvent C (0.1% formic acid). The composition of solvents A and B were 0.1% (vol/vol) formic acid in 2% (vol/vol) and 80% (vol/vol) MeCN, respectively. After washing of the precolumn, the ten-port valve was switched, allowing delivery of the MeCN gradient at 180 to 200 nl/min onto the analytical column (Waters Atlantis dC18 m, 3 μm 75 μm × 150 mm) by back-flushing the precolumn. The gradient was from 5 to 60% B in 75 min. The mass spectrometer was calibrated using the CID spectrum of Glu-fibrinopeptide (Sigma) at 500 fmol/μl in 0.1% formic acid. The Q-TOF was operated in DDA mode with a 1-s MS survey scan on three different precursor ions. CID spectrum acquisition was allowed for up to a total of 10 s on each precursor ion or stopped when the signal intensity fell below 2 counts per second, before a new MS to MS/MS cycle was started. Precursors were excluded from any further MS/MS experiment for one minute, and singly charged ions were excluded as precursors for MS/MS. Protein identification was carried out using the Mascot MS-MS ion search software from Matrix Science, against the NCBInr database, allowing one trypsin missed cleavage and using a ±0.2-Da mass tolerance, a fixed carbamidomethyl Cys modification and a variable Met oxidation.

RT-PCR analysis.

The RNA of a pool of approximately 30 roots stored in liquid nitrogen was extracted according to the hot-phenol method (Reinhold-Hurek et al. 1993) with following modifications. An initial incubation step of 60 min in proteinase K buffer (200 mM Tris-HCl, pH 8.5, 250 mM NaCl, 25 mM EDTA, 0.5 % SDS, and 0.2 mg/ml proteinase K) was followed by two phenol (pH 5.5) extractions, and RNA was precipitated in the presence of Pellet Paint (Novagen, Darmstadt, Germany). DNase I (Roche) treatment was followed by one chloroform-phenol-isoamylalcohol extraction before precipitation. RNA was resuspended in 1× RNA-secure (Ambion, Austin, TX, U.S.A.), RNA concentrations were estimated spectrophotometically, and then, were carefully normalized against the actin 1-specific control RT-PCR. Scanning of ethidium bromide-stained gels (Typhoon 8600 variable mode imager) and the corresponding software ImageQuant (Amersham) were used for the accurate estimation of the band intensities. Forward and reverse primers (0.5 μM) were applied for the RT-PCR reactions based on Ready•To•Go RT-PCR beads (Amersham). The RT step was done for 30 min at 42°C, followed by inactivation for 5 min at 95°C and by cycling for 1 min at 95°C, 1 min at the specific annealing temperature, and 1 min at 72°C, followed by 5 min of extension at 72°C. The specific annealing temperature for the primers amplifying Bowman-Birk trypsin inhibitor (gi7619798), SalT (gi256637), or actin 1 were 62°C, 60°C, and 56°C, respectively. The absence of DNA template was controlled by inactivation of the RT prior to the reaction. Primer sequences were: SalT fw, TCGGTGTGGATGGACAGGAA; SalT rev, AAGTGAGCG GACCAGGTGCT; Bowman-Birk fw, GGCGACACCATGA TCCGTCT; Bowman-Birk rev, CCCAGTAGACGTCCTCG CAGA; Actin1 fw, CGCAGTCCAAGAGGGGTATC; Actin 1 rev, TCCTGGTCATAGTCCAGGGC (for Actin1, Yamaguchi et al. 2002).

ACKNOWLEDGMENTS

We thank D. Brar, International Rice Research Institute (IRRI), Los Banos, Philippines, for supplying rice seeds. We also thank Bruker Daltronics, Bremen, for MS/MS analyses of protein spots 46, 102, 107, and 108, K. Niehaus, University Bielefeld, for MS analysis of protein 55 and K. Rischka, Fraunhofer Institut IFAM, Bremen, for permission of mass spectrometer use. We would also like to thank T. Hurek for critical reading of the article. This work was supported in part by grant RE756/6-1 from the Deutsche Forschungsgemeinschaft, and by grants from the BMBF (0311946 and GenoMik 031U213D) to B. Reinhold-Hurek.

LITERATURE CITED

Agrawal, G. K., Jwa, N. S., and Rakwal, R. 2000. A novel rice (*Oryza sativa* L.) acidic PR1 gene highly responsive to cut, phytohormones, and protein phosphatase inhibitors. Biochem. Biophys. Res. Comm. 274:157-165.

Bantignies, B., Seguin, J., Muzac, I., Dedaldechamp, F., Gulick, P., and Ibrahim, R. 2000. Direct evidence for ribonucleolytic activity of a PR-10-like protein from white lupin roots. Plant Mol. Biol. 42:871-881.

Barraquio, W. L., Revilla, L., and Ladha, J. K. 1997. Isolation of endophytic diazotrophic bacteria from wetland rice. Plant Soil 194:15-24.

Berger, S. 2002. Jasmonate-related mutants of *Arabidopsis* as tools for studying stress signaling. Planta 214:497-504.

Claes, B., Dekeyser, R., Villarroel, R., Van den Bulcke, M., Bauw, G., Van Montagu, M., and Caplan, A. 1990. Characterization of a rice gene showing organ-specific expression in response to salt stress and drought. Plant Cell 2:19-27.

Czernic, P., Visser, B., Sun, W., Savoure, A., Deslandes, L., Marco, Y., Van Montagu, M., and Verbruggen, N. 1999. Characterization of an *Arabidopsis thaliana* receptor-like protein kinase gene activated by oxidative stress and pathogen attack. Plant J. 18:321-327.

de Souza Filho, G. A., Ferreira, B. S., Dias, J. M., Queiroz, K. S., Branco, A. T., Bressan-Smith, R. E., Oliveira, J. G., and Garcia, A. B. 2003. Accumulation of SALT protein in rice plants as a response to environmental stresses. Plant Sci. 164:623-628.

Egener, T., Hurek, T., and Reinhold-Hurek, B. 1999. Endophytic expression of *nif* genes of *Azoarcus* sp. strain BH72 in rice roots. Mol. Plant-Microbe Interact. 12:813-819.

Glazebrook, J. 1999. Genes controlling expression of defense responses in *Arabidopsis*. Curr. Opin. Plant Biol. 2:280-286.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100.

Hashimoto, M., Kisseleva, L., Sawa, S., Furukawa, T., Komatsu, S., and Koshiba, T. 2004. A novel rice PR10 protein, RSOsPR10, specifically induced in roots by biotic and abiotic stresses, possibly via the jasmonic acid signaling pathway. Plant Cell Physiol. 45:550-559.

Hause, B., Maier, W., Miersch, O., Kramell, R., and Strack, D. 2002. Induction of jasmonate biosynthesis in arbuscular mycorrhizal barley roots. Plant Physiol. 130:1213-1220.

Hurek, T., and Reinhold-Hurek, B. 2003. *Azoarcus* sp. strain BH72 as a model for nitrogen-fixing grass endophytes. J. Biotechnol. 106:169-178.

Hurek, T., Reinhold-Hurek, B., Van Montagu, M., and Kellenberger, E. 1994. Root colonization and systemic spreading of *Azoarcus* sp. strain BH72 in grasses. J. Bacteriol. 176:1913-1923.

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

Iniguez, A. L., Dong, Y., and Triplett, E. W. 2004. Nitrogen fixation in wheat provided by *Klebsiella pneumoniae* 342. Mol. Plant-Microbe Interact. 17:1078-1085.

Iniguez, A. L., Dong, Y., Carter, H. D., Ahmer, B. M. M., Stone, J. M., and Triplett, E. W. 2005. Regulation of enteric endophytic bacterial colonization by plant defense. Mol. Plant-Microbe Interact. 18:169-178.

James, E. K., and Olivares, F. L. 1998. Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. Crit. Rev. Plant Sci. 17:77-119.

James, E. K., Reis, V. M., Olivares, F. L., Baldandi, J. I., and Döbereiner, J. 1994. Infection of sugar cane by the nitrogen-fixing bacterium *Acetobacter diazotrophicus*. J. Exp. Bot. 45:757-766.

James, E. K., Olivares, F. L., Baldani, J. I., and Döbereiner, J. 1997. *Herbaspirillum*, an endophytic diazotroph colonizing vascular tissue in leaves of *Sorghum bicolor* L. Moench. J. Exp. Bot. 48:785-797.

Kim, S. T., Cho, K. S., Yu, S., Kim, S. G., Hong, J. C., Han, C. D., Bae, D. W., Nam, M. H., and Kang, K. Y. 2003. Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension-cultured rice cells. Proteomics 3:2368-2378.

Konishi, H., Kitano, H., and Komatsu, S. 2005. Identification of rice root proteins regulated by gibberellin using proteome analysis. Plant Cell Environ. 28:328-339.

Kunkel, B. N., and Brooks, D. M. 2002. Cross talk between signaling pathways in pathogen defense. Curr. Opin. Plant Biol. 5:325-331.

Lane, B. G. 2002. Oxalate, germins, and higher-plant pathogens. IUBMB Life 53:67-75.

Lange, J., Xie, Y.-P., Broughton, W. J., Vögeli-Lange, R., and Boller, T. 1999. A gene encoding a receptor-like protein kinase in the roots of common bean is differentially regulated in response to pathogens, symbionts and nodulation factors. Plant Sci. 142:133-145.

Liu, J., Blaylock, L. A., Endre, G., Cho, J., Town, C. D., VandenBosch, K. A., and Harrison, M. J. 2003. Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. Plant Cell 15:2106-2123.

Majoul, T., Chahed, K., Zamiti, E., Ouelhazi, L., and Ghrir, R. 2000. Analysis by two-dimensional electrophoresis of the effect of salt stress on the polypeptide patterns in roots of a salt-tolerant and a salt-sensitive cultivar of wheat. Electrophoresis 21:2562-2565.

McGee, J. D., Hamer, J. E., and Hodges, T. K. 2001. Characterization of a *PR-10* pathogenesis-related gene family induced in rice during infection with *Magnaporthe grisea*. Mol. Plant-Microbe Interact. 14:877-886.

Montesano, M., Koiv, V., Mae, A., and Palva, E. T. 2001. Novel receptor-like protein kinases induced by *Erwinia carotovora* and short oligogalacturonides in potato. Mol. Plant Pathol. 2:339-346.

Moons, A., Prinsen, E., Bauw, G., and Van Montagu, M. 1997. Antagonistic effects of abscisic acid and jasmonates on salt stress-inducible transcripts in rice roots. Plant Cell 9:2243-2259.

Passardi, F., Penel, C., and Dunand, C. 2004. Performing the paradoxical: How plant peroxidases modify the cell wall. Trends Plant Sci. 9:534-540.

Qu, L. J., Chen, J., Liu, M., Pan, N., Okamoto, H., Lin, Z., Li, C., Li, D., Wang, J., Zhu, G., Zhao, X., Chen, X., Gu, H., and Chen, Z. 2003. Molecular cloning and functional analysis of a novel type of Bowman-Birk inhibitor gene family in rice. Plant Physiol. 133:560-570.

Rakwal, R., and Komatsu, S. 2000. Role of jasmonate in the rice (*Oryza sativa* L.) self-defense mechanism using proteome analysis. Electrophoresis 21:2492-2500.

Rakwal, R., Agrawal, G. K., and Yonekura, M. 1999. Separation of proteins from stressed rice (*Oryza sativa* L.) leaf tissues by two-dimensional polyacrylamide gel electrophoresis: Induction of pathogenesis-related and cellular protectant proteins by jasmonic acid, UV irradiation and copper chloride. Electrophoresis 20:3472-3478.

Rakwal, R., Agrawal, G. K., and Jwa, N. S. 2001. Characterization of a rice (*Oryza sativa* L.) Bowman-Birk proteinase inhibitor: Tightly light regulated induction in response to cut, jasmonic acid, ethylene and protein phosphatase 2A inhibitors. Gene 263:189-198.

Reinhold, B., Hurek, T., Niemann, E.-G., and Fendrik, I. 1986. Close association of *Azospirillum* and diazotrophic rods with different root zones of Kallar grass. Appl. Environ. Microbiol. 52:520-526.

Reinhold-Hurek, B., and Hurek, T. 1998a. Interactions of gramineous plants with *Azoarcus* spp. and other diazotrophs: Identification, localization and perspectives to study their function. Crit. Rev. Plant Sci. 17:29-54.

Reinhold-Hurek, B., and Hurek, T. 1998b. Life in grasses: Diazotrophic endophytes. Trends Microbiol. 6:139-144.

Reinhold-Hurek, B., Hurek, T., Claeyssens, M., and Van Montagu, M. 1993. Cloning, expression in *Escherichia coli*, and characterization of cellulolytic enzymes of *Azoarcus* sp., a root-invading diazotroph. J. Bacteriol. 175:7056-7065.

Salekdeh, G. H., Siopongco, J., Wade, L. J., Ghareyazie, B., and Bennett, J. 2002. A proteomic approach to analyzing drought- and salt-responsiveness in rice. Field Crops Res. 76:199-219.

Sevilla, M., Burris, R. H., Gunapala, N., and 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*$^-$ mutant strains. Mol. Plant-Microbe Interact. 14:358-366.

Shen, S., Jing, Y., and Kuang, T. 2003. Proteomics approach to identify wound-response related proteins from rice leaf sheath. Proteomics 3:527-535.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68:850-858.

Silverman, P., Seskar, M., Kanter, D., Schweizer, P., Metraux, J. P., and Raskin, I. 1995. Salicylic acid in rice biosynthesis, conjugation, and possible role). Plant Physiol. 108:633-639.

Turner, J. G., Ellis, C., and Devoto, A. 2002. The jasmonate signal pathway. Plant Cell 14:S153-S164.

Van Damme, E. J., Zhang, W., and Peumans, W. J. 2004. Induction of cytoplasmic mannose-binding jacalin-related lectins is a common phenomenon in cereals treated with jasmonate methyl ester. Comm. Appl. Biol. Sci. 69:23-31.

van Loon, L. C., and van Strien, E. A. 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol. Mol Plant Pathol. 55:85-97.

van Loon, L. C., Bakker, P. A., and Pieterse, C. M. 1998. Systemic resistance induced by rhizosphere bacteria. Annu. Rev. Phytopathol. 36:453-483.

Vernoud, V., Horton, A. C., Yang, Z., and Nielsen, E. 2003. Analysis of the small GTPase gene superfamily of *Arabidopsis*. Plant Physiol. 131:1191-1208.

Wang, K. L., Li, H., and Ecker, J. R. 2002. Ethylene biosynthesis and signaling networks. Plant Cell 14 Suppl:S131-151.

Wang, Y., Ohara, Y., Nakayashiki, H., Tosa, Y., and Mayama, S. 2005. Microarray analysis of the gene expression profile induced by the endophytic plant growth-promoting rhizobacteria, *Pseudomonas fluorescens* FTP9601-T5 in *Arabidopsis*. Mol. Plant-Microbe Interact. 18:385-396.

Yamaguchi, T., Nakayama, K., Hayashi, T., Tanaka, Y., and Koike, S. 2002. Molecular cloning and characterization of a novel beta-1,3-glucanase gene from rice. Biosci. Biotech. Biochem. 66:1403-1406.

Yanni, Y. G., Rizk, R. Y., Corich, V., Squartini, A., Ninke, K., Philip-Hollingsworth, S., Orgambide, G., De Bruijn, F., Stoltzfus, J., Buckley, D., Schmidt, T. M., Mateos, P. F., Ladha, J. K., and Dazzo, F. B. 1997. Natural endophytic association between *Rhizobium leguminosarum* bv. *trifolii* and rice roots and assessment of its potential to promote rice growth. Plant Soil 194:99-114.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L., and Yang, H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92.

AUTHOR-RECOMMENDED INTERNET RESOURCES

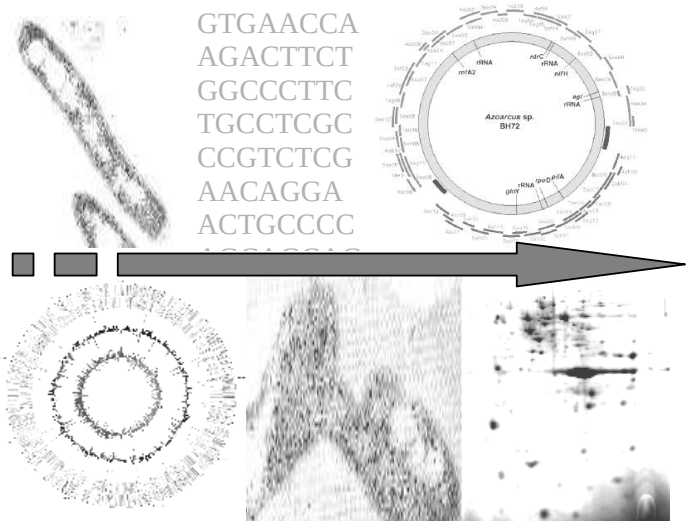
Technical University of Denmark, Center for Biological Sequence Analysis, NetPhos 2_0: www.cbs.dtu.dk/services/NetPhos

ProteoMetrics ProFound Advanced software: 65.219.84.5/service/prowl/profound/profound_E_adv.html

Matrix Science website: www.matrixscience.com

CHAPTER G

Discussion



The elucidation of an organism's genome is an important step towards understanding its biology**.** Data created by whole-genome sequencing are a pre-requisite for fields outside those genomics and bioinformatics, like transcriptomics or proteomics. In the present work three main topics from genomics to functional genomics studies in the model N_2-fixing grass endophyte *Azoarcus* sp. BH72, were covered.

Azoarcus sp. BH72 provides a high biotechnological potential. This bacterium is able to supply fixed nitrogen to its natural host, Kallar grass and additionally, it is capable to colonize the interior of rice roots, one of the globally most important crops (Hurek and Reinhold-Hurek, 2003). Moreover, since the rice genome has been sequenced, the *Azoarcus*-rice system represents a great model for bacteria-grass interaction studies. Understand the mechanisms of interaction and of host specificity are important and crucial knowledge for a larger agriculture application of these systems. In this sence for comparative genomics of the lifestyle, the genome of *Azoarcus* sp. strain EbN1 is available (Rabus *et al.*, 2005). This strain belongs to the branch of *Azoarcus* species that typically occur in soils and sediments but not in association with plants (Reinhold-Hurek, *et al* 2005). In contrast, the plant-associated strain BH72 has not been detected in root-free soil (Reinhold-Hurek and Hurek, 1998b). Genome comparison of both strains will give some hightlight of features that may contribute to the endophytic lifestyle of *Azoarcus* strain BH72, which could be use for developing biotechnological applications.

1. *Azoarcus* sp. BH72 genome: sequence assembly validation and physical map.

Strain BH72 genome was sequenced by using a WGS approach. For contig assembly validation and gap closure of the genome shot gun library, a chromosomal BAC-library was used. Statistics on clone size based on a large number of clones indicated an average BAC insert size of 90 kb for the strain BH72 library. This is comparable to other prokaryotic genomes, in particular those with similar GC content and similar genome size such as *Mycobacterium* (Brodin *et al.*, 2002; Brosch *et al.*, 1998). Through a combination of BAC-library screening, BAC clone end-sequencing as well as computational tools, a minimal tiling path of BAC clones, which cover almost the whole chromosome, were assembled (Chapter B, manuscript Figure 2). As it is shown, two gaps were not closed by screening the entire BAC library. This indicates that certain regions appeared to be underrepresented, either due to problems in cloning of particular fragments, or to statistical reasons. An additional cause could be the low *Hin*dIII site cut frequency per 100 kb (1 or 2 cuts), within these gaps. In contrast to that, other parts of the chromosome, which were well covered by BAC clones, possess 4 cuts/100 kb. This problem was overcome by the building of a genome fosmid library and the construction of a genome fosmid map (Chapter C, manuscript Figure 1).

For additional chromosome assembly validation, BAC clones from the minimal tiling path (Chapter B, manuscript Figure 2) were subjected to RFLP analysis using the restriction endonuclease *Not*I. Additionally, the *in gel* and *in silico* size determination was compared using the program pDRAW32. In visual inspections of the restriction patterns, they correlated very well (Chapter B, Figure 1). Restriction fragment sizes were determined from gels and *in silico* digests, and were considered to be in agreement when at least 60% of the fragments were identical. 100% of the analyzed clones showed agreement with the draft genome assembly according to these criteria. These analyses provide strong evidence for a correct assembly of the genome sequence from the shotgun library and support the observation that the genome of strain BH72 consists only of one replicon.

Using specific and general probes targeting different positions in the *Azoarcus* sp. BH72 16S rRNA sequence, four clusters of BAC clones were identified that contain four different *rrn* operons according to RFLP (Chapter B, manuscript Table 1). Thus, *rrn* operons were well represented in the BAC- library. The number of *rrn* operons identified in *Azoarcus* is similar to copy numbers in other relatively fast-growing, related Proteobacteria such as *N. meningitidis, R. solanacearum, P. aeruginosa* and *A. tumefacien* (Gäher *et al.*, 1996; Goodner *et al.*, 2001; Rabus *et al.*, 2005). In contrast, the genome of *B. japonicum* that grows extremely slowly harbors only one copy despite a larger genome size (Kaneko *et al.*, 2002). This is according to the supposition that the amount of 16S rDNA copies may correlate with ecological strategies to respond more or less rapidly to changing environmental conditions (Klappenbach *et al.*, 2000). Strain EbN1 genome also encodes for four 16S rRNA, but their relative position in the chromosome is different in comparison with the relative position in strain BH72 (Rabus *et al.*, 2005). End-sequences of BAC clone inserts carrying 16S rRNA genes enabled us to map the clones onto the draft of the *Azoarcus* genome sequence assembled from shotgun cloning (Chapter B, manuscript Figure 2). *In gel* size determination of inserts and RFLP, validated the expected position of BAC clones on the assembled shotgun sequence and confirmed the suggested correlations of *in silico* and *in gel* restriction patterns. These results give additional evidence for a correct assembly of the genome sequence from the shotgun library.

Based on the BAC clone library a representative physical map of *Azoarcus* sp. strain BH72 genome was constructed by contig assembly via BAC clone end-sequencing and hybridization. A set of probes for marker genes and for BAC clone end-sequences was used for library screening and was positioned in the physical map. Additionally the end sequences of 109 BAC inserts that cover the entire genome were compared with the NCBI database. Assignment of these sequences to the genome assembly provided a high-density map with marker genes of the *Azoarcus* genome (Chapter B, manuscript Figure 3B). Additionally, protein or nucleic acid similarities to known genes were assigned to different protein function categories (Chapter B, manuscript Figure 3). Data obtained in this analysis show a large group (29%) of proteins

similar to conserved hypothetical proteins in bacteria. Most of those are similar to proteins found in the β-or δ-subgroup of *Proteobacteria* (37% and 24% respectively) as to be expected according to the taxonomic affiliation of *Azoarcus* sp. (Reinhold-Hurek *et al.*, 1993b). In particular, similarities were found with the soil-borne species *Azoarcus* sp. EbN1 and plant-pathogenic bacteria, suggesting that *Azoarcus* sp. BH72 may share some unknown functions with bacteria of this life style (Hurek and Reinhold-Hurek, 2003).
A similar percentage (30%) was obtained in a similar analysis carried out for the N_2-fixing symbiont *B. japonicum* (Tomkins *et al.*, 2001). As in strain BH72, also in *B. japonicum* a large group of proteins were related to bacterial metabolism (41%) (Tomkins *et al.*, 2001). On the other hand, *Azoarcus* sp. BH72 physical map shows that the localization of marker genes related to nitrogen metabolism and nitrogen fixation, such as *nifH, ntrC, glnY* or *rnfA2*, are distributed in the chromosome and not necessarily physically linked (Chapter B, paper Figure 2).

1.1 Genome comparison

With the aim to elucidate if the plant-associated bacterium strain BH72 shares some similarity in genome composition with the related soil-borne bacterium strain EbN1 further genomes comparison were done. Comparison of the BAC-end sequences (from clones which cover almost the whole genome), with the recently published genome of *Azoarcus* sp. EbN1 (Rabus *et al.*, 2005) revealed considerable differences. Only 81% of the analyzed genes had homologues in the genome of EbN1. Out of those, only two third had highest hits in BLAST analyses with genes of EbN1, while one third had highest similarities to genes of other bacteria. Important marker genes such as nitrogenase genes (*nifHDK*) were not present in EbN1, which is not diazotrophic. Comparison of this map with the gene arrangements in the genome of the soil strain *Azoarcus* sp. EbN1 revealed a relatively low amount of syntheny. Moreover, the relative position of the initially known common marker genes (Chapter B, manuscript Figure 2) in the genome was different in strain EbN1. These observations are in agreement with results obtained from whole genome comparisons which revealed a surprisingly low degree of syntheny between genomes of strains BH72 and EbN1, although they are two species of the same genus (Chapter C, Figure 2). At a low cut-off e-value of e^{-30}, the majority of genes in strain BH72 have some counterparts in strain EbN1 (Chapter C, manuscript Figure 1b, circle 1). However, only 39% of the translated protein-coding genes share highest amino acid sequence identity to proteins of EbN1. Taking in mind the different lifestyle of both strains, these differences detected may give important hints on genes required specifically for the respective lifestyle.

2. From genomic to functional genomic: proteomics studies and metabolic pathways reconstruction of N_2-fixing *Azoarcus* sp. BH72 cells.

2.1 Comparison of 2D-proteome pattern of strain BH72 growth in pure- or co-culture N2-fixing conditions.

In post-genomic era, proteomics is one of the best strategies used to reveal the dynamic expression of whole proteins in cells and their interactions. High resolution 2D-PAGE combined with mass spectrometry and bioinformatics analysis is widely used for a large-scale study of protein identification based on peptide mass fingerprint (PMF) (Mann *et al.*, 2001; Pandey and Mann, 2000). A proteomic analysis of an organism under different conditions, can contribute to understand of how a microbe adapts its metabolism to different environments (Bestel-Corre *et al.*, 2004a). Identification of differential displayed proteins could be used to determine the genes responding to relative physiological actions and clarify the function of these genes.

In this work for the first time a high-throughput functional analysis using a differential-display proteomic approach was conduced in *Azoarcus* sp. BH72 grown under differents N_2-fixing conditions (Chapter E). This approach was used with the aim to identify proteins of strain BH72 cells differentially expressed in pure culture or in co-culture N_2-fixing conditions growths, to reconstruct metabolics pathways, as well as for identified putatives proteins involved in diazosomes structure or function. For that, *Azoarcus* sp. BH72 membrane and cytoplasmic proteins were isolated from cells grown on both N_2-fixing conditions. Different 2D-PAGE gels protein profiles were compared and spots of interest were analyzed by MALDI-TOF for its identification using an unreleased *Azoarcus* sp. BH72 genome data (Chapter E).

Proteome patterns comparison of membrane and cytoplasmic fractions, from strain BH72 grown under both N_2-fixing conditions were done (Chapter E, Figure 1 and Table1). Both fractions proteome patterns of strain BH72 pure culture cells revealed significant differences from co-culture ones (Chapter E, Figure 1). Gel image analysis reveals that in both fractions most of the spots were down regulated (53 and 46% in the membrane and cytoplasmic fraction respectively). This down regulation of co-culture proteins indicates that under these conditions, cells must have a more specialized metabolism to cover their metabolic requirements with fewer enzymes. New proteins were detected in the proteome patterns of co-culture cells. In particular, 9 novel proteins were detected in the wild type proteome (Chapter E, Figure 1F), which most probably are related to the presence of specific internal membranes formed under co-culture conditions.

It is well known that the methodology used in this work has several limitations for integral membrane protein resolution and identification. Hydrophobic proteins are not easily solubilized

in the non-ionic detergents used for isoelectric focusing, and the hydrophobic fragments obtained after trypsin treatment are difficult to ionize for mass spectrometric analysis (Aebersold and Mann, 2003; Pandey and Mann, 2000; Santoni *et al.*, 1999; Santoni *et al.*, 2000b). The establishment of several 2D-bacterial reference map and posterior indentification by MS shows that only few membrane proteins were able to identify in comparation with the number of membrane proteins annotated on those genomes sequences. In these cases few numbers of transmembrane motifs in the proteins identified were available (Santoni *et al.*, 2000b). Proteomics studies of membrane proteins have been mostly carried out using carrier amphlytes or inmobilized pH gradients IEF strips as a first dimension (Santoni *et al.*, 2000b), or large 1D SDS-PAGE gels (Lai *et al.*, 2004; Rhomberg *et al.*, 2004; Santoni *et al.*, 2000b). In these studies several detergents were used with the aim to improve the solubilization of membrane proteins in the extraction and sample buffer (Santoni *et al.*, 2000b). In the present proteomic study, a capillar gel with ampholites isoelectric focused was used as a first dimension, while a combination of different detergents (ASB-14 and CHAPS) were applied in the sample buffer and in the capillar gel buffer (Chapter E, manuscript Materials and Methods). Results obtained using this methodology shows a large number of identified membrane proteins, including outer membrane ones, from where transmembrane domains were detected (Chapter E, manuscript Table 1). This means that the methodology for membrane protein extraction and separation was successfully (Chapter E, manuscript Figure 1, panel E, F and G). However a degree of under-representation of membrane proteins is expected. This is reflected in the fact that from the 5 ABC-transporter systems identified (Chapter E, manuscript Table 1), only the ATP-binding protein or the periplasmic binding proteins subunits were detected. This was also observed in proteomics studies of membrane fractions in *S. meliloti* and *A. borkumensis* (Djordjevic *et al.*, 2003; Sabirova *et al.*, 2006).

PMF of approximately 800 proteins spots were evaluated from several gels in the pH range of 3-10. The protein products of 179 unique genes were identified using this approach, which represent 4.5% of the genome coding genes. From these group 140 genes products were assigned EC number in the *Azoarcus* sp. BH72 database. These numbers were used to search the KEGG database for pathways reconstruction (Chapter E, Table 1 and Figure 2).

2.2. Carbon metabolism

C4-dicarboxylates like succinate, fumarate and malate are transported and metabolized by bacteria under aerobic or anaerobic conditions. In aerobic growth, the substrates serve as carbon sources and become oxidized to CO_2 in the citric acid cycle (Janausch *et al.*, 2002). Major carbon sources for strain BH72 are dicarboxylic acids and ethanol (Reinhold-Hurek *et al.*,

2005). In agreement with this, four complete and two incomplete copies of the TRAP-transport system (*dctPQM*) were detected in the genome, which are a family of high-affinity transporters for C4-dicarboxylates (Kelly and Thomas, 2001) (Chapter D, Figure 10). As in symbiotic rhizobia this C4-dicarboxylate transporters might be of vital importance during the association with the host plant (Yurgel and Kahn, 2004). C4-dicarboxylates are most probably metabolized via the glyoxylate shunt pathway to the level of malate, which is then decarboxylated to pyruvate by the malic enzyme (MaeB) from which two copy are present in strain BH72 genome (Chapter D, Figure 11). Terminal oxidation of pyruvate may then be performed by the pyruvate dehydrogenase complex and the TCA-cycle. Proteomic studies on strain BH72 grown under different N_2-fixing conditions demonstrated this hypothesis. Using this approach one component of a C4-dicarboxylate transporter system (DctP2), a malic enzyme (Maeb2), as well as several enzymes involved in the glyoxylate shunt were identified (Chapter E, table 1 and Figure 2). The malic enzyme was detected constitutively expressed, while the DctP2 protein was found down-regulated in co-cultures cells. The presence of these proteins in pure culture is related with the fact that strain BH72 was grown in SM medium with L-malate as sole carbon source (Reinhold *et al.*, 1985). Pyruvate formed by the malic enzyme, is probably converted to acetyl-CoA by the pyruvate dehydrogenase complex, from which a subunit (LpdA) was also identified (Chapter E, Figure 2). In co-culture the added concentration of L-malate in the FU medium was chosen to be 1% from those taken under pure culture conditions. This was done with the aim to promote the growth of strain BH72 in the beginning and to bypass the lag-phase (Hurek *et al.*, 1995). By HPLC-analysis it was ensured that the L-malate was completely metabolized, reflecting that the expression in co-culture of MaeB2 is due to its role in the pyruvate metabolism. In rhizobia as well as in *R. capsulatus*, C4-dicarboxylate transport systems and the corresponding assimilatory pathways, are induced only in the presence of external C4-dicarboxylates, where a two component regulatory system is involved (Forward *et al.*, 1997; Hamblin *et al.*, 1993; Janausch *et al.*, 2002; Yurgel and Kahn, 2004). The genes *dctRS* coding for a sensor and a regulator, are also present in strain BH72 genome, located upstream of one set of the C4-dicarboxylate transporter operon (*dctP2Q2M2*) from where the DctP2 subunit was identified (Chapter D, Figure 10). These results show that in *Azoarcus* sp. BH72 the *dctP2Q2M2* operon codes for a malate transport system, which is likely to be regulated by external amount of malate via the two component system *dctRS*.

As mentioned, mayor carbon sources for strain BH72 are dicarboxylic acids and ethanol (Reinhold-Hurek *et al.*, 2005). Ethanol might be important for association with flooded plants like rice, which accumulate ethanol under anoxic conditions, especially at root tips - one primary site of colonization of strain BH72. In agreement to that, strain BH72 genome has a high number of genes encoding putative alcohol dehydrogenases (10) (Chapter C, manuscript supplementary Table 1 online), five of which are PQQ-dependent enzymes that are not present

in strain EbN1. On the other hand, in co-culture with *A. alternatum* strain 2003, strain BH72 was totally dependent on fungal exudates since no other carbon source than glucose was included in the FU medium except for a minimal amount of malate to promote the beginning of strain BH72 growth and overcame the lag phase (Hurek *et al.*, 1995). Exudates detected in the supernatant fungual pure culture were ethanol, lactate, acetate and formiate. In co-culture growth acetate, ethanol and lactate were detected in less amount but not formiate. These results are in agreement with the fact that strain BH72 can growth on acetate, ethanol and lactate as a sole carbon sources but not in formiate. Moreover, these results are in correlation with the identification of several enzymes, up-regulated in co-culture growth, involved in the assimilation of these compounds (Chapter E, Figure 2).

Like the other plant-associated species of *Azoarcus* spp., strain BH72 has a strictly respiratory type of metabolism and does not grow on carbohydrates (Reinhold-Hurek *et al.*, 2005). The genome of strain BH72 does not contain any of the genes required for a functional Entner-Doudoroff pathway or the oxidative branch of the pentose phosphate pathway, but the non-oxidative pentose phosphate pathway is complete (Chapter D, Figure12). However, all the enzymes required for glycolysis via the Embden-Meyerhoff pathway, gluconeogenesis (Chapter D, Figure 13), and the TCA cycle (Chapter D, Figure 14) are present in *Azoarcus* sp. BH72 genome. According to that, in pure- or co-culture N_2-fixing cells, several TCA enzymes were identified constitutively expressed. This confirms the key role of this pathway in the *Azoarcus* sp. BH72 metabolism for energy production as well as for a source of different intermediate precursors for other metabolic pathways. In this sense, numerous enzymes involved in the amino acid, fatty acid or cofactor biosynthetic pathways as lysine, glutamate, propanoate or porphyrins, which use TCA intermediate compounds as precursors, were also identified in both conditions studied (Chapter E, table 1). The glycolysis-gluconeogenesis pathways were also active in the N_2-fixing conditions studied, most probably for anabolism. This hypothesis is supported by the mentioned fact that strain BH72 can not use glucose and other carbohydrates as carbon sources (Reinhold-Hurek *et al.*, 2005; Reinhold-Hurek and Hurek, 2000) and additionally, by the up-regulated identification of proteins involved in the synthesis and transport of exopolysacharides (Chapter E, table 1). In this study was found that in strain BH72, the TCA and glyoxylate cycles are linked to the gluconeogenesis pathways by the phosphoenolpyruvate carboxykinase Pck enzyme, which was identified up-regulated in co-culture.

This functional genomic analysis reflects the fact that under the N_2-fixing conditions *Azoarcus* sp. BH72 adapts its carbon metabolism to scope with the sources available. Particulary, in co-culture growth conditions, the carbon metabolism was adapted with the aim to utilize the exudates secreted by the fungus into the culture medium.

2.3 N_2-fixation and nitrogen assimilation metabolism

The genome sequence of *Azoarcus* sp. strain BH72 revealed that in contrast to the soil strain EbN1, structural genes for the Mo-dependent nitrogenase complex and all genes required for cofactor synthesis and maturation of the nitrogenase are present in strain BH72 (Chapter C, manuscript supplementary Table 1 online). Additionally, several putative low-potential electron donors for N_2-fixation were identified in the genome sequence. These include flavodoxin-encoding genes, genes for ferredoxin-like proteins as well as clusters encoding putative membrane-bound electron transport systems which could be involved in the electron supply to ferredoxin during nitrogen fixation (Chapter C, manuscript supplementary Table 1 online). By proteomic studies, several proteins involved in the nitrogen metabolism were detected in cells grown in pure or in co-culture under conditions of N_2-fixation. The three subunits of the nitrogenase complex NifHDK, as well as enzymes related to its synthesis (NifSU, NifE and NifN) and activation (NifM), were identified. Interesting was the detection of the NifH and NifM proteins, being down-regulated in the cytoplasm fraction of co-culture cells, while constitutive and up-regulated in the membrane fraction respectively. This results give more evidence to the observation that the NifH protein is strongly associated to the membrane fraction in co-culture, most probably to diazosome membranes as was described (Hurek *et al.*, 1995).

A mobile electron donor ferredoxin V FdxD protein was identified to be constitutively expressed in strain BH72 cytoplasmic fractions. Previous work in *Azoarcus* sp. BH72 has demonstrated that the *nifHDK* genes are co-transcribed in one operon together with a ferredoxin gene *fdxN*. Mutational analysis showed that FdxN play an important but not essential role in N_2-fixation (Egener *et al.*, 2001). To posses an extra ferredoxin, which most probably is also involved in the electron transport flux to nitrogenase, could be the explanation of the residual nitrogenase activity in $FdxN^-$ mutants.

One of the key regulators of *Azoarcus* sp. BH72, involved in the nitrogen metabolism, was also identified under the N_2-fixing conditions studied in this work (Chapter E, Table 1). The NtrBC two-component system plays an important role in sensing and inducing N-regulatory cascades involved in N_2-fixation and assimilation (Ninfa *et al.*, 1995). In *Azoarcus* sp. BH72 it was demonstrate that NtrBC activates *nifLA* gene expression, moreover that NifA acts as an essential transcription activator for the *nifHDK* genes (Egener *et al.*, 2002). These previous results are in agreement with results obtained here in which the constitutive expressions of NtrC as well as of the NifHDK proteins were shown during growth on N_2-fixing conditions.

In the context of N-assimilation, the strain BH72 genome reveals only the presence of the high-affinity ATP-consuming assimilation system (GS-GOGAT). From this system, two copies of the glutamine synthetase gene (*glnA* and *glnT*) were annotated, which encode for a glutamine synthetase type I and III protein respectively (Chapter C, supplementary Table 1 online).

Proteomics studies reveal that the glutamine synthetase type I protein (GlnA) was constitutively expressed under both N_2-fixing conditions studied.

Additional proteins probably involved in the N-assimilation, were identified in strain BH72 N_2-fixing cells. These included the two subunits CarA and CarB of a carbamoyl phosphate synthase (glutamine-hydrolyzing) protein. The Car protein is involved in the glutamine assimilation by forming glutamate and carbamoyl-phosphate as a part of the glutamate and pyrimidine metabolism. Since no GOGAT protein was detected under the conditions studied, most probably the glutamine assimilation to glutamate could occur in *Azoarcus* sp. BH72 by using this pathway.

Further proteins related to the N-metabolism were identified in the cytoplasm fraction. These include the two subunits constitutively expressed of the glycine cleavage multienzyme complex (GCV), the up-regulated PurD protein, as well as the alpha subunit of the urease protein, (UreC). This last protein was detected novel in the wild type strain. Interestingly the GCV complex as well as the PurD protein use glycine as a substrate. The PurD protein, form N1-(5-phospho-D-ribosyl) glycinamide, an intermediate of the purines metabolism. More probably this protein is regulated by glycine and its up-expression could be as a consequence of a high amount of glycine produced by an active N_2-fixing process in co-culture conditions. On the other hand, the GCV complex catalyses the oxidative cleavage of glycine into: CO_2, NH_3 and a C1 unit. In *E.coli* the GCV encoding genes (*gcvTHP*), are positively regulated by glycine (Heil *et al.*, 2002). *Azoarcus* sp. BH72 genome encodes for a *gcvTHP* operon which has the same genetic arrangement as in *E. coli*, which means that most probably the expression of this complex is due to a glycine induction. Moreover, the C1 units generated as a product of the reaction are not end products. Like in *E.coli*, these compounds could be used as substrates in other metabolic pathways like the biosynthesis of purines, thymidine, histidine, coenzyme A and methionine (Heil *et al.*, 2002), from where proteins were also identified in strain BH72 (Chapter E, Table 1).

The UreC protein, which is part of the urease complex, catalyzes the hydrolysis of urea to form ammonia and CO_2 as a part of the urea cycle. The regulation of these proteins is diverse in rhizobia. Subunits from this protein complex were detected as down-regulated in *S. meliloti* in protein extract of bacteroids (Natera *et al.*, 2000), while urease activity were detected in *R. leguminosarum* vegetative cells as well as in pea bacteroids (Toffanin *et al.*, 2002). These observations show no clear regulation of the urease expression and activity in rhizobia under N_2-fixing conditions. Results obtained in *Azoarcus* sp. BH72 suggests a urease ammonium-independent regulation since both conditions studied were no rich in ammonium. The question to we addressed in this point is from where the urea come from.

2.4 Energy production

The process of N_2-fixation is highly energy demanding as nitrogenase requires 16 molecules of ATP for the reduction of 1 molecule of N_2 to ammonium. During this process, which occurre under extremely low oxygen concentrations, a respiratory chain with high affinity to oxygen closely coupled to ATP production is required (Saeki, 2004). This hypothesis is valid for *Azoarcus* sp. BH72 and it is supported by results obtained from the functional genomic approach used on N_2-fixing cells. In particular, several subunits of the NADH-ubiquinone oxidoreductase complex (Nuo), involved in the respiratory chain, as well as a cytochrome c-type biogenesis protein CycH, and an ubiquinol-cytochrome-c reductase iron-sulfur protein PetA1, were detected membrane-associated (Chapter E, table 1). The PetA1 protein was found constitutively expressed, while the CycH was detected novel in co-culture wild type cells. This specific up-regulation is correlated with the presence of specialized internal membranes in strain BH72 co-culture cells where this protein might locate.

Additionally, in the cytoplasmic fraction three electron transfer flavoproteins EtfA1, Etf1 and IsiB were identified (Chapter E, table 1). In particular, the Etf1 protein was detected up-regulated in correlation with an intensive, high energy demanding, N_2-fixation process in co-culture conditions. Finally, several subunits of the ATP synthase were detected membrane-associated and in some cases up-regulated in co-culture growths conditions (Chapter E, table1). In a context these results are in agreement with the hypothesis in which diazosomes formed in strain BH72 under co-culture conditions are not only involved in the N_2-fixation, but also in the respiration process (Hurek *et al.*, 1995).

2.5. Protein synthesis and folding

An active protein synthesis in both N_2-fixing conditions studied is reflected by the large number of proteins identified related to this process particulary in co-culture (Chapter E, manuscript Table 1). This suggests a differentiation process during co-cultures growths and is in concordance with the detection of smaller amount of spot proteins in co-culture 2D-proteome patterns.

Several proteins involved in chaperoning were identified under the N_2-fixing conditions studied including the constitutive expressed GroES1, DnaK, Tig, PpiB, Hsp18 and ClpX proteins (Chapter E, manuscript Table 1). The first three appeared to be the major spot in the gels and their role in protein folding, degradation and in the assembly of large protein complexes are well known in bacteria (Bukau and Horwich, 1998). Additionally, PpiB and ClpX are chaperones which accelerates the folding of proteins (Herrler *et al.*, 1994) while Hsp18 belongs to the small heat shock protein family, which are involved in the prevention of the accumulation of unfolded intermediates during stress periods (Veinger *et al.*, 1998). Members of these chaperone families

were reported to be important during the establishment of the rhizobia-host symbiotic association and expressed in bacteroids cells, as well as induced under nutrient stress conditions (Djordjevic *et al.*, 2003; Djordjevic, 2004; Munchbach *et al.*, 1999; Yeh *et al.*, 2002). Results obtained from *Azoarcus* sp. BH72 suggest that the N_2-fixation process could be a stressful event for the bacterium, as many of the chaperones identified play a key role during stress response. This hypothesis is supported also by the presence of additional co-culture up-regulated proteins in strain BH72 with chaperoning function like GroeL1 and a conserved hypothetical peptidyl-prolyl cis-trans isomerase protein (Chapter E, manuscript Table1).
Differentially expression of chaperone sets in pure- or co-culture N_2-fixing conditions indicates a shift in the cellular status and moreover, where an active specific protein synthesis is taking place. This observation is correlated with the presence of diazosomes in co-culture cells, and is supported by the presence of additional enzymes related to the processing of proteins.

2.6. Iron transport and metabolism in *Azoarcus* sp. BH72

In both COG categories analyzed in this work (Chapter D), strain BH72 genome possesses double amount of genes than the closest relative *Azoarcus* sp. strain EbN1, providing additional evidence that these two strains have different genome compositions (Rabus *et al.*, 2005) (Chapter D, Figure 2). Moreover the strain BH72 genome has one of the highest numbers of genes belonging to the Ion transport and Metabolism category in comparison with the other studied genomes. In particular, a high amount of genes within this category coded for proteins related with iron transport and metabolism (Chapter D, Figure 3). Within this set, remarkable was the presence of a large number of encoding genes for putative TonB-dependent outer membrane (OM) receptors (Chapter D, Table 5). This is twice the number of genes described for *Azoarcus* sp. EbN1 and even more in comparison to N_2-fixing symbionts, such as *B. japonicum* (13), *S. meliloti* (9) and *M. loti* (1). Interestingly from this set of genes, four are not present in *Azoarcus* sp. EbN1 genome. Additionally two of these genes (azo2156, azo3836) are not even present in the *P. fluorescens* Pf5 genome, a plant associated bacterium known for its capacity to produce and up take a wide range of siderophores (Loper and Henkels, 1999; Paulsen *et al.*, 2005). Blast analysis of the TonB-dependent OM receptors sequences shows that strain BH72 has several putative receptors for the general classes of hydroxamates and catechol-type siderophores, ferricitrate, vitamin B12 and colicins (Chapter D, Table 5). However, no evidence for biosynthetic pathways for known hydroxamate or catecholate siderophores in its genome was found. Moreover, no positive reaction was observed when Chromeazurol S (CAS) assay was used to detect siderophores independent of their structure (Chapter D, Figure 5). Results obtained suggest that strain BH72 is highly adapted to obtaining chelated iron from other sources, since fungi and monocotyledonous plants also produce siderophores (Crowley *et*

al., 1991). In this sense interesting was the detection of siderophore production by the fungus partner when strain BH72 was streaked in contact with the fungus mycelium (Chapter D, Figure 6). These results shows that most probably in its natural environment, strain BH72 use as iron source the heterologous ferric-siderophore synthesized and excreted by the fungal partner as was reported for other bacteria (Jurkevitch *et al.*, 1993; Llamas *et al.*, 2006; Loper and Henkels, 1999).

Proteomic analysis of *Azoarcus* sp. BH72 cells, demonstrated the expression of five of those OM TonB-dependent receptors (Chapter D, Table 5). Interestingly one of them (azo3556), which has high homology with TonB-dependent OM receptors for vitamin B12, was found to be up-regulated in co-culture cells. This gene is located within a genomic region which contains genes related with the vitamin B12 metabolism. Since vitamin B12 was supplied in the FU medium used for co-culture growths, these results suggested a putative function of this receptor in the vitamin B12 uptake.

In rhizobia the expression of OM TonB-dependent siderophore receptors was found negatively regulated by iron. Additionally, mutants in these receptors were not affected in their capacity to fix-nitrogen which means that they are not necessary during symbiosis (Lynch *et al.*, 2001; Yeoman *et al.*, 2000). Both growth media used in this study were iron rich, as this element was supplied in the form of Fe^{3+}-EDTA. Results obtained in strain BH72 suggest an iron-independent regulation of these OM receptor proteins under both N_2-fixing conditions studied and moreover, that these receptors are involved in other process independently of iron acquisition from siderophores.

In *R. solanacearum* it was described that the binding of the PrhA protein (which has high homology to an OM TonB-dependent siderophore receptor) to a non-diffusible structure of the cell wall, initiated a regulatory cascade that finally induced the *hrp* (hypersensitive response and pathogenicity) genes. These genes are crucial for the interaction with host and non-host plants (Brito *et al.*, 2002; Marenda *et al.*, 1998). Concordantly, it was reported that OMP TonB-dependent receptors are also important for perception of environmental signals and are associated with pathogenicity of plant pathogens (Clarke *et al.*, 2001; Koebnik, 2005). Taking in mind these observations we can speculate that in strain BH72 some OM TonB-dependent receptors could be involved in the perception of environmental signals or are playing a role in the bacteria-fungus interaction or even in plant-microbe interaction. Nevertheless, more experiments are necessary with the aim to demonstrate this hypothesis and the real function of these receptors in *Azoarcus* BH72.

An alternative function of these set of OM-TonB receptors might be a role in biocontrol. Was reported that several rhizobacteria, which produce various types of siderophores to chelated the scarcely available iron, can deprive pathogens from acquiring iron (Van Loon and Bakker, 2003). Mutants defective in siderophore biosynthesis in the plant growth-promoting *P. putida*

WCS358, were unable to increased potato root growth and tuber field in pot and field experiments, while the wild-type was able to such effect (Marugg *et al.*, 1985). In these experiments with potato, the increased plant growth was due to suppression of deleterious rhizosphere microorganism (Schippers *et al.*, 1987). The involvement of siderophore production in disease supression by strain WCS358, was further studied on carnation, radish, and flax using different strains of *Fusarium* (Van Loon and Bakker, 2003). In the case of *Azoarcus* sp. BH72, is suitable to speculate that the fact to posses different types of receptors for siederophores, could give and advantage with respect to others microorganism for iron acquisition including pathogens.

In addition to the OM receptor proteins described, strain BH72, harbors genes encoding encode for all the proteins necessary for ferric-siderophore uptake. These include two complete *tonB-exbBD* operons (Chapter D, Figure 7), as well as two complete and one incomplete ABC-iron transport systems operons (Chapter D, table 6). The TonB-exbBD complex is involved in transducing the proton-motive force energy from the cytoplasmic membrane to the outer membrane where it is needed (Postle and Kadner, 2003), whereas the ABC-iron transport systems is involved in the transport across the cytoplasmic membrane (Koster, 2001). Additionally, strain BH72 genome has two genes encoding a putative Fur regulator protein. This protein was described in several bacteria as a key negative regulator for genes related to the iron uptake and siderophore biosynthesis (Escolar *et al.*, 1999)**.** This set of genes speaks in the favour of that strain BH72 being able to up take different types of heterologous-siderophore.

2.7. Cellular protection and detoxification

Production of reactive oxygen species (ROS), are an intrinsic property of aerobic metabolism. Particularly in N_2-fixing bacteria, ROSs are additionally generated by the high requirement for iron as a key component in the nitrogenase and several electron-transport proteins(Santos *et al.*, 2001a; Santos *et al.*, 2001b). Additionally, nitrogenase is extremely oxygen sensitive which means that the inactivation of ROSs is very important for a correct N_2-fixation process. According to that several proteins involved in the detoxification of ROSs were identified in both N_2-fixation conditions studied. Moreover, some of them were up-regulated in co-culture (Chapter E, table 1). Several proteins includong the two major spots SodB and KatA, involved in the detoxification of ROSs, were found to be up-regulated in *S. meliloti* bacteroids. Mutants from both proteins were drastically affected in their symbiotic capacity reflecting a key protective role in their symbiotic process (Djordjevic *et al.*, 2003; Djordjevic, 2004). These observations show that the main importance of the detoxification proteins appeared to be in highly N_2-fixing conditions, a feature also shown by *Azoarcus* sp. BH72 in co-culture growths.

2.8. Membrane proteins

Membrane proteins are of particular interest in proteomic studies. In bacteria they play important functions in signal transduction pathways, transport, cell interactions and other processes. A set of three major constitutive spots in the membrane fraction were identified as outer membrane porins (Chapter E, table 1). Interestingly isoforms of the porins mentioned were detected to be up-regulated in co-culture (Chapter E, Figure 1F and Table 1). In many host-microbe interactions, porin proteins are virulence factors (Massari *et al.*, 2003), the presence of new isoforms in co-culture possibly indicates that these porins play a role in strain BH72-fungus interaction.

Proteins related to the cell envelope biogenesis, integrity and shape were detected in both conditions studied. These include the OstA and the Pal proteins which were reported to be involved in the cell envelope biogenesis and integrity (Braun and Silhavy, 2002). Interestingly, the Pal lipoprotein, a member of the OmpA family, was down regulated in the wild type strain and constitutive in the strain BH1599 (Chapter E, Table 1). Additionally, a rod shape-determining protein MreB, was down-regulated in co-culture cells. Previous microscopical observations in strain BH72 co-culture cells detected frequently dumbbell-shaped bacteria which can apparently not divide anymore (Hurek *et al.*, 1995). These observations are in correlation with the absence in co-culture of the MreB protein and with the down-regulation in the wild-type of the Pal protein. Moreover, theses results show that during co-culture growths strong changes are taking place in the wild-type outer membrane.

Under the growth conditions studied two proteins related to the transport and/or efflux of compound were detected. Interestingly, different type of expression was found in the wild-type or in the mutant strain, suggesting different efflux process in both strains (Chapter E, Table 1, n° 140 and 141).

Finally, two different outer membrane proteins (OprM3 and OmpW), involved in drug or antibiotic resistance were up-regulated (Chapter E, table 1). The OprM3 protein, which is a probable outer membrane efflux protein, was detected as novel in co-culture wild-type cells. Both proteins were reported as involved in intrinsic multiple antibiotic resistance (Hancock and Brinkman, 2002; Xu *et al.*, 2006). The induction of these proteins in co-culture is most probably related to the resistance against antibiotics synthesized by *A. alternatum* in co-culture. To possess this set of antibiotic resistance proteins might allow strain BH72 to establish a compatible interaction with *A. alternatum*.

2.9. Motility and adhesion

Bacterial flagella play a crucial role in motility, adhesion, biofilm formation and colonisation of the host. *Azoarcus* sp. strain BH72 is highly motile by means of a polar flagellum. In the

genome at least 48 genes were found that are generally required for biosynthesis and function of flagella and chemotaxis (Chapter C, Figure 3a). Interesting is the presence of several encoding genes for flagellins (*fliC1, fliC2, fliC3*) and flagellar motor proteins (two), suggesting an important role of motility for the plant-associated lifestyle. However it is not clear yet which one is the major flagellin. Two flagellins copies (FliC2 and FliC3), were identified in the membrane fraction under both N_2-fixation conditions suggesting a quite unusual role for both proteins.

On the other hand, a putative PilP protein which is essential for the biogenesis of type IV pili system (Filloux, 2004), was identified as novel in the mutant strain. As strain BH1599 is a pilus mutant, maybe the overexpression of this protein is because no PilA is present to be assembly.

3.0. Diazosome related proteins

As mentioned, unusual internal membrane stacks (diazosomes), are formed in *Azoarcus* sp. BH72 during co-culture with *A. alternatum* and at low O_2 concentration (Hurek *et al.*, 1995). The approach used in the last part of this work allows studying specific proteins which are expressed only in diazosome forming cells which could most probably be involved in the diazosome structure or function. In this sense several membrane and membrane associated proteins were identified up-regulated in the wild-type strain (Chapter E, Table1). From this set two of them (Chapter E Table 1, n°147 and 148) were annotated as conserved hypothetical membrane proteins. In both proteins sequences, a transmembrane motif was found. In particular, the protein azo3667 was also present in the mutant BH1599 strain, but significantly in less amount (wt/mut: 0.59± 0.29/0.2±0.17). Additionally proteins probably located in these specialized structures are the mentioned NifH and CycH proteins involved in the N_2-fixation and respiration process. Moreover, proteins like Pnp and PrlC were detected as well up-regulated in the membrane fraction of the wild-type cells (Chapter E, table1). These proteins are involved in RNA processing and messenger RNA degradation, as well as in oligopeptide cleavage, respectively. Results obtained are in correlation with the presence of diazosomes in strain BH72 co-culture cells, where supposedly a process such as N_2-fixation, respiration and ATP-synthesis are very active.

3. Concluding remarks

In the present thesis the genomic sequence information of the N_2-fixing bacterium *Azoarcus* sp. BH72 was the starting point for studies. Part of the *Azoarcus* sp. BH72 genome sequence was annotated and further genome comparisons were done revealing several highlights of endophyte life style, which can be used as a starting point in functional genomics studies. Availability of complete genomes sequences has now made it possible to analyze and monitor multiple changes

in gene expression patterns under a variety of conditions. However, gene expression alone does not reveal the biochemical function associated with its gene (Pandey and Mann, 2000). Proteomic analysis using high-resolution 2D-PAGE provides a unique method to monitor global changes of protein expression pattern occurring in a tissue or an organism. Recently studies using this technology revealed a large amount of information on different aspects of the plant-microbe interaction, including bacterial pathways reconstruction in pure culture growths or in simbiotic association, protein expression of plant root in respones the bacteria partner, as well as the identification of protein putative involved in the bacteria-plant interacion (Rolfe et al., 2003, Bestel-Corre et al 2004). In this work for the first time a proteomic functional genomic analysis in *Azoarcus* sp. BH72 grown under different N_2-fixing conditions, was successfully conduced. Using this approach, main pathways for carbon and nitrogen metabolism were reconstructed in both conditions studied. Additionally, the biological model used allowed the possibility to study proteins involved in diazosome structure and function. These novel internal stack membranes formed in co-culture conditions were propose to be involved in the respiration and nitrogen-fixation process (Hurek *et al.*, 1995). This hypothesis was demonstrated by the identification of several up-regulated and membrane associated proteins involved in these process. Moreover, membrane proteins probably involved in diazosomes structure were also identified. Results obtained suggests that these specialized membranes give a suitable support for a coordinate and effective respiration and ATP production process, as well as for an effective electron transfer to the nitrogenase protein. Since the nitrogenase complex is very sensitive to oxygen, to posses these specialized structures give to strain BH72 an advantage in their natural environment for an effective nitrogen-fixation process.

Results obtained in this thesis show how the proteomic approach successfully applied can complement data obtained form the genome sequence. Additionally, this analysis also represents the basis for additional functional proteomic analysis of *Azoarcus* sp. BH72 and provides a complementary strategy for other high-throughput global analysis approaches. Moreover, the conditions-specific proteins identified, as well as the exclusive genes detected in strain BH72 genome, provide new targets for genome-wide mutagenesis with the aim to understand its role in *Azoarcus*-fungus and plant interaction.

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yell, M.D., Zhang, Q., Chen, L.X., Bron, R.C., Rogers, Y.H., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L., Abril, J.F., Agbayani, A., An, H.J., rews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chra, I., Cherry, J.M., Cawley, S., Dahlke, C.,

Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissenbach, J., Williams, S.M., WoodageT, Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M., and Venter, J.C. (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185-2195.

Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**: 198-207.

Aguena, M., Yagil, E., and Spira, B. (2002) Transcriptional analysis of the pst operon of *Escherichia coli. Mol Genet. Genomics* **268**: 518-524.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Anders, H.J., Kaetzke, A., Kämpfer, P., Ludwig, W., and Fuchs, G. (1995) Taxonomic position of aromatic-degrading denitrifying pseudomonad strains K 172 and KB 740 and their description as new members of the genera *Thauera*, as *Thauera aromatica* sp. nov., and *Azoarcus*, as *Azoarcus evansii* sp. nov., respectively, members of the beta subclass of the *Proteobacteria. Int J Systc Bacteriol* **45**: 327-333.

Andersen, J.B., Heydorn, A., Hentzer, M., Eberl, L., Geisenberger, O., Christensen, B.B., Molin, S., and Givskov, M. (2001) *gfp*-based *N*-acyl homoserine-lactone sensor systems for detection of bacterial communication. *Appl Environ Microbiol* **67**: 575-585.

Anderson, N.G., and Anderson, N.L. (1996) Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* **17**: 443-453.

Andrews, S.C., Robinson, A.K., and Rodriguez-Quinones, F. (2003) Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**: 215-237.

Arnow, L.E. (1937) Colorimetric determination of the components of 3,4-dihydroxyphenylalanine-tyrosine mixtures. *J.Biol. Chem.* **118**: 531-537.

Arumuganathan, K.a., and Earle, E.D. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.*: 208-218.

Atkin, C.L., Neilands, J.B., and Phaff, H.J. (1970) Rhodotorulic acid from species of *Leucosporidium, Rhodosporidium, Rhodotorula, Sporidiobolus, and Sporobolomyces*, and a new alanine-containing ferrichrome from Cryptococcus melibiosum. *J Bacteriol* **103**: 722-733.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., and Yeh, L.S. (2005) The Universal Protein Resource (UniProt). *Nucleic. Acids Res.* **33**: D154-159.

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

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic. Acids Res.* **30**: 276-280.

Battistoni, F., Bartels, D., Kaiser, O., Reamon-Buettner, S.M., Hurek, T., and Reinhold-Hurek, B. (2005) Pysical map of the *Azoarcus* sp. strain BH72 genome based on a bacterial artificial chromosome (BAC) library as a platform for genome sequencing and functional analysis. *FEMS Microbiol. Lett.* **249**: 233-240.

Behrens, S., Maier, R., de Cock, H., Schmid, F.X., and Gross, C.A. (2001) The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *Embo J* **20**: 285-294.

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T., Hyman, L.J., Holeva, M.C., Thomson, N.R., Bentley, S.D., Churcher, L.J., Mungall, K., Atkin, R., Bason, N., Brooks, K., Chillingworth, T., Clark, K., Doggett, J., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Norbertczak, H., Ormond, D., Price, C., Quail, M.A., Sanders, M., Walker, D., Whitehead, S., Salmond, G.P., Birch, P.R., Parkhill, J., and Toth, I.K. (2004) Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. atroseptica and characterization of virulence factors. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 11105-11110.

Bernal, A., Ear, U., and Kyrpides, N. (2001) Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res* **29**: 126-127.

Bestel-Corre, G., Dumas-Gaudot, E., Poinsot, V., Dieu, M., Dierick, J.F., van, T.D., Remacle, J., Gianinazzi-Pearson, V., and Gianinazzi, S. (2002) Proteome analysis and identification of symbiosis-related proteins from *Medicago truncatula* Gaertn. by two-dimensional electrophoresis and mass spectrometry. *Electrophoresis* **23**: 122-137.

Bestel-Corre, G., Dumas-Gaudot, E., and Gianinazzi, S. (2004a) Proteomics as a tool to monitor plant-microbe endosymbioses in the rhizosphere. *Mycorrhiza* **14**: 1-10.

Bieber, D., Ramer, S.W., Wu, C.Y., Murray, W.J., Tobe, T., Fernandez, R.a., and Schoolnik, G.K. (1998) Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. *Science* **280**: 2114-2118.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1474.

Braun, M., and Silhavy, T.J. (2002) Imp/OstA is required for cell envelope biogenesis in *Escherichia coli*. *Mol Microbiol* **45**: 1289-1302.

Braun, V., Mahren, S., and Ogierman, M. (2003) Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Curr Opin. Microbiol* **6**: 173-180.

Brito, B., Aldon, D., Barberis, P., Boucher, C., and Genin, S. (2002) A signal transfer system through three compartments transduces the plant cell contact-dependent signal controlling *Ralstonia solanacearum hrp* genes. *Mol Plant Microbe Interact* **15**: 109-119.

Brodin, P., Eiglmeier, K., Marmiesse, M., Billault, A., Garnier, T., Niemann, S., Cole, S.T., and Brosch, R. (2002) Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Imm* **70**: 5568-5578.

Brosch, R., Gordon, S.V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B.G., and Cole, S.T. (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect Imm* **66**: 2221-2229.

Bukau, B., and Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92**: 351-366.

Busch, W., and Saier, M.H., Jr. (2002) The transporter classification (TC) system, 2002. *Crit Rev Biochem Mol Biol* **37**: 287-337.

Carson, K.C., Meyer, J.M., and, and Dilworth, M.J. (2000) Hydroxamate siderophores of root nodule bacteria. *Soil Biol Biochem* **32**: 11-21.

Carter, R.A., Worsley, P.S., Sawers, G., Challis, G.L., Dilworth, M.J., Carson, K.C., Lawrence, J.A., Wexler, M., Johnston, A.W., and Yeoman, K.H. (2002) The *vbs* genes that direct synthesis of the siderophore vicibactin in *Rhizobium leguminosarum*: their expression in other genera requires ECF sigma factor RpoI. *Mol Microbiol* **44**: 1153-1166.

Cascales, E., Bernadac, A., Gavioli, M., Lazzaroni, J.C., and Lloubes, R. (2002) Pal lipoprotein of *Escherichia coli* plays a major role in outer membrane integrity. *J Bacteriol* **184**: 754-759.

Celestino, P.B., de Carvalho, L.R., de Freitas, L.M., Dorella, F.A., Martins, N.F., Pacheco, L.G., Miyoshi, A., and Azevedo, V. (2004) Update of microbial genome programs for bacteria and archaea. *Genet Mol Res* **3**: 421-431.

Chao, T.C., Becker, A., Buhrmester, J., Puhler, A., and Weidner, S. (2004) The *Sinorhizobium meliloti* fur gene regulates, with dependence on Mn(II), transcription of the *sitABCD* operon, encoding a metal-type transporter. *J Bacteriol* **186**: 3609-3620.

Chhabra, S.R., He, Q., Huang, K.H., Gaucher, S.P., Alm, E.J., He, Z., Hadi, M.Z., Hazen, T.C., Wall, J.D., Zhou, J., Arkin, A.P., and Singh, A.K. (2006) Global analysis of heat shock response in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* **188**: 1817-1828.

Clarke, T.E., Tari, L.W., and Vogel, H.J. (2001) Structural biology of bacterial iron uptake systems. *Curr Top Med Chem* **1**: 7-30.

Colebatch, G., Trevaskis, B., and, and Udvardi, M. (2002a) Symbiotic nitrogen fixation research in the postgenomics era. *New Phytologist* **153**: 37-42.

Colebatch, G., Trevaskis, B.a., and Udvardi, M. (2002b) Functional genomics: tools of the trade. *New Phytologist* **153**: 27-36.

Crosa, J.H., and Walsh, C.T. (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **66**: 223-249.

Crowley, D.E., Wang, Y.C., Reid, C.P.P., and, and Szaniszlo, P.J. (1991) Mechanisms of iron acquisition from siderophores by microorganisms and plants. *Plant Soil* **130**: 179-198.

Davidson, A.L., and Chen, J. (2004) ATP-binding cassette transporters in bacteria. *Annu Rev Biochem* **73**: 241-268.

Dawes, E.A., and Senior, P.J. (1973) The role and regulation of energy reserve polymers in micro-organisms. *Adv Microb Physiol* **10**: 135-266.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999) Improved microbial gene identification with GLIMMER. *Nu Acids Res* **27**: 4636-4641.

Diaz-Mireles, E., Wexler, M., Sawers, G., Bellini, D., Todd, J.D., and Johnston, A.W. (2004) The Fur-like protein Mur of *Rhizobium leguminosarum* is a Mn(2+)-responsive transcriptional regulator. *Microbiology* **150**: 1447-1456.

Djordjevic, M.A., Chen, H.C., Natera, S., Van Noorden, G., Menzel, C., Taylor, S., Renard, C., Geiger, O.a., and Weiller, G.F. (2003) A global analysis of protein expression profiles in *Sinorhizobium meliloti:* discovery of new genes for nodule occupancy and stress adaptation. *Mol Plant Microbe Interact* **16**: 508-524.

Djordjevic, M.A. (2004) *Sinorhizobium meliloti* metabolism in the root nodule: a proteomic perspective. *Proteomics* **4**: 1859-1872.

Dörr, J., Hurek, T.a., and Reinhold-Hurek, B. (1998) Type IV pili are involved in plant-microbe and fungus-microbe interactions. *Mol Microbiol* **30**: 7-17.

Egener, T., Hurek, T.a., and 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.

Egener, T., Hurek, T.a., and Reinhold-Hurek, B. (1999) Endophytic expression of *nif* genes of *Azoarcus* sp. strain BH72 in rice roots. *Mol Plant Microbe Interact* **12**: 813-819.

Egener, T., Martin, D.E., Sarkar, A., and Reinhold-Hurek, B. (2001) Role of a ferrodoxine gene cotranscribed with the *nifHDK* operon in N_2 fixation and nitrogenase "switch off" of *Azoarcus* sp. strain BH72. *J Bacteriol* **183**: 3752-3760.

Egener, T., Sarkar, A., Martin, D.E., and Reinhold-Hurek, B. (2002) Identification of a NifL-like protein in a diazotroph of the beta-subgroup of the *Proteobacteria*, *Azoarcus* sp. strain BH72. *Microbiology* **148**: 3203-3212.

Eichacker, L.A., Granvogl, B., Mirus, O., Muller, B.C., Miess, C., and Schleiff, E. (2004) Hiding behind hydrophobicity. Transmembrane segments in mass spectrometry. *J Biol Chem* **279**: 50915-50922.

Escolar, L., Perez-Martin, J., and de Lorenzo, V. (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**: 6223-6229.

Expert, D. (1999) Witholding and exchange iron: Interactions Between *Erwinia* spp. and Their Plant Hosts. *Annu Rev Phytopathol* **37**: 307-334.

Filloux, A. (2004) The underlying mechanisms of type II protein secretion. *Biochim Biophys Acta* **1694**: 163-179.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., and et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.

Forward, J.A., Behrendt, M.C., Wyborn, N.R., Cross, R., and Kelly, D.J. (1997) TRAP transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J Bacteriol* **179**: 5482-5493.

Franza, T., Michaud-Soret, I., Piquerel, P., and Expert, D. (2002) Coupling of iron assimilation and pectinolysis in *Erwinia chrysanthemi* 3937. *Mol Plant Microbe Interact* **15**: 1181-1191.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J.L., Nguyen, D.T., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., and Venter, J.C. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397-403.

Fraser, C.M., Eisen, J.A., and Salzberg, S.L. (2000) Microbial genome sequencing. *Nature* **406**: 799-803.

Frijters, A.C.J., Zhan, Z., van Damme, M., Wang, G.L., Ronald, P.C.a., and Michelmore, R.W. (1997) Construction of a bacterial artificial chromosome library containing large *Eco*RI and *Hin*D III genomic fragments of lettuce. *Theo Appl Genet* **94**: 390-399.

Fuhrmann, D.R., Krzywinski, M.I., Chiu, R., Saeedi, P., Schein, J.E., Bosdet, I.E., Chinwalla, A., Hillier, L.W., Waterston, R.H., McPherson, J.D., Jones, S.J., and Marra, M.A. (2003) Software for automated analysis of DNA fingerprinting gels. *Genome Res* **13**: 940-953.

Gadda, G.a., and McAllister-Wilkins., E.E. (2003) Cloning, expression and purification of choline deydrogenase from the moderate halophile *Halomonas elongata*. *Appl Environ Microbiol* **69**: 2126-2132.

Gäher, M., Einsiedler, K., Crass, T., and Bautsch, W. (1996) A physical and genetic map of *Neisseria meningitidis* B1940. *Mol Microbiol* **19**: 249-259.

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

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**: 3784-3788.

Gevaert, K., Demol, H., Martens, L., Hoorelbeke, B., Puype, M., Goethals, M., Van Damme, J., De Boeck, S., and Vandekerckhove, J. (2001) Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry. *Electrophoresis* **22**: 1645-1651.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**: 92-100.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S.G. (1996) Life with 6000 genes. *Science* **274**: 563-567.

Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B.S., Cao, Y., Askenazi, M., Halling, C., Mullin, L., Houmiel, K., Gordon, J., Vaudin, M., Iartchouk, O., Epp, A., Liu, F., Wollam, C., Allinger, M., Doughty, D., Scott, C., Lappas, C., Markelz, B., Flanagan, C., Crowell, C., Gurson, J., Lomo, C., Sear, C., Strub, G., Cielo, C., and Slater, S. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**: 2323-2328.

Gosset, G. (2005) Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microb Cell Fact.* **4:** 14.

Guerreiro, N., Djordjevic, M.A., and Rolfe, B.G. (1999) Proteome analysis of the model microsymbiont *Sinorhizobium meliloti*: Isolation and characterisation of novel proteins. *Electrophoresis* **20**: 818-825.

Hamblin, M.J., Shaw, J.G., and Kelly, D.J. (1993) Sequence analysis and interposon mutagenesis of a sensor-kinase (DctS) and response-regulator (DctR) controlling synthesis of the high-affinity C4-dicarboxylate transport system in *Rhodobacter capsulatus*. *Mol Gen Genet* 237: 215-224.

Hancock, R.E., and Brinkman, F.S. (2002) Function of pseudomonas porins in uptake and efflux. *Annu Rev Microbiol* **56**: 17-38.

Hansen, A.M., Qiu, Y., Yeh, N., Blattner, F.R., Durfee, T., and Jin, D.J. (2005) SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli. Mol Microbiol* **56**: 719-734.

Harder, J. (1997) Anaerobic degradation of cyclohexane-1,2-diol by a new *Azoarcus* species. *Arch Microbiol* **168**: 199-204.

Harland, D.N., Garmory, H.S., Brown, K.A., and Titball, R.W. (2005) An association between ATP binding cassette systems, genome sizes and lifestyles of bacteria. *Res Microbiol* **156**: 434-442.

Harris, R.M., Webb, D.C., Howitt, S.M., and Cox, G.B. (2001) Characterization of PitA and PitB from *Escherichia coli. J Bacteriol* **183**: 5008-5014.

Heil, G., Stauffer, L.T., and Stauffer, G.V. (2002) Glycine binds the transcriptional accessory protein GcvR to disrupt a GcvA/GcvR interaction and allow GcvA-mediated activation of the *Escherichia coli gcvTHP* operon. *Microbiology* **148**: 2203-2214.

Hellman, U. (2000) *Sample preparation by SDS-PAGE and in-gel digestion.* Basel, Switzerland.: Birkhäuser Verlag.

Henzel, W.J., Billeci, T.M., Stults, J.T., and , and Wong, S.C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of pepride fragments in protein sequences databases. *Proc. Natl. Acad. Sci. U S A* **90**: 5011-5015.

Herbert, B.R., Harry, J.L., Packer, N.H., Gooley, A.A., Pedersen, S.K., and Williams, K.L. (2001) What place for polyacrylamide in proteomics? *Trends Biotechnol* **19**: S3-9.

Herrler, M., Bang, H., and Marahiel, M.A. (1994) Cloning and characterization of ppiB, a *Bacillus subtilis* gene which encodes a cyclosporin A-sensitive peptidyl-prolyl cis-trans isomerase. *Mol Microbiol* **11**: 1073-1083.

Higgins, C.F. (2001) ABC transporters: physiology, structure and mechanism--an overview. *Res. Microbiol.* **152**: 205-210.

Holland, I.B., and Blight, M.A. (1999) ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol* **293**: 381-399.

Hong, G. (1997) A rapid and accurate strategy for rice contig map construction by combination of fingerprinting and hybridization. *Plant Mol Biol* **35**: 129-133.

Hoskins, R.A., Nelson, C.R., Berman, B.P., Laverty, T.R., George, R.A., Ciesiolka, L., Naeemuddin, M., Arenson, A.D., Durbin, J., David, R.G., Tabor, P.E., Bailey, M.R., DeShazo, D.R., Catanese, J., Mammoser, A., Osoegawa, K., de Jong, P.J., Celniker, S.E., Gibbs, R.A., Rubin, G.M., and Scherer, S.E. (2000) A BAC-based physical map of the major autosomes of *Drosophila melanogaster. Science* **287**: 2271-22714.

Hurek, T., Reinhold, B., Fendrik, I.a., and Niemann, E.G. (1987a) Root-zone-specific oxygen tolerance of *Azospirillum* spp. and diazotrophic rods closely associated with Kallar grass. *Appl Environ Microbiol* **53**: 163-169.

Hurek, T., Reinhold, B., and Niemann, E.G. (1987b) Effect of oxygen on ammonium-grown continuous cultures of *Azospirillum*-spp. and diazotrophic rods closely associated with Kallar grass. *Canadian J Microbiol* **33**: 919-922.

Hurek, T., Reinhold-Hurek, B., Van Montagu, M., and Kellenberger, E. (1991) Infection of intact roots of Kallar grass and rice seedlings by *Azoarcus*. In *Nitrogen fixation*. Polsinelli, M., Materassi, R. and Vincenzini, M. (eds). Dordrecht, Netherlands: Kluwer academic publishers, pp. 235-242.

Hurek, T., Burggraf, S., Woese, C.R., and 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., Reinhold-Hurek, B., Turner, G.L., and Bergersen, F.J. (1994a) Augmented rates of respiration and efficient nitrogen fixation at nanomolar concentrations of dissolved O_2 in hyperinduced *Azoarcus* sp. strain BH72. *J Bacteriol* **176**: 4726-4733.

Hurek, T., Reinhold-Hurek, B., Van Montagu, M.a., and Kellenberger, E. (1994b) Root colonization and systemic spreading of *Azoarcus* sp. strain BH72 in grasses. *J Bacteriol* **176**: 1913-1923.

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

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

Hurek, T., Wagner, B.a., and Reinhold-Hurek, B. (1997b) Identification of N_2-fixing plant- and fungus-associated *Azoarcus* species by PCR-based genomic fingerprints. *Appl Environ Microbiol* **63**: 4331-4339.

Hurek, T., and Reinhold-Hurek, B. (1999) Interactions of *Azoarcus* sp. with rhizosphere fungi. In *Mycorrhiza*. Varma, A. and Hock, B. (eds). Berlin: Springer Verlag, pp. 595-614.

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

Hurek, T., and Reinhold-Hurek, B. (2003) *Azoarcus* sp. strain BH72 as a model for nitrogen-fixing grass endophytes. *J Biotech* **106**: 169-178.

Hurek, T.a., and Reinhold-Hurek, B. (1995) Identification of grass-associated and toluene-degrading diazotrophs, *Azoarcus* spp., by analyses of partial 16S ribosomal DNA sequences. *Appl Environ Microbiol* **61**: 2257-2261.

James, E.K.a., and Olivares, F.L. (1998) Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. *Crit Rev Plant Sci* **17**: 77-119.

Janausch, I.G., Zientz, E., Tran, Q.H., Kroger, A., and Unden, G. (2002) C4-dicarboxylate carriers and sensors in bacteria. *Biochim Biophys Acta* **1553**: 39-56.

Jurkevitch, E., Hadar, Y., Chen, Y., Chino, M., and Mori, S. (1993) Indirect utilization of the phytosiderophore mugineic acid as an iron source to rhizosphere fluorescent *Pseudomonas. Biometals* **6**: 119-123.

Kagi, J.H., and Vallee, B.L. (1960) The role of zinc in alcohol dehydrogenase. V. The effect of metal-binding agents on thestructure of the yeast alcohol dehydrogenase molecule. *J Biol Chem* **235**: 3188-3192.

Kaiser, O., Bartels, D., Bekel, T., Goesmann, A., Kespohl, S., Puhler, A., and Meyer, F. (2003) Whole genome shotgun sequencing guided by bioinformatics pipelines--an optimized approach for an established technique. *J Biotechnol* **106**: 121-133.

Kammler, M., Schon, C., and Hantke, K. (1993) Characterization of the ferrous iron uptake system of *Escherichia coli. J Bacteriol* **175**: 6212-6219.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32**: D277-280.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* **7**: 331-338.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., and Tabata, S. (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* **9**: 189-197.

Karg, T., and Reinhold-Hurek, B. (1996) Global changes in protein composition of N_2-fixing *Azoarcus* sp. strain BH72 upon diazosome formation. *J Bacteriol* **178**: 5748-5754.

Kelly, D.J., and Thomas, G.H. (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol Rev* **25**: 405-424.

Khan, M.D. (1966) `Kallar grass`, a suitable grass for saline lands. *Agriculture Pakistan* **17**: 375.

Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., and Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**: 213-218.

Klappenbach, J.A., Dunbar, J.M., and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* **66**: 1328-1333.

Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R., and Venter, J.C. (1997) The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**: 364-370.

Koebnik, R. (2005) TonB-dependent trans-envelope signalling: the exception or the rule? *Trends Microbiol* **13**: 343-347.

Koonin, E.V., Mushegian, A.R., and Rudd, K.E. (1996) Sequencing and analysis of bacterial genomes. *Curr Biol* **6**: 404-416.

Koonin, E.V. (1997) Genome sequences: genome sequence of a model prokaryote. *Curr Biol* **7**: 656-659.

Koonin, E.V., and Galperin, M.Y. (1997) Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr Opin Genet Dev* **7**: 757-763.

Koster, W. (2001) ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Res Microbiol* **152**: 291-301.

Krause, A., Ramakumar, A., Bartels, D., Battistoni, F., Bekel, T., Boch, J., Böhm, M., Friedrich, F., Hurek, T., Krause, L., Linke, B., McHardy, A., Sarkar, A., Schneiker, S., Syed-Ali, A., Thauer, R., Vorhölter, F., Weidner, S., Pühler, A., Reinhold-Hurek, B., and, O.K., and Goesmann., A. (2006) Life inside cereals: Genomic insights into the lifestyle of the mutualistic, N_2-fixing endophyte *Azoarcus* sp. strain BH72. *Nat Biotech in press.*

Kurata, N., Umehara, Y., Tanoue, H., and Sasaki, T. (1997) Physical mapping of the rice genome with YAC clones. *Plant Mol Biol* **35**: 101-113.

Kuster, B., and Mann, M. (1998) Identifying proteins and post-translational modifications by mass spectrometry. *Curr Opin Struct Biol* **8**: 393-400.

Kustu, S.G., McFarland, N.C., Hui, S.P., Esmon, B., and Ames, G.F. (1979) Nitrogen control of *Salmonella typhimurium*: co-regulation of synthesis of glutamine synthetase and amino acid transport systems. *J Bacteriol* **138**: 218-234.

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

Lai, E.M., Nair, U., Phadke, N.D.a., and Maddock, J.R. (2004) Proteomic screening and identification of differentially distributed membrane proteins in *Escherichia coli*. *Mol Microbiol* **52**: 1029-1044.

Lane, D.J. (1991) 16S / 23S rRNA sequencing. In *Nucleic acid techniques in bacterial systematics*. Vol. 115-175. Stackebrandt, E. and Goodfellow, M. (eds). Chichester: Wiley.

Lay, O.J.J. (2001) MALDI-TOF mass spectrometry of bacteria. *Mass Spectrom Rev* **20**: 172-194.

Llamas, M.A., Sparrius, M., Kloet, R., Jimenez, C.R., Vandenbroucke-Grauls, C., and Bitter, W. (2006) The heterologous siderophores ferrioxamine B and ferrichrome activate signaling pathways in *Pseudomonas aeruginosa*. *J Bacteriol* **188**: 1882-1891.

Lombardot, T. (2004) The genome of the free-living, marine *Plantomycete Pirellula* sp.strain 1 ("*Rhodopirellula baltica")*: Bioinformatics and Biology. In *Max Planck for Marine Microbiology. Molecular Ecology*. Bremen: Bremen, pp. 154.

Loper, J.E., and Henkels, M.D. (1999) Utilization of heterologous siderophores enhances levels of iron available to *Pseudomonas putida* in the rhizosphere. *Appl Environ Microbiol* **65**: 5357-5363.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955-964.

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

Malik, K.A., Zafar, Y., and Hussain, A. (1980) Nitrogenase activity in the rhizosphere of Kallar grass (*Diplachne fusca* (Linn.) Beauv.). *Biologica* **26**: 107-111.

Mann, M., Hendrickson, R.C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **70**: 437-473.

Mann, M., and Jensen, O.N. (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**: 255-261.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., and

Bryant, S.H. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33**: D192-196.

Marenda, M., Brito, B., Callard, D., Genin, S., Barberis, P., Boucher, C., and Arlat, M. (1998) PrhA controls a novel regulatory pathway required for the specific induction of *Ralstonia solanacearum hrp* genes in the presence of plant cells. *Mol Microbiol* **27**: 437-453.

Marouga, R., David, S., and Hawkins, E. (2005) The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal Bioanal Chem* **382**: 669-678.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., and Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res* **7**: 1072-1084.

Martin, D., Hurek, T., and Reinhold-Hurek, B. (2000) Occurrence of three P_{II}-like signal transmitter proteins in the diazotroph *Azoarcus* sp. BH72. *Mol Microbiol* **38**: 276-288.

Martin, D.E.a., and Reinhold-Hurek, B. (2002) Distinct roles of P_{II}-like signal transmitter protein 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.

Martinez, J.L., Herrero, M., and de Lorenzo, V. (1994) The organization of intercistronic regions of the aerobactin operon of pColV-K30 may account for the differential expression of the *iucABCD iutA* genes. *J Mol Biol* **238**: 288-293.

Marugg, J.D., van Spanje, M., Hoekstra, W.P., Schippers, B.a., and Weisbeek, P.J. (1985) Isolation and analysis of genes involved in siderophore biosynthesis in plant-growth-stimulating *Pseudomonas putida* WCS358. *J Bacteriol* **164**: 563-570.

Massari, P., Ram, S., Macleod, H., and Wetzler, L.M. (2003) The role of porins in neisserial pathogenesis and immunity. *Trends Microbiol* **11**: 87-93.

Mathesius, U., Keijzers, G., Natera, S.H., Weinman, J.J., Djordjevic, M.A., and Rolfe, B.G. (2001) Establishment of a root proteome reference map for the model legume **Medicago truncatula** using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics* **1**: 1424-1440.

Mathesius, U., Imin, N., Chen, H., Djordjevic, M.A., Weinman, J.J., Natera, S.H., Morris, A.C., Kerim, T., Paul, S., Menzel, C., Weiller, G.F., and Rolfe, B.G. (2002) Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. *Proteomics* **2:** 1288-1303.

McHugh, J.P., Rodriguez-Quinones, F., Abdul-Tehrani, H., Svistunenko, D.A., Poole, R.K., Cooper, C.E., and Andrews, S.C. (2003) Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis. *J Bio. Chem* **278**: 29478-29486.

Meesters, T.M. (1987) Localization of nitrogenase in vesicles of *Franckia* sp. Cc1.17 by inmunogoldlabbelling on ultrathin cryosections. *Arch Microbiol* **146**: 327-331.

Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., and Koornneef, M. (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* **282**: 679-682.

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O., Giegerich, R., and Puhler, A. (2003) GenDB-an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**: 2187-2195.

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

Moens, S., and Vanderleyden, J. (1996) Functions of bacterial flagella. *Crit Rev Microbiol* **22**: 67-100.

Moller, S., Croning, M.D., and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**: 646-653.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R.R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S.E., Pagni, M., Peyruc, D., Ponting, C.P., Selengut, J.D., Servant, F., Sigrist, C.J., Vaughan, R., and Zdobnov, E.M. (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res* **31**: 315-318.

Munchbach, M., Dainese, P., Staudenmann, W., Narberhaus, F., and James, P. (1999) Proteome analysis of heat shock protein expression in *Bradyrhizobium japonicum*. *Eur J Biochem* **264**: 39-48.

Nandi, B., Nandy, R.K., Sarkar, A., and Ghose, A.C. (2005) Structural features, properties and regulation of the outer-membrane protein W (OmpW) of *Vibrio cholerae*. *Microbiology 151*: 2975-2986.

Natera, S.H., Guerreiro, N., and Djordjevic, M.A. (2000) Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. *Mol Plant Microbe Interact* **13**: 995-1009.

Neilands, J.B. (1981) Microbial iron compounds. *Annu. Rev. Biochem.* **50**: 715-731.

Neilands, J.B. (1995) Siderophores: structure and function of microbial iron transport compounds. *J Biol Chem* **270**: 26723-26726.

Neumann, S., Wynen, A., Truper, H.G., and Dahl, C. (2000) Characterization of the *cys* gene locus from *Allochromatium vinosum* indicates an unusual sulfate assimilation pathway. *Mol Biol Rep* **27**: 27-33.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* **8**: 581-599.

Nikaido, H. (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67**: 593-656.

Ninfa, A.J., Atkinson, M.R., Kamberov, E.S., Feng, J., and Ninfa, E.G. (1995) Control of bacterial nitrogen assimilation by the NRI-NRII two component system of enteric bacteria. In *Two Component Systems of Bacteria.* Silhavy, T.J. and Hoch, J.A. (eds). Washington D.C.: American Society for Microbiology, pp. 67-88.

Nohno, T., and Saito, T. (1987) Two transcriptional start sites found in the promoter region of *Escherichia coli* glutamine permease operon, *glnHPQ*. *Nucleic Acids Res* **15**: 2777.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**: 4007-4021.

Oppenheim, J.a., and Marcus, L. (1970) Correlation of ultrastructure in *Azotobacter vinelandii* with nitrogen source for growth. *J Bacteriol* **101**: 286-291.

Osteras, M., O'Brien, S.A., and Finan, T.M. (1997) Genetic analysis of mutations affecting *pckA* regulation in *Rhizobium (Sinorhizobium) meliloti*. *Genetics* **147**: 1521-1531.

Osteras, M., E. Boncompagni, N. Vincent, M.-C. Poggi, a., and Rudulier, D.L. (1998) Presence of a gene encoding choline sulfatase in *Sinorhizobium meliloti bet* operon: choline-*O*-sulfate is metabolized into glycine betaine. *Proc Natl Acad Sci U S A* **95**: 11394-11399.

Pallen, M.J. (1999) Microbial genomes. *Mol Microbiol* **32**: 907-912.

Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405**: 837-846.

Panter, S., Thomson, R., de Bruxelles, G., Laver, D., Trevaskis, B., and Udvardi, M. (2000) Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. *Mol Plant Microbe Interact.* **13**: 325-333.

Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. (1998) Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**: 1-34.

Parsons, L.M., Lin, F., and Orban, J. (2006) Peptidoglycan recognition by Pal, an outer membrane lipoprotein. *Biochemistry* **45**: 2122-2128.

Pate, J.L., Shaj, V.K., and, and Brill, W.J. (1973) Internal membrane control in *Azotobacter vinelandii*. *J Bacteriol* **114**: 1346-1350.

Patriarca, E.J., Chiurazzi, M., Manco, G., Riccio, A., Lamberti, A., De Paolis, A., Rossi, M., Defez, R., and Iaccarino, M. (1992) Activation of the *Rhizobium leguminosarum glnII* gene by NtrC is dependent on upstream DNA sequences. *Mol Gen Genet* **234**: 337-345.

Paulsen, I.T., Press, C.M., Ravel, J., Kobayashi, D.Y., Myers, G.S., Mavrodi, D.V., DeBoy, R.T., Seshadri, R., Ren, Q., Madupu, R., Dodson, R.J., Durkin, A.S., Brinkac, L.M., Daugherty, S.C., Sullivan, S.A., Rosovitz, M.J., Gwinn, M.L., Zhou, L., Schneider, D.J., Cartinhour, S.W., Nelson, W.C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E.A., Pierson, L.S., 3rd, Thomashow, L.S., and Loper, J.E. (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nature Biotech* **23**: 873-878.

Persuhn, D.C., Souza, E.M., Steffens, M.B., Pedrosa, F.O., Yates, M.G., and Rigo, L.U. (2000) The transcriptional activator NtrC controls the expression and activity of glutamine synthetase in *Herbaspirillum seropedicae*. *FEMS Microbiol Lett* **192**: 217-221.

Platero, R., Peixoto, L., O'Brian, M.R., and Fabiano, E. (2004) Fur is involved in manganese-dependent regulation of mntA (sitA) expression in *Sinorhizobium meliloti*. *Appl Environ Microbiol* **70**: 4349-4355.

Post, E., Golecki, J.R., and , and Oelze, J. (1982) Morphological and ultrastructural variations in *Azotobacter vinelandii* grwoing in O_2-controlled continuous cultured. *Arch. Microbiol* **133**.

Post, E., Vakalopoulou, E., and , and Oelze, J. (1983) On the relationship of intracytoplasmic to cytoplasmic membranes in nitrogen-fixing *Azotobacter vinelandii*. *Arch Microbiol* **134**: 265-269.

Postle, K., and Kadner, R.J. (2003) Touch and go: tying TonB to transport. *Mol Microbiol* **49**: 869-882.

Postma, P.W., Lengeler, J.W., and Jacobson, G.R. (1993) Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57**: 543-594.

Rabus, R., Jack, D.L., Kelly, D.J., and Saier, M.H., Jr. (1999) TRAP transporters: an ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. *Microbiology* **145**: 3431-3445.

Rabus, R., Kube, M., Heider, J., Beck, A., Heitmann, K., Widdel, F.a., and Reinhardt, R. (2005) The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Arch Microbiol* **183**: 27-36.

Rabus, R.a., and Widdel, F. (1995) Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria. *Arch Microbiol* **163**: 96-103.

Ratledge, C., and Dover, L.G. (2000) Iron metabolism in pathogenic bacteria. *Annu Rev Microbiol* **54**: 881-941.

Reinhold, B., Hurek, T., and Fendrik, I. (1985) Strain-specific chemotaxis of *Azospirillum* spp. *Journal of Bacteriology* 162: 190-195.

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

Reinhold-Hurek, B., and Shub, D.A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* **357**: 173-176.

Reinhold-Hurek, B., Hurek, T., Claeyssens, M., and Van Montagu, M. (1993a) Cloning, expression in *Escherichia coli*, and characterization of cellulolytic enzymes of *Azoarcus* sp., a root-invading diazotroph. *JBacteriol* **175**: 7056-7065.

Reinhold-Hurek, B., Hurek, T., Gillis, M., Hoste, B., Vancanneyt, M., Kersters, K.a., and De Ley, J. (1993b) *Azoarcus* gen. nov., nitrogen-fixing proteobacteria associated with roots of Kallar grass (*Leptochloa fusca* (L.) Kunth) and description of two species *Azoarcus indigens* sp. nov. and *Azoarcus communis* sp. nov. *Int J Syst Bacteriol* **43**: 574-584.

Reinhold-Hurek, B., and Hurek, T. (1998a) Interactions of gramineous plants with *Azoarcus* spp. and other diazotrophs: Identification, localization and perspectives to study their function. *Crit Rev Plant Sci* **17**: 29-54.

Reinhold-Hurek, B., Tan, Z., and Hurek, T. (2005) In *Bergey's Manual of Systematic Bacteriology*. Vol. 2. Garrity, G.M. (ed). New York: Springer Verlag, pp. 890-901.

Reinhold-Hurek, B., Maes, T., Gemmer, S., Van Montagu, M., and Hurek, T. (2006) An endoglucanase is involved in infection of rice roots by the not-cellulose-metabolizing endophyte *Azoarcus* sp. strain BH72. *Mol Plant Microbe Interact* **19**: 181-188.

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

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

Rhomberg, T.A., Karlberg, O., Mini, T., Zimny-Arndt, U., Wickenberg, U., Rottgen, M., Jungblut, P.R., Jeno, P., Andersson, S.G.a., and Dehio, C. (2004) Proteomic analysis of the sarcosine-insoluble outer membrane fraction of the bacterial pathogen *Bartonella henselae. Proteomics* **4**: 3021-3033.

Riedel, K., Hentzer, M., Geisenberger, O., Huber, B., Steidle, A., Wu, H., Hoiby, N., Givskov, M., Molin, S., and Eberl, L. (2001) N-acylhomoserine-lactone-mediated communication between *Pseudomonas aeruginosa* and *Burkholderia cepacia* in mixed biofilms. *Microbiology* **147**: 3249-3262.

Robson, L. (1979) Characterization of an O_2-stable nitrogenase complex isolated from *Azotobacter vinelandii*. *Biochm J* **181**: 569-575.

Rolfe, B.G., Mathesius, U., Djordjevic, M., Weinman, J., Hocart, C., Weiller, G., and , and Bauer, W.D. (2003) Proteomic analysis of legume–microbe interactions. *Comp Funct Genom* **4**: 225-228.

Rosen, R., and Ron, E.Z. (2002) Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom Rev* **21**: 244-265.

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

Ruiz-Argüeso, T., Palacios, J.M., and Imperial, a.J. (2001) Regulation of the hydrogenase system in *Rhizobium leguminosarum. Plant and Soil* **230**: 49–57.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.

Sabirova, J.S., Ferrer, M., Regenhardt, D., Timmis, K.N.a., and Golyshin, P.N. (2006) Proteomic insights into metabolic adaptations in *Alcanivorax borkumensis* induced by alkane utilization. *J Bacteriol* **188**: 3763-3773.

Saeki, K. (2004) Electron transport to nitrogenase: diverse routes for a common destination.

Sandhu, G.R., and Malik, K.A. (1975) Plant succession - A key to the utilization of saline soils. *Nucleus* **12**: 35-38.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA-sequencing with chain-terminating inhibitors. *Procs Nat Acad SciUSA* **74**: 5463-5467.

Santoni, V., Rabilloud, T., Doumas, P., Rouquie, D., Mansion, M., Kieffer, S., Garin, J.a., and Rossignol, M. (1999) Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. *Electrophoresis* **20**: 705-711.

Santoni, V., Kieffer, S., Desclaux, D., Masson, F., and Rabilloud, T. (2000a) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis* **21**: 3329-3344.

Santoni, V., Molloy, M., and Rabilloud, T. (2000b) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**: 1054-1070.

Santos, R., Franza, T., Laporte, M.L., Sauvage, C., Touati, D., and Expert, D. (2001a) Essential role of superoxide dismutase on the pathogenicity of *Erwinia chrysanthemi* strain 3937. *Mol Plant Microbe Interact* **14**: 758-767.

Santos, R., Herouart, D., Sigaud, S., Touati, D., and Puppo, A. (2001b) Oxidative burst in alfalfa-*Sinorhizobium meliloti* symbiotic interaction. *Mol Plant Microbe Interact* **14**: 86-89.

Schaffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., and Altschul, S.F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* **29**: 2994-3005.

Schippers, B., Bakker, P.A.a., and Bakker, P. (1987) Interactions of deleterious and beneficial rhizosphere microorganism and effects of cropping practices. *Ann Rev Phyto* **25**: 339-358.

Schmitt, L., and Tampe, R. (2002) Structure and mechanism of ABC transporters. *Curr Opin Struc Biol* **12**: 754-760.

Schultz, J., Milpetz, F., Bork, P.a., and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**: 5857-5864.

Schwyn, B.a., and Neilands, J.B. (1987) Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160**: 47-56.

Sevilla, M., Burris, R.H., Gunapala, N., and 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*⁻ mutant strains. *Mol Plant Microb Int* **14**: 358-366.

Shaw, J.G., Hamblin, M.J., and Kelly, D.J. (1991) Purification, characterization and nucleotide sequence of the periplasmic C4-dicarboxylate-binding protein (DctP) from *Rhodobacter capsulatus. Mol Microbiol* **5**: 3055-3062.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chemy* **68**: 850-858.

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Nat Acad Sci U S A* **89**: 8794-8797.

Shizuya, H., and Kouros-Mehr, H. (2001) The development and applications of the bacterial artificial chromosome cloning system. *The Keio J Med* **50**: 26-30.

Springer, N., Ludwig, W., Philipp, B.a., and Schink, B. (1998) *Azoarcus anaerobius* sp. nov., a resorcinol-degrading, strictly anaerobic, denitifying bacterium. *Int J Syst Bacteriol* **48**: 953-956.

Sun, J., Peng, X., Van Impe, J., and Vanderleyden, J. (2000) The *ntrB* and *ntrC* genes are involved in the regulation of poly-3-hydroxybutyrate biosynthesis by ammonia in *Azospirillum brasilense* Sp7. *Appl Environ Microbiol* **66**: 113-117.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278**: 631-637.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28**: 33-36.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29**: 22-28.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., and Natale, D.A. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**: 41.

Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Buttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., Kirchner, O., Lanz, C., Linke, B., McHardy, A.C., Meyer, F., Mittenhuber, G., Nies, D.H., Niesbach-Klosgen, U., Patschkowski, T., Ruckert, C., Rupp, O., Schneiker, S., Schuster, S.C., Vorholter, F.J., Weber, E., Puhler, A., Bonas, U., Bartels, D., and Kaiser, O. (2005) Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. vesicatoria revealed by the complete genome sequence. *JBacteriol* **187**: 7254-7266.

Toffanin, A., Cadahia, E., Imperial, J., Ruiz-Argueso, T., and Palacios, M. (2002) Characterization of the urease gene cluster from **Rhizobium leguminosarum** bv. viciae. *Arch Microbiol* **177**: 290-298.

Tomb, J.-F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H.G., Glodek, A., McKenney, K., Fitzegerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Karp, P.D., Smith, H.O., Fraser, C.M., and Venter, J.C. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539-547.

Tomii, K., and Kanehisa, M. (1998) A comparative analysis of ABC transporters in complete microbial genomes. *Genome Res* **8**: 1048-1059.

Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J.a., and Wing, R.A. (2001) A marker-dense physical map of the *Bradyrhizobium japonicum* genome. *Genome Research* **11**: 1434-1440.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**: 377-396.

Tseng, T.T., Gratwick, K.S., Kollman, J., Park, D., Nies, D.H., Goffeau, A., and Saier, M.H., Jr. (1999) The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. . *Mol Microbiol Biotechnol* **1**: 107-125.

Van Loon, L.C.a., and Bakker, P.A. (2003) Signalling in rhizobacteria-plant interactions. In *Ecological studies*. Vol. 168. Visser., H.d.K.a.E.J.W. (ed). Heidelberg: Springer-Verlag Berlin Heidelberg., pp. 297-330.

Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E., Menck, C.F., Da Silva, A.C., Ferro, J.A., Oliveira, M.C., Setubal, J.C., Kitajima, J.P., and Simpson, A.J. (2002) Comparative genomic analysis of plant-associated bacteria. *Annu Rev Phytopathol* **40**: 169-189.

Vande Broek, A., Keijers, V.a., and Vanderleyden, J. (1996) Effect of oxygen on the free-living nitrogen fixation activity and expression of the *Azospirillum brasilense nifH* gene in various plant-associated diazotrophs. *Symbiosis* **21**: 25-40.

Vasil, M.L., and Ochsner, U.A. (1999) The response of *Pseudomonas aeruginosa* to iron: genetics, biochemistry and virulence. *Mol Microbiol* **34**: 399-413.

Veinger, L., Diamant, S., Buchner, J., and Goloubinoff, P. (1998) The small heat-shock protein IbpB from *Escherichia coli* stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. *J Biol Chem* **273**: 11032-11037.

Venter, J.C., Smith, H.O., and Hood, L. (1996) A new strategy for genome sequencing. *Nature* **381**: 364-366.

Viguier, C., P, O.C., Clarke, P., and O'Connell, M. (2005) RirA is the iron response regulator of the rhizobactin 1021 biosynthesis and transport genes in *Sinorhizobium meliloti* 2011. *FEMS Microbiol Lett* **246**: 235-242.

Wandersman, C., and Delepelaire, P. (2004) Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol* **58**: 611-647.

Washburn, M.P., and Yates, J.R., 3rd (2000) Analysis of the microbial proteome. *Curr Opin Microbiol* **3**: 292-297.

Weber, J.L., and Myers, E.W. (1997) Human whole-genome shotgun sequencing. *Genome Res* **7**: 401-409.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., and Hochstrasser, D.F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *BioTechnology* **14**: 61-65.

Wilkins, M.R., Gasteiger, E., Gooley, A.A., Herbert, B.R., Molloy, M.P., Binz, P.A., Ou, K., Sanchez, J.C., Bairoch, A., Williams, K.L., and Hochstrasser, D.F. (1999) High-throughput mass spectrometric discovery of protein post-translational modifications. *J Mol Biol* **289**: 645-657.

Winkelmann, G. (2002) Microbial siderophore-mediated transport. *Biochem Soc Trans* **30**: 691-696.

Xu, C., Lin, X., Ren, H., Zhang, Y., Wang, S., and Peng, X. (2006) Analysis of outer membrane proteome of *Escherichia coli* related to resistance to ampicillin and tetracycline. *Proteomics* **6**: 462-473.

Yates, J.R., 3rd (2000) Mass spectrometry. From genomics to proteomics. *Trends Genet* **16**: 5-8.

Yeh, K.C., Peck, M.C., and Long, S.R. (2002) Luteolin and GroESL modulate in vitro activity of NodD. *J Bacteriol* **184**: 525-530.

Yeoman, K.H., Wisniewski-Dye, F., Timony, C., Stevens, J.B., deLuca, N.G., Downie, J.A., and Johnston, A.W. (2000) Analysis of the *Rhizobium leguminosarum* siderophore-uptake gene *fhuA*: differential expression in free-living bacteria and nitrogen-fixing bacteroids and distribution of an *fhuA* pseudogene in different strains. *Microbiology* **146**: 829-837.

Young, J.P., Crossman, L.C., Johnston, A.W., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., Wexler, M., Curson, A.R., Todd, J.D., Poole, P.S., Mauchline, T.H., East, A.K., Quail, M.A., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., Davis, P., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., Whitehead, S., and Parkhill, J. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* **7**: R34.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L., and Yang, H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79-92.

Yurgel, S.N., and Kahn, M.L. (2004) Dicarboxylate transport by rhizobia. *FEMS Microbiol Rev* **28**: 489-501.

Zhang, H.B., and Wing, R.A. (1997) Physical mapping of the rice genome with BACs. *Plant Mol Biol* **35**: 115-127.

Zhang, H.B.a., and Wu, C. (2001) BAC as tools for genome sequencing. *Plant Physiol Biochem* **39**: 195-209.

Zhou, J., Fries, M.R., Chee-Sanford, J.C., and Tiedje, J.M. (1995) Phylogenetic analyses of a new group of denitrifiers capable of anaerobic growth on toluene and description of *Azoarcus tolulyticus* sp. nov. *Int J Syst Bacteriol* **45**: 500-506.