Quorum sensing-dependent expression of small proteins and structural analysis of new class of quorum quenching enzymes

Dissertation

for obtaining the degree Doctor rerum naturalium (Dr. rer. nat)

at the Department of Microbiology and Biotechnology

Subdivision of the Faculty of Mathematics, Informatics and Natural Sciences at the University of Hamburg

by

Katrin Petersen

from Hamburg

Hamburg 2019

The following evaluators recommend the admission of the dissertation:

Prof. Dr. Wolfgang R. Streit

Prof. Dr. Christian Betzel

Date of Defense (Disputation): 30th April 2020

Declaration on oath

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated or listed.

Hamburg, 29. December 2019

_____

Katrin Petersen

English Language Declaration

I hereby declare as a native English speaker that I have checked the thesis “Quorum sensing-dependent expression of small proteins and structural analysis of new class of quorum quenching enzymes” by Katrin Petersen for grammatically correct English and the scientific accuracy of the language. I also confirm that I am a native English speaker.

Sincerely,

_____

Mathew Davis

Publications

The results of this research have been published in the following journals:

Katrin Petersen, Hilke Duin, Christel Vollstedt, Anja Poehlein, Simon Güllert and Wolfgang Streit 2020. **Quorum sensing-dependent expression of small proteins in *Sinorhizobium fredii* NGR234.** *In preparation*

- Planning, realization, data analysis and interpretation of the experiments
- Writing manuscript

Nadine Werner*, **Katrin Petersen***, Laura Fernandez-Lopez, Christel Vollstedt, Markus Pernandt, Manuel Ferrer, Wolfgang R Streit and Christian Betzel (2020). **The *Komagataeibacter europaeus* CECT 8546 GqqA protein resembles the paradigm of a novel type of acylase acting on N-Acyl homoserine lactones.** *In preparation*
*contributing equal authors

- Planning, data analysis and interpretation of MS analysis with cooperation partner
- Recombinant protein expression of GqqA
- Planning and performance of GqqA mutagenesis
- *In vivo* Assay
- Writing manuscript

Vera Bettenworth, Benedikt Steinfeld, Hilke Duin, **Katrin Petersen**, Wolfgang R Streit, Ilka Bischofs and Anke Becker (2019). Mini review. **Phenotypic Heterogeneity in Bacterial Quorum Sensing Systems**. 30.04.2019 in Journal of Molecular Biology. DOI: 10.1016/j.jmb.2019.04.036

- Writing and proof-reading parts of the manuscript

Jessica Grote, Dagmar Krysciak, **Katrin Petersen**, Simon Güllert, Christel Schmeisser, Konrad U Förstner, Hari B Krishnan, Harald Schwalbe, Nina Kubatova an Wolfgang R Streit (2016). **The Absence of the N-Acyl-Homoserine-Lactone Autoinducer Synthase Genes *TraI* and *NgrI* Increases the Copy Number of the Symbiotic Plasmid in *Sinorhizobium fredii* NGR234.** Frontiers in Microbiology Volume **7**: Article 1858. DOI: 10.3389/fmicb.2016.01858

- Planning and performance of the root hair curling assay
- Structure prediction of the small proteins RepX and RepA0
- aa alignment of RepX and RepA0
- Assistance with writing parts of the article
- Adjustments of the article

Table of contents

1. Abstract

Many bacterial processes such as pathogenicity or symbiosis are controlled by an autoinducer (AI) based communication known as quorum sensing (QS). The plant symbiont *Sinorhizobium fredii* NGR234 shows a unique and diverse tool to fix nitrogen in more than 120 plant genera. All genes that are important for establishing the symbiosis are located on the symbiotic plasmid pNGR234*a*. A deletion of both AI synthases in *Sinorhizobium fredii* NGR24 results in an upregulation of nearly all genes on pNGR234*a*, which can be used as a fantastic tool to analyze the expression of small proteins on pNGR234*a* in a QS dependent manner. A combination of an ORF search and mapping these resulting new ORFs to the transcriptomic profile resulted in an identification of 251 additional small ORFs with a size between 33 nts (10 aa) and 183 nts (60 aa). Additionally, an operon containing three smORFs located in-between *traI* and *repA* was identified and showed impact to symbiotic plasmid maintenance. The corresponding protein were designated RepX (51 aa), RepY (57 aa) and RepA0 (143 aa). However, a fourth small protein locating on pNGR234*a* is coded by the smORF NGR_a01725 (78 nts, 25 aa). Mutagenesis approaches, immunoblotting with specific polyclonal antibodies and studies with translation fusion verified the expression of these smORFs to functional proteins and the essential to plasmid replication and stability.

Bacterial communication via QS can be interrupted by inactivating AI molecules that was catalyzed by quorum quenching (QQ) enzymes. The QQ protein, designated GqqA, from *Komagataeibacter europaeus* CECT 8546 shows highly similarity to prephenate dehydratase (PDT) and strongly interferes with N-acyl-homoserine lactone (AHL) QS signals. A previous study demonstrated that GqqA cannot complement an *E. coli* PDT enzyme but affects QS dependent biofilm formation in *K. europaeus* CECT 8546. Here, ESI-MS/MS measurements using a GqqA *in vitro* enzyme assay with 3-oxo-C8-HSL as substrate imply that GqqA cleaves the amide bond and releases a lactone ring and the corresponding acyl acid. Due to the structure similarity to PDT enzymes and the acylase activity GqqA represents the first member of a novel type of acylases involved in bacterial QQ.

Zusammenfassung

Wie Pathogenität oder Symbiose werden viele bakterielle Prozesse durch eine chemisch basierte Kommunikation gesteuert, die als Quorum Sensing (QS) bekannt ist. So ist der Pflanzensymbiont *Sinorhizobium fredii* NGR234 in der Lage in einer QS gesteuerten Symbiose in mehr als 120 Pflanzenarten Stickstoff zu fixieren. Alle Gene, die in der Stickstofffixierung involviert sind, liegen auf dem symbiotischen Plasmid pNGR234a. Da durch eine Deletion beider AI Synthasen, sämtliche Gene auf pNGR234a hochreguliert werden, kann NGR234 als ein fantastischer Untersuchungsorganismus verwendet werden, um die Expression von kleinen Proteinen auf pNGR234a in einer QS-abhängigen Weise zu analysieren. Durch eine Kombination einer ORF-Suche mit der Zuordnung der resultierenden kleinen ORFs zum Transkriptom von NGR234, konnten 251 zusätzliche kleine ORFs mit einer Größe zwischen 33 nts (10 aa) und 183 nts (60 aa) identifiziert werden. Zudem konnte gezeigt werden, dass zwischen den Genen *traI* du *repA* ein Operon liegt, welches aus drei kleine ORFs besteht und Auswirkungen auf die Plasmiderhaltung hat. Die entsprechenden Proteine wurden als RepX (51 aa), RepY (57 aa) und RepA0 (143 aa) bezeichnet. Ein weiteres kleines Protein, welches identifiziert wurde, wird durch den kleinen ORF NGR_a01725 kodiert und weist eine Größe von 25 aa auf. Mittels Mutagenese, Immundetektion mit spezifischen Antikörpern und Translationfusionen konnten die Expressionen dieser kleinen ORFs zu funktionellen Proteinen und deren Einfluss auf die Plasmidreplikation und -stabilität nachgewiesen werden.

Die bakterielle Kommunikation mittels QS kann durch Inaktivierung von den AI Molekülen gestört werden, was mittels Quorum Quenching (QQ) Enzymen katalysiert wird. Das QQ Protein, genannt GqqA, aus *Komagataeibacter europaeus* CECT8546 zeigt eine große Ähnlichkeit zu Prephenat-Dehydratasen (PDT) und zeigt Störungen der N-Akylhomoserin Laktone QS Signale. Frühere Studien zeigten, dass GqqA ein *E. coli* PDT-Enzym nicht komplementieren kann, sondern die QS abhängige Biofilmbildung in *K. europaeus* CECT 8546 beeinflusst. ESI-MS/MS-Messungen mit einem GqqA *in vitro* Enzymassay mit 3-oxo-C8-HSL als Substrat zeigt, dass GqqA in der Lage ist die Amid Bindung zu spaltet und einen Laktonring und die zugehörige Akylsäure freisetzt. Aufgrund der Strukturähnlichkeit zu PDT-Enzymen und der Acylasen Aktivität stellt GqqA das erste Mitglied einer neuartigen Klasse von Acylasen da, die bakterielle QQ Eigenschaft besitzt.

2. Introduction

Bacterial cell-to-cell communication

Quorum sensing (QS) describes the cell- cell communication within a bacterial community in dependence of cell density. The bacteria cells produce, release and detect small mostly diffusible molecules called autoinducers (AI). With these AI molecules the bacteria sense the population densities and coordinate the gene expression for gene products, which are beneficial only for the population but non-essential for the single cell (Nealson, 1977; Ng and Bassler, 2009). At low cell density these AI disperse in the environment before being absorbed by the other cells. When the population density increases to a threshold value the AI are detected by the neighboring cells by binding to a transcriptional regulator (Darch et al., 2012; Papenfort and Bassler, 2016). The complex regulates gene expression of target genes responsible for biofilm formation, virulence factor production, bioluminescence, sporulation and symbiosis (Cao and Meighen, 1989; Davies et al., 1998; Zhu et al., 2002; Williams and Cámara, 2009). QS was first described in the marine bacterium *Vibrio fischeri* (now reclassified as *Aliivibrio fischeri*) regulating the production of bioluminescence. The regulator LuxR binds to the bioluminescence operon l*uxICDABEG* and thereby activates the transcription of these genes (Nealson and Hastings, 1979; Engebrecht and Silverman, 1984).
QS is common in both Gram-negative and Gram-positive bacteria (Li and Tian, 2012). In the course of evolution, a variety of different mechanisms with various QS signal molecules have been developed. The most common AI molecules are the *N*-acyl-homoserine lactones (AHLs) in Gram-negative bacteria and consist of an acyl chain linked to a homoserine lactone ring via an amide bond (Figure 3; B, D; Fuqua and Greenberg, 1998). Due to the different variants of these compounds with different acyl chain length and substitutions, the acyl-HSL belongs to the intraspecific signals and is used for species-specific communication. In contrast to the AHLs, the autoinducer-2 (AI-2) has evolved an interspecific AI molecule. Since it is used in both Gram-negative and Gram-positive bacteria, AI-2 is considered to be a universal language (Xavier and Bassler, 2003). Structurally this molecule is a furanosyl borate diester which is catalyzed by an S-ribosyl homocysteinase (e.g. LuxS; Chen et al., 2002). The receptors for AI-2 molecules have been identified in numerous bacterial species, reinforcing the presumption of using interspecific communication (Xavier and Bassler, 2003). For example, the bioluminescence in *Vibrio harveyi* or the virulence factor production in pathogenic *E. coli* bacteria are regulated by AI-2 molecules (Cao and Meighen, 1989; Ren et al., 2004). In addition to the common AI molecule Acyl-HSL and AI-2, there are some more

specific AI molecules like the diffusible signal factor (DSF) in the plant pathogen *Xanthomonas campestris* (Barber et al., 1997), Pseudomonas Quinolone Signal (PQS) in the *Pseudomonas* genus (Pesci et al., 1999) or the Diketopiperzine (DKP) in bacteria, fungi or other higher organisms (Belin et al., 2012; de Carvalho and Abraham, 2012). DKPs can activate or inhibit LuxR-mediated QS systems by competing DKPs with AI-1 molecules for the AI-binding site of LuxR regulators. It is common that most bacteria encode for more than one QS system.

The genus Sinorhizobium

The genus *Sinorhizobium* is a group of rods-shaped and nitrogen-fixing bacteria belonging to the *Rhizobiaceae* family in the class of α-Proteobacteria (Figure 1, A). Some rhizobia have established mechanisms that allow them to nodulate a wide range of host. There are only a few *Sinorhizobium* strains that have this broad host range. These are namely the closed related *S. fredii* NGR234 (hereafter NGR234), *S. fredii* USDA257, *S. fredii* HH103 and *S. fredii* GR4 (Marie et al., 2004; Krysciak et al., 2015). Whereas HH103 can only get in symbiosis with ≥10 legumes and is more classified to the narrow host range, USDA257 is able to fix nitrogen in 79 different plant genera, but none of them has such a broad host range as it is known for NGR234. Especially, NGR234 can form nodules in more than 120 legume genera and the non-legume *Parasponia andersonii* (Pueppke and Broughton, 1999). Because of this ability and the fact, that genome of NGR234 is complete sequenced, NGR234 is a great model bacterium for the investigation of the symbiosis between rhizobia and legumes and the QS dependent gene expression (Viprey et al., 2000; Perret et al., 2003;

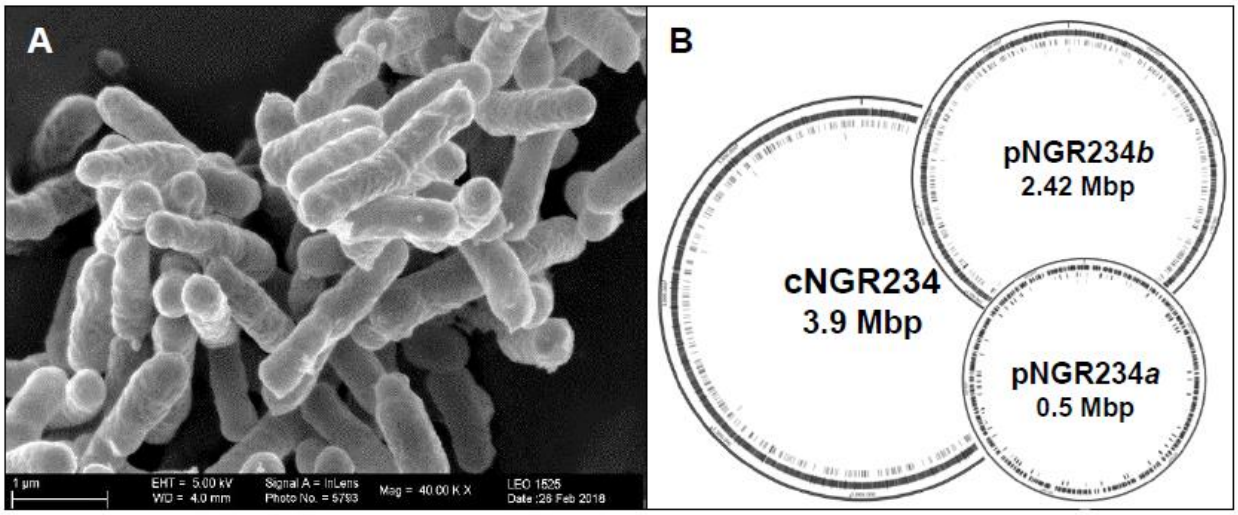

Figure 1: Scanning electron microscope image of NGR234 cells (A) and the three replicons of NGR234 with the corresponding genome sizes (B).
A: The scanning electron microscope image was taken with the SEM Leo 1525 Gemini. **B:** The genome of NGR234 consists of three replicons- the chromosome (3.9 Mbp), the megaplasmid pNGR234*b* (2.42 Mbp) and the symbiotic plasmid pNGR234*a* (0.5 Mbp).

Streit et al., 2004; Schmeisser et al., 2009).

The genome of the rod shaped NGR234 consists of three replicons with a total size of 6.891 Mbp (Figure 1, B). The main replicon is the chromosome with a size of 3.9 Mbp that encodes for 3,633 genes. All essential genes important for bacterial survival are on the chromosome. The second replicon is the megaplasmid pNGR234*b* harboring 2,342 genes and has a size of 2.42 Mbp. Many of genes involved in the production of exopolysaccharides and secretion of cell components are located on the megaplasmid (Schmeisser et al., 2009). On the second plasmid pNGR234*a*, all genes necessary for establishing the symbiosis and for nitrogen fixation are located on this replicon and the replicon is often mentioned as symbiotic plasmid. So, here the symbiotic plasmid harbored the genes for the Nod factor production (nod*ABCDEF*), for conjugal transfer (*trb* and *tra* genes) and for nitrogen fixation (*fix* and *nif* genes). The fact that the symbiotic plasmid lacks all essential genes necessary for growth and survival is shown by the NGR234 derivate ANU265 which was deprived of pNGR234*a* by heat curing. This derivate showed no impairment in growth but lost the ability to form nodules (Morrison et al., 1983; Freiberg et al., 1997; Perret and Broughton, 1999). Both megaplasmids of NGR234 belong to *repABC* plasmid family using RepABC proteins for plasmid replication (Freiberg et al., 1997; Streit et al., 2004). These RepABC type of plasmids are common in Rhizobiales and play an important role in the infection of the host (Higashii, 1967).

The symbiosis between legumes and rhizobia

The growth of all organisms is limited by the presence of fixed nitrogen. But only a few bacteria and archaea are able to fix nitrogen into ammonia that can be assimilated by plants or other organisms. So, the nitrogen-fixing bacteria have an environmental importance to the global nitrogen cycling balance and the agriculture economy (Sessitsch et al., 2002; Jensen et al., 2012). The soil bacteria that reach symbiosis with the legumes are collectively referred to as rhizobia. Most of these rhizobia are classified to the α-Proteobacteria and the β-Proteobacteria (Peter et al., 1996; Moulin et al., 2001) including the genera *Rhizobium*, *Bradyrhizobium*, *Mesorhizobium*, *Azorhizobium*, *Photorhizobium* and *Sinorhizobium.*

The symbiosis is initiated by a signal exchange between the legume plant and the compatible *Rhizobium* bacteria in the soil (Denarie et al., 1992). The legume plant releases secondary metabolites like flavonoids or other compounds (Figure 2, A), which were detected by the bacterial transcriptional regulator NodD. NodD upregulates rhizobial *nodABC* genes, which produce lipo-chitin oligosaccharides (LCOs) that act as nodulations factors (Nod factors; Firmin et al., 1986). These Nod factors are released to the soil and detected by the plant Nod

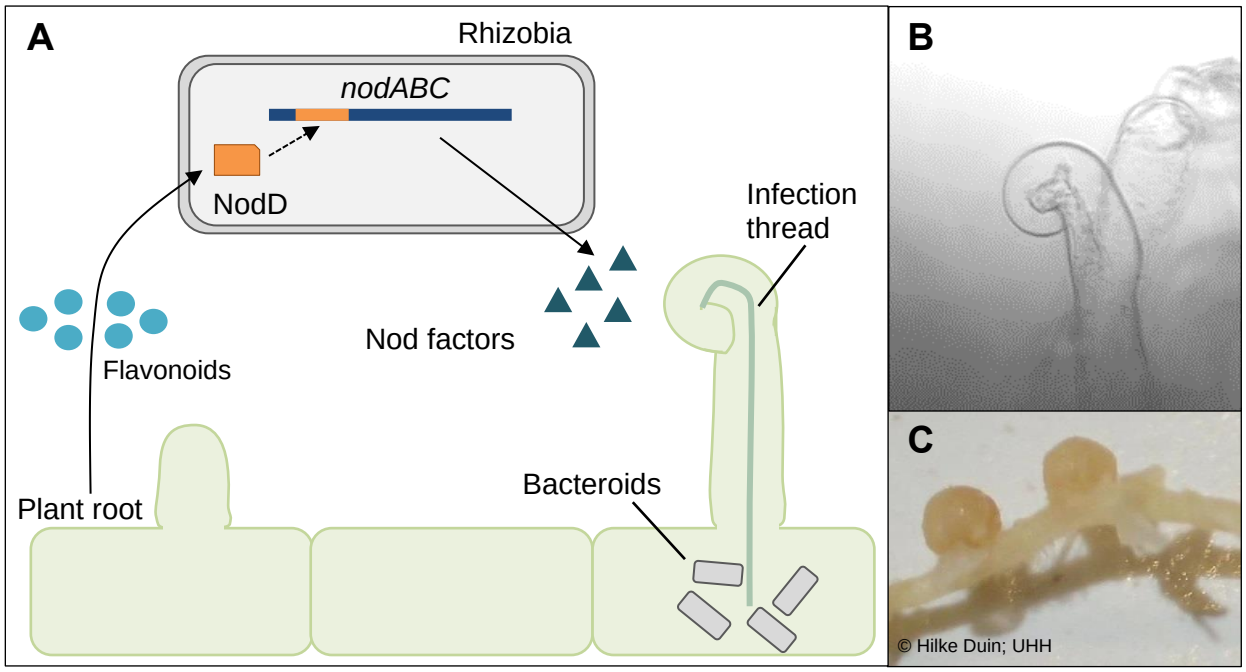

Figure 2: Symbiosis between Rhizobia and plants.
A: schematic model of the initiation of the symbiosis between the Rhizobia cell and the plant root. The plant releases flavonoids (circles), which are detected by the bacterium. The bacteria produce Nod factors (triangle) and release them to the soil. Thereupon the plant roots start to curl the root hair and the bacterial cells penetrate the plant cell wall and move to the plant cortex, where the cells start the differentiation to bacteroides (grey rectangle). **B**: microscopy image of the root hair curling in *Vigna unguiculata* incubated with Nod factors of NGR234 supplemented with flavonoids (Grote et al 2016). **C**: Root with nodules from *Vigna unguiculata* incubated with NGR234 (photo of nodules was taken by Hilke Duin, UHH).

factor receptors (NFRs) at the root hair. Particular residues and modifications at LCOs are important for the host specificity. Many species produce more types of Nod factors to increase the host range number (Geurts and Bisseling, 2002; Perret et al., 2003). As a result of the attached LCOs, the plant root hairs begin to curl (Figure 2, B). The rhizobia cells are entrapped and penetrate the plant cell wall where they induce the formation of an infection thread that brings the cells to the cortex (D'Haeze and Holsters, 2002). At the cortex the bacteria cells multiply extremely and differentiate to bacteroides and trigger the formation of specialized root structures called nodules (Figure 2, C; Freiberg et al., 1997; Esseling et al., 2003; Parniske and Downie, 2003; Jones et al., 2007). The nodules are the location where the nitrogen fixation takes place.

Quorum sensing in *Sinorhizobium fredii* NGR234

QS systems enable bacteria to sense their neighboring environment to coordinate various gene expressions like the infection of a host (Gurich et al., 2009; Hartmann et al., 2014). The symbiotic plasmid pNGR234*a* shows high similarity to the plant pathogen *Agrobacterium*

tumefaciens Ti- plasmid which is involved in the infection of the plant and this infection is correlated to QS (Piper et al., 1993; Hwang et al., 1994).

The genome of NGR234 encodes for two QS systems. The genomic organization of both QS systems and the mechanisms of these systems are displayed (Figure 3). The NgrI/R- QS system is located on the chromosome (Figure 3, A). The autoinducer synthase NgrI produces probably a 3-oxo-dodecanoyl homoserine lactone (Figure 3, B; 3-oxo-C12-HSL) as AI

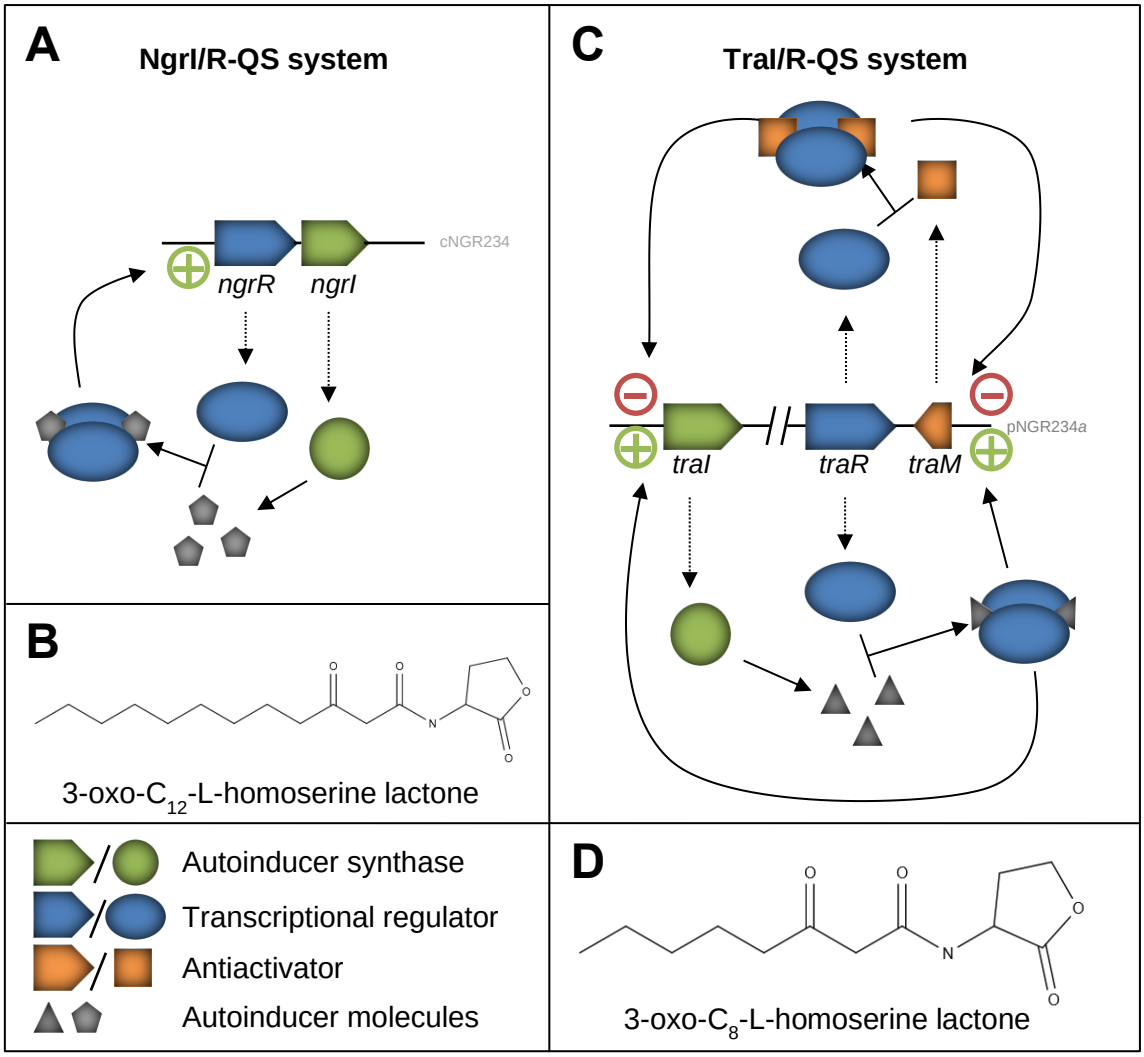

Figure 3: Quorum sensing systems in NGR234.
A: The NgrI/R- QS system is located on the chromosome and consists of a transcriptional regulator NgrR (green) and an autoinducer synthase NgrI (blue) producing a probably a 3-oxo-C12-HSL as AI molecule (grey pentagon). **B:** The chemical structure of the AI molecule produced by NgrI. The AI molecule is classified to the AHL and consists of HSL core with a 3-oxo-dodecanoyl (C12) side chain. **C:** The symbiotic plasmid encodes for the TraI/R QS system. This system is highly similar to the *A. tumefaciens* QS system and consists of the AI synthase TraI (green), the TraR transcriptional regulator (blue) and the antiactivator TraM (orange). TraI produces 3-oxo-C8-HSL as AI molecules (grey triangle). These AHLs interact with TraR and initiate a positive feedback loop for the expression of *traI* and *traM*. TraM is an antagonist of TraR and inhibits the attachment of the AHLs to TraR initiate a negative feedback loop. **D:** The chemical structure of 3-oxo-C8-HSL produced by TraI. The figure is constructed and modified according to information of He et al., 2003 and Fuqua et al. 2007.

molecule (He et al., 2003). When a threshold concentration of AI molecules is reached, the 3-oxo-C12-HSL binds to the transcriptional regulator NgrR. The whole NgrI/R QS system is not clarified yet and must be characterized in the future (He et al., 2003; Schmeisser et al., 2009). The second QS system TraI/R is located on the symbiotic plasmid pNGR234*a* (Figure 3; C) and consists of an autoinducer synthase TraI, a transcription regulator TraR and an antiactivator TraM (He et al., 2003).

The mechanism is similar to that of the plant pathogen *A. tumefaciens* (Hwang et al., 1994; Fuqua and Winans, 1996). At a low cell density, the autoinducer synthase TraI produces AI molecules, verified to be 3-oxo-octanyl-homoserine lactone (3-oxo-C8-HSL). These molecules are diffused out of the cells. If cell density increases, AI molecules accumulate in the soil until a threshold value is reached. Afterwards, the AHL interacts with the transcriptional regulator TraR. This interaction leads to a confirmation change and the complex can bind to the promoter region of target genes and initiates a positive feedback loop of the expression of *traI* and conjugal plasmid transfer (*trb*) genes (Figure 4, C; Gargioli et al., 2002; Gonza and Marketon, 2003; He et al., 2003). The antiactivator TraM is an antagonist of TraR and competes with the AI molecules for attaching to TraR and thus inhibits the binding of TraR to the target promoter regions (Swiderska et al., 2001; Jeffrey et al., 2007).

In NGR234, 186 QS-regulated genes were identified by using RNA-seq. The NgrI/R system regulates in total 466 genes (7.3% of all predicted genes), which are responsible for the expression of the type-III- secretion system, type-IV-attachment pili or biotin biosynthesis, for example. The TraI/R system regulates 361 genes, including genes associated with pNGR234*a* replication and cytochrome *c* synthesis. All regulated genes were distributed among the three replicons, with most genes observed on the chromosome and only a small amount of differently regulated genes located on the symbiotic plasmid (Krysciak et al., 2014). A mutant, in which both autoinducer synthase genes *ngrI* and *traI* were deleted, showed especially for the symbiotic plasmid pNGR234*a* that nearly all genes were differently regulated and that most of them showed an increased expression of the genes (Grote et al., 2016). This switching on of almost all genes on the symbiotic plasmid is a useful tool to investigate the expression of small proteins located on this replicon.

Small proteins in the world of bacteria

Small proteins are proteins whose small size results directly from the translation of a small open reading frame (ORF). These proteins differ from peptides that have been cleaved by larger proteins (Storz et al., 2014). Traditionally in the past, proteins were set to a minimum of

100 aa and everything below neglected. The purpose that most algorithms of gene annotation or protein prediction used a cutoff of more than 300 nts (100 aa) was to reduce the likelihood of false positive genes (Boekhorst et al., 2011). Although, this was certainly the easiest way of genome annotation in the past, it has led to the problem that mostly ORFs with a smaller size were overseen and hence are not on the database in order to use these smaller ORFs as templates for functional analysis and reannotations (Storz et al., 2014). In recent years, the sequencing technique has improved. With this deeper and faster throughput in combination with Ribosome profiling and LC-MS/MS analysis, it has become clear, that all genomes code for a higher number of smaller proteins and that these small proteins are involved in various cellular functions (Wang et al., 2008a). Shell and her co-workers identified 2,166 transcriptional start sites as well as hundreds of possible small proteins in

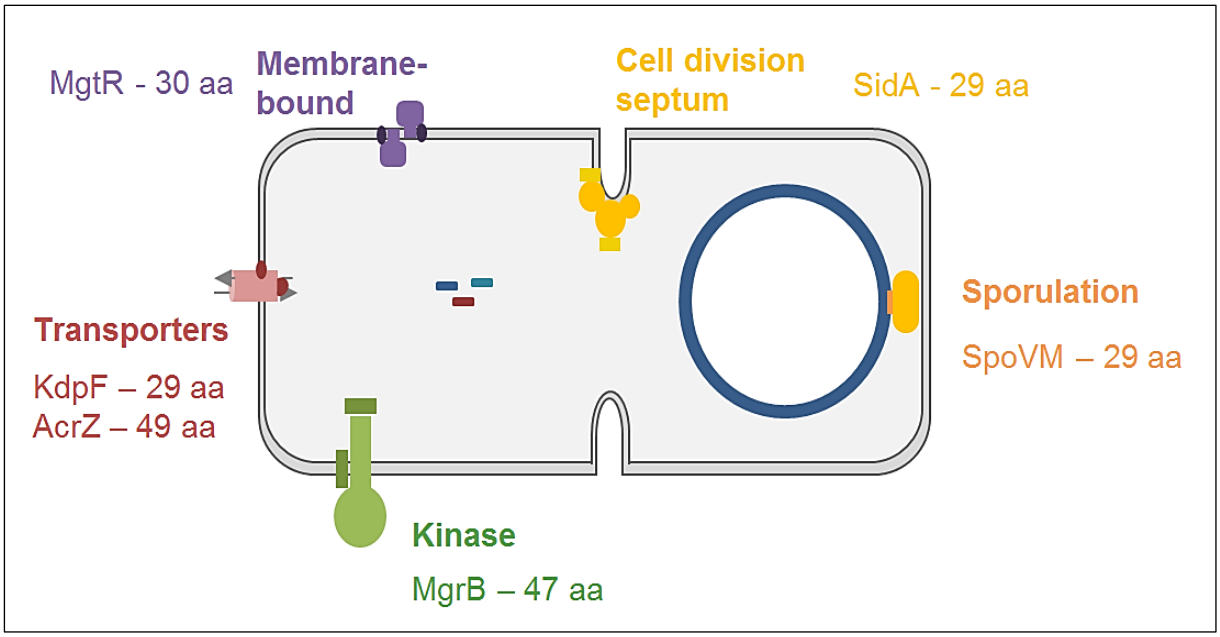

Figure 4: Functional role of small proteins in bacteria.
A bacterium cells is shown. The black lines represent the bacterium plasma membrane, which enclose the cytosol (light grey). The forespore membrane is shown in blue. The different cell functions where proteins are identified are displayed in colored in following order: membrane-bound proteins (violet), proteins involving in cell division (yellow), sporulation proteins (orange), transport proteins (red) and small proteins associated to kinase (green). Examples that are explained in the text are listened closed to the location. (According to Storz et al. 2014)

Mycobacterium tuberculosis (Shell et al., 2015). Similar to this study in various other bacteria, researchers began to analyze the expression and the functions of small proteins (Figure 4). First examples of small proteins were published showing the important and physiological role of small proteins in bacteria (Hobbs et al., 2011; Storz et al., 2014; Cheng and Ma, 2017; Yin et al., 2019).

At current state, small proteins are known to be involved in morphogenesis (SpoVM, 29 aa; Levin et al., 1993), cell division (SidA, 29 aa; Modell et al., 2011), regulatory networks (MgtR, 30 aa; Choi et al., 2012), transport (KdpF, 29 aa; Gaßel et al., 1999), enzymatic activity, and stress response. Small proteins can therefore not only provide information on how biological functions can be performed with very few aa, but also serve as tools to investigate how their larger interacting partners are involved in different cellular processes (Storz et al., 2014).

One of the first reported small proteins is the monocistronic locus that encodes for a 26 aa protein and is totally conserved in the genome of the Gram-positive endospore forming bacteria species like *Bacillus* sp. (Levin et al., 1993). The protein SpoVM is a membrane-associated protein only produced in the mother cell and a deletion of the *spoVM* allows the formation of fore spore but impairs the formation and assembly of the spore cortex. So the deletion resulted in decreased sporulation ability (Levin et al., 1993; Cutting et al., 1997). This example shows an effect of a small protein in the sporulation and the ability to adapt to changing environmental conditions and is significant for the survival of the bacterial cells.

A small protein involved in the stability of a virulence factor is MgtR, a 30 aa small protein identified in the genome of *Salmonella* species (Alix and Blanc-Potard, 2008). MgtR modulates the stability of the MgtC virulence factor. The small protein MgtR binds directly to MgtC at the plasma membrane and promotes MgtC degradation by the FtsH protease. An over-expression of MgtR in macrophages led to a decreased replication rate of *Salmonella* bacteria in the mammalian cells. That MgtC is a common virulence factor in multiple human pathogens and that the small protein MgtR acts a natural antagonist of MgtC with downregulation of the MgtC expression underlines the importance of small proteins in the regulatory mechanism of pathogenic bacteria (Alix and Blanc-Potard, 2008; Wang et al., 2017; Yin et al., 2019). A second small protein affecting the transporter activity of an antibiotic efflux pump is the 49 aa protein AcrZ. It was first described in *E. coli* and is highly conserved among *Enterobacteria* sp. like *Salmonella*, *Klebsiella* or *Erwinia* species (Papenfort et al., 2009; Hobbs et al., 2012). Structurally the protein has a N-terminal transmembrane helix and is associated to the AcrAB-TolC efflux pump. The interaction partner of AcrZ is AcrB, which is the inner membrane component of this efflux pump. The *acrZ* gene expression is induced by a lot of antibiotics and other detergents. A lack of this gene results in sensitivity to some antibiotics that are exported by the AcrAB-TolC pump. It seems possible that the effect of ArcZ on the antibiotic sensitivity might be initiated from the allosteric conformation of AcrB by AcrZ and that antibiotics can’t bind to AcrB (Hobbs et al., 2012).

The hydrophobic small protein KdpF is a part of the Kdp potassium pump in *E. coli* and also shows the impact of small proteins to bacterial transporter complexes. The protein has a size

of 29 aa and is responsible for the stabilization of the complex (Gaßel et al., 1999) A deletion of *kdpF* has no influence to the growth of *E. coli* with low potassium concentration in the medium but Gaßel and colleagues showed that KdpF is indispensable for a functional enzyme complex *in vitro* and so they suggested that the gene *kdpF* is a main part of *the kdpFABC* operon. This operon was also identified in the pathogen *Mycobacterium bovis,* which is the causative agent of tuberculosis in cattle (Gannoun-Zaki et al., 2013; Rosas Olvera et al., 2017). The small protein can affect the stability of proteins interacting with KdpF. Olvera and her colleagues suggested that KdpF acts as a regulatory molecule and interferes with bacterial virulence. The protein has also been identified in other pathogens (Gannoun-Zaki et al., 2013; Gannoun-Zaki et al., 2014; Rosas Olvera et al., 2017). Small proteins may also be involved in toxin-antitoxin (TA) systems consisting of two or more closely linked genes coding for a toxin and the corresponding antitoxin. There are 6 different classes of TA systems that are based on the nature of the antitoxin. Small proteins are mostly found in type 1 TA systems, where the toxin is often a hydrophobic protein and the antitoxin a noncoding small RNA that inhibits the translation of the toxin protein (Page and Peti, 2016). The *hok* (host killing) gene encodes for a 52 aa toxic protein that causes cell death by depolarization of the cell membrane. The translation of the *hok* mRNA is regulated by *sok* (suppressor of killing) RNA. The binding to the hok mRNA inhibits translation and the toxin protein cannot be synthesized. These Hok/*sok* systems are often located on the *parB* region on R1 plasmid in *E. coli* and play a major role in plasmid maintenance and stability (Gerdes et al., 1986; Rasmussen et al., 1987; Gerdes et al., 1997).

The examples named show that small proteins are involved in various important, cellular functions and in the virulence of some bacteria. Since it is known that the molecular mechanisms that mediate communication and cellular modulation are very similar in pathogens and symbionts, it can be assumed that small proteins also play a role in symbiosis (Nelson and Sadowsky, 2015).

Interruption of bacterial communication

In the past many enzymes have been discovered, which are able to interrupt the bacterial communication by blocking the receptor of the regulator proteins or by degrading the AI molecules. This interruption is called quorum quenching (QQ) and can affect cellular function like virulence factor production or biofilm formation (Lin et al., 2003). Due to this inhibition of QS-dependent virulence of bacteria, QQ is becoming more and more interesting for the development of new anti-infecting therapies. QS may be interrupted firstly by inhibition of the signal biosynthesis, secondly by inhibition of the signal detection or lastly by inactivation of

the signal molecule. For the latter, various enzymatic degradations processes are possible (Figure 5). The most common enzymatic degradation is performed by AHL lactonases which have been found in a wide range of bacteria species and different protein families. AHL lactonases catalyze the opening of the lactone ring by addition of a water molecule or

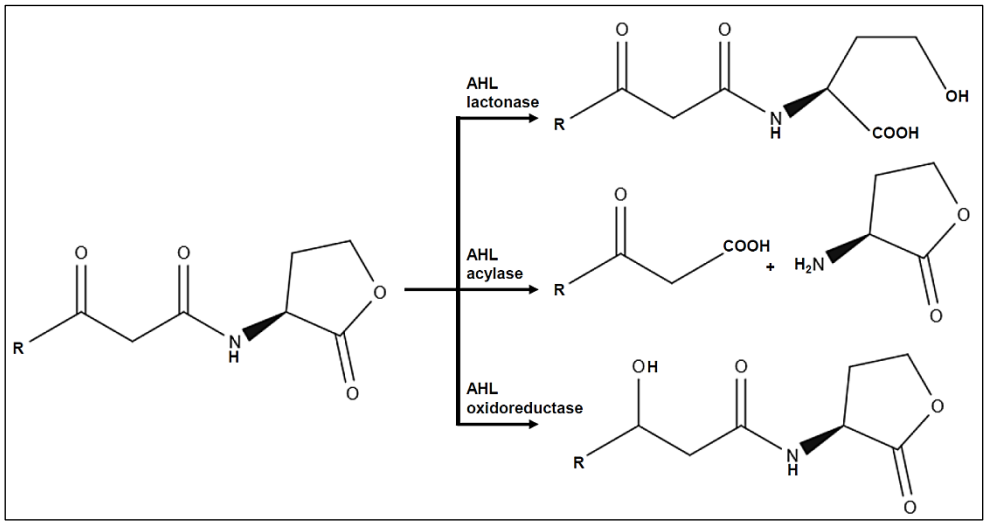

Figure 5: Enzymatic degradation of AHLs catalyzed by QQ enzymes.
Three different enzyme classes are known to catalyze the enzymatic degradation of acyl-homoserine lactones. The most common class is AHL lactonase which open the lactone ring by hydrolytic cleavage. AHL acylase cleave hydrolytic the acyl chain and the lactone ring at the amide bond. The reduction of the 3-oxo substitute is catalyzed by the AHL oxidoreductase.

spontaneously in aqueous solutions. The AHL acylases cleave AI signals at the amide bond and release fatty acid and homoserine lactone ring. The third class of inactivating enzymes is the class of reductases class converting the 3-oxo substituted AHL to their cognate 3-hydroxyl substituted AHL (Dong et al., 2000; Byers et al., 2002; Lin et al., 2003; Bijtenhoorn et al., 2011). Whereas lactonases reaction can be reversed to AHL in acidic pH solution, the acylase products cannot regenerate to a functional AHL. In addition, the fatty acid produced by the acylase is usually rapidly metabolized for growth and the HSL is used as nitrogen source (Fetzner, 2014).

In the past many QQ enzymes have been identified and characterized. AHL lactonases have been found in a wide range of organisms and are first and best characterized as a group of QQ enzymes. They can be divided into different protein families: aminohydrolases [EC 3.5.4], paraoxonases [EC 3.1.8.1] and metallo-β-lactamases [EC 3.5.2.6]. One of the first analyzed QQ enzymes is AiiA from *B. subtilis* spp. The gene *aiiA* encodes for 250 aa protein containing a HXHXDH zinc-binding domain that is conserved in metallo-β-lactamases. These residues are required for AiiA activity. It was shown that AiiA can inhibit the virulence of the pathogen *Erwinia carotovora* if the gene *aiiA* is expressed in this strain (Dong et al., 2000). The

lactonase QsdR1 from NGR234 shows also the typical HXHXDH zinc-binding domain and can reduce the motility and biofilm formation in *Pseudomonas aeruginosa*. QsdR1 also showed a decreasing effect of the rhizosphere colonization capability of cowpea roots. This indicated the importance of QS for host infection (Krysciak et al., 2011). The second group of QQ enzymes is AHL acylases. Some of these AHL acylases are members of the N-terminal nucleophile hydrolase superfamily [EC 3.5.1.97]. In this group of QQ enzymes, the best characterized enzyme is the PvdQ from *P. aeruginosa* (Huang et al., 2003). Due to a signal peptide in the sequence, the precursor PvdQ is exported directly into the periplasm and undergoes two auto processing events. PvdQ is divided into an N-terminal and a C-terminal peptide, which are clustered together via non-covalent binding to form the mature heterodimeric enzyme. Due to this new formation, the enzyme receives its activity. It was shown, that PvdQ catalyzes the inactivation of long-acyl AHLs and that this enzyme was the first whose AHL degradation activity has been shown in *γ*-proteobacteria (Huang et al., 2003). The third group of QQ enzyme is the Oxidoreductases catalyzing the reduction of 3-oxo-AHL to the corresponding 3-hydroxy-AHL and thereby inactivates the signal molecules (Figure 5). The converted products of the reductase still act as AI molecules, but often they are less active than the initial substrate (Chowdhary et al., 2007). Oxidoreductase activity was found in some bacteria species like in *Rhodococcus erythropolis* (Uroz et al., 2005) *Burkholderia* sp. GG4 (Chan et al., 2011) or in *P. aeruginosa* (Bijtenhoorn, 2012) and also the aldehyde dehydrogenase AldR shows QQ activity in NGR234 (Krysciak et al., 2011). In general, oxidoreductases have not yet been extensively investigated and further analyses are required. Often the role of these enzymes in their native environment is not clear, but the ability to inhibit the virulence of some bacteria by using QQ enzymes shows a high potential in industrial and medical applications with lower susceptibility for resistance, because the survival of the bacterium is not affected by QQ enzymes (Fast and Tipton, 2012; Grandclément et al., 2015).

GqqA- a novel QQ enzyme from *Komagataeibacter europaeus*

Komagataeibacter europaeus CECT 8546 (formerly *Gluconacetobacter europaeus*) belongs to the group of acetic acid bacteria (AAB) and this group is known for its ability to produce acetic acid on ethanol-containing material that leads to vinegar. In the last years it became clear that the acetic acid and gluconic acid production and the antifoam activity is regulated in a N-AHL-dependent QS mechanism designated as GinI/R system (Iida et al., 2008; Iida et al., 2009; Valera et al., 2015). Until three years ago, nothing was known about QS interfering enzymes. However, in 2016, a gene designated as *gqqA* was identified in the genome of the

cellulose overproducer and biofilm-forming bacterium *K. europaeus* CECT 8546 by using a screening protocol for QS interfering enzymes (Schipper et al., 2009; Valera et al., 2016). The gene *gqqA* encodes for a 281 aa protein with a molecular mass of 30 kDa. The aa sequence of the protein shows high similarity to predicted prephenate dehydratase (PDT; EC 4.2.1.51), which are involved in the metabolic pathways of aromatic amino acids and convert prephenate to phenylpyruvate for the biosynthesis of L-phenylalanine (L-phe). GqqA consists of the N-terminal periplasmatic binding and the C-terminal regulatory domain. But a genetically complementation of *E. coli* Δ*pheA* mutant showed no effect suggesting GqqA is not involved in the L-phe biosynthesis and has different function (Valera et al., 2016). Due to the QQ screening they could verify the quenching effect of GqqA in *A. tumefaciens* NTL4 and *Chromobacterium violaceum* CV026 reporter strains. When *gqqA* is present in *P. aeruginosa* PAO1, the expressed enzyme had a strong effect in reducing the motility and the pyocyanin production. *K. europaeus* CECT 8546 cells tend to form aggregates in cellulose biofilm but if the cells are supplemented with purified GqqA the formation of aggregates is reduced. This effect was also observed in related AAB. It is curiously that GqqA has a low similarity to known QQ and a high similarity to PDT enzymes which have been verified functionally, but GqqA does not have a PDT function itself. These data indicate a QS-dependent cellulose formation in AAB which is biotechnologically for interest. However, the mechanism behind GqqA is not clear yet and needs to be analyzed.

Intention of this work

This thesis is composed of two sections:

I. The role of small proteins in a quorum sensing dependent expression
II. Structural analysis of a novel quorum quenching hydrolase designated as GqqA

In the first section of this thesis, we want to explore the possible role of small proteins in the complex QS regulatory circuits of the model organism *Sinorhizobium fredii* NGR234. Therefore, a protein search will be performed, which will be mapped to the previously published transcriptome of NGR234 strains. The expression of selected candidates will be verified and analyzed for first biochemical functions. A new AI double deletion mutant will be constructed, and the small proteins will be examined for their QS dependent expression.

The second intention of this research is to characterize a novel QQ enzyme that was identified 2015 in the acetic acid bacterium *Komagataeibacter europaeus* CECT 8546. It is known that this enzyme shows no similarities to QQ enzymes described so far, but QQ

enzyme activity was detectable in *in vivo* assays. Thus, the hypothesis is that GqqA belongs to a new group of AI degradation hydrolases.

For this purpose, the protein crystal structure of the enzyme will be identified in cooperation with the working group of Prof. Dr. Betzel (University of Hamburg, DESY). Various mutations are to be constructed to test the structure and function of GqqA. With ESI-MS/MS approach the reaction mechanism behind this enzyme is to be identified.

3. Materials and methods

All genetic work was performed by standard techniques (Sambrook and Russell, 2001) or by following the manufacture's manuals. All mediums and thermostable solutions, glass vessels and other instruments were autoclaved at 121 °C for 20 min. All thermosensitive solutions were filtered sterile and all devices were sterilized with 70% (vol/vol) EtOH and flamed.

Bacterial strains

Bacterial strains used in this study are listed in Table 1.

Table 1: Bacterial strains used in this study

Strain	Characteristics [1]	Reference/Source
	Escherichia coli	
DH5α	Cloning strain *supE*44 Δ*lacU*169 (Φ80 lac*Z* ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1	Hanahan and Harbor, 1983
S17-1 λpir	*pro thi hsdR* $^+$ Tp^r Sm^r; chromosome::RP4-2 Tc::Mu-Kan::Tn7/λpir	Simon, Priefer, and Pühler 1983
WM3064	*thr*B1004 pro *thi rpsL hsdS* lacZΔM15 RP4-1360 Δ(*ara*BAD)567 Δ*dap*A1341::[erm pir]	William Metcalf, UIUC
BL21 (DE3)	F$^-$ *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S)	Studier and Moffatt, 1986
T7 Shuffle	F´ *lac*, *pro*, *lacI*^{q} / *Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt*::pNEB3-r1-*cDsbC* (Spec^R, *lacI*^{q}) *ΔtrxB rpsL150*(Str^R) *Δgor Δ(malF)3*	New England BioLabs
	Rhizobiales	
S. fredii NGR234 wt	Wild type New Guinea isolate, Rif^R	Trinick, 1980
S. fredii NGR234 Δ*ngrI*Δ*traI*_copy+	The QS AI synthase genes *ngrI* and *traI*$_{627}$ (534,836 – 535,462) were deleted; the 5'UTR and the translations start of *repX* are missed on pNGR234*a, gen*^{R}	Krysciak et al., 2014, Grote et al. 2016
S. fredii NGR234 Δ*traR*	Gene for the QS transcriptional regulator TraR is deleted, *gen*^{R}	Grote et al., 2014
S. fredii NGR234 Δ*ngrI*Δ*traI*	Δ*ngrI*Δ*traI*$_{425}$ (534,260 - 535,260), only the first 202 nts of the 5' *traI* were retained. The gene *repX* and the promoter region are unimpaired	This work
S. fredii NGR234 Δ*repX*	The smORF-249 (*repX*) is knockout between the positions 535,545 and 535,616 on pNGR234*a*.	This work

Strain	Characteristics [1]	Reference/Source
	Rhizobiales	
S. fredii NGR234 Δ*repY*	The smORF-249 (*repY*) is deleted between the positions 535,637 and 535,738 on pNGR234*a*.	This work
S. fredii NGR234 Δ*a1725*	The smORF-104 (NGR_a01725) is deleted between the position 218,028 and 218,105 on pNGR234*a*.	This work
S. fredii ANU265	a Muc$^+$ Sym of strain NGR234; Spec^R	Morrison et al., 1983
S. fredii NGR234 Δ*traR* ::gTF*repX*::mCherry	Genomic translation fusion: The red fluorescence protein mCherry is fused to the smORF-249 (*repX*) on pNGR234*a.*	This work
	other	
Chromobacterium violaceum CV026	Reporter strain for autoinducer I; mini-Tn5 in cviI	McClean et al., 1997
Agrobacterium tumefaciens NTL4	Reporter strain for AHL detection, *traI*::*lacZ*, Tet^R, Spec^R	Fuqua and Winans, 1996
Komagataeibacter europaeus CECT 8454	Strain recovered from vinegar. Cellulose producer	Valera et al., 2016

[1] Abbreviations describing geno- and phenotypes were made according to Bachmann 1983

Vectors and constructs

All vectors and constructs used and designed in this work are summarized in the following table (Table 2).

Table 2: Vectors and constructs used in this study

Vector/Construct	Characteristics	Reference/Source
pBBR1MCS-2	Broad host range vector, *kan*^{R}	Kovach et al. 1995
pBBR1MCS-5	Broad host range vector, *gen*^{R}	
pNPTS138-R6KT	Suicide vector, *sacB*, *kan*^{R},	Lassak et al. 2010
pET-21a	Expression vector, T7Prom, Amp^R, C-His	Novagen, Darmstadt Germany
pMALc2x	Expression vectorT7 Prom, Amp^R, N-MBP;	New England Biolabs GmbH, Frankfurt on Main, Germany
pc*ngrIctraI*	The genes *ngrI* and *traI* were inserted to the broad host range vector pBBR1MCS-2	Krysciak et. al. 2014

Vector/Construct	Characteristics	Reference/Source
pP_{repX}mCherry	Promoter fusion: the promoter of *repX* was fused to the red fluorescence protein mCherry and cloned into the vector pBBR1MCS-2.	This work
pP_{a1725}mCherry	Promoter fusion: the promoter of NGR_a1725 was fused to the red fluorescent protein mCherry and cloned into the vector pBBR1MCS-2.	This work
pP_{repY}mCherry	Promoter fusion: the promoter region of repY was fused to the red fluorescent protein mCherry and cloned into the vector pBBR1MCS-2.	This work
pTF*repX*::mCherry	Translation fusion: the promoter region and the gene repX were fused to the red fluorescent protein mCherry and cloned into the vector pBBR1MCS-2.	This work
pTF*a1725*::eGFP	Translation fusion: the promoter region and the gene NGR_a01725 were fused to the green fluorescent protein eGFP and cloned into the e vector pBBR1MCS-2.	This work
p-2::*repX*	pBRR1MCS-2 containing the gene *repX* cloned into *Sal*I and *Hind*III restriction sites	This work
p-2::*a1725*	pBBR1MCS-2 containing the gene a1725 cloned into *Xho*I and *Hind*III restriction sites	This work
p-5::*repX*	pBRR1MCS-5 containing the gene *repX* cloned into *Sal*I and *Hind*III restriction sites	This work
p-5::*a1725*	pBBR1MCS-5 containing the gene a1725 cloned into *Xho*I and *Hind*III restriction sites	This work
pET-21a::*gqqA*	pET-21a containing the *gqqA* gene from *Komagataeibacter europaeus CECT 8454*	Valera et al., 2016
pET-21a::*gqqA*-M1	pET-21a::*gqqA* was modified by using site directed mutagenesis (T782C)	This work
pET-21a::*gqqA*-M2	pET-21a::*gqqA* was modified by using site directed mutagenesis (Δ835- 843)	This work
pET-21a::*gqqA*-M3	pET-21a::*gqqA* was modified by using site directed mutagenesis (C73T;G74C)	This work
pET-21a::*gqqA*-M4	pET-21a::*gqqA* was modified by using site directed mutagenesis (C559G; Δ562- 570; C571T)	This work
pET-21a::*gqqA*-M5	pET-21a::*gqqA* was modified by using site directed mutagenesis (A352G; C353T; C354C)	This work
pET-21a::*qsdR1*	pET-21a containing the *qsdR1* gene cloned into *Nde*I and *Xho*I restriction sites	Krysciak et al., 2011

Primers

Primers used and designed in the present research are summarized in the appendix (Table 21). Restriction sites are underlined. Primers were designed with the tool Clone Manager Professional CMSuite9 (© Scientific & Education Software) and ordered at Eurofins Genomics GmbH (Ebersberg, Germany).

Bacterial cultivation

E. coli strains were grown in lysogeny broth (LB) medium (Bertani, 1951) at 37°C overnight. Strains of the genus *Sinorhizobium* were cultivated in tryptone-yeast (TY) medium (Streit et al., 2004) or yeast-extract-mannitol (YEM) medium (Allen and Allen, 1950) at 28°C for 2 to 3 days. All required antibiotics and supplements were added after autoclaving. For solid medium 1.5% agar was added to the medium and the medium was poured into petri dishes. The different stock solutions and final concentrations of the antibiotics and supplements are listed in the following Table 3.

Agrobacterium tumefaciens NTL4 was cultivated in AT Medium (Tempe et al., 1977) with Tetracycline and Spectinomycin. The cells grew over night at 28°C. *Chromobacterium violaceum* grew in LB medium at 28°C.

Table 3: Antibiotics and other supplements used in this study

Supplement	Stock solution	Final concentration	Solvent
Antibiotics			
Ampicillin	100 mg/ml	100 µg/ml	H_2O
Kanamycin	25 mg/ml	25 µg/ml	H_2O
Rifampicin	25 mg/ml	25 µg/ml	MetOH
Gentamycin	30 mg/ml	10 µg/ml	H_2O
Tetracycline	5 mg/ml	2.5 – 5.0 µg/ml	EtOH (70%)
Spectinomycin	50 mg/ml	50 mg/ml	H_2O
Other supplements			
DAP	60 mg/ml	30 µg/ml	H_2O
Apigenin	1mM	1 µM	KOH
IPTG	1 M	1 µM	H_2O
X-Gal	40 mg/ml	40 µg/ml	H_2O

Sequence alignment

The sequence alignment was performed with T-coffee server (Notredame et al., 2000). The phylogenetic tree was created using Molecular Evolutionary Genetics Analysis X (MEGA X) software (Kumar et al., 2018) and default parameters.

I. Quorum sensing dependent expression of small proteins

In this part all the methods that were used for the identification and characterization of small proteins in the plant symbiont NGR234 are described.

Small ORF search on the symbiotic plasmid

The smORF search was done by using the platform Artemis (Carver et al., 2012). First all possible ORFs in a range between 30 nts and 180 nts were searched. All possible ORFs were trimmed to the start methionine. Only intergenic ORFs were used for the comparison with the transcriptomic data of the NGR234 wt and NGR234 Δ*ngrI*Δ*traI*_copy+ strain (Grote et al., 2016).

(The next part was conducted with the help of my colleague Dr Simon Güllert)

The raw, de-multiplexed reads available under the NCBI Gene Expression Omnibus accession number GSE78039 were mapped to the symbiotic plasmid pNGR234*a* using bowtie2 2.2.9 (Langmead and Salzberg, 2012) and allowing one mismatch in seed alignment. Subsequently, the htseq-count script from HTSeq 0.6.1 package (Anders et al., 2014) was applied in strand-specific mode to count the reads which map to putative small ORFs. After this, the calculation of log2FoldChange values, as well as differential gene expression analysis, was conducted using DeSeq2 (Love et al., 2014) of the Bioconducter 3.4 package in program R. Finally, the log2FoldChange values were visualized using Circos 0.67 (Krzywinski et al., 2009).

RNA-seq and transcriptomic analysis

For RNA-seq and transcriptomic analysis the states of growth of the tested strains were analyzed. NG234 strains were cultivated in 50 ml YEM media. The cells to be tested should be harvest both in the exponential phase and in the stationary phase. Thus, the cells were centrifuged after 24 h and after 55 h (6,000 x g, 10 min, 4°C). The supernatants were discarded carefully, the pellets were shock frozen with liquid nitrogen and stored at -70 °C. The RNA-sequences were performed at Genomics Laboratory of George-August-University of Goettingen.

Ribo-seq and translation analysis

The Ribo-seq was performed in cooperation with the group of Cynthia Sharma (University of Würzburg). NGR234 wt and NGR234 Δ*ngrI*Δ*traI*_copy+ were shipped to Wuerzburg. Then the cells were lysed, and the macromolecules were isolated by using sucrose gradient centrifugation. The RNA footprint was isolated, and the synthesized cDNA library should have been deep sequenced.

Operon analysis

Due to the narrow localization of the three ORFs, the question arises whether the three genes will be expressed as an operon or not. Therefore, the RNA of NGR234 Δ*ngrIΔtraI*_copy+ was isolated according to manufacturer's protocol (Direct-zol RNA-Miniprep Plus; Zymo Research, Freiburg, Germany). A cDNA library was constructed (SuperScript™ Vilo™ cDNA Synthesis Kit, Invitrogen™, Carlsbad, USA). This cDNA was used as template for qPCR with different primers (Table 4). The received amplificates should gave exclusions about the expression of this genetic region. The qPCR reactions were set up according to manufacturer's instruction using the SYBR® Select Master Mix for CFX (Applied Biosystems®, Life technologies, Texas, USA). The qPCR was running with CFX96 Touch™ Realtime PCR Detection System (BioRad Laboratories, Munich, Germany) with the following conditions: 1. 50°C for 2 min; 2. 95°C for 2 min; 3. 95°C for 15 s; 4. 55°C for 15 s: 5. 72°C 1 min 20 s; step 3 to step 5 was repeated 39 times, 6. melting curve 65°C – 95 °C. Data were analyzed by CFX Manager™ software (BioRad Laboratories) and agarose gel.

Table 4: Primer and product overview for qPCR

Product	Size [nts]	Primer forwards	Primer revers
repX	172	oKP42	oKP43
repY	192	oKP199	oKP200
repA0	448	oKP66	oKP67
repXYA0	825	oKP42	oKP67
repXY	367	oKP42	oKP200
repYA0	650	oKP199	oKP67

Transfer of plasmid DNA to bacteria cells

The plasmid DNA was transferred in bacteria cells by using different approaches. For transferring DNA to *E. coli* cells, the heat shock approach was used with chemical competent *E. coli* strains. The steps were conducted by previously published protocols (Sambrook and Russell, 2001). For the transformation of plasmids into competent rhizobia strains electroporation and conjugation were carried out.

Electroporation of Rhizobia bacteria

First rhizobia bacteria cells had to prepare for the electroporation. Therefore, 50 ml TY medium were inoculated by a fresh overnight culture of rhizobia cells with a start density of 0.05. The culture was incubated at 28 °C for round approximately 16 h to an OD_{600} of 0.4 -

0.6. The cells were chilled for 30 min on ice and then harvested by centrifugation for 20 min at 4,500 x g at 4 °C. All following steps were done by keeping the cells on ice. The supernatant was discarded, and the pellet was washed four times with 1 ml cold sterile H_2O_{bidest}. Lastly, the pellet was washed with 10 % glycerol. The cell suspension was distributed in aliquots of 90 µl and storage at -70 °C until required for electroporation.

For the electroporation, competent NGR234 cells were thawed on ice for 5 min. After adding of the target plasmid DNA to the cells, the suspension was mixed for 10 sec and incubated on ice for 30 min. A pre chilled electroporation cuvette (Biolabs products GmbH; Bebensee, Germany) was loaded with the mixture and subjected to a single pulse of high voltage (2,400 V, 200 Ω, 25 µF). After pulsing the cells, the cell-DNA mixture was then resuspended with 1ml TY broth. After incubation at 28 °C overnight, the cells were plated on selective TY agar plates and were incubated at 28 °C for 3 to 4 days.

Conjugation with Rhizobium sp. cells

For the conjugation of NGR234, *E. coli* WM3064 (Table 1) functioned as a donor strain which transferred the single stranded plasmid DNA into the recipient NGR234 cells via direct cell contact. For this first the donor strain and the acceptor strains had to be prepared. The evening before conjugation 30 ml TY medium were inoculated with a fresh colony of NGR234 and incubated at 30 °C overnight until an OD_{600} of 3 to 4 was reached. Cells were kept on ice until conjugation. For the donor strain, 30 ml LB broth was inoculated with a start OD_{600} of 0.05 and then incubated at 37 °C until an OD_{600} of 0.9 was reached. The cells were kept on ice until conjugation.

The conjugation started with the harvesting of the donor strain. Therefore, 5 ml of *E. coli* WM3064 were centrifuged in a falcon tube for 20 min at 4,500 x g and the supernatant was pipetted off accurately. 30 ml of the acceptor strain were pelleted in the same falcon tube for 20 min. The supernatant was pipette off carefully and the pellet was washed twice with 1 mL TY broth. The pellet was resuspended in 150 µl TY broth and the cell suspension was pipetted on a TY agar plate containing 0.3 mM DAP. After one day incubation at 28 °C, the cells were washed off from the plate with 1 ml TY medium and harvested by centrifugation (5 min; 4,500 x g; RT). The pellet was washed twice with 1 ml TY broth and a series of dilution was produced (10^{-1}- 10^{-7}). The five highest dilutions were streak out on selective TY agar plates, which were incubated at 28 °C for several days. The colonies were controlled by colony PCR.

Mutagenesis of *Sinorhizobium fredii* NGR234 by heterologous recombination

Different mutants of NGR234 were constructed in this work. The methods for constructing the mutants were carried out by the principles of the homologous recombination (Sambrook and Russell, 2001; Lassak et al., 2010). Therefore, the flanking regions of the target genes were amplified with Phusion polymerase according to the manufacture's specifications and were cloned into the suicide vector pNPTS138-R6KT. The primers that were used for construction of the mutants are listed in the following Table 5.

Table 5: Overview of primers that were used for the construction of the mutants

mutant	Flanking region	Primer forwards	Primer revers
NGR234 Δ*repX*	Upstream (468 nts)	oKP112b	oKP166
	Downstream (217 nts)	oKP116	oKP117
NGR234 Δ*repY*	Upstream (485 nts)	oKP178	oKP179
	Downstream (396 nts)	oKP180	oKP181
NGR234 Δa1725	Upstream (404 nts)	oKP108	oKP109
	Downstream (263 nts)	oKP110	oKP111
NGR234 Δ*ngrI*Δ*traI*	Upstream 369 nts)	oKP155	oKP156
	Downstream (418 nts)	oKP157	oKP158
NGR234 Δ*traR* gTF*repX*::mCherry	Upstream (404 nts)	oKP112b	oKP113
	Downstream (219 nts)	oKP116	oKP117
	mCherry (727 nts)	oKP114	oKP115

The final constructs were transferred to *E. coli* WM3064 cells by heat shock transformation. This *E. coli* strain was used as donor strain for the conjugation of the mutant constructs to the recipient NGR234 strains (0). Single recombinants carrying the mutant construct were selected on selective TY plates. In order to obtain double recombinant mutants, the heterogenetic strain was streaked on TY in the presence of 15 % (w/v) sucrose. The verification of the obtained mutants was done by PCR using different primer pairs (Table 6) and by sequencing the target mutant regions with specific primers (Table 6).

Table 6: Primers that were used for verification of the mutants

Mutant	Control region (Product size wt)	Product size mutant [nts]	Primer forwards	Primer revers
NGR234 Δ*repX*	Inner (1,191 nts)	--	oKP155	oKP158
	Outer (1,704 nts)	1641	oKP158	oKP142
NGR234 Δ*repY*	Inner (600 nts)	--	oKP178	oKP242
	Outer (1,063 nts)	967	oKP178	oKP181
NGR234 Δa1725	Flanking (736 nts)	667	oKP108	oKP111
	Outer (1,038 nts)	958	oKP149	oKP150
NGR234 Δ*ngrI*Δ*traI*	Inner (185 nts)	--	oKP164	oKP165
	Outer (2,257 nts)	1,873	oKP141	oKP142
NGR234 gTF*repX*::mCherry	Outer (1,438 nts)	2,149	oKP146	oKP147

The new constructed deletion mutants were used for further approach like confocal microscopy, copy number verification or transcriptomic analysis.

Confocal imaging of the genomic translation fusion

The genomic fusion of the red fluorescence protein mCherry was observed with confocal microscopy. TY medium was inoculated with the strain NGR234 Δ*traR::*gTF*repX*::mCherry. The start OD_{600} was set to 0.05. The culture grew for 48 h at 28 °C. The cells were harvested and washed twice with 1x PBS and then resuspended in 100 µl 1x PBS. The cells were incubated for 2 h at room temperature in the dark with ProLong Antifade Reagents for Live Cells (# P36975; Thermo Scientific) for preventing the loss of fluorescent signal. The cells were also stained with DAPI for observing the cells with the airyscan mode. Therefore, the cell solution was mixed with 0.1 % (vol/vol) of a 100 mM DAPI solution and then 10 µl of the microscope sample was spotted on a microscope slide and mixed with 15 µL 0.8 % (wt./vol) agarose. A cover glass was put on the slides and the cells were observed by using the confocal laser scanning microscope LSM800 (Zeiss; Jena Germany). The fluorescence was excited with laser wavelength for mCherry (516 nm) and DAPI (405 nm). The complete setting for microscopy of genomic translations fusion is listed below (Table 7).

Table 7: Setting for confocal laser scanning microscopy of genomic translation fusion

settings	mCherry	DAPI
Objective	Plan- Apochromat 63x/1.40 Oil DIC M27	
Reflector	None	None
Contrast method	Fluorescence	Fluorescence
Pinhole	266 µm	210 µm
Laser wavelength	516 nm (35.00 %)	405 nm (2.00 %)
Excitation wavelength	587 nm	353 nm
Emission wavelength	610 nm	465 nm
Detection wavelength	570-700 nm	400- 600 nm
Detector gain	850 V	850 V
Digital detector gain	1.0	1.0

Copy number analysis of pNGR234*a*

To analyze the effect of smORFs on symbiotic plasmid copy number a quantitative PCR was performed. Various cultures of NGR234 strains were inoculated with an OD_{600} of 0.05 in YEM media and then getting to grow for 96 h at 28°C. The gDNA of these cultures were isolated with the NucleoSpin Microbial DNA kit (# 740235; Macherey Nagel GmbH & Co. KG, Düren, Germany) according to manufactures' protocol for gram negative bacteria. The gDNA concentrations were measured and set to a final concentration of 2.4 ng/µl for qPCR approach. Quantitative PCR was performed with the CFX96 Touch™ Real-Time OCR Detection System (Bio-Rad Laboratories Inc., CA, USA) and the applied Biosystems™ 2x SYBR™ Select Master Mix according to supplied manual. The efficiency tests for each pair of primers (Table 21) were performed by establishing standard curves of 10-fold gDNA series dilution. For each gDNA sample four replicates were used for at least three times. The genes *rec*A and NGR_c03800 were served as reference genes. For determining the symbiotic plasmid copy number, the genes *repA* and *nifB* were chosen as target genes on pNGR234*a*. The copy number was calculated with the Bio-Rad CFX Manager 3.1 software.

Analysis of gene expression using mCherry-based promoter fusion

The promoter fusions were constructed by amplifying the target promoter region and fusing to the red fluorescent protein mCherry by using the Phusion polymerase (Table 21). All promoter fusions were inserted to the broad-host-range cloning vector pBBR1-MCS-2 and electroporated into the different NGR234 strains. The NGR234 strains harboring the construct grew for 24, 48, 55 and 72 h. Culture aliquots of 200 µl were transferred into a blackened

microtiter plate and analyzed with the Synergy™ HT plate reader (BioTek Instruments Inc., Winooski, VT, USA). The red fluorescence of mCherry was measured with an excitation filter (590/20 nm filter set) and an emission filter (645/40 nm filter set) by consideration of the absorbance at 600 nm. The amount of fluorescence units (FU) were corrected by the optical density at 600 nm (OD_{600}) of the culture. The data were analyzed with the supplied Gen5 software.

Analysis of protein expression using fluorescence-based translations fusion

For the translation fusion the promoter region and the target gene region were amplified and fused to the red fluorescent protein mCherry or to the green fluorescent protein eGFP. The translation fusion was integrated to the broad-host-range cloning vector pBBR1MCS-2 (Table 2) and electroporated to the different NGR234 strains. Also, negative controls of the translation fusion were constructed. Therefore, the start sites of the target genes were deleted by using PCR and the plasmid-based translations fusions as template. The NGR234 strains harboring the plasmid-based translation fusion and NGR234 cells containing the negative control grew for 48 h. The fluorescence was measured with the Synergy™ HT plate reader.

To verify that the fluorescence comes from the bacterial cells, the cultures containing the translational fusions were examined by confocal microscopy. Therefore 1 ml of the cultures were washed with 1 x PBS and resuspended in 100 µL 1 x PBS. 10 µL of the cells were spotted on a microscope slide and mixed with 15 µL 0.8 % (w/v) agarose. A cover glass was put on the slides and the cells were observed by confocal microscopy LSM800 (Zeiss; Jena Germany). The settings for the microscope are listed in Table 8.

Table 8: Confocal microscope settings for fluorescence imaging

	settings	mCherry	eGFP
fluorescence	**Objective**	Plan-Apochromat 100x/1.4 Oil DIC M27	
	Reflector	none	none
	Contrast method	fluorescence	fluorescence
	Pinhole	54 µm	54 µm
	Laser wavelength	561 nm; 0.10 %	488 nm; 0.14 %
	Excitation wavelength	587 nm	488 nm
	Emission wavelength	610 nm	509 nm
	Detection wavelength	570 - 700 nm	480 - 700 nm
	Detector gain	800 V	750 V
	Digital detector gain	1.0	1.0
bright field	**Detection wavelength**	ESID	ESID
	EM gain	2	2
	Detector	Photo diode	Photo diode
	Detector offset	3.041	3.720
	Digital detector gain	15.0	15.0

Phenotypic analysis of NGR234 strains

To compare the different NGR234 mutants with the NGR234 wt strain various phenotypic analyses were performed.

Growth curve

For identifying difference in the growth and assimilation between mutants and parental strain a growth curve was generated. Therefore, the cultures grew in TY media with a start OD_{600} of 0.05. The cultures were incubated under shaking conditions (120 rpm) at 28°C. Every two hours 500 µl sample was taken and the cell density was measured at 600 nm with the NanoPhotometer® NP80 (Implen GmbH, Munich, Germany). The cell density was plotted against time. Each strain grew in triplicates and the whole experiment was performed three times.

Thin-layer chromatography and *A. tumefaciens* soft agar overlay assay

AI molecule could be separated using thin-layer chromatography (TLC) and could be detected by using the reporter strain *A. tumefaciens* NTL (Shaw et al., 1997). Therefor crude extracts of acyl-HSL were generated from whole cultures of different NGR234 strains. The NGR234 derivatives grew for 48 h with a start density (OD_{600}) of 0.05 in 300 ml of TY

medium. The cells were harvest by centrifugation and the supernatants were transferred in a new flask. The acryl-HSL were extracted twice with 0.5 vol. ethyl acetate, were concentrated in rotary evaporator Syncore® Polyvap (BÜCHI Labortechnik GmbH, Essen, Germany), and resuspended in 1 mL ethyl acetate. Samples volume of 2 µL was applied to cellulose TLC plates. The run was developed with methanol- water (60:40, vol/vol), and air dried. The detection was carried out as previously described with the *A. tumefaciens* soft- agar screening (Zhu and Winans, 1998). A 10^{-8} M solution of 3-oxo-C8-HSL, a 10^{-6} M solution of 3-oxo-C10-HSL were dissolved in ethyl acetate and used as standards. The overlaid TLC-plates were incubated for two nights at 28°C.

Root hair curling Assay with *Vigna unguiculata*

To analyze the nodulation effect of the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant, supernatant of 48 h cultures cultivated in TY medium of either uninduced or induced cultures of NGR234 wt or NGR234 Δ*ngrI*Δ*traI*_copy+ were collected. The Nod factors were extracted with 0.4 volumes of *n*-butanol according to earlier published protocols (Lopez-lara et al., 1995). The organic layer was collected, and the butanol removed by rotary evaporation. The residue was resolved in H_2O overnight to a final volume of 1/100 of the initial culture.

The extracted Nod factors have been tested for the biological activity to the plant. Therefore, *V. unguiculata* seedlings were sterilized and germinated as previously published (Krysciak et al., 2011). The germinated seedlings were transferred into glass containers that were supplemented with glass beads and Hoagland medium (Hoagland 1950). The seedlings grew for additional 24 h under following conditions: day/night 24°C/19°C; 16/8 h; 60 % relative humidity. The small plants were transferred into a small plastic container and coated with 500 µL Hoagland medium and 500 µl supernatant extracts or 1 ml supernatant extracts of the NGR234 cultures. After 24 h incubation in the dark, the root hairs were analyzed using a Zeiss AxioCam microscope.

Nodulation of *Vigna unguiculata*

(The nodulation assay was conducted by Hilke Duin; UHH)

V. unguiculata sterilized seeds (p. 28) were germinated on 0.5 % (W7V) TY plates or onto NOD-plates covered with filter paper. The seeds were incubated in the dark at 28°C until germination. The germinated beans were replaced in glass containers containing NOD-medium (pH 6.5; Irmer *et al.*, 2015) and grown in a 16/8 h light/dark incubator and 24°C with an angle of 45°C (Barbulova and Chiurazzi, 2005). NGR234 strains were grown as described to an optical density of 0.8- 1.0. The cells were harvested, washed with NOD medium and resolved in the equal volume of NOD medium. This infection solution was used for incubation

with the five to seven days old seedlings. The infected seedlings were further incubated on NOD plates for five days and then transferred into glass container with solid NOD medium (1.2 % agar). Germinated seedlings were replaced into a 16/8 h light/dark regime and 24 °C with an angle of 45° degree for further growth (Barbulova and Chiurazzi, 2005). 52 days past infection (dpi) the roots were examined for the presence of nodules.

Recombinant protein expression and immunoblot analysis

For the recombinant protein expression, the pET-vector system with HIS-tag purification (The QIAexpressionist™; Qiagen GmbH; Hilden, Germany), or the pMAL-vector system with MBP-tag purification (New England BioLabs, Frankfurt, Germany) were used. Proteins were transferred to the membrane using the Western blot approach. The membrane-bound proteins were detected with different antibodies and subsequently visualized by enzymatic activity with antibody-bound enzymes.

Overexpression and purification of small proteins

For the expressional and functional analysis several small proteins were overproduced in *E. coli* and subsequent purified. For each small protein various tools were used and are described here.

The small protein RepX was recombinant produced in *E. coli* BL21 (DE3) containing the pMALc2x::*repX* overexpression construct. The primers that were used for construction are listed in following Table 9. For amplifying of the target gene region, the Phusion polymerase was used under manufacture instruction. The amplified gene *repX* was ligated into *Sal*I and *Hind*III restriction site of the pMALc2x expression vector. *E. coli* BL21 (DE3) harboring the expression construct pMALc2x::*repX* was grown in 1 l LB supplemented with 0.2 % (w/v) glucose and ampicillin. The expression of the protein was induced with 0.1 mM and was performed at 22 °C for 16 h. The cells were harvested by centrifugation, resuspended in column buffer (New England BioLabs) and lysed three times with French pressure cell (6.895 x 10^6 Pa). The cell lysate was incubated with Amylose Resin (New England BioLabs) and then the MBP fused protein was purified according to pMAL™ Protein Fusion & Purification System manual (New England BioLabs).

The small protein RepY was also purified using the pMALc2x vector system. The target gene *repY* was amplified and ligated into *Eco*RI and *Sal*I restriction sites of pMALc2x. The overexpression was performed similar to the RepX expression.

The protein RepA0 was recombinant produced in *E. coli* T7 SHuffle harboring the pET-21a::*repA0* expression construct. The amplified gene *repA0* was ligated into the *Eco*RI and *Hind*III restriction sites of pET-21a. The expression strain containing the construct pET-

21a::*repA0* was overproduced according to the handbook “The QIAexpressionist” chapter No.12 (QIAGEN, Hilden, Germany). The cells grew in 1 l LB supplemented with ampicillin. The protein expression was induced with 0.1 mM IPTG and the expression was carried out at 22°C for 16 h. The cells were harvested and lysed three time with French pressure cell. The cell lysate was mixed with Protino Ni-NTA Agarose (Macherey-Nagel, Düren, Germany) and together incubated at 4°C for 2 h. The His-tagged protein RepA0 was purified using the protocol no. 12 of the QIAexpressionist manual.

For overproducing the small protein A1725, the target gene a1725 was amplified and ligated into the *Eco*RI and *Hind*III restriction sites of the pMALc2x vector. This construct was then transformed to *E. coli* BL21 (DE3). The small protein was recombinant expressed similar to RepX and RepY expression.

Table 9: Recombinant expression of small proteins

Target gene	Primer forward	Primer revers	T_{ann} [°C]	Size gene [nts]	Size protein [kDa]
repX	oKP68	oKP82	54	172	5.5
repY	oKP239	oKP240	59	228	6.0
repA0	oKP76	oKP77	57	445	16.1
a1725	oKP120	oKP121	57	92	3.0

The buffers were changed using Vivaspin™ sample concentrator (GE Healthcare GmbH, Solingen, Germany) with a to the protein size matching molecular weight cutoff.

SDS-Gels, western blot and immune detection

Polyacrylamide SDS- Gels

Purified proteins were controlled with SDS- polyacrylamide gel electrophorese (SDS-PAGE; Laemmli, 1970). Usually, discontinuous Tris Tricine polyacrylamide gels with a 7% stacking gel and 16 % separating gel were used for small proteins analysis and Tris Glycine SDS-PAGE with 12 % separating gel for QQ enzyme electrophorese. The solutions that were used for gel preparation are listed in the following table.

Table 10: Solutions that were used for SDS-PAGE

Solution	Component	Concentration
Separating gel stock solution pH 8.8	Tris	1.5 M
	SDS	0.4 % (w/v)
	H_2O_{bidest}	*ad* 250 ml
Stacking gel stock solution pH 6.8	Tris	0.5 M
	SDS	0.4 % (w/v)
	H_2O_{bidest}	*ad* 100 ml
Acrylamide stock solution	Rotiphorese® Gel 40 bisacrylamide (37,5:1; Carl Roth GmbH, Karlsruhe, Germany)	
Ammonium persulfate (APS)		10 % (w/v)

For native polyacrylamide analysis the proteins were loaded on Mini-PROTEAN® TGX™ Precast Gels (BioRad, Munich, Germany). The electrophorese ran under the same Laemli conditions expect that denaturing SDS is lacking in running buffer, gel solutions and loading dye. If the protein gel is not used for downstream processing, the gel is stained with Coomassie staining solution (Sambrook and Russell, 2001).

Western blotting (Towbin et al., 1979)

Western blotting was accomplished in order to detect the purified small proteins with specific antibodies or Tag antibodies. The proteins and the prestained molecular weight marker (Figure 54) were transferred from the SDS-Gel to a nitrocellulose membrane (0.2 µm) by using the Trans-Blot Turbo transfer system (Bio-Rad Laboratories Inc.). After gel electrophorese, all components were equilibrated in transfer buffer (125 mM Tris, 192 mM Glycine, 20% (vol/vol) Methanol, pH 8.6) before using. Then the components were assembled in the following order: bottom cassette electrode (anode) - Bottom ion reservoir stack - blotting membrane - gel - Top ion reservoir stack - Top cassette electrode (cathode). The roller was used to remove any air bubbles between the layers. The Western blot ran for 30 minutes with the following settings: 30 min, up to 1.0 A; 25 V constant. After the run the membrane was used for immune detection.

Immune detection

For the immune detection anti-His or specific antibodies were used. The specific antibodies were ordered from Davids Biotechnology GmbH (Regensburg, Germany). Either a synthesized peptide (RepX, RepY and A1725) or a purified protein (RepA0) was used as

antigens. The RepX peptide was synthesized from GeneCust (Dudelange, Luxembourg) and the peptides RepY and a1725 were synthesized from Davids Biotechnology GmbH (Regensburg, Germany). Polyclonal antibody production was done by Davids Biotechnology GmbH. The antigens were used for the immunization of the rabbits and the immunization was done for 63 days (Table 11). The antibodies were purified by affinity purification and depletion.

Table 11: List of specific first antibodies with the corresponding antigen sequence

Name	Antigen
Anti-RepX-AB	MNFSVVVGPRGNQHKSESGGSCRILQGSLE
Anti-RepY-AB-1	DGRDPRLTLSG
Anti-RepY-AB-2	ESNPSAHLSQAGSFQPIVNPNLLDG
Anti-RepA0-AB	Purified protein
Anti-a1725-AB	LRLDRWRRFSFSALIMRHAVGASL

After the western blotting first, the membrane was washed two times with 1x TBS and then incubated in blocking solution (5 % milk powder in TBST buffer) overnight at 4°C. After washing the membrane three times with TBST the first antibody (Table 11) was placed to the membrane and incubated for one to two hours with gentle rocking. After primary antibody solution the membrane was washed three times with TBST. The secondary anti-rabbit antibody that was conjugated either with a horse radish peroxidase (HRP) or an alkaline phosphatase (AP) was incubated together with membrane for one hour. The detection of the protein bands was performed depending on the enzyme with which the second antibody was conjugated according to the manufacturer.

II. Structure analysis and characterization of GqqA- a Quorum Quenching hydrolase

In this part all methods that were used for structure analysis and characterization of the novel QQ enzyme GqqA are described.

Mutagenesis of GqqA

For the mutagenesis of specific aa in GqqA divers quick exchange mutagenesis on nucleotide level were performed. For this purpose, the mutations were done by using the Phusion Polymerase, the mutation specific primers (Table 12) and the construct pET-21a::*gqqA* as template (Table 2). The PCR products were controlled on an Agarose gel and

the charging stock was digested with *Dpn*I for removing the template DNA (30 min 37 °C, 10 min heat inactivation at 75 °C). After digestion, a phosphate group was added to the amplicon by using the polynucleotide phosphokinase under using manufacture protocol (Thermo Scientific, Bremen, Germany). With the added phosphate group, the linear amplicon could be ligated with the T4 ligase with the addition of 1 µL 10 mM ATP. Ligation took place at alternating 10 °C and 30 °C for 60 minutes. After a heat inactivation step a heat shock transformation in *E. coli* DH5α was done. The correctness of mutation was done with sequencing by using plasmid specific pET- primer (Table 21). The mutation construct with the right sequence was transformed in the overexpression strain *E. coli* BL21(DE3). The GqqA-Mut4 construct was generated by using a synthesized gene fragment (appendix, p. CXXXIII) containing the substitution and deletion side was used as primer.

Table 12: Overview of GqqA mutagenesis

Name	Mutagenesis	Primer I	Primer II
GqqA-Mut1	F261S	oKP221	oKP222
GqqA-Mut2	ΔR279; ΔK280; ΔP281	oKP223	oKP224
GqqA-Mut3	R25S	oKP225	oKP226
GqqA-Mut4	P187A; ΔP188; ΔP189, ΔG190; P191S	--*	--*
GqqA-Mut5	T118V	oKP229	oKP230

* a synthesized gene fragment was used as primer for the amplification of GqqA-Mut4 (p. CXXXIII, appendix)

Overexpression and purification of GqqA

The overexpression of GqqA was done like previously published (Valera et al., 2016) with some modifications. The strain harboring the expression plasmid pET-21::*gqqA* was grown in 1 l LB medium supplemented with specific antibiotic. When the culture reached an OD_{600} of 0.6 – 0.8, the culture was induced with 1 µM IPTG and the expression of the target protein was carried out for 16 h at 22°C. The cells were harvested and resuspended in lysis buffer (QIAexpressionist™) and were disrupted three times with a French pressure cell. The purification was performed according to the QIAexpressionist™ protocol no. 12. The protein eluates were buffered and concentrated using Vivaspin™ sample concentrator and 100 mM potassium phosphate buffer (pH 8.0).

Structure analysis of GqqA

For identifying the structure of the novel class of QS degrading enzyme different approach were used. First the aa sequence of GqqA was compared to known QQ enzymes and for secondary structure and protein structure prediction. The purified protein was shipped to Collaboratory partner for crystallization and data analysis.

For similarity analysis of GqqA to other QQ enzymes, 22 sequences of characterized QQ enzymes were used and compared to GqqA sequence by using T- COFFEE multiple sequence alignment server (http://tcoffee.crg.cat/). The resulted alignment of these 23 sequences was phylogenetic analyzed using Mega X (https://www.megasoftware.net/) and the Maximum Likelihood method with JTT matrix-based model. The bootstrap was set to 500 replicates.

(Crystallization experiments and data analysis of GqqA were performed in cooperation with the working group of Prof. Dr. Christian Betzel from the Institute of Biochemistry and Molecular Biology of the University of Hamburg. Nadine Werner performed the following experiments and established the best conditions for GqqA crystallization).

Purified protein (10 mg/ml) was used for crystallization. For an initial screening, the sitting-drop vapor diffusion method was used in 96-well plates. Small crystals were identified after three days. These seedings were used to get larger crystals that can be used for X-ray analysis. Therefore, the hanging-drop vapor diffusion method was used.
The X-ray diffraction data were collected at the P11 beamline at Petra III Synchrotron (DESY Hamburg, Germany). The data were analyzed using XDS software package (Kabsch, 2010).

***In vivo* QQ enzyme assays**

For a functional AHL- degrading assay the reporter strain *C. violaceum* CV026 was used. Therefore, the *E. coli* BL21 (DE3) strains harboring pET-21a::*gqqA* wt or once of pET-21a::*gqqA* M1-M5 mutant constructs grew in a preculture. Fresh LB medium was inoculated with 1 % (vol/vol) of the precultures and cultured at 37 °C until an OD_{600} of 0.6 was reached. Then the cells were induced with 0.1 mM IPTG. The proteins were now expressed recombinant for 16 h. Fresh LB medium with ampicillin and 10 µM of oxo-C8-HSL was inoculated with the expression strains in a ratio of 2 % (vol/vol) and incubated at 37 °C for 6 h. The culture supernatant was collected by centrifugation and 30 µL were used for the bioassay. Therefore, a grown CV026 culture was mixed with 24-fold volume of LB agar and poured in plates. Sterile filter papers were placed on the plates. The CV026 plates containing the supernatants of the cultures were incubated for 24 h at 28 °C. When CV026 cells become

violet, AI molecules are present, when CV026 cells remain colorless, the AI molecules are degraded.

In vitro enzyme assay and ESI-MS/MS analysis

(The enzyme assay and ESI-MS/MS measurements were performed in cooperation with Manuel Ferrer and Laura Fernandez-Lopez from Department of Applied Biocatalysis of the Institute of Catalysis (CSIC) in Madrid (Spain).

The *in vitro* enzyme assay was performed in triplicates and corrected for non-enzymatic transformation at 30°C. The substrate oxo-C8-HSL was added to a final concentration of 0.5 mg/ml (stock solution 100 mg/ml in DMSO) in one milliliter of 5 mM EPPS buffer, pH 8.0. A final concentration of 10 µg/ml GqqA (5 mg/ml stock solution in 40 mM HEPES, pH 7.0) was added immediately to the batch. Aliquots of the samples with and without protein were taken at different time intervals and the reaction products were analyzed by Mass spectrometry.
The MS analysis was performed in a hybrid quadrupole-time of flight mass spectrometer (QTOF, QSTAR pulsar I, ABSciex) supplied with a micro electrospray ion source for measurements in positive and negative mode. The assay products were dissolved in methanol in a 1:10 dilution and then introduced in the spectrometer. For identifying the products, the measurements were recorded in TOFMS mode. The parameters have been set as shown below: mass range 50-2000 Dalton; ion spray voltage (IS): 4500 V; ion source gas pressure (GS1): 10 psi; curtain gas pressure (Cur): 20 psi; declustering potential 30 V; focusing potential: 210 V; declustering potential 2: 15 V; collision gas: 3.

4. Results

NGR234 is able to fix nitrogen with more than 120 plant genera and is therefore a great model organism to study host-microbe interactions and rhizosphere processes. The genome of NGR234 consists of two genes coding for QS AI synthases (Schmeisser et al., 2009). In a previous work, several AI synthases mutants were constructed, including a NGR234 Δ*ngrI*Δ*traI*_copy+ double deletion mutant (Krysciak et al., 2014; Grote et al., 2016). Interestingly, the NGR234 Δ*ngrI*Δ*traI*_copy+ double deletion mutant showed upregulation of almost all genes on the symbiotic plasmid. It has already been demonstrated that not only the transcription is higher, but also the translation of proteins is increased. For example, this mutant produces a higher number of NOPs proteins and secretes them (Grote et al., 2016).

Root hair curling in the Absence of Apigenin

Due to the upregulation of almost all genes on pNGR234*a*, the question arises whether this upregulation is sufficient to enable the production of Nod factors that are necessary for the establishment of symbiosis (Figure 2) and root hair curling of the host. Therefore, the Nod factors were extracted from culture supernatants of the NGR234 wt strain and the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant, which were incubated with and without apigenin. The cultures cultivated with apigenin were selected as a positive control known to produce Nod factors and

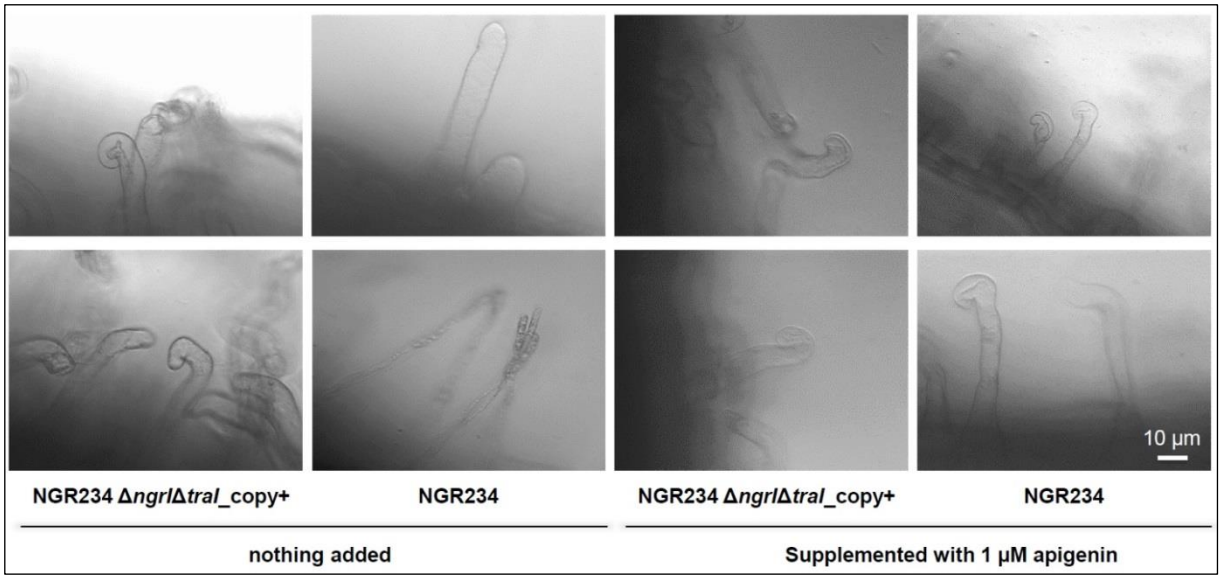

Figure 6: Root hair curling assay using Nod factor extracts of NGR234 wt strain and the NGR234Δ*ngrI*Δ*traI*_copy+ mutant.
Root hairs of *V. unguiculata* were incubated with extracted supernatants of NGR234 cultures treated either with 1 µM apigenin (right) or the non-treated cultures (left). The curved root hairs of the NGR234Δ*ngrI*Δ*traI*_copy+ where nothing was added are clearly recognizable, whereas the root hairs of the wt strain (NGR234) are straight. The control cultures, which were grown with apigenin, showed both the typical root hair curling. The images were taken with a Zeiss microscope and the provide software AxioVision.

trigger root hair curling. The extracted supernatants were incubated together with a germinated *Vigna unguiculata* bean for 24 h and then the root hairs were examined for their curling using a microscope. The root hairs incubated with the non-treated supernatant of the mutant show a distinct root hair curling while the wt supernatant without apigenin shows non-curled root hairs (Figure 6, left). Both extracted supernatants of the wt and the mutant strain, in which the cultures grew in the presence of 1 µM apigenin, trigger root hair curling after 24 h. (Figure 6; right).

I. Quorum Sensing dependent expression of small proteins

Small proteins are characterized as proteins much smaller than 100 aa synthesized directly by ribosomes and have been neglected in the past. The fact that in the absence of AI molecules almost all genes on the symbiotic plasmid were upregulated in NGR234 Δ*ngrI*Δ*traI*_copy+ mutant could be used as an advantageous tool to study the expression of small proteins in QS regulatory circuits. For this purpose, an ORF search was carried out and the received smORFs were compared with supposedly homologous proteins. The expression of selected smORFs was verified and a functional characterization for some of the resulting smORFs was started.

Prediction of smORFs on the symbiotic plasmid pNGR234*a*

With an Artemis six-frame translation ORF search (Carver et al., 2008) and settings to obtain entries with a minimum size of 30 nts, a total of 10,730 possible entries were found on the replicon. By focusing on new proteins below 60 aa, all entries longer than 183 nts were neglected and 4,074 entries remained. Due to this high number of possible new smORFs, only intergenic smORFs were considered. These 1,330 entries were mapped to the transcriptomic profile of the wt and the AI double deletion mutant strain (Grote et al., 2016). The entries which fit better to the transcript were analyzed more in detail. All sense/antisense or other frames ORFs were deleted and the amount of new possible small ORFs was reduced to 680 ORFs with a range of 30 to 180 nts. With the transcriptomic profile the possible smORFs could be evaluated for the fact that these ORFs would be really transcribed. The transcriptome profile was conducted with NGR234 cells grew 48 h in YEM medium (Grote et al., 2016). Because of the deep sequencing event only smORFs with a transcript hit level of ≥ 10 denote a real transcript. So, all ORFs without any transcript hit level and all ORFs with a transcript hit level under ten were neglected. 30 smORFs that showed no transcript in the wt strain either in the mutant strain and 489 smORFs in the wt and 453 smORFs in the mutant strain showed only background or no transcript (Table 13).

Table 13: Transcript hit number of the new small ORFs in NGR234 wt and NGR234Δ*ngrI*Δ*traI*_copy+ mutant

strain	total	No transcript	transcript hit levels (normalized reads)					
			total	≥ 10	≥ 50	≥ 100	≥ 500	≥ 1,000
wt	680	489	191	135	23	25	4	4
Δ*ngrI*Δ*traI*_copy+	680	453	227	163	31	25	4	4

All these ORFs, which have no transcript either in the wt or in the mutant could be ignored because these ORFs appeared to be false-positive. A total of 191 expressed smORFs with different transcript hits were found in the NGR234 wt strain (Table 13) and 227 expressed smORFs in the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant. So, a higher number of expressed smORFs were observed in the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant. The most smORFs showed a lower transcript hit rate between 10 to 50 readings. Some few ORFs were transcribed at high level with a reading size of more than 500 or 1,000 reads (Table 13).

When the transcriptomic profile of the wt strain was compared to the deletion strain (Table 14), the majority of smORFs had no differences in the transcription levels. These smORFs were not regulated and had a fold change level of -1 to +1 and were listened as zero in the summary table (Table 14). A total of 37 smORFs showed altered transcription rates in the mutant strain. Of these smORFs, 13 ORFs showed a lower and 24 a higher transcription rate. Most of the differently regulated genes were regulated with a lower fold change rate of +/- 2 or +/- 3 (Table 14).

Table 14: Transcription analysis of NGR234 wt and NGR234Δ*ngrI*Δ*traI*_copy+ in comparison to each other

	total	Fold change (log2) of small ORFs								
		-3	-2	0	+2	+3	+4	+5	+6	+7
Wt vs Δ*ngrI*Δ*traI*_copy+	251	2	11	214	16	3	2	0	2	1
+/- regulated		13			24					

The new smORFs were distributed on the whole symbiotic plasmid. All smORFs (smORF-1 to smORF-251) were listed with start and stop sites in appendix (Table 25) and were displayed in the circular diagram (Figure 7). The dots scattered over the light green, light red and white circles represented the transcripts of NGR234Δ*ngrI*Δ*traI*_copy+ compared to the parental strain. The smORFs were indicated by the green (leading strand) and blue (lagging strand) lines. All dots, which were located on the green circles, had an increased transcription in comparison to the parental strain. The dots on the red circles had a decreased transcription level. All dots in the white circles were not differently regulated. Three ORFs stood out when

locking at the circular diagram. Surprisingly, these ORFs were located together in the intergenic region in-between the AI synthase gene *traI* and the replication operon *repABC (*Figure 7, blue scare). All of them had a drastically fold change in the transcriptomic profile of the mutant. Due to this location close to the replication operon *repABC* and the autoinducer synthase gene *traI*, these ORFs would be interesting for further investigations. Because of the location proximity to the replication operon, these smORFs were referred as *repX*, *repY* and *repA0*.

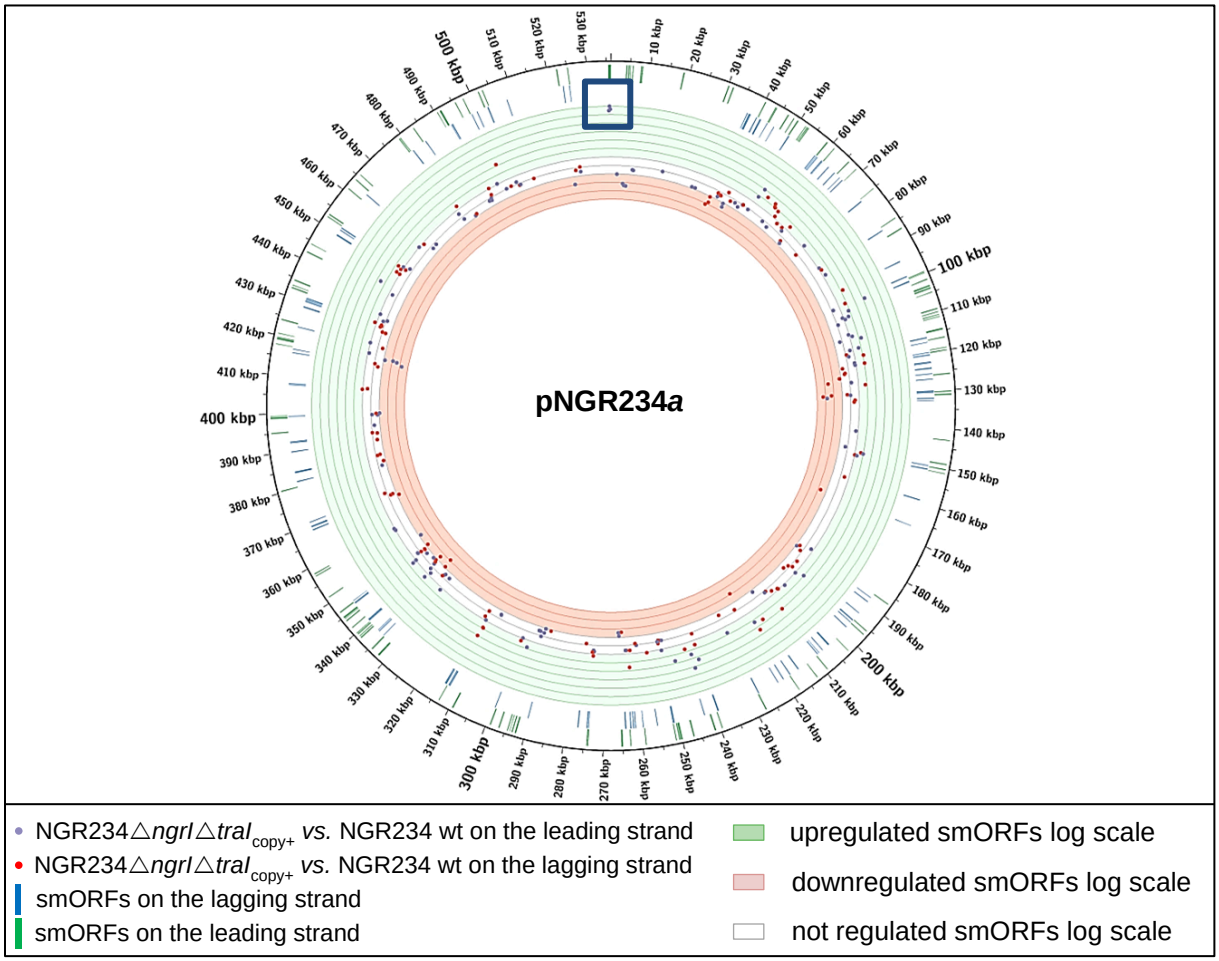

Figure 7: Circular presentation of the small ORF search mapped to the RNA-seq based transcriptome data for pNGR234*a* of the NGR234ΔngrIΔtraI_copy+ mutant vs. the wt strain.
The µ-ORF search was done with Artemis with the settings describe in material and methods. The circular diagram was calculated by using Circos software 0.67. The dots scattered over the light green, light red and white circles represent the transcripts of NGR234Δ*ngrI*Δ*traI*_copy+ in comparison to the wild type strain. The lines indicate the µ-ORFs on the leading (green) and the lagging (blue) strand. The circles in light green indicate log2 8; 7; 6; 5; 4; 3; 2, the white circles in-between the colored once represent log2 1/−1 and the light red circles indicate log2 −2; −3 for the sense transcripts.

All these new features (smORF-1 – smORF-251) are submitted to NCBI in an updated version of NGR234 symbiotic plasmid pNGR234*a* (Refseq: NC_000914.2; Genbank U00090.2).

Blast search of discovered smORFs

For further analysis of the discovered new small ORFs a Basic Local Alignment Search Tool (BLAST) search was performed. The Blast tool from NCBI (https://www.ncbi.nlm.nih.gov/) was used to find similarity between the new smORFs and sequences of another organism. For this purpose, the nts and the corresponding aa sequences were compared with sequences databases and the statistical significance was calculated. These possible similarities should give first indications of a possible gene or protein function.

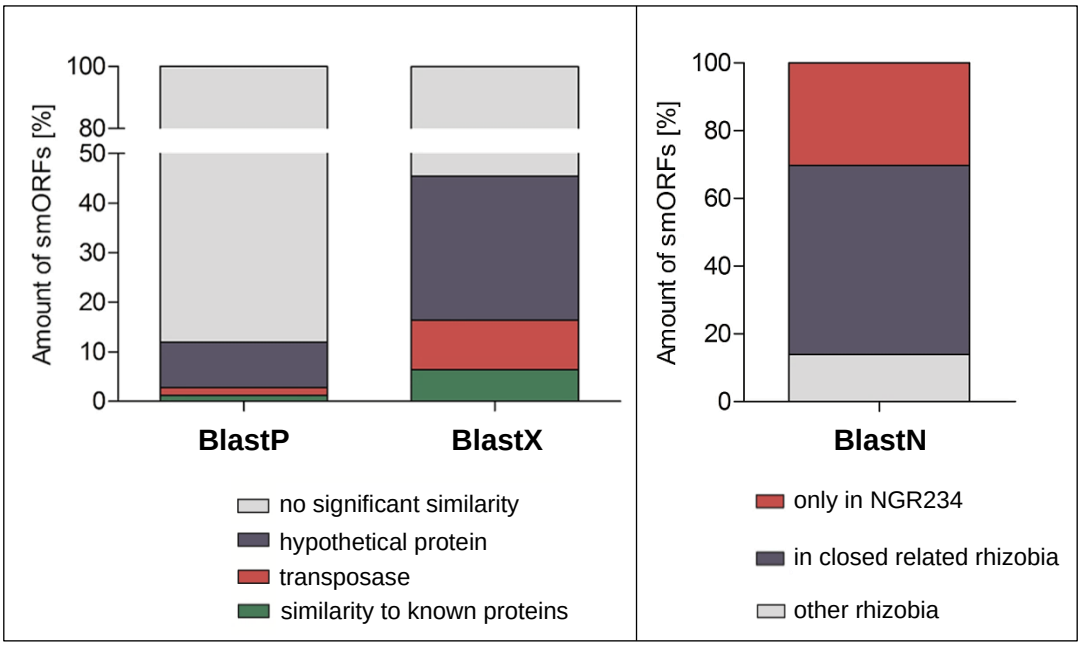

Figure 8: Blast search of smORFs.
The hits were classified in different groups and the relative amounts of smORFs in these groups were calculated. For BlastP and the non-redundant protein sequences database and the BlastP algorithm were selected. BlastX used also the non-redundant protein sequences databases. Both resulted hits were grouped in no significant similarity, hypothetical proteins, transposases and similarity to known proteins. For BlastN the nucleotide collection database (nr/nt) was used. The hits were splitted in the groups: “only found in NGR234, found in closed related strains and other rhizobia”.

First, the corresponding aa- sequences were compared with the BlastP non-redundant protein databases. The majority of the smORFs (221 smORFs; 88 %) showed no significant similarity to known proteins (Figure 8). 23 small ORFs (9 %) had similarities to hypothetical proteins found in closely related rhizobia strains. But most of the resulted hypothetical proteins were not small proteins and showed only a small percentage identity to the query sequences. Some few small ORFs (2 %) had similarities to different transposase families and

only two genes (1 %) showed similarities to known proteins. These proteins were encoded by the already annotated genes *repX* and *repA0* (smORF-249 and smORF-251; BK009410). A BlastX search that compares translates nts sequences to biological sequences showed that 137 of the 251 smORFs (55 %) had no similarities to other translated nts sequences. A third of the sequences could be assigned to hypothetical proteins, which occur mainly in closely related rhizobia and usually code for slightly larger proteins. Thus, these smORFs could be part of some larger genes, approximately 180 to 240 nts in size. Similar to BlastP, BlastX found some similarity to transposases. One tenth of smORFs had these similarities to this kind of enzyme class. 16 translated sequences have similarity to known proteins (6 %). Most of them are slightly larger proteins with a size of 100 to 200 aa. The last similarity search was performed with BlastN, comparing the pure nts sequences with similarities in the databases. Interestingly, only 30 % of these smORFs were found only in the genome of NGR234 (NC_00914.1) and 56 % are found also in closed related rhizobia strains. The smORFs were also present in *S. fredii* CCBAU 25509 (CP029453.1), *S. fredii* CCBAU45436 (CP029233.1), *S. fredii* CCBAU 83666 (CP023072.1), *E sojae* CCBAU 05684 (CP023069.1), *S. fredii* USDA257 (CP003567.1), *S. fredii* HH103 (HE616897.1) and *S.* sp. CCBAU 05631 (CP023065.1). Further 35 smORFs (13 %) were also identified in other rhizobia strains like *S. meliloti* or *A. tumefaciens*.

All Blast tools that had been used show no similarity of the smORFs to a known gene or protein. This means that none of these ORFs was described in earlier studies and a characterization of these new ORFs started at the beginning.

Verification of the expression to small proteins by using Ribo-seq analysis

Ribosome profiling (also known as Ribo-seq) was a technique that used deep sequencing to investigate *in vivo* translation. This allowed new insights into the identification and the expression of proteins synthesized by bacteria. This approach should be used to identify which of the new smORFs were expressed as small proteins. The Ribo-seq was performed in cooperation with the lab of Cynthia Sharma (University of Würzburg). The NGR234 wt strain and the NGR234 Δ*ngrI*Δ*traI*_copy+ strains grew in YEM medium for 55 h. The cells were harvested, and the polysome of the wt strain and the mutant strain were isolated. The ribosome-protected mRNA fragments that were known as ribosome footprints, were recovered and converted into DNA libraries, which were analyzed by deep sequencing.

NGR234 is a bacterium that is not easy to handle, and *E. coli* protocols can often not be adopted simply. Each step of the Ribo-seq must be optimized. The Sharma lab found the conditions for cell lysis and the sucrose gradient fractionation of macromolecules. But

previously, it was not possible to extract the RNA from the footprint fractions because the RNAs were degraded. The problem with RNA degradation must be solved and some other RNA extraction methods will be tested in the coming weeks and months. So that in the future the Ribosome profiling of NGR234 offered information, which smORF would be translated into small proteins.

Revaluation of the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant and construction of a new AI double deletion mutant

In the past a double deletion of both autoinducer synthases genes *traI* and *ngrI* was constructed (Grote et al., 2016). For the deletion of *traI*, a gentamycin cassette was integrated in place of the deleted *traI* gene. With the identification of the small ORF *repX*, it was clear that the promoter region of the new smORF *repX* that is located on the first 200 nucleotides of the autoinducer synthase gene *traI* was lost in the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant. A second fact was that the translation start ATG of *repX* was also changed to GAG. Thus, it was possible that the new smORF *repX* was not expressed correctly or was expressed uncontrollably. Because of these findings a new AI double deletion mutant (NGR234 Δ*ngrI*Δ*traI*) was constructed in this work, where the promoter region and the translation start ATG of *repX* were preserved (Figure 9, A). In this AI deletion mutant only the last 427 nts of *traI* were deleted and a stop codon after the first preserved 200 nts was inserted. The newly created NGR234 Δ*ngrI*Δ*traI* AI double deletion mutant was constructed by amplifying new flanking region of *traI* with maintenance of the *repX* promoter region and ligation of these fragments to the suicide vector pNTPTS138-R6KT (Table 2). The new deletion of *traI* was verified either with PCR products of different primer pairs (Figure 9; B) or with the sequencing of the PCR product of the outer primer pair oKP141/oKP142 (Table 21). Due to the insertion of the gentamycin cassette (NGR234 Δ*ngrI*Δ*traI*_copy+) or the deletion of the 427 nts of the 3' end of *traI* in the newly constructed AI mutant, the PCR should resulted in different product sizes compared to the wt strain (Figure 9, C). As a control, also the AI double deletion mutant NGR234 Δ*ngrI*Δ*traI*_copy+ was verified with the same primer pairs.

The PCR product amplified with outer primers (PCR 1, Figure 9) showed in the wt a DNA band with a size of 2, 287 nts. The PCR product of the mutant gDNA had a PCR product with a size of 1,873 nts (Figure 9; C). In contrast, the AI double mutant NGR234 Δ*ngrI*Δ*traI*_copy+ with an inserted gentamycin cassette showed a larger PCR product (2,648 nts). A second verification was conducted with inner primer (oKP164/oKP165). In the wt strain, the PCR product had a size of 185 nts, even as in the mutant strains no PCR product could be identified, because the primer could not bind to isolated gDNA (Figure 9; C; PCR 2). A last

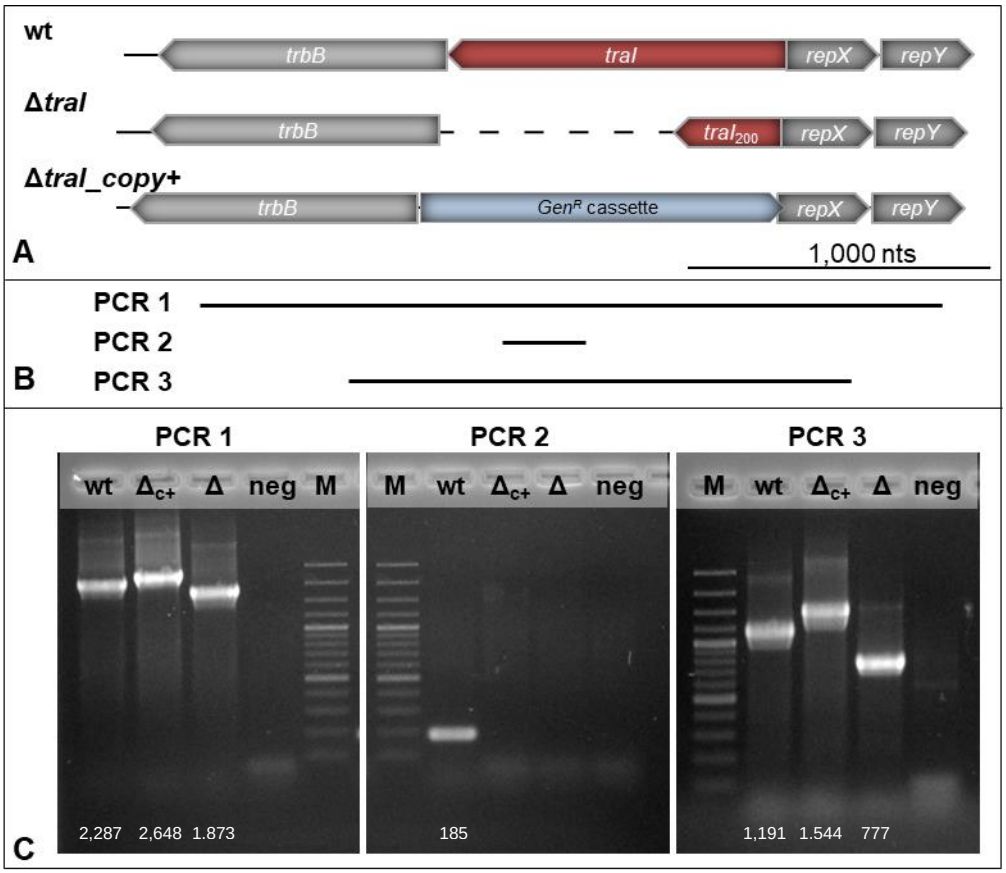

Figure 9: Verification of the in this work constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI*.
A: Genomic organization of the AI synthase gene *traI* in the wt strain, the in this work constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI* (Δ*traI*) and the first AI double mutant NGR234 Δ*ngrI*Δ*traI*_copy+ (Δ*traI*_copy+/Δc+). The deletion of the 3'-end of *traI* is displayed with separated lines (Δ*traI*). **B:** Overview about the received PCR products mapped to the genomic organization of *traI*. The lines shown here represent the different PCR products in wt strain. PCR 1 amplified a XX nts product. PCR2 resulted in a XX nts product and PCR product has a size of XX nts in the wt strain. **C:** 2 % (w/V) agarose gels for all three PCRs. The running conditions were 120 V for 27 min. The 100 bp plus DNA Ladder (#SM0321;Figure 53) was used as marker (M). H_2O_{bidest} was used as negative control in each PCR (neg). The samples were isolated gDNA of NGR234 wt (wt), NGR234 Δ*ngrI*Δ*traI* (Δ) and NGR234 Δ*ngrI*Δ*traI*_copy+ (Δ_{c+}). The PCR product size (in nts) were shown for each sample at the bottom of gel.

PCR based verification was conduct with primers of the flanging regions (oKP155/oKP158). The deletion of *traI* was shown by the different PCR products sizes. Whereas the wt PCR product had a size of 1,191 nts and the NGR234 Δ*ngrI*Δ*traI*_copy+ PCR product had a size of 1,544 nts, the newly constructed AI double deletion mutant showed a product with a size of 777 nts. In total, all PCRs resulted in a smaller PCR product for the gDNA of the newly constructed AI double deletion mutant compared to the parent strain or the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant. This suggested that the mutant was constructed correctly.

In addition to the genetic examination of the mutant, functional tests had been carried out. These functional tests were intended to show that the mutant was no longer able to produce AI molecules. For this propose, the AI indicator strain *A. tumefaciens* NTL4 (Table 1) was used for testing the potential of producing AI molecules. The parent strain and the new mutant strain were cultured in TY for 48 h. The AI molecules were extracted (p. 25) and were loaded on a thin layer plate. The AI molecules were separated on the plates by capillary action. The TLC plates were coated with *A. tumefaciens* supplemented with specific antibiotics and X-gal. When NTL4 turns blue, the strain was able to produce AI molecules. If NTL4 remained colorless, no AI molecules were produced (Figure 10, A).

The side of the TLC plate that contained extract of the wildtype, clearly showed four blue spots. Consequently, it seemed that NGR234 wt was able to produce four variants of AHL molecules. The upper spot corresponded to the 3-oxo-C8-HSL, which was synthetized by TraI. The second spot was assigned to oxo-C12-HSL. The spots below were unknown. The AI molecule that was produced by NgrI is not yet known, but it seemed that NgrI produced a mix of AI molecules. In contrast to the wt extract, the extract of the in this work constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI* showed no blue stain at the corresponding positions for AI molecules.

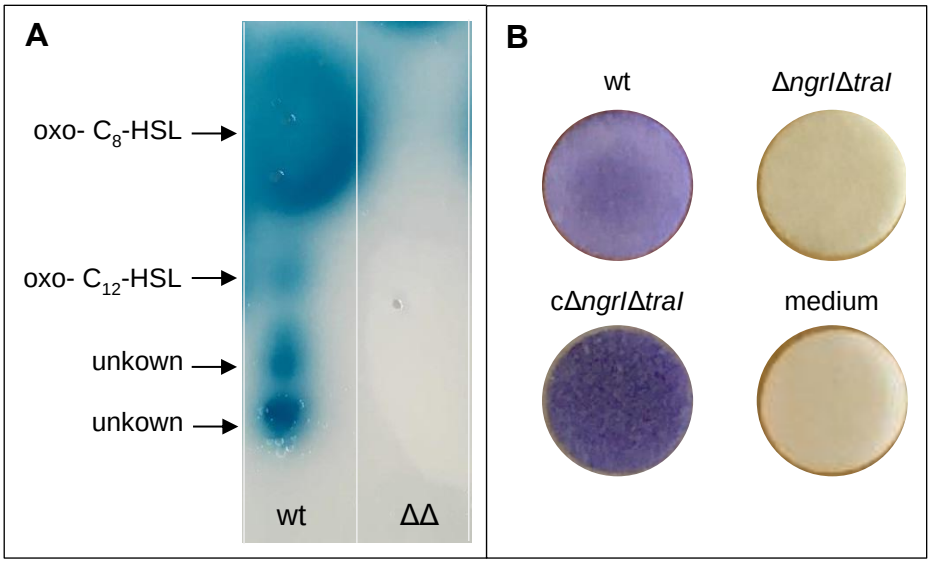

Figure 10: Detection of extracted AI from NGR234 wt and NGR234 Δ*ngrI*Δ*traI*.
A: TLC of the extracted AI from NGR234 wt and NGR234 Δ*ngr*Δ*traI*. The acryl-HSL were extracted twice with 0.5 vol. ethyl acetate, were concentrated in rotary evaporator and then resuspended in 1 mL ethyl acetate. Samples volume of 2 µL was applied to cellulose TLC plates. The run was developed with methanol- water. The TLC plate was coated with NTL4. A blue spot indicates am acyl-HSL. **B:** Detection of AI molecules with the reporter strain CV026. The AI of NGR234 wt, NGR234 Δ*ngrI*Δ*traI* and of a genetically complemented mutant were extracted as described above. 30 µl of these extracted AI molecules were pipetted onto a filter paper located on a with CV026 inoculated agar plate. When the filter paper turns purple, AI molecules were produced from the strains

For verifying that this effect turned from den deletion of the synthases, a genetically complementation on the broad host range-vector was constructed and transformed to NGR234 Δ*ngrI*Δ*traI*. A second test was performed with wt, mutant and complemented deletion strain. Therefore, the AI reporter strain *C. violaceum* CV026 was used to detect produced AI molecules. Extracted AI molecules were pipet on a filter paper placed on a LB agar plate inoculated with CV026. The violet color of the CV026 strain that was inoculated with the extracted supernatant of NGR234 parent strain indicated the production of AI molecules. Here a clear purple stain was observable (Figure 10, B). Whereas in the sample treating supernatant of mutant strain no purple color was visible. But if the mutant was genetically complemented with both AI synthases, the CV026 became purple again. The effect that on the TLC plates no AHL molecules were detectable in the mutant was a result of the deletion of both AI synthase genes *ngrI* and *traI*.

Growth properties of the mutant strain

NGR234 Δ*ngrI*Δ*traI*_copy+ showed a different growth property compared to the wt strain. This mutant could not reach the same cell density as the wt if both strains were in the stationary phase (Grote et al., 2016). This raises the question of whether the newly produced mutant had a similar growth characteristic. The newly produced AI double mutant, the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant and the wt were inoculated with a start OD_{600} of 0.04 and

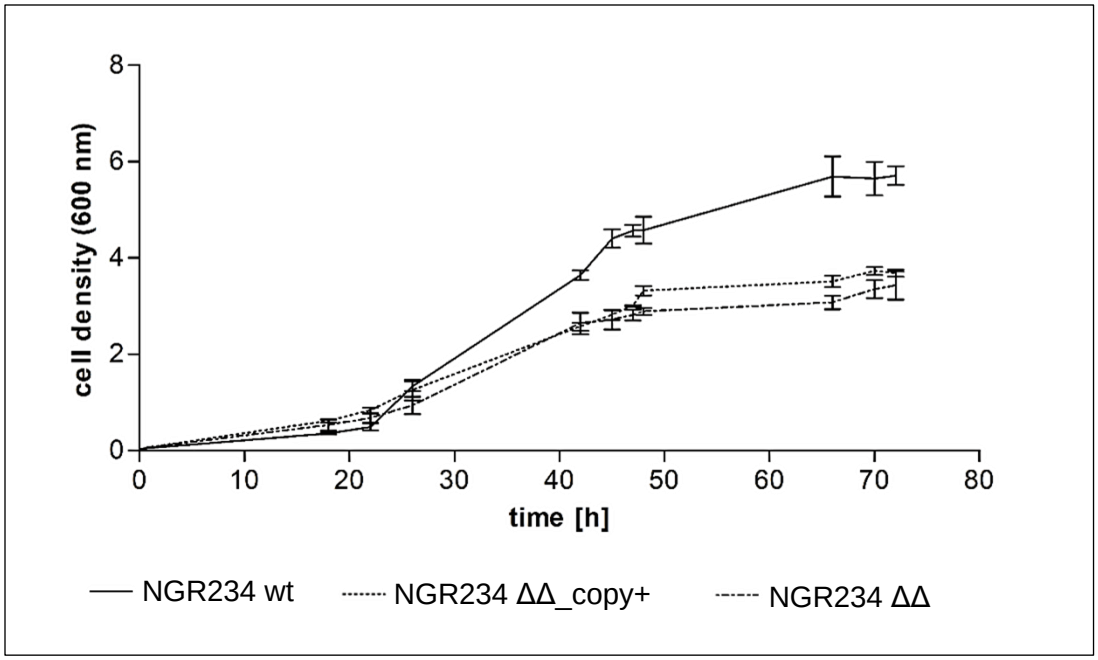

Figure 11: Growth curve of the newly constructed AI double deletion mutant
The cell densities of the NGR234 wt strain (black line) and both AI double deletion mutants are applied against time and are compared to each other. The NGR234 Δ*ngrI*Δ*traI*_copy+ are displayed with the dotted line and the newly constructed mutant NGR234 Δ*ngrI*Δ*traI* is shown in broken line.

incubated shaking at 28°C in YEM medium for 72 h. Every four hours always 50 µl culture were taken and the OD_{600} was measured, and the cell densities were applied against time (Figure 11).

All three strains reached the exponential growth phase after 20 h and remained in this phase for approximately 30 h. Only the wt achieved a cell density of more than 4 x 10^9 cells per ml. Both mutants did not exceed the cell density of 4 x 10^9. NGR234 Δ*ngrI*Δ*traI*_copy reached an OD_{600} of 3.68 ± 0.11 x 10^9 cells per ml and the newly constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI* attained a cell density of 3.43 x 10^9 ± 0.41 x 10^9 cells per ml. This growth curve displayed the result that were published previously (Grote et al., 2016). Without the production of AI molecules, NGR234 could not achieve the same cell density as wt, but they were in the same growth phase after the same time interval and could be used for transcriptomic analysis.

The effect of lacking AI molecules on gene expression

For testing the effect of both AI on gene expression, first a RNA-seq was performed. Therefore, the wt strain of NGR234 and the new AI double deletion mutant NGR234 Δ*ngrI*Δ*traI* grew shaking at 28°C in YEM medium for 24 and 55 h. All cells reached the stationary phase after 55 h. The cells were harvested and frozen with dry ice. The RNA-seq was performed in Goettingen in our co-laboratory (p. 20). The reads were normalized and mapped to the genome of NGR234. The fold-change and the p-value were calculated.

Table 15: Number of up- down regulated genes in NGR234Δ*ngrI*Δ*traI* compared to parent strain

time	Replicon	Total	down regulated	not regulated	up regulated
55 h	**cNGR234**	3,699	72 (2 %)	3,465 (94 %)	162 (4 %)
	pNGR234*b*	2,351	73 (3 %)	2,236 (95 %)	42 (2 %)
	pNGR234*a*	422	6 (1 %)	398 (94 %)	18 (5 %)

In total, 373 genes were differently regulated in stationary growth phase (Table 15). This suggested that only 6 % off all encodes genes in NGR234 Δ*ngrI*Δ*traI* were differently regulated. On the chromosome 6 % of the genes were differently regulated. The relative number of regulated genes for the symbiotic plasmid was similar to the relative amount of the chromosome. On the second megaplasmid pNGR234b only 5 % of the genes are regulated in a different manner. The comparison of the gene expression in the parent and the AI double deletion strain after 55 h is displayed in the circus diagrams (Figure 12). All three replicons were represented. The dots scattered over the light green, light red and white circles

represented the transcripts of the NGR234 AI double deletion strain in comparison to the wildtype strain. The dark green and blue bars represented the genes on the replicons. When the dots are in the light red circles, the transcription was downregulated. Dots on the light green circles indicated an upregulation of genes. All not regulated transcripts were displayed with the dots in the white circles. This figure clearly showed that most genes were not regulated differently, since 95 % of all red dots were in the white range. In addition, there were entire gene clusters that were regulated up or down. The gene cluster for the formation of flagella (NGR_c02610 - NGR_c03010) was upregulated on the chromosome. Interestingly a gene cluster was observed on the chromosome that was downregulated in the newly constructed AI synthase double deletion mutant. The gene cluster was located in-between the position 1,038,789 and 1,065,775 and included 30 genes that were involved in nitric oxide reduction. On pNGR234*b,* a gene cluster that is involved in formation of an ABC transporter (NGR_b21730- NGR_b21780) was downregulated in the mutant compared to the parent strain. On the symbiotic plasmid there were some single genes that were differently regulated like the mutated gene encoding the AI synthase TraI was no longer transcribed. Three genes encoding for hypothetical proteins was found in the transcriptomic date of the symbiotic plasmid that were transcribed in a QS dependent manner (NGR_a02340, NGR_a02760; NGR_a03270). There were also two genes encoding LysR family transcriptional regulator (NGR_a00810 and NGR_a01980) that are upregulated when both AI synthases genes are deleted.

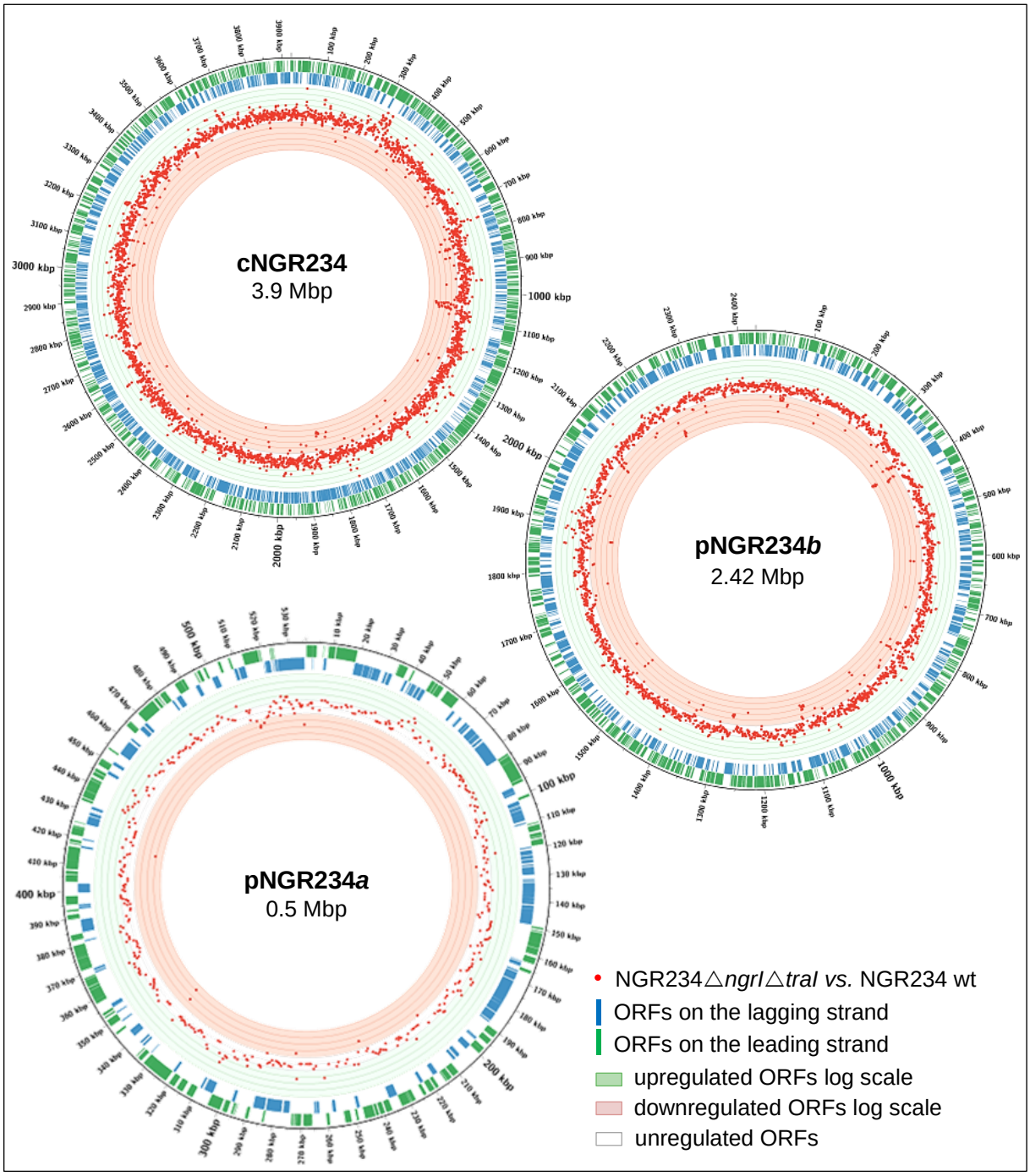

Figure 12: Transcriptomic analysis of the newly constructed NGR234 Δ*ngrI*Δ*traI* double deletion mutant

Circular diagram was calculated by using Circos software 0.67. The dots scattered over the light green, light red and white circles represent the transcripts of the in this work constructed NGR234Δ*ngrI*Δ*traI* mutant in comparison to the wild type strain. The lines indicate the ORFs on the leading (green) and the lagging (blue) strand. The circles in light green indicate log2 8; 7; 6; 5; 4; 3; 2, the white circles in-between the colored once represent log2 1/−1 and the light red circles indicate log2 −2; −3 for the sense transcripts.

The effect of QS to the novel smORFs on pNGR234*a*

Within this research project 251 new smORFs were identified on the symbiotic plasmid (p. 37). For comparing the expression of these smORFs in NGR234 Δ*ngrI*Δ*traI*_copy+ and the newly constructed AI double deletion mutant, the transcriptomic profile of the new AI double deletion mutant was also mapped to the symbiotic plasmid GenBank file containing all new smORFs.

The majority of these smORFs (90.03 %) are not differently regulated. Thus, in total there are 25 smORFs being differently regulated in the new AI double deletion mutant compared to the parent strain. Of these differently regulated smORFs, six smORFs (smORF-22, smORF-19, smORF-20, smORF-23, smORF-62, smORF-149) showed an increased and 19 a decreased transcription rate (Figure 13,Table 25 in appendix).

The smORFs smORF-112, smORF-98 and smORF-111 had an increased transcription in NGR234 Δ*ngrI*Δ*traI*_copy+ compared to parent strain. But in the newly constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI*, these genes were downregulated when the transcription rate is compared to the parent strain. In the transcriptomic data, there were found two genes that were upregulated in both variants of AI double deletion mutant. The transcription rate of smORF-149 and smORF-23 were upregulated in the newly constructed NGR234 Δ*ngrI*Δ*traI* and in NGR234 Δ*ngrI*Δ*traI*_copy+ mutant.

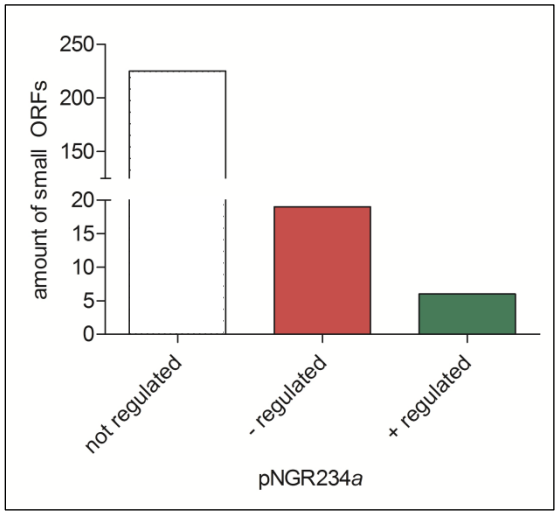

Figure 13: QS dependent expression of small proteins on pNGR234*a*
Bar diagram with amount of small ORFs on pNGR234a that were not regulated (white), down regulated (red) and up regulated (green) in NGR234Δ*ngrI*Δ*traI* compared to NGR234 wt.

However, in the data were found seven smORFs that were downregulated in both mutants. These seven smORFs had a higher negative fold2change in the newly constructed AI double deletion mutant than in the NGR234 Δ*ngrI*Δ*traI*_*copy*+ mutant.

In summary, the newly constructed mutant confirmed the QS dependent expression of small proteins in NGR234.

Functional characterization of selected small proteins

The small ORFs with the highest transcript rate were characterized in more detail. Therefore, the smORFs smORF-249, smORF-250, smORF-251 and smORF-104 were chosen and the expression and first hints for the function of these small proteins should be identified by using mutagenesis approach, immunoblotting with specific polyclonal antibodies and studies with translational fusions.

The smORF *repX*

The first smORF, that was characterized here, was the smORF-249 that was named as *repX.* This smORF was located at position 535,461 to 535,616 in the intergenic region in-between

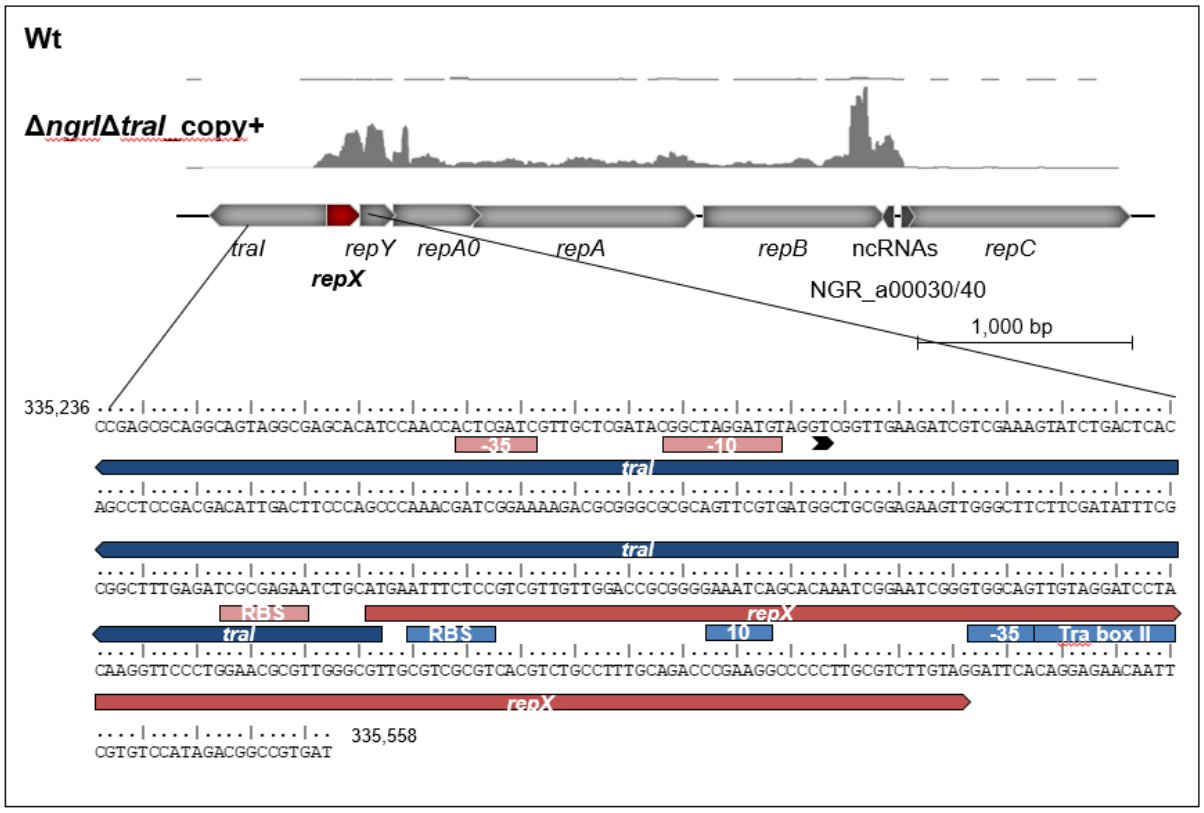

Figure 14: Genetic organization of the smORF *repX* with the corresponding transcriptomic profile and the coding and leader sequence of *repX*.
The genomic organization of the smORF *repX* (red box) is shown with the corresponding transcriptomic profile of the parent strain and the NGR234 Δ*ngrl*Δ*traI*_copy+ mutant. The grey shadows represent the reads. Both graphs were set to 100 reads. The lower part of this image illustrates the coding and the 153 nts leader sequence of *repX*. The promoter regions of *repX* (light red boxes) and of *traI* (light blues boxes) are shown. Some nts before the starts of *repX* and *traI* are located the RBS of respective genes. The transcription start of *repX* (black arrow) was calculated with the transcriptomic data.

traI and *repABC* on pNGR234*a*. SmORF-249 consisted of 156 nts and encoded for a probably 51 aa small protein RepX. That the expression of *repX* was increased to a 6.49-fold higher value in NGR234 Δ*ngrI*Δ*traI*_copy+ made this protein to a nice candidate for further characterization. A closer look at the nucleotide level revealed the overlapping of *traI* and *repX* genes and their promoter regions (Figure 14). Both genes pointed in the opposite directions and their promoter regions were located in the other gene. The *traI* ribosome binding site (RBS, also known as Shine Dalgarno sequence, SD), the general promoter elements (-10 and - 35 region) and TRA box II, which is the binding site for the transcriptional factor TraR, were shown in blue rectangles and was directly within the first 85 nts of *repX*. The *repX* general promoter elements and the RBS were illustrated in red rectangles and were located at the first 200 nts of the gene *traI*. A RBS was detectable 5 nts upstream of translation start ATG. Also, the general promoter region -35 and -10 could be observed in the sequence and were predicted with the Softberry internet tool (Solovyev and Salamov, 2010). A transcription start site could be verified in the transcriptomic data of the NGR234 Δ*ngrI*Δ*traI*_copy+ and is shown with a black arrow (black arrow; Figure 14). These findings led the transcription of a 154 nts leader sequence which function was not yet clear. The smORF *repX* started with an ATG as translation start and stopped with TAG as translation termination.

The nucleotide level of *traI* and *repX* demonstrated the contrary direction of both genes and the challenges with the transcription of both genes at the same time. If the *traI* gene was expressed, the binding of RNA polymerase to the *traI* promoter inhibited the expression of *repX* due to contrary direction of these both genes (Figure 14).

Similarity to other strains

The smORF was present in eight different rhizobia strains: *S*. *fredii* CCBAU 25509, *S. fredii* CCBAU45436, *S. fredii* CCBAU 83666, *E sojae* CCBAU 05684, *S. fredii* USDA257, *S. fredii* HH103, *S.* sp. CCBAU 05631 and *S. fredii* SMH13. Here, the nts sequence and the corresponding aa sequence of *repX* from NGR234 and from the closely related strains were aligned by using the similarity search tool BlastN and the ClustalW alignment tool. All eight ORFs showed a high similarity to *repX* (96.15 % identity). In total six nts of the query sequences differed to the homologue target sequences. Five nucleotide exchanges (red boxes, Figure 15) led to the alteration of the corresponding aa- sequence. The nts exchange G96T (green box, Figure 15) was a silent mutation and had no effect to the corresponding aa sequence of the small protein RepX. Since the function of RepX was still unclear, a possible

consequence of the aa changes was not known. A structural prediction of the small protein RepX could provide the first indications of the protein function.

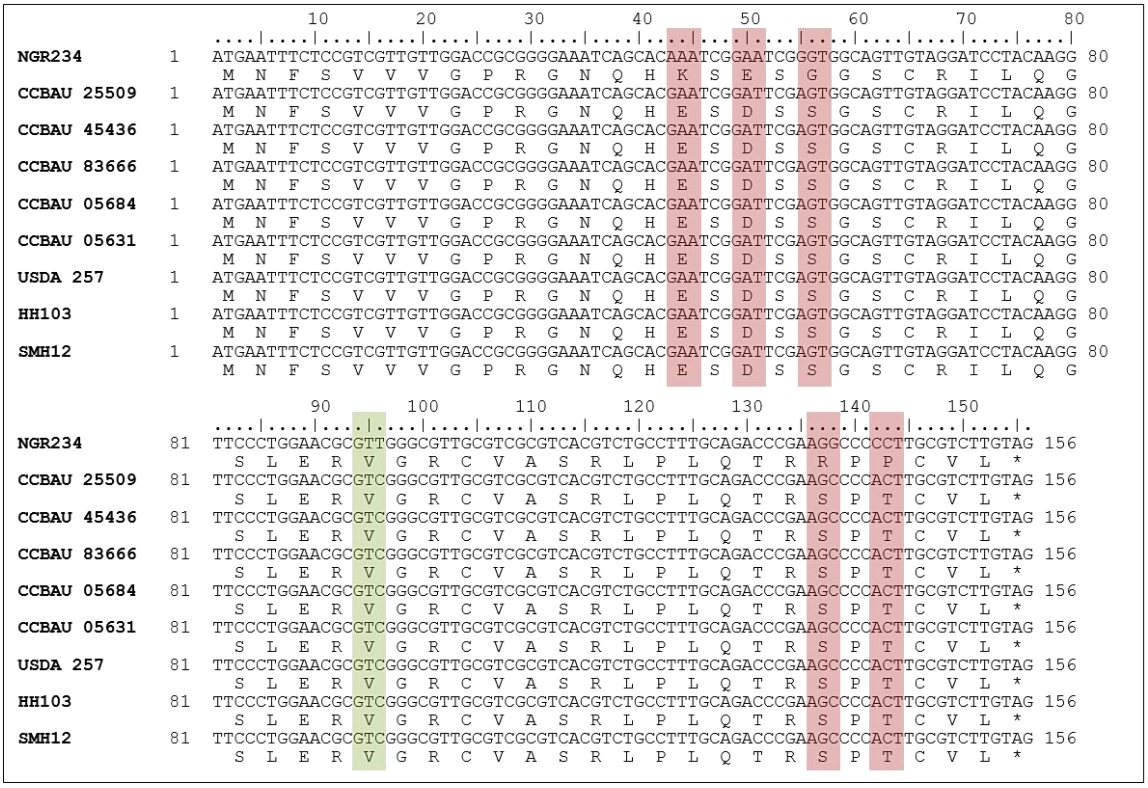

Figure 15: Sequence homology of repX to other rhizobial strains.
The nts-sequence of repX was blasted to the nucleotide collection (nr/nt) database of NCBI. The resulted homologous sequences were aligned to *repX* by using ClustalW alignment. the red boxes indicate differences in the nts sequence that result in an aa- exchange. The green box highlighted the difference T96C that has no effect of the aa sequence.

Homology model of RepX

The secondary structure of RepX was predicted using the Phyre2 online prediction tool. All in all, the protein was largely unstructured. Only 27 % of the peptide chain had a β-sheet structure and 23 % of the peptide chain had a helix structure. Thus, four secondary structures were predicted for the 51 aa peptide chain (Figure 16, A). Two β-sheets and two α-helices were identified in this peptide chain. A model of RepX protein structure was calculated with the prediction server Robetta (http://robetta.bakerlab.org; Kim et al., 2004) with the Ginzu PDB template identification and domain prediction protocol. The resulting PDB file was visualized with the program Chimera (Pettersen et al., 2004). The tertiary structure was represented from two sides to represent the entire structure. Both terms were displayed with the corresponding letters for the N- and C-terminus. The secondary structures were clearly visible. The interactions and bonds of the aa residues determined the tertiary structure of a protein. The structure begins with a β-sheet, followed by a large loop and the second β-sheet.

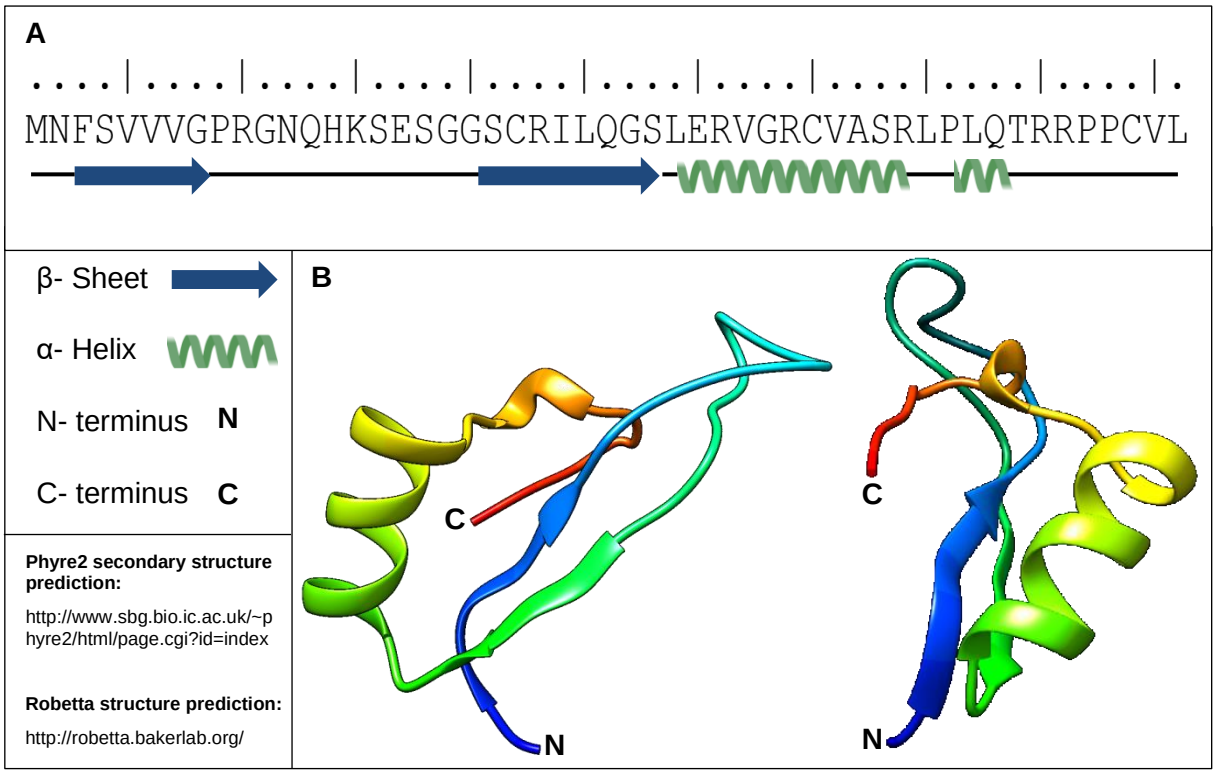

Figure 16: Structure prediction of RepX.
A: Secondary protein structure prediction of RepX is shown. The β-sheets are displayed in blue arrows and the α-helices are shown in green helix. The prediction of the secondary structure was calculated with Phyre2. **B:** Three-dimensional structure of RepX is represented from two sides. The N-terminus and the C-terminus are indicated with the letters N and C. The three-dimensional structure was predicted by Robetta structure prediction using Ginzu PDB template identification and domain prediction protocol.

After that the helices followed directly. Both β-sheets were connected by hydrogen bonds and were available as an antiparallel two strands of a β-sheet (Figure 16).

Regulation of the *repX* expression by using promoter fusion studies

Promoter fusion studies were performed to investigate the effects of *repX* gene expression. Therefore, the promoter region of *repX* was fused to the red fluorescence protein mCherry and both were inserted in the broad host range vector pBBR1MCS-2 and electroporated into various NGR234 strains. The strains harboring the promoter fusions grew in TY medium supplemented with selective antibiotics for 48 h at 28°C. 200 µl of these cultures were used for fluorescence and optical density measurements with Synergy™ HT plate reader. For red fluorescence, the emission filter 645/49 nm and the excitation filter 590/20 nm were used. The fluorescence of the cells was adjusted to an OD_{600} of 1.0 for better comparability. All fluorescence's were presented in the bar diagram (Figure 17). The parent strain showed a fluorescence of 1,251 ± 32.35 (n=15) fluorescence units. Compared to this fluorescence, the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant (1,194 ± 34.19, n=18), NGR234 Δ*a1725*+ mutant (1,177 ±

42.9, n=17), NGR234 Δ*traR* mutant (1,208 ± 25.15, n=18 and both single AI synthase deletion mutant NGR234 Δ*traI* (1,177 ± 35.64, n=18) and NGR234 Δ*ngrI* (1,116 ± 34.29, n=18) showed no different mCherry expression. There was no significant change in expression. The *p*- values were clearly above 0.05 between all named strains (Table 22, appendix). In the newly constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI,* a significant higher expression of the red fluorescent protein mCherry (p-value < 0.0001) was observed. Here, the fluorescence reached a 20 % higher fluorescence value than the wt strain with 1,589 ± 57.68 (n= 15) fluorescence units. When the symbiotic plasmid was lost as in strain ANU265, expression of mCherry was reduced to almost 54 % and fluorescence reached a value of 674 ± 44.85 (n= 16) fluorescence units. By adding *repX* on the broad host range vector pBBR1MCS-5 to ANU265 harboring the promoter fusion, a further significant reduction of mCherry expression could be seen (412 ± 67.03, n=12). Also, the fluorescence was decreased, if the smORF *a1725* (smORF-104) was added to the ANU265 cells. The

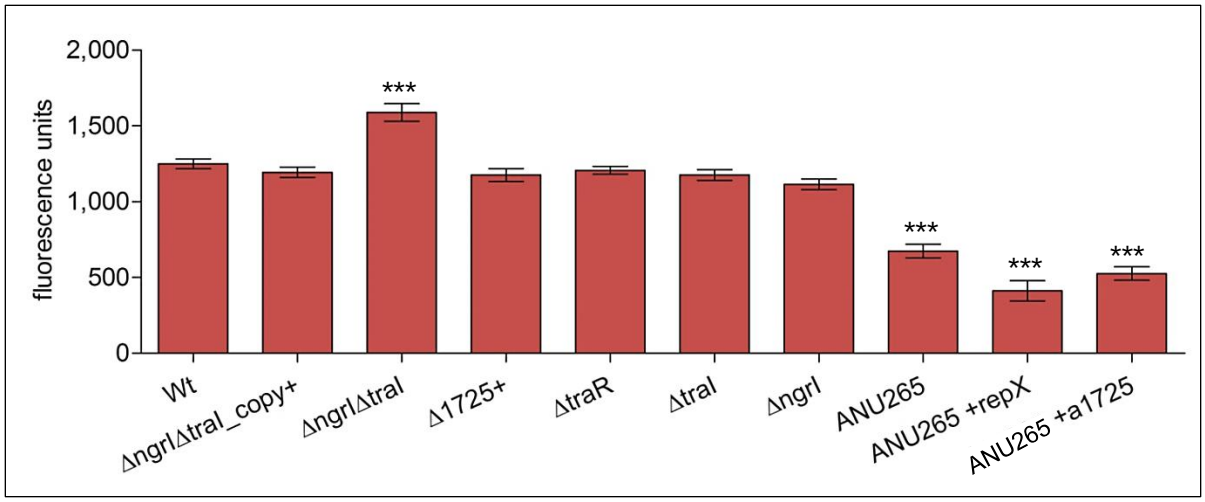

Figure 17: Promoter fusion studies of *repX* promoter fused to mCherry.
The bar diagram represented the promoter fusion pBBR1MCS-2::P_{repX}::mCerry in various NGR234 strains. The samples grew in TY medium for 48 h and the fluorescence and the optical density were measured with the Synergy™ HT plate reader. For better comparability, the fluorescence of each sample was calculated to an OD_{600} of 1.0. The stars indicated the significance of different fluorescence compared to the wt strain.

expression of mCherry that was under the control of the *repX* promotor was decreased to 527 ± 5.62 fluorescence units. A significant difference to the expression of mCherry in ANU265 (p-value: 0.0287) was observed.

Summarized the red fluorescence protein mCherry expression was higher in the newly constructed AI double deletion mutant and lower in ANU265 where the symbiotic plasmid is absence. Thus, the transcription of *repX* was dependent on QS and something that was located on pNGR234*a* that was not yet known.

Translation fusion to verify the protein expression

For the verification of the *repX* expression to a functional protein, translation fusions with the red fluorescence protein mCherry were constructed. Therefore, the promoter region, the gene *repX* itself and the gene for the fluorescence protein mCherry were fused and integrated to the broad host range vector pBBR1MCS-2. Translation fusions with the translation start site ATG and without the ATG were created. The final constructs were sequenced and were controlled of their correctness. The sequences were shown in the appendix (Table 24). The vector carrying the translation fusions were electroporated to NGR234 cells. The cells grew for 48 h in TY medium, the fluorescence of the cells was measured, and the cells were obtained under the confocal microscope (Figure 18).

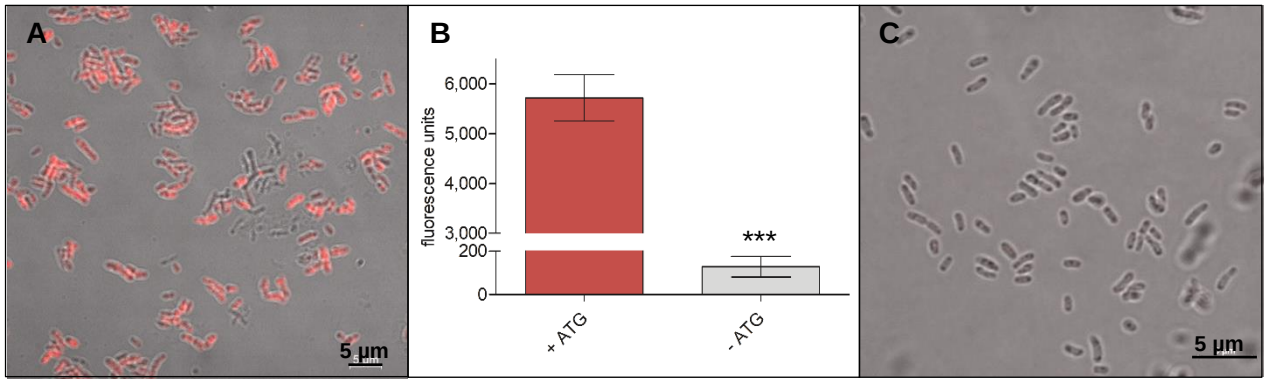

Figure 18: Translation fusion of RepX with the red fluorescence protein mCherry with and without ATG as translation start

For the translation fusion the promoter region and the gene *repX* were fused to the red fluorescence protein mCherry and these fragments were ligated into the broad host range vector pBBR1MCS-2. The translation fusion was electroporated into NGR234 cells. The fluorescence of the RepX::mCherry translation fusion was compared to the RepX::mCherry fusion without the translation start site ATG of *repX*. **A:** Microscope imaging of the translation fusion of RepXmCherry harboring the translation start ATG. This image was taken with the confocal microscope LSM800 airyscan (Carl Zeiss). The settings for the RepX::mCherry fusion are listed in Table 8. **B:** Fluorescence measurements of the both RepXmCherry translation fusion. Fluorescence was measured with the Synergy™ HT plate reader (Biotek Instruments GmbH) and the provided software Gen5. The excitation filter 590/20 and the emission filter 645/40 were used for measurements. **C:** Microscope imaging of the RepX::mCherry translation fusion lacking the translation start site ATG. The settings for taking this image were listed in Table 8.

The NGR234 cells including the plasmid-based translation fusion with translation start site ATG showed high fluorescence (5,718 ± 810) compared to the strains without translation start site (128 ± 83). So here a significant (p-value: < 0.0001) difference in the fluorescence was observed (Figure 18; B). Visualizations of the cells with the confocal microscope LSM800 airyscan (Carl Zeiss, Oberkochen, Germany) showed a different fluorescence in the cells containing the translation fusions. The cells that carried the translation fusion with ATG showed clear red fluorescence in the NGR234 cells (Figure 18; A), whereas the cells lacking

translation start ATG did not have any fluorescence (Figure 18; C). These measurements suggested that the gene *repX* encodes for a protein.

Genomic translation fusion of RepX

For localization and verification of the small protein RepX in the NGR234 cells a genomic translation fusion was constructed by using the recombinant homology and integrated into the genome of NGR234 Δ*traR.* The resulted genomic fusion of gTFrepX::mCherry was verified by sequencing that was shown in the appendix (Table 24). The cells grew in TY medium for 48 h and then harvested by centrifugation (RT, 6,000 x g, 5 min). The cells were washed with PBS and resuspended in 1/100 vol of starter culture size. For saving the fluorescence the cells were incubated with ProLong Antifade Reagents for Live Cells for two hours. By using LSM800 airyscan detection method the DAPI stained cells and the red fluorescence translation fusion could be observed. With this genomic translation fusion, the expression of *repX*::*mCherry* to the protein RepX::mCherry could be observed by microscope imaging. In the overview a heterologous expression of the fusion could be observed. Red fluorescence was not visible in every NGR234 cell (Figure 19, bottom).

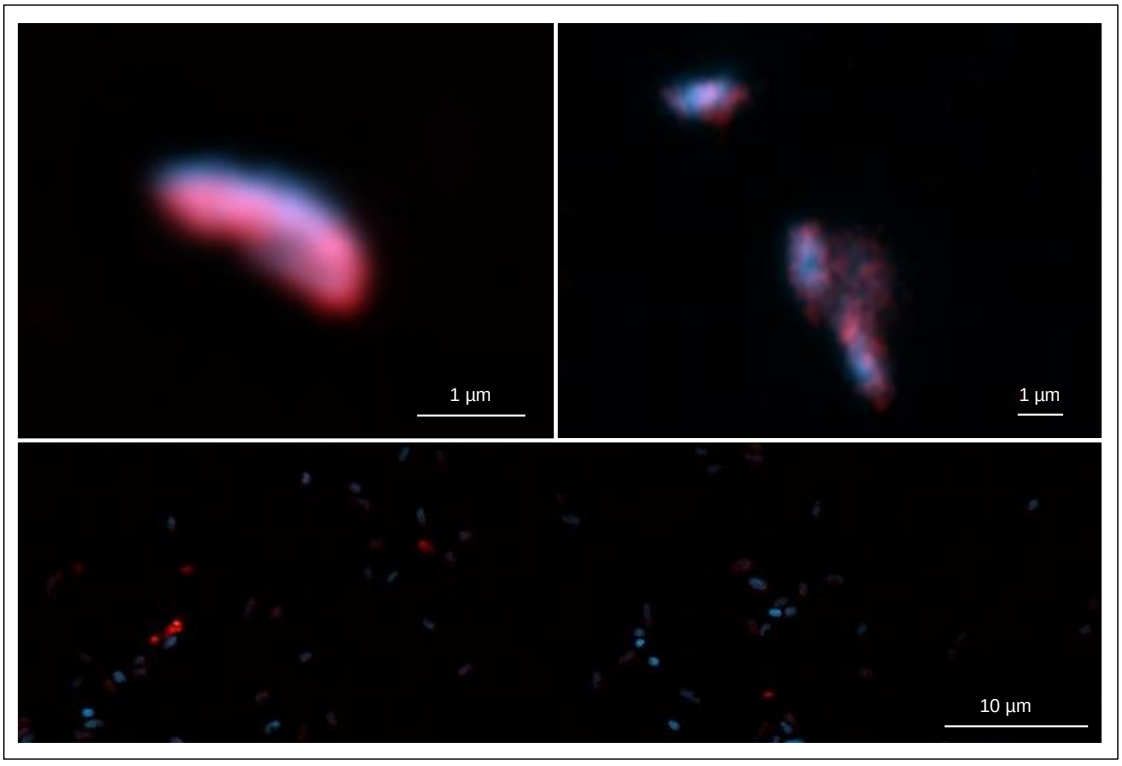

Figure 19: Confocal microscope imaging of the RepX::mCherry genomic translation fusion in NGR234 Δ*traR*.
The gene repX was fused to the red fluorescence protein mCherry and integrated into the genome of NGR234 Δ*traR*. For microscopy, the cells grew for 48 h and prepared for fluorescence imaging. The settings for the microscope LSM800 airyscan were listed in material and methods (Table 7). Here two close-up and one overview exposures were shown.

Overexpression of RepX in *E. coli*

For proteins characterization assay a fast purification system in *E. coli* was of advantage. Therefore, different approaches were tested. First the pET-vector system was applied. But in different *E. coli* expression strains no RepX was overproduced (data not shown). So, the pMAL-vector system was tried out. The gene *repX* was amplified and ligated in the pMALc2x vector. The final construct was sequenced (p. CXXXIV, appendix) and transformed to the expression strain *E. coli* BL21 (DE3). A culture of one liter was induced with 0.1 mM IPTG and the cells were incubated over night at 22 °C. The harvested cells were resuspended with column buffer and lysed with pressure by using the French press. The cleared cell lysate was incubated with amylose (New England BioLabs GmbH, Frankfurt/Main) at 4 °C and after the incubation of 2 hours the protein was purified by manufacture protocol New England BioLabs GmbH). After purification, the single elution fractions were pooled by using the Vivaspin 6 concentrator columns with a 30 kDa cut off. The protein concentrations of each fraction were measured, and 20 µg of each fraction were loaded on 16 % (v/v) Tris Tricine SDS- Gel. The proteins were transferred to PVDF membrane and the purified protein RepX was detected by using the specific anti-RepX-antibody (Figure 20). The primary antibody was detected with an anti-rabbit-antibody. MBP-RepX (Figure 20, A) was detected by using alkaline phosphatase and NBT and BCIP as substrates. A synthesized RepX peptide was load on gel as control (see Table 11 for peptide sequence). A clear violet protein band was observed in all wells excepted the control well with size greater than 40 kDa. These protein bands had a molecular weight that indicated the recombinant expressed fused protein MBP-RepX. Additional to the more than 40 kDa protein band in the sample, which was treated with the protease Factor Xa (New England Biolabs), a protein band was seen less than 10 kDa. This protein band showed a similar size such as in the control sample and indicated the single protein RepX. The second membrane was treated with HRP and the corresponding Enhanced chemiluminescence Western Blotting substrates (Figure 20, B). This method was highly sensitive for detecting proteins. Compared to the colorimetric method (Figure 20, A), a clearer protein band that indicated the single RepX protein was observed. This protein showed a molecular weight smaller than 10 kDa and fitted well to the calculated weight of RepX with 6 kDa.

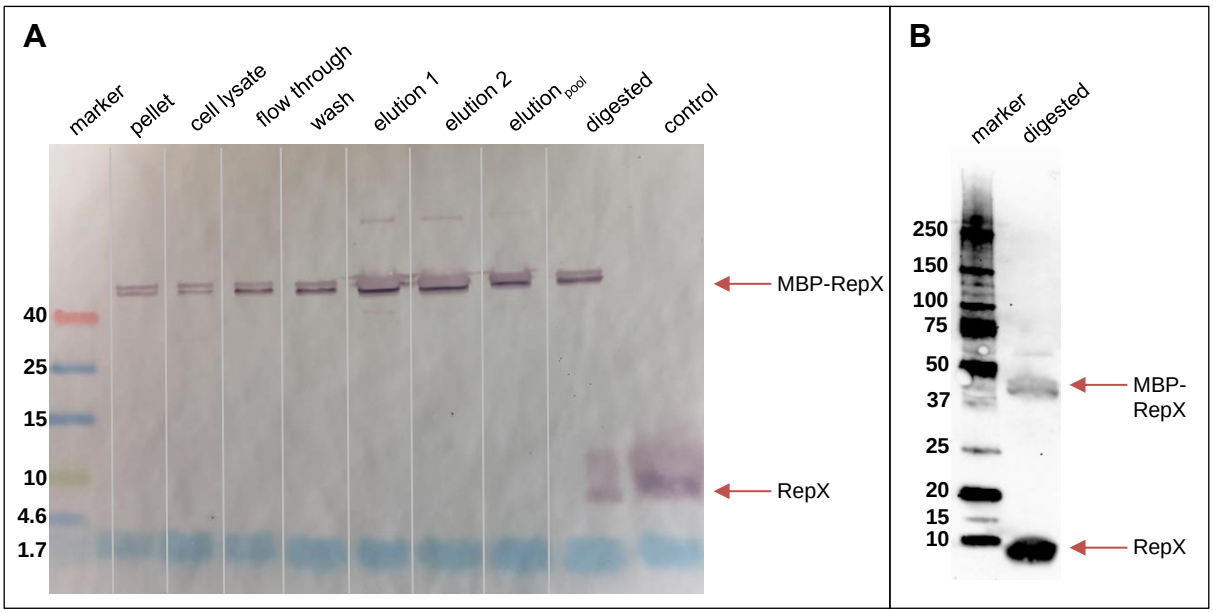

Figure 20: Recombinant protein purification of MBP-RepX.
A: 0.2 µm Nitrocellulose membrane with RepX purification. RepX was detected with specific RepX-antibody. The detected proteins were shown in purple protein bands. In addition to the purification samples, the sample that has been treated with Factor Xa (digested) has also been applied. The synthesized peptide RepX was used as control. The protein ladder Spectra Multicolor Low Range Protein Ladder was used as standard. 20 µg protein samples were loaded in each well. **B:** 0.2 µm PFDV membrane that was loaded with the factor Xa treated sample of MBP-RepX. The Precision Plus Protein™ WesternC™ Protein Ladder was used as standard.

The recombinant expression of RepX in *E. coli* could be verified with both immunoassayed membranes. *E. coli* could synthesize the small protein RepX that is encoded by a smORF and it was possible to purify RepX from *E. coli* cell lysate. The fact that the specific anti-RepX antibody was able to bind to the target antigen showed the correct expression of RepX in *E. coli*.

Pulldown assay with purified MBP-RepX and cell lysate of NGR234 wt

The purified protein MBP-RepX was used for a binding test with NGR234 wt cell lysate. Therefore, the MBP-RepX was incubated with the cell lysate and then purified with Amylose resin. All cell lysate components binding to RepX were co-purified with amylose resin. All purifications steps were loaded onto SDS gel and then blotted to a nitrocellulose membrane. The cell lysate of NGR234 wt and of NGR234 Δ*ngrI*Δ*traI*_copy+ were used as control. The membrane was incubated with the specific anti-RepX antibody solution and then detected by chemiluminescence (Figure 21, A). The cell lysate was also treated with anti-RepA0 antibody solution (Figure 21, B).

In each purification step of the pulldown assay, a protein was observed with a size between 40 and 55 kDa and that fitted with purified protein MBP-RepX. There were also some larger

protein bands visible. One protein band had a size of more than 70 kDa and could be a dimerization of MBP-RepX. Also, the highest detected protein that was seen in the eluates (e1, e2 and e3) at a size of more than 130 kDa, could be MBP-RepX that bound to other MBP-RepX proteins. Interestingly, a protein was detected with the specific anti-RepX antibody in the flow-through and had a size of approximately (approx.) 30 kDa. This protein was also observed in the pure cell lysates of NGR234 wt and NGR234 Δ*ngrI*Δ*traI*_copy+ (wt, c+). The cell lysates that were treated with the specific anti-RepA0 antibody solution had also this protein band, with a molecular weight of around 30 kDa (anti-RepA0, Figure 21, B). In summary, the pulldown assay showed only different protein bands with three different molecular weights showing the purified MBP-RepX. An LC-MS/MS analysis files of these protein bands and mapping these data to NGR234 genome did not yield any results on possible binding partners.

But when the cell lysates were incubated either with anti-RepX antibody or with anti-RepA0 antibody, a protein band with a molecular weight of around 30 kDa indicates that RepX could bind to RepA0. Since RepX-RepA0 has molecular weight of 22 kDa, a third binding partner with a molecular weight of around 8 kDa could be exist.

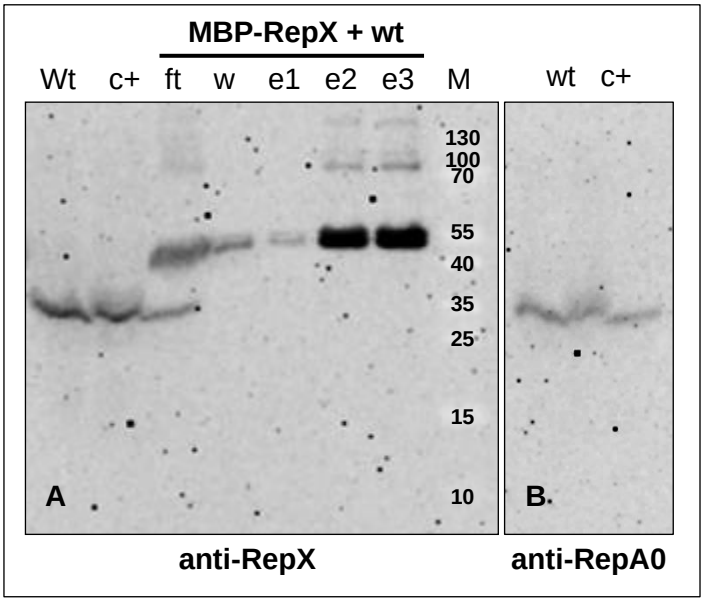

Figure 21: PVDF membrane with the samples of the RepX pulldown assay
A: PVDF membrane was loaded with the both cell lysates NGR234 wt (wt) and NGR234Δ*ngrI*Δ*traI*_copy+ (c+). Also, all purification steps of the pulldown assay performing with purified MBP-RepX and NGR234 wt cell lysate, were loaded on membrane. This part of the membrane was treated with anti-RepX-antibody solution. **B:** PVDF membrane was loaded with cell lysate of NGR234 wt and NGR234Δ*ngrI*Δ*traI*_copy+ (c+). The membrane was treated with anti-RepA0 antibody solution. Here, the pure cell lysates were loaded on the membrane.

Confirmation of RepX with specific Anti-RepX antibodies in NGR234 strains

The small protein RepX should also be verified in NGR234 cells by using the specific anti-RepX antibody. Therefore, the cells grew in YEM medium for 48 h, harvested carefully and washed twice with PBS buffer. The cells were fixed with paraformaldehyde, embedded in resin and were cut into thin sections. The sections were incubated with specific antibody to detect RepX in these cell sections. A gold labelled secondary antibody was used to visualize the specific antibody with Transmission electron microscopy (TEM). NGR234 wt, NGR234 Δ*ngrI*Δ*traI*_copy+ and the new constructed NGR234 Δ*ngrI*Δ*traI* mutant were taken for TEM imaging. The *E. coli* RepX overexpression strain was used as positive control and an *E. coli* strain containing an empty expression vector was used as negative control. The gold labeled antibodies become visible in TEM imaging as black dots. After imaging the black dots were counted (Figure 22).

The Micrographs (Figure 22) showed differences in morphology between NGR234 cells and *E. coli* control cells. The NGR234 cells have a more clearly delineated cell membrane with less EPS around the cells. In contrast, the *E. coli* cells show a larger EPS layer around the cells. And in the NGR234 cells, there are more white granules structures These granules were missing in *E. coli.* Micrographs of NGR234 wt cells incubating with anti-RepX antibodies

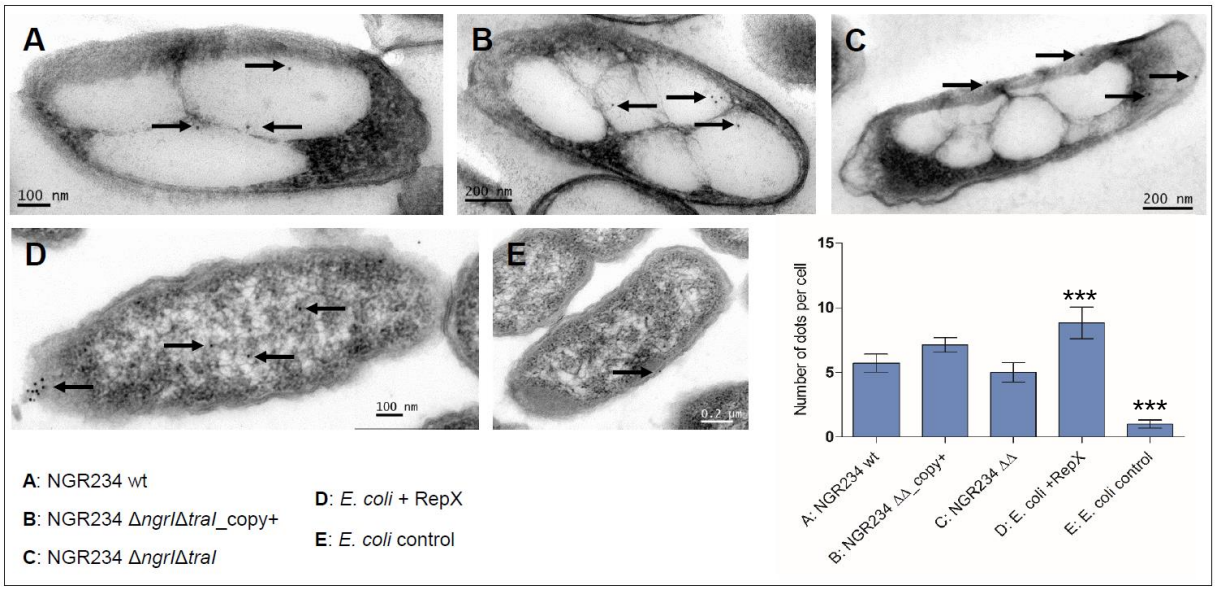

Figure 22: TEM imaging of NGR234 cells incubated with specific anti-RepX antibodies
Transmission electron micrographs of ultrathin sections through NGR234 cells incubated with specific anti-RepX antibodies that were visualized with gold-labeled secondary antibodies. A: NGR234 wt **B:** NGR234 Δ*ngrI*Δ*traI*_copy+ **C:** NGR234 Δ*ngrI*Δ*traI* **D:** E. coli BL21 (DE3) containing RepX expression construct **E:** E. coli BL21 (DE3) containing empty expression vector. The gold labeled antibodies become visible in TEM imaging as black dots (arrows). **Bar diagram:** Number of dots per cell in the different strains. The three stars indicated the significance for the different number of dots in the bacterial cells compared to NGR234 strains.

showed 5.72 ± 0.70 (n=18) dots per cell. It was shown that NGR234 Δ*ngrI*Δ*traI*_copy+ mutant 7.13 ± 0.57 (n=16) dots were observed. The newly constructed AI double deletion mutant had approx. 5.00 ± 0.76 (n=12) dots per cell. Both AI mutants showed no significant differences compared to NGR234 wt. This could be observed differently in the controls. While 8.83 ± 1.23 (n=18) black points were visible in the positive control, only 1 ± 0.32 (n=13) RepX protein per cell was observed on average in the negative control.

In summary, it could be said that RepX could be detected in NGR234 cells using the specific antibody and that there was no difference in the number of expressed RepX proteins using the ultrathin sections. In the overproducing *E. coli* Strain a significant higher number of RepX protein was observed.

Deletion of *repX* via homologous recombination

For analyzing the putative function of the small protein RepX, a *repX* deletion mutant was constructed by using homologous recombination. The flanking region of *repX* were amplified and inserted into the suicide vector pNTPTS138-R6KT using *EcoR*I and *Sal*I restrictions sites. The construct was conjugated to NGR234 wt. The first crossing over event in which the NGR234 cells carried the parent and the deletion fragment was controlled by colony PCR.

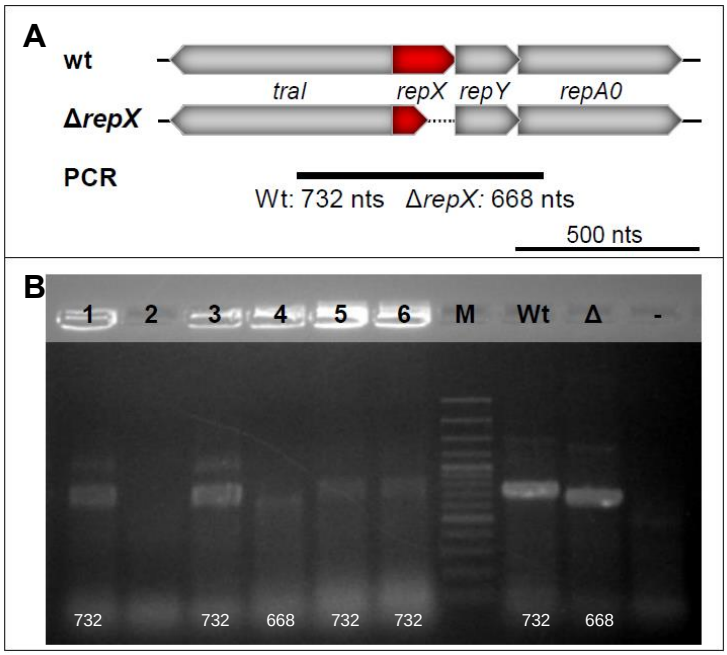

Figure 23: Confirmation of NGR234 Δ*repX*.
A: Genomic organization of the smORF *repX* (red) in the wt strain and the in this work constructed deletion mutant NGR234 Δ*repX* with the corresponding PCR product when the resulted product was verified by using the primers oKP112b and oKP117. **B:** 2 % agarose gels loaded with testes colonies that grew after the second crossing over event. gDNA of NGR234 wt was used as positive control, whereas pNPTS138-R6KT::Δ*repX* was used as positive control for a putative mutant. H_2O was used as negative control. Six colonies were tested, and the resulted PCR products sizes were listed at the bottom of gel (in nts). The GenRuler™ 100 bp+ served as DNA standard.

A clone harboring both fragments was spread out on agar plates supplemented with 15 % (w/v) sucrose. The second crossing over event resulted in NGR234 parent cells or mutant cells. Which event happened was verified by PCR (Figure 23). The PCR was performed with the primers oKP112b and oKP117. The PCR product for a resulting mutant had a size of 668 nts, while a resulting wt had a larger product with 732 nts (Figure 23, A). The performed PCR resulted in 4 samples that had products with 732 nts corresponding for the wt strain (well no.: 1, 3, 5 and 6). Only one sample showed a smaller amplificated product that had the same PCR product size such as the *repX* deletion control (Δ; Figure 23, B).

Since the *repX* gene was within the direct neighborhood of the plasmid replication operon and origin, it might influence directly or indirectly the plasmid copy number. Therefore, the symbiotic plasmid copy numbers of different NGR234 strains were verified. As previously published, the symbiotic plasmid copy number of the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant was already measured and showed a 4- 6-fold higher copy number of the symbiotic plasmid (Grote et al., 2016). With the identification of the smORF *repX* it became clear that it was not exactly known whether the missing AI alone influenced higher copy number in NGR234 Δ*ngrI*Δ*traI*_copy+ or repX also effected copy number. For the copy number calculation, the genes *recA* and NGR_c03800 were served as reference genes for controlling the number of chromosomes.

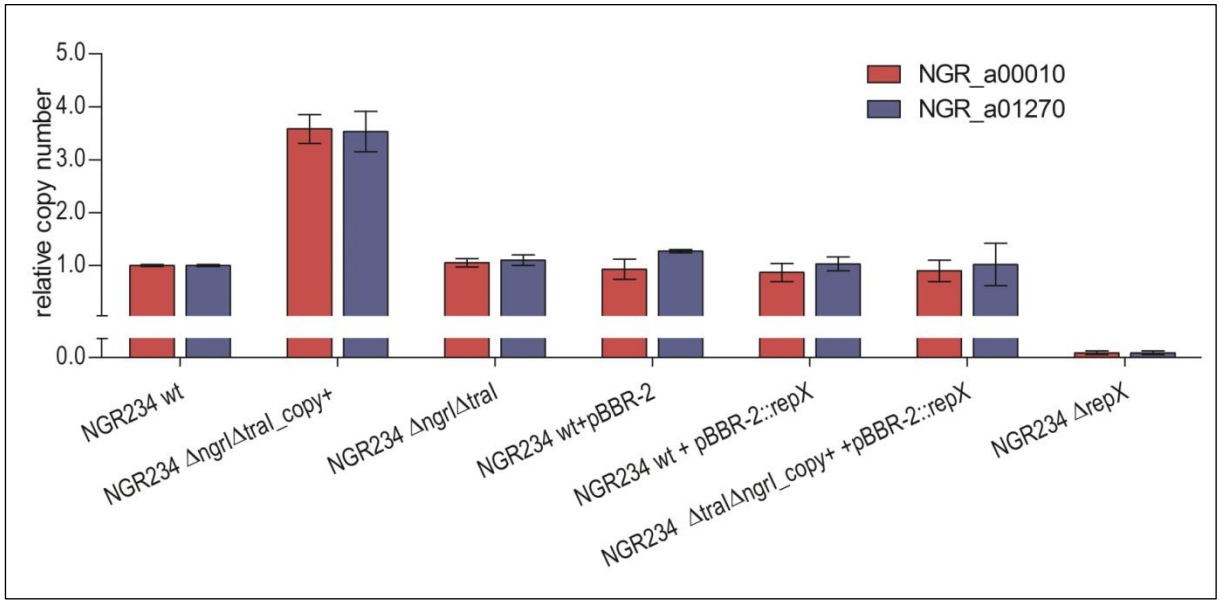

Figure 24: Copy number verification of pNGR234*a*
The bar diagram shows the results of the qPCR for verification of symbiotic plasmid number. The copy number of the parent strain (NGR234 wt) was set to 1.0. Thus, the relative copy number of the tested strains can be read off. The tested genes lying on the plasmid are *repA* (NGR_a00010, red bars) and *nifB* (NGR_a01270, blue bars).

The genes *repA* (NGR_a00010) and *nifB* (NGR_a01270) were used for determining the symbiotic plasmid copy number. The values of the parent strain were set to 1.0 (100 %). Similar to the previously published copy number data of NGR234 Δ*ngrI*Δ*traI*_copy+, the verification that was performed in this work showed also a significant increased symbiotic plasmid copy number compared to NGR234 wt (Figure 24). Interestingly, the new constructed AI mutant with preserved promoter region of *repX* quenched the copy number to wt level. By complementing the NGR234 Δ*ngrI*Δ*traI*_copy+ genetically with *repX* on the broad host range vector pBBR1MCS-2 the copy number was also quenched to wt level. Controls with empty vector did not influence copy number. For further attempt, the question is what happens when the genes *repX* is completely deleted. Thus, also the symbiotic plasmid number of the constructed NGR234 Δ*repX* deletion mutant was verified. When the smORF was deleted on the genome, no symbiotic plasmid was observed in the isolated gDNA. The copy number decreased to zero compared to NGR234 wt (Figure 24).

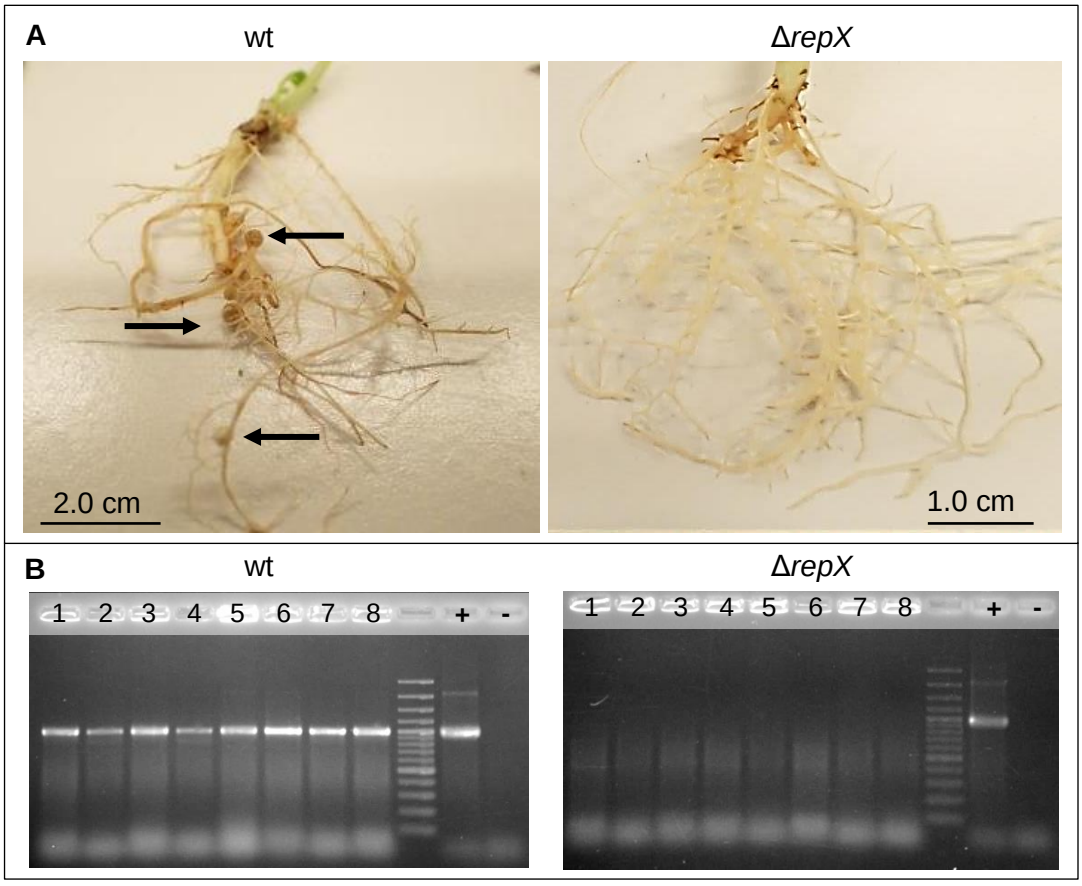

Figure 25: Nodulation assay and colony PCR for verifying the presence of pNGR234*a*
A: Nodulation assay with *V. unguiculate*. The plant roots are shown 50 days after infection with NGR234 wt or NGR234 Δ*repX*. The black arrows point the nodules at the plant roots. **B:** Colony PCR with specific symbiotic plasmid primer (oKP169 and oKP170) for detecting the plasmid in the cells. Eight different clones were tested. Genomic DNA of NGR234 wt was used as positive and H_2O as negative control. The GeneRuler™ 100 bp+ was used as standard.

That the symbiotic plasmid copy number was lost in the mutant could be confirmed with nodulation assay because all genes that are necessary for establishing the symbiosis are located on it. For this assay, the host *Vigna unguiculata* was infected with NGR234 wt and NGR234 Δ*repX*. 50 days after infection, the first nodules were observed at the roots of the plant that was infected with NGR234 wt (Figure 25, A). No nodules were observed at the plant roots, which were infected with the NGR234 Δ*repX*. Without symbiotic plasmid the mutant was not able to establish a symbiosis with the plant. For a second verification for the loss of pNGR234*a,* the mutant was spread out on a plate and single colonies were tested for the presence of symbiotic plasmid by using colony PCR with specific primers (oKP169 and oKP170) that resulted in a 1,000 nts PCR product when pNGR234*a* was present in the cells. For the wt strain in all tested colonies a PCR product was observed with the corresponding size of 1,000 nts. But in NGR234 Δ*repX* mutant in none of the tested colonies a PCR product was detectable (Figure 25, B). Both experiments confirmed the results of the qPCR. The NGR234 Δ*repX* mutant lost the symbiotic plasmid and showed same genotype as ANU265 strain, which had lost the pNGR234*a* through a complex heat treatment. The results also showed that the smORF *repX* influences additional to the non-coding RNAs in-between *repB* and *repC* the plasmid copy number and could be import for plasmid maintenance

The smORF *repY*

The second smORF (smORF-250) in the intergenic-region in-between *traI* and *repA* had a size of 178 nts and started at position 535,637 and ended at position 535,810 on pNGR234*a*. Because of the location directly downstream of the smORF *repX,* this gene is designated as *repY* and encoded probably for a 57 aa small protein with molecular weight of 6 kDa. The smORF was found in the transcriptomic data of NGR234 Δ*ngrI*Δ*traI*_copy+ due to the 7.8-fold higher transcription level compared to the parent strain (Figure 26). The smORF *repY* was only present in a few other *Sinorhizobium* sp. strains and had a percentage identity of 98.85 % to these strains (e-value: 6e-84) determined by using BlastN similarity search. Two nts were differing to the other sequences but both nts (T79C, A87G) exchanges led to amino acid exchange (L27L, Q29Q).

The smORF *repY* started with GTG as translation start (Figure 26) and had a GC content of 49 %. A potential RBS was located three nts upstream of the translation start. The intergenic region in-between *repX* and *repY* contained 20 nts and the ORF *repA0* started 27 nts downstream of *repY.*

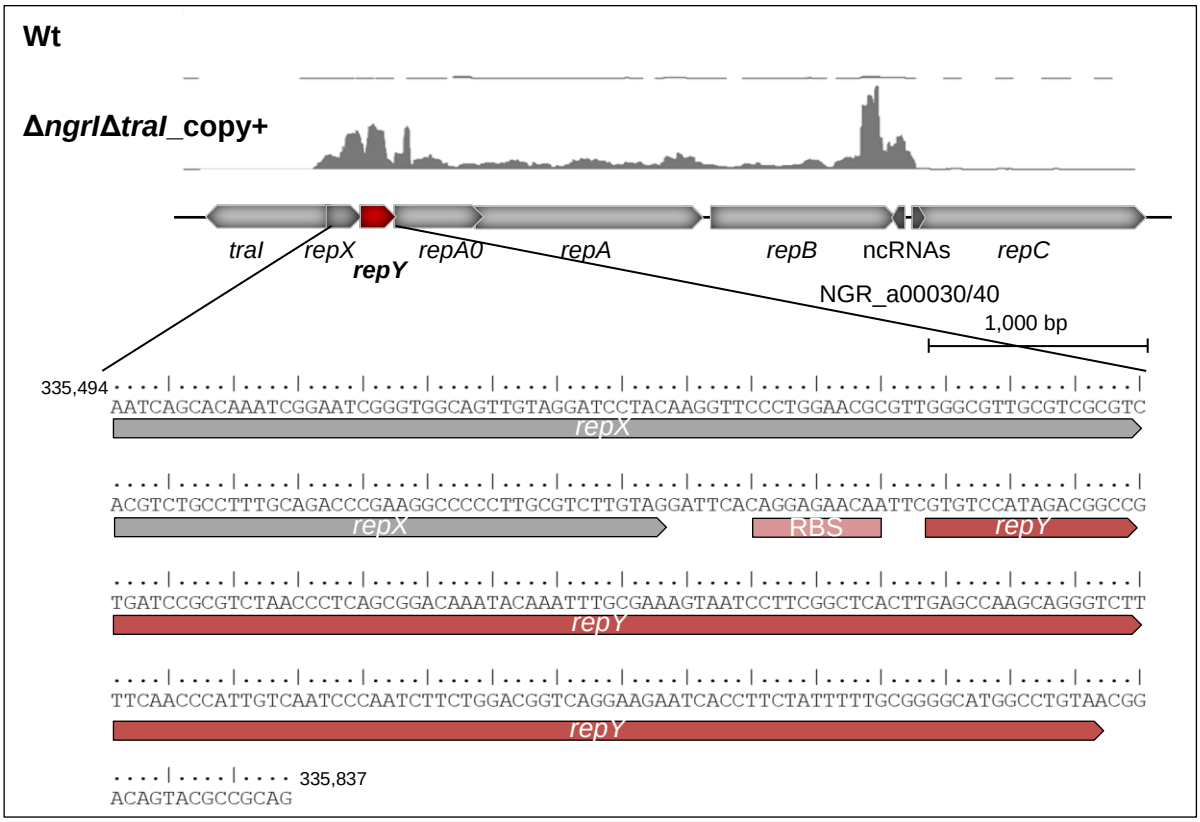

Figure 26: Genomic organization and coding sequence of the smORF *repY*
The genomic organization of the smORF *repY* (red box) was shown with the corresponding transcriptomic profile of the wt and the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant. The grey shadows represented the RNA-seq reads. Both graphs were set to 100 reads. The lower part of this figure showed the coding sequences of *repY* (red) and the upstream located smORF *repX* (grey). A RBS of the respective gene was located three nts upstream of the *repY* translations start GTG.

Homology model of RepY

The secondary structure of RepY was predicted using the Phyre2 online prediction tool. For the 74 aa peptide chain, five secondary structures were predicted (Figure 27; A). All in all, 53 % of the aa chain did not have a secondary structure and were uncoiled. 28 % of the amino acids were included in α- helices and 19 % in β-sheets. The chain started with two β-sheets and three α-helices were identified in this peptide chain. And then followed by an α-helix. Between the first and the second α-helix were ten uncoiled amino acids. The peptide chain terminated with two short α-helices. A model of RepY protein tertiary structure was calculated with the prediction server Robetta using the Ginzu PDB template identification and domain prediction protocol. The resulting PDB file was visualized with the program Chimera. The tertiary structure was represented from two sides to represent the entire structure. Both terms were displayed with the corresponding letters for the N- and C-terminus (Figure 27; B). The interactions and bonds of the amino acid's residues determined the tertiary structure of a protein. The structure started with an unfold area, followed by two antiparallel β-sheets that

were connected by hydrogen bonds. Then after a larger α-helix, an uncoiled region of ten amino acids followed and afterwards a smaller α-helix was connected by a loop with a second small α-helix.

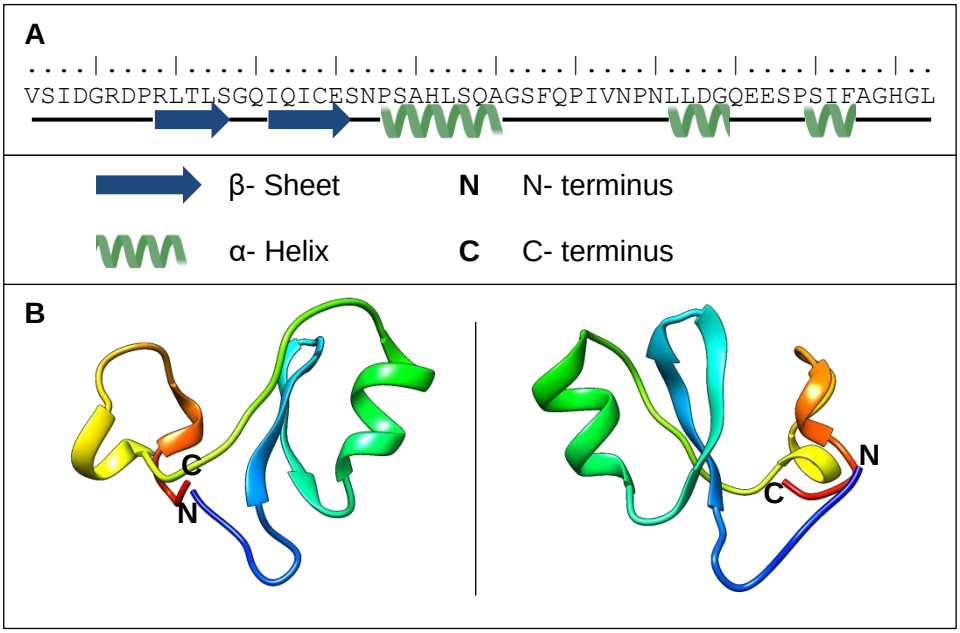

Figure 27: Secondary structure and homology model of RepY
A: Secondary protein structure prediction of RepY using Phyre2 secondary structure prediction (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index). The β-sheets were displayed in blure arrows and the α-helices highlighted with a green helix. **B:** Three-dimensional structure of RepY was represented from two sides. The N- termini and C- termini were indicated by the letters N and C. The three-domensional structure was predicted by Robetta structure prediction using Ginzu PDB template identification and domain prediction protocol (http://robetta.bakerlab.org/).

Recombinant protein expression of RepY in *E. coli*

For further characterization of the smORF *repY* the recombinant expression and purification system in *E. coli* was used. The gene *repY* was amplified and ligated into *Eco*RI and *Sal*I restriction site of the expression vector pMALc2x. The final construct pMALc2x::*repY* was sequenced and transformed into the expression strain *E. coli* BL21 (DE3). The cells of a 300 ml culture were harvested and lysated by using French pressure. The recombinant expressed protein MBP-RepY was purified using Amylose resin according to manufacturer's protocol. All eluates were concentrated with Vivaspin 6 concentrator columns having a 30 kDa cut off. All steps of the purification were loaded on a Tris-Tricine-SDS gel and then blotted to a 0.2 µm nitrocellulose membrane for immune detection (Figure 28).

The purified protein MBP-RepY was detected in the washing and elution steps by the anti-MBP antibody with a size of approximately 45 kDa (Figure 28; A). No other bands were detectable. It seems that the antibody binds specific to MBP fusion protein.

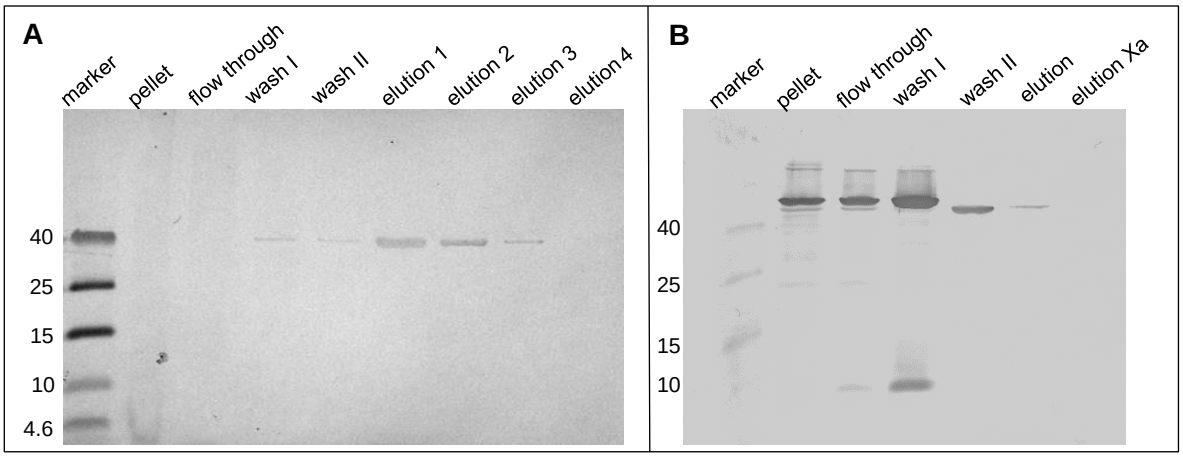

Figure 28: Nitrocellulose membrane containing all steps of MBP-RepY Purification
A: 0.2 µm Nitrocellulose membrane containing all steps of MBP-RepY purification. The membrane was treated with anti-MBP antibody as primary and anti-rabbit- IgG-AP as secondary antibody. **B:** 0.2 µm Nitrocellulose membrane treated with specific anti-RepY antibody. The protein ladder Spectra Multicolor Low Range Protein Ladder was used as standard for both membranes.

By using the specific antibody anti-RepY-AB (Table 11) for detecting RepY on the nitrocellulose membrane a band with a size of approx. 45 kDa was visible in the pellet, flow through, washing and elution steps (Figure 28; B). In addition to this band a smaller protein band with a size of less than 10 kDa was observed in the flow through and washing I step but not in the elution step. In the last well the eluted protein MBP-RepY was cut with factor Xa. But neither a protein band with a size of around 45 kDa nor a protein band with less than 10 kDa could be observed.

Expression of RepY in NGR234 cells

NGR234 cells were incubated with specific Anti-RepY antibodies (Table 11) to detect the expression of this small protein in the cells. The NGR234 Δ*ngrI*Δ*traI*_copy+ and NGR234 Δ*ngrI*Δ*traI* cells were fixed with paraformaldehyde, embedded in resin and were cut into ultra-thin sections. These sections were treated with the specific anti-RepY antibodies that bind to RepY proteins in the cells. A gold marked secondary antibody could visualized the primary antibody so that the antibodies could be detected with TEM LEO 906 (Zeiss, Oberkochen, Germany) as black dots (arrow, Figure 29). The TEM images showed a bacteria cell of the NGR234 Δ*ngrI*Δ*traI*_copy+ (Figure 29, A). Many black dots were visible in the bacteria cell that represented the detected small protein RepY. Some of them were located closed to the bacteria membrane but some dots were also observed in cytosol or granule structures. Some dots were also visible outside the cell, which would be false-positive that might have happened during the preparations. The second micrograph showed a bacteria

cells of the newly constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI* (Figure 29, b). In this bacteria cell, only five dots were observed.

In summary, it could be said, that RepY was expressed in NGR234 cells. It could speculate if the number of RepY proteins NGR234 Δ*ngrI*Δ*traI*_copy+ mutant was higher than in the newly constructed AI double deletion mutant NGR234 Δ*ngrI*Δ*traI* because of higher plasmid copy number.

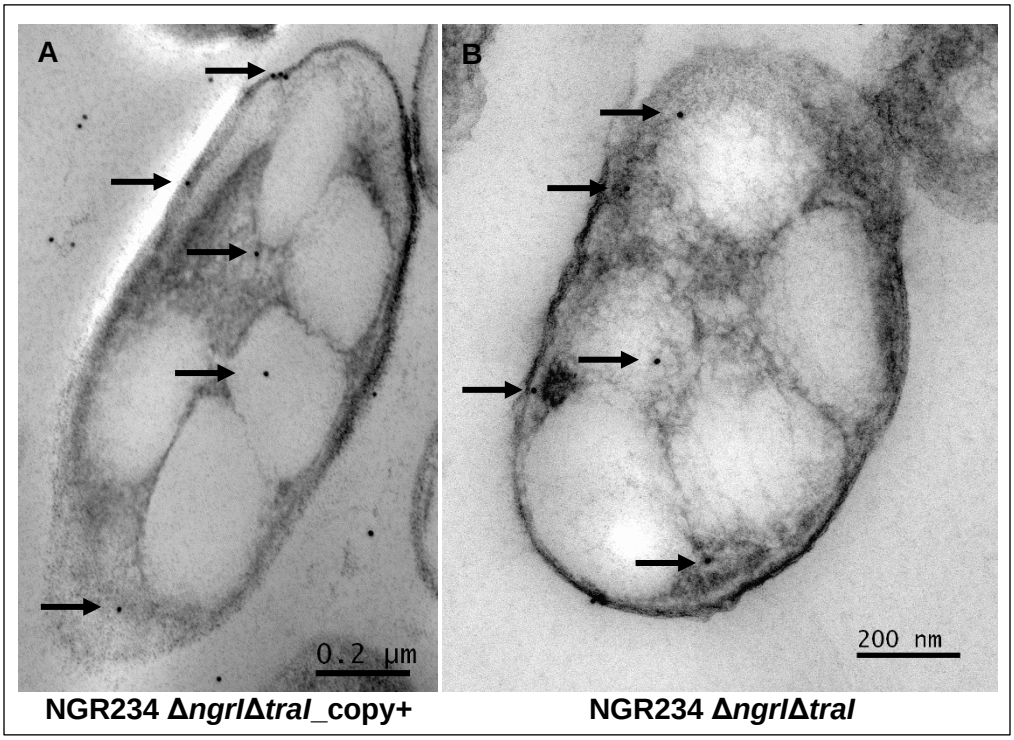

Figure 29: TEM imaging of NGR234 cells incubated with specific anti-RepY antibodies
Transmission electron micrographs of ultrathin sections through NGR234 cells incubated with specific anti-RepY antibodies that were visualized with gold-labeled secondary antibodies. The gold labeled antibodies became visible in TEM imaging as black dots (arrows). Here, the both AI double deletion mutants were shown.

Deletion of *repY* via homologous recombination

The design of a NGR234 Δ*repY* deletion mutant could gave hints on the function of the smORF-250. Therefore, the flanking regions of repY were amplified and ligated into the suicide vector pNPTS138-R6KT.For preserving the beginning of *repA0*, the last 72 nts of *repY* remained unchanged and only the first 102 nts of *repY* were deleted. The final construct pNPTS138-R6KT::Δ*repY* was integrated into NGR234 wt genome by homologue recombination. The second crossing-over event was driven by the *sacB* selection marker and results either in the wt or in the mutant genotype (Figure 30; A). Two resulted mutants were verified by running two PCRs with different primers. Mutant no. 17 (Δ_{17}) and mutant no. 43

(Δ_{43}) showed both a smaller PCR product than the gDNA of NGR234 wt (wt, Figure 30, B and C). The second PCR was done with the primer oKP178 and oKP242. By using gDNA as template the PCR resulted in 507 nts PCR product. Either by using gDNA of NGR234 Δ*repY* no. 17 or NGR234 Δ*repY* no. 43, no PCR Products were observed. Both PCRs indicated that the second crossing over event of mutant no. 17 and mutant no. 43 resulted in the expected mutant and the sequencing of both mutants showed the correctly construction of NGR234 Δ*repY*. The sequence is attached (Table 24).

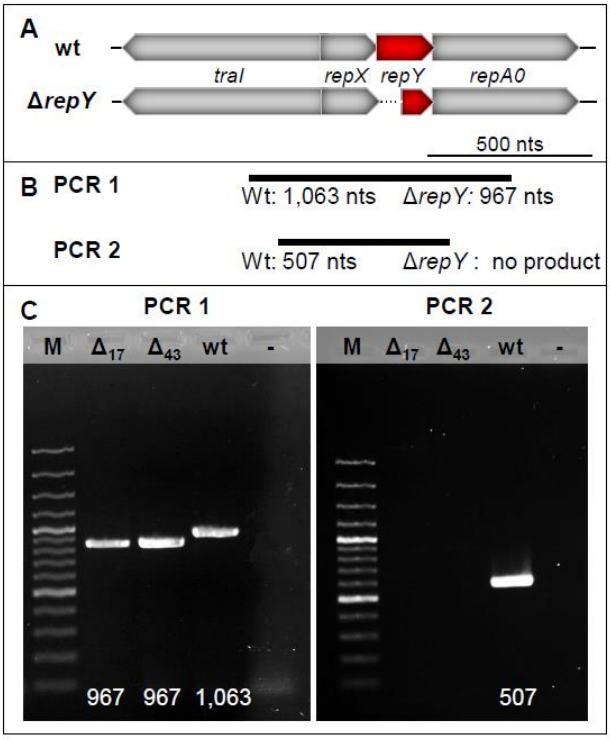

Figure 30: Construction of a NGR234 Δ*repY* deletion mutant
A: Genomic organization of the gene *repY* in the wt strain and in the newly constructed NGR234 Δ*repY* deletion mutant. **B:** Corresponding PCR Products for the verification of the mutant. The products in the mutant were smaller by the corresponding mutant range of 102 nts. In PCR 2 no product is expected for the mutant. C: The 2.0 % agarose gels for PCR 1 and PCR. Wt gDNA was used as positive control. The GeneRuler™ 100 bp + DNA ladder was loaded to estimate the product sizes.

Since the smORF was a part of the new identified ORFs *repX, repY and repA0* that were located in intergenic region and *repX* influenced the copy number, it might be possible that *repY* had some similar effects. Therefore, the copy number of pNGR234a of both newly constructed NGR234 Δ*repY* mutants should be checked. The gDNA of NGR234 wt and both NGR234 Δ*repY* mutants were isolated after 96 h and qPCR was performed (Figure 31, A). The symbiotic plasmid copy number was set to 1.0. Compared to this number, the NGR234

Δ*repY* no. 17 showed no changing copy number. But interestingly, the copy number of mutant no. 43 decreased to 0.3 for both tested genes.

To verify this result, single colonies of both mutants were tested with pNGR234a specific primer for the presence of the symbiotic plasmid (Figure 31, B). For a better affirmation, two regions of the plasmid were amplified. Eight different colonies were tested to give first hints. Whereas in mutant no. 17 all colonies showed products for PCR 1 and PCR 2., only five colonies had a product for PCR 1 and six colonies had a amplified product for PCR 2 in

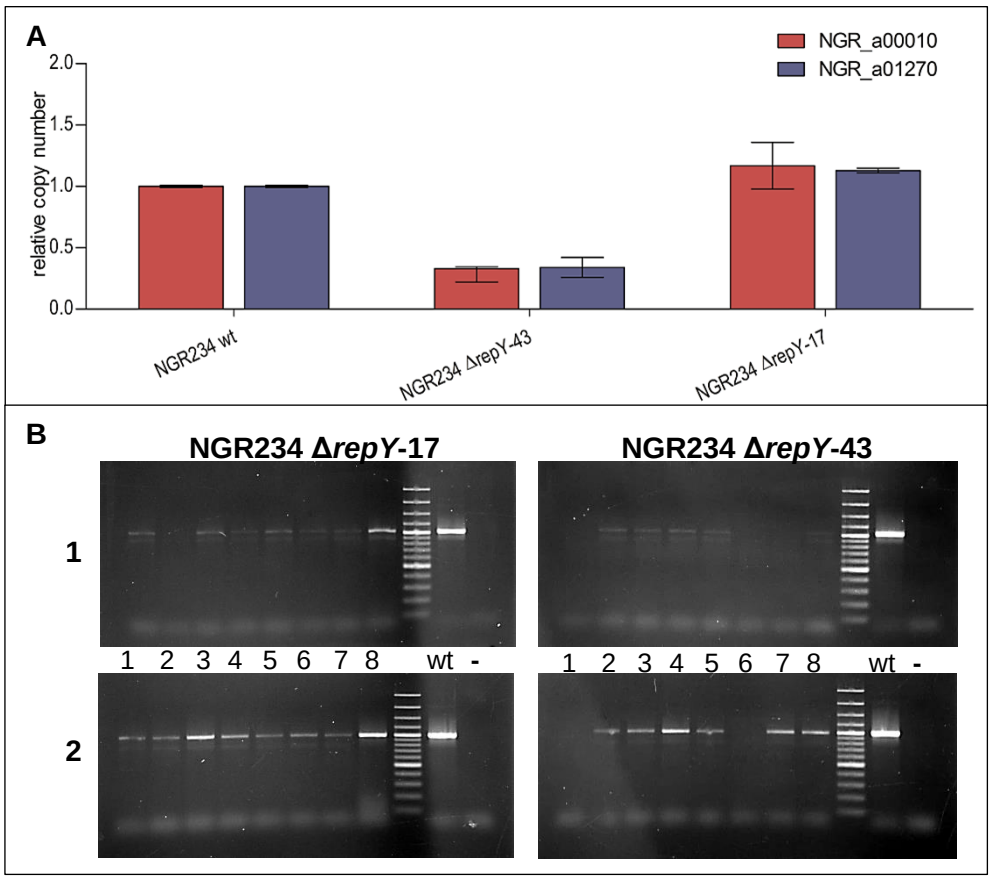

Figure 31: Copy number verification of pNGR234*a*
A: The bar diagram shows the result of the performed qPCR for the verification of symbiotic plasmid number. The copy number of the wt strain was set to 1.0. Thus, the relative copy number of the mutant strains can be read off. The tested genes lying on pNGR234a are *repA* (NGR_a00010) and *nifB* (NGR_a01270). **B:** Colony PCR with specific symbiotic plasmid primer (1: oKP149 and oKP150; 2: oKP169 and oKP170) for detecting the plasmid in single colonies. Eight different clones were tested. Genomic DNA of NGR234 wt was used as positive and H_2O as negative control. The GeneRuler™ 100 bp+ was loaded as DNA standard.

mutant no.43 (Figure 31, B). It seemed that 25 % of these colonies had lost their symbiotic plasmid and that in 75 % of the tested colonies the plasmid was stable.

That in one of the NGR234 Δ*repY* mutant the plasmid was not stable could give hints for the importance of *repX* and *repY* for the plasmid maintenance or partitioning.

The ORF *repA0*

The ORF *repA0* with its size of 432 nts did not necessarily encode for a small protein but it was the last gene that was located in the intergenic region and was first identified in the transcriptomic data of NGR234 (Grote et al., 2016). The gene rep*A0* started a position 535,838 and terminated at position 104 on pNGR234*a*. It was previously shown that the ORF *repA0* was not co-transcribed with the replication gene *repA* (Grote et al., 2016). The gene showed based on the nts sequence 98.15 % similarity to other closed related *Sinorhizobium* sp. Strains (Table 16).

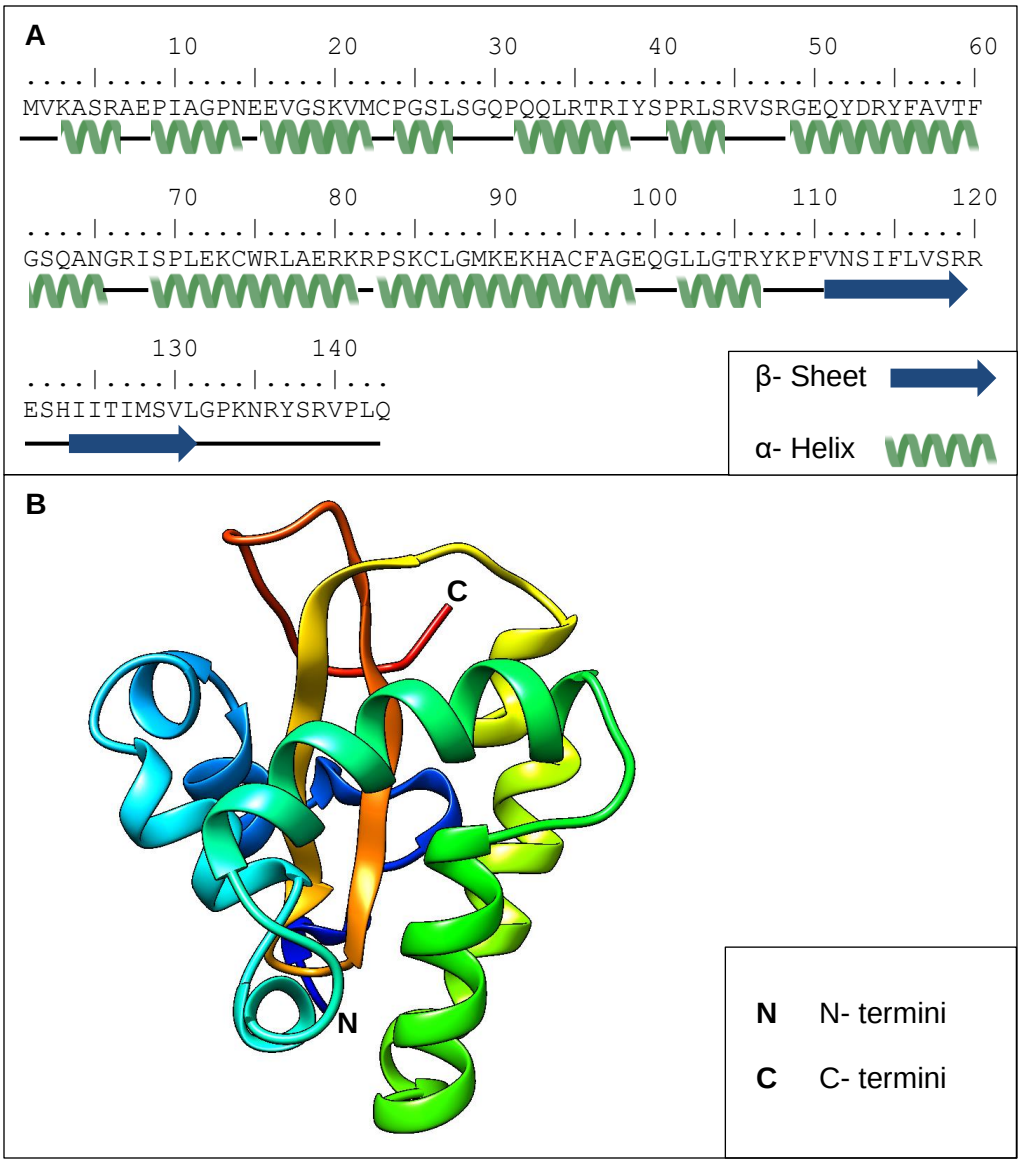

Figure 32: Structure prediction of RepA0
A: Secondary protein structure of RepA0 is shown. The β-sheets are displayed in blue arrows and α-helices are shown in green helices. The prediction of the secondary structure was calculated with Phyre2. **B:** Three-dimensional structure of RepA0 is displayed. The N-terminus and C-terminus are indicted with the letters N and C. The three-dimensional structure was predicted by Robetta structure using Ginzu PDB template identification and domain prediction.

The peptide chain consisted of 143 aa. A prediction of the secondary structure by using the Phyre2 online tool (Kelley et al., 2015) resulted in a peptide chain full of α-helices (Figure 32; A). In total, there were found 10 α-helices that made up 58 % of the peptide chain. Two β-sheets were predicted at the C-termini of RepA0.

A model of RepA0 protein tertiary structure was calculated with the prediction server Robetta using the Ginzu PDB template identification and domain prediction protocol. The resulting PDB file was visualized with the program Chimera. Both terms were displayed with the corresponding letters for the N- and C-termini (Figure 32; B). The interactions and bonds of the amino acid's residues determined the tertiary structure of the RepA0 protein. The protein structure consisted of predominantly α-helices. The β-sheets at the end of the peptide chain were antiparallel to each other and were surrounded by the α-helices.

Recombinant protein expression of RepA0 in *E. coli*

To obtain specific antibodies and for proteins investigation, the protein RepA0 was produced recombinantly in *E. coli* and purified. The gene *repA0* was cloned into pET-21a vector using the restriction sites *Eco*RI and *Hind*III and transformed in *E. coli* T7 SHuffle. A culture of 300 ml containing the expression was induced with 0.1 mM IPTG. The cells were harvested, and the recombinant protein was purified using Qiaexpressionist™ protocol no. 12 according to manufacture protocol (Qiagen GmbH; Hilden, Germany). The purification of RepA0 was controlled with SDS-PAGE and the resulted specific antibody was tested by western blot and immune detection (Figure 33).

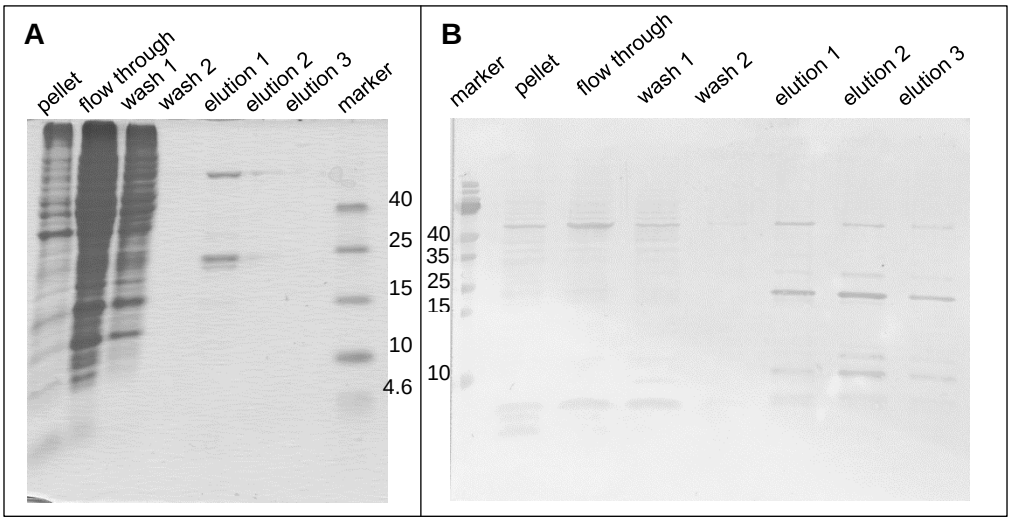

Figure 33: Recombinant RepA0 protein expression and purification
A: 16 % (w/v) Tris-Tricine- SDS gel that was loaded with the purification steps of RepA0-His. The Spectra multicolor Low Range protein ladder was used as protein standard. **B:** PVDF membrane (0.2 µm) containing all steps of RepA0 purification and that was incubated with specific anti-RepA0 antibody. Here, the PageRuler Prestained protein ladder was used as protein standard.

By using the pET-vector expression system a recombinant expression of RepA0 was possible. In the elution step a protein band with a size of less than 25 kDa was observed and displayed the calculated protein size of approx. 20 kDa with the fused His-tag. A second protein band with a size of more than 40 kDa was also seen on protein gel. A membrane that was treated with anti-His antibody (not shown here) verified the expression of His-RepA0. The purified protein was used for immunization of the rabbits for antibody production.

The resulted antibody was used for further investigation. First the protein was tested by using this antibody for the detection of RepA0 after recombinant *E. coli* expression. All purification steps were loaded on SDS gel and were than blotted to a PVDF membrane. The membrane was first incubated with the specific anti-RepA0 antibody. As secondary antibody, the anti-rabbit antibody fused to alkaline phosphatase was used for visualization of the primary antibody anti-RepA0. There were a lot of protein bands visible on the PVDF membrane (Figure 33; B). But in the eluates the strongest protein bands with a size of less than 25 kDa represented the recombinant protein RepA0. All other visible protein bands of this *E. coli* purification were unspecific bindings of the antibody. By using this specific antibody for cell lysates of NGR234 cells as it was used for the pulldown assay (Figure 21), the specificity of this antibody was confirmed using NGR234 cell lysate.

Expression of RepA0 in NGR234 cells

The specific anti-RepA0 antibody was also used for confirmation of the RepA0 expression in NGR234 cells. Cells of NGR234 Δ*ngrI*Δ*traI*_copy+ were fixed with paraformaldehyde,

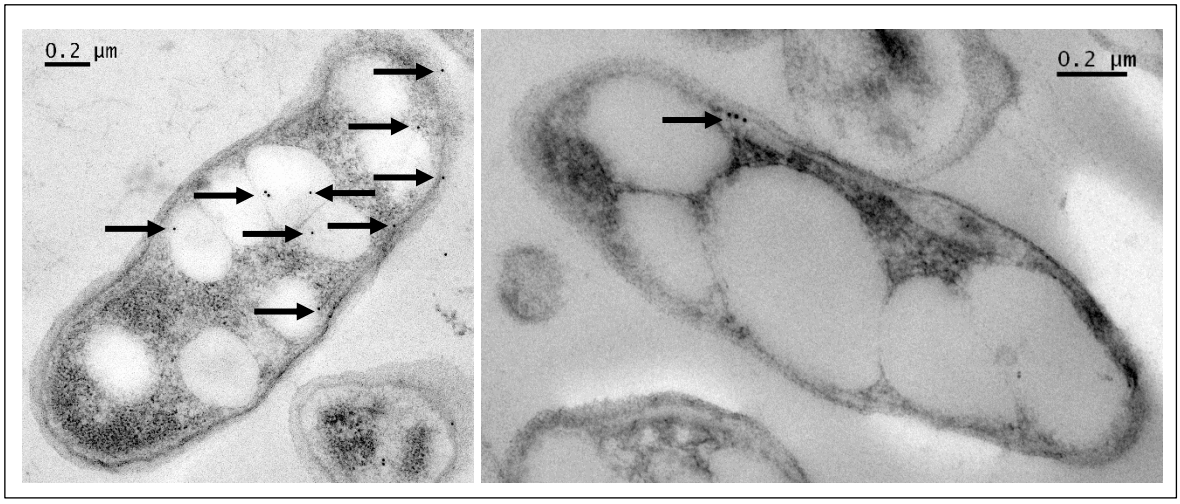

Figure 34:TEM imaging of NGR234 cells incubated with specific anti-RepA0 antibodies
Transmission electron micrographs (LEO 906, Zeiss) of ultrathin sections through NGR234 Δ*ngrI*Δ*traI*_copy+ cells incubated with specific anti-RepA0 antibodies that were visualized with gold-labeled secondary antibodies. The gold labeled antibodies become visible in TEM imaging as black dots (arrows) Here, the both AI double deletion mutants are shown.

embedded in resin and were cut into ultra-thin sections, which were treated with specific anti-RepA0 antibody. For visualization of the specific antibody in TEM imaging, a second gold-labelled anti-body was used. The sections were used for TEM imaging (Figure 34).
Many black dots were seen in the NGR234 Δ*ngrI*Δ*traI*_copy+ cell representing the protein RepA0. Some of them were located near the bacterial membrane, but most of them are observed in granule structures. The second micrographs showed a further NGR234 Δ*ngrI*Δ*traI*_copy+, where only three dots could be observed.

In summary, RepA0 would be expressed in NGR234 cells. The expression could be detected with a specific antibody. The fact that *E. coli* can expressed RepA0 recombinant also showed that the gene *repA0* encoded for a protein RepA0 with a size of 143 aa.

Characterization of a potential operon with small proteins that is located in-between the AI synthase gene *traI* and the replication operon *repABC*

The intergenic region in-between the autoinducer synthase gene *traI* and the replication operon *repABC* had a size of 830 nts (Grote et al., 2016). In this region, three ORFs could be identified by using the NCBI and the artemis ORF search tool. These ORFs were designate as *repX*, *repY* and *repA0*. (Figure 35, A) Due to this location of the three ORFs and the fact, that *repX*, *repY* and *repA0* had an increased transcription in the NGR234 Δ*ngrI*Δ*traI*_copy+ mutant, these three ORFs should be analyzed if they were also present in other strains and if they were expressed as an operon.

The presence of *repXYA0* in other strains

As previously published the intergenic region with the same size of 830 nts was also observed in two other rhizobia strains (Grote et al., 2016). Meanwhile, there were further bacteria identified, where the intergenic region was also present with the same size as NGR234. With a NCBI BlastN search all sequences with high similarity to the intergenic region of NGR234 were identified and were listed in the Table 16. The intergenic region of the strains *S. fredii* USDA257 and *S. fredii* HH103 showed the highest similarity to NGR234 with 98.91 % and had already been identified in earlier studies. Now this 830 nts intergenic region was identified on the symbiotic plasmids of five other rhizobia strains: *S. fredii* CCBAU 25509, *S. fredii* CCBAU 45436, *S. fredii* CCBAU 83666, *S. sp*. CCBAU 05631 and *Ensifer sojae* CCBAU 05684. All these intergenic regions were similar to the intergenic region on pNGR234*a* with 98.39%, 98.43% and 98.53% identity, respectively. The strain *S.* sp. M14 consisted of a symbiotic plasmid harboring an intergenic region of 377 nts that showed only similarity in the genes *traI* and *repA*.

Table 16: BlastN result of the *traI-repA* intergenic region on pNGR234*a*

Description	Max score	Total Score	Query cover	E value	Per. Ident	Accession	intergenic region size
S. fredii NGR234 pNGR234*a*	4902	4902	100 %	0.0	100 %	BK009410.1	830 nts
S. fredii USDA257 pUSDA257 fragment 4	4741	4741	100 %	0.0	98.91 %	CP003567.1	830 nts
S. fredii HH103 pSfHH103*d*	4741	4741	100 %	0.0	98.91 %	HE616897.1	803 nts
S fredii CCBAU 25509 pSf25509*a*	2586	4742	100 %	0.0	98.43 %	CP029453.1	830 nts
S fredii CCBAU 45436 pSf45436*a*	2586	4742	100 %	0.0	98.43 %	CP029233.1	830 nts
S fredii CCBAU 83666 pSf83666*a*	2525	4748	100 %	0.0	98.53 %	CP023072.1	830 nts
S. sp. CCBAU 05631 pSS05631*a*	2525	4748	100 %	0.0	98.53 %	CP023065.1	830 nts
E. sojae CCBAU 05684 pSJ05684*a*	2514	4742	100 %	0.0	98.39 %	CP023069.1	830 nts
S. sp. M14 pSinB	1042	1136	49 %	0.0	82.14 %	KU140623.1	377 nts

To visualize the homology to the other strains, a phylogenetic tree based on the nucleotide sequences of the intergenic region in-between *traI* and *repA* (Figure 35, A) was calculated using the MEGA X neighbor joining tool. The number of bootstrap replications was set to 100. The tree could be grouped into two clusters (Figure 35, B). The first cluster included all strains that had an intergenic region of 830 nts in-between the autoinducer synthase gene *traI* and the plasmid replication associated gene *repA*. In addition to the *S. fredii* strain NGR234, HH103 and USDA257, this groups also contained the strains *S. fredii* CCBAU 25509, CCBAU 45436 and CCBAU 83666. The intergenic regions of the *S.* sp. CBBAU 05684 and *Ensifer sojae* CCBAU 05684 showed also high similarity to NGR234 intergenic region and were classified to this group. Only the *S.* sp. M14 had no large intergenic region and was grouped into the second cluster.

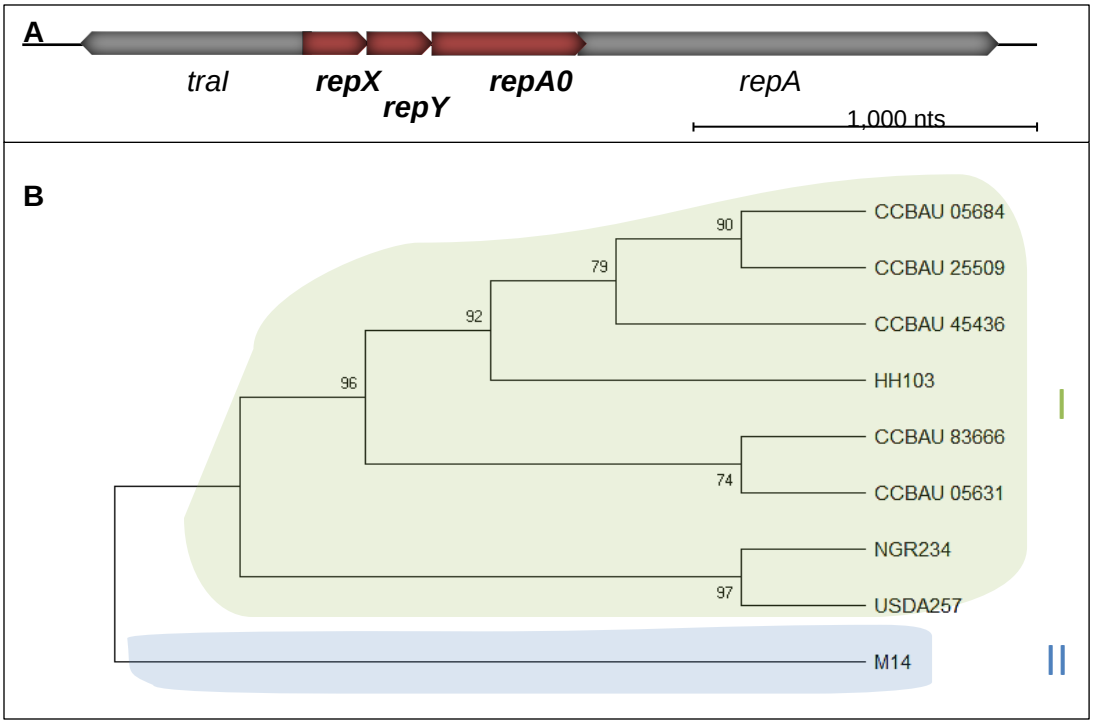

Figure 35: Intergenic region in-between *traI* and *repA*
A.: Genomic organization at the region from *traI* to *repA* (534,836 - 1,324) containing the new ORFs *repXYA0* (red) in the intergenic region. **B.:** Phylogenetic tree of the intergenic regions *traI* - *repA* of various rhizobia (TAB 9) that were found by similarity search (BlastN). The sequenced were aligned with T-coffee and the tree was calculated with Mega X neighbor joining tool. Number of Bootstrap Replication: 1,000. The two clusters are highlighted in green (I) and blue (II).

The expression of *repXYA0* as an Operon

Because of the neighborly arrangement of the genes *repX*, *repY* and *repA0* the question arose whether the three genes would be expressed as an operon or not. Therefore, the RNA of NGR234 Δ*ngrI*Δ*traI*_copy+ mutant was isolated, and reverse transcribed to cDNA that was used as template for qPCR. To verify the expression of *repX*, *repY* and *repA0* as an operon, each single gene and different combinations (*repXYA0*, *repXY* and *repYA0*) of the genes were amplified with specific primers (Table 4). The PCR products were analyzed by electrophorese (Figure 36). For each PCR, a positive control with gDNA (+) and a negative control with H_2O (-) as template were performed. The single gene expression could be observed for all three genes (Figure 36, left). On the left side all single genes were displayed. The gene *repX* has a size of 172 nts. The product of the sample showed the same size as the positive control where the qPCR was performed with gDNA. The gene *repY* had 192 nts. The positive control as well as the RNA sample of *repY* showed the expected band at 192 nts. For the single gene expression of *repA0* it was the same result. All combined gene products were shown on the right side. First all three genes *repX*, *repY* and *repA0* should be tested for

expression as operon. The whole operon had a size of 825 nts that was observed in the gDNA sample (+). The cDNA sample also showed a band at the desired height with the primers oKP42 and oKP67. With the primer combination of oKP42 and oKP200 the expression of *repXY* should be verified and resulted in a 367 nts PCR product. A clear PCR product was visible in the well for the cDNA sample. The third gene combination that was tested was *repYA0*. Therefore, the primers oKP199 and oKP67 were used for verification of the joined mRNA for *repY* and *repA0*. With these primers a region of 650 nts could be amplified and a clear PCR product for the control gDNA was observed. An amplified product was also seen in the cDNA sample. All these PCR products suggested that *repX*, *repY* and *repA0* were co-transcribed to a single mRNA and were clustered to an operon.

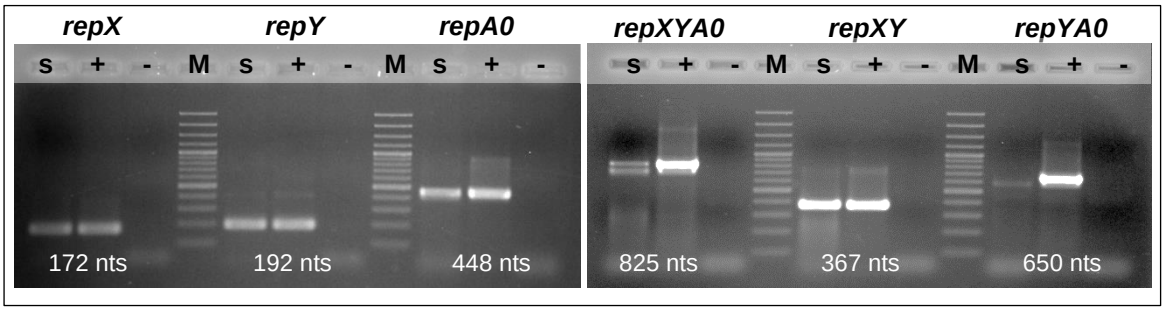

Figure 36: Verification of the *repXYA0* expression as an operon
The qPCR products were analyzed by electrophoresis. The reverse transcribed cDNA samples (s) were loaded together with the positive control (+) and negative control (-) onto a 2 % agarose gel. The expected product sizes were shown at the bottom of the gels. The GeneRuler™ 100 bp + DNA ladder was used as DNA standard (Figure 53).

Characterization of the smORF-104

Another small ORF that is characterized in detail is the snORF-104. It is located in the intergenic region of two hypothetical proteins NGR_a01720 and NGR_a1730 (Figure 37) and designated as NGR_a01725. This small ORF was noted in the transcriptomic analysis of the wt and the AI double deletion NGR234 Δ*ngrI*Δ*traI*_copy+, because of the significantly higher transcription rate in the mutant. In NGR234 Δ*ngrI*Δ*traI*_copy+ mutant, the transcription of this smORF-104 was 2.31-fold higher compared to the wt strain (Table 25, appendix).

The fact that this smORF-104 was so highly upregulated and has this small size made it an interesting candidate for further investigations.

The neighborhood of NGR_a01725 was not characterized in detail. Both genes, which delimit the intergenic region, were not functionally described. But a conserved domain prediction of both suggests that NGR_a01720 had conserved domains to a transferase that could be involved in amino acid transport or metabolism. And NGR_a1730 had a conserved domain that showed similarity to an efflux pump involving in transport of a variety of substrates.

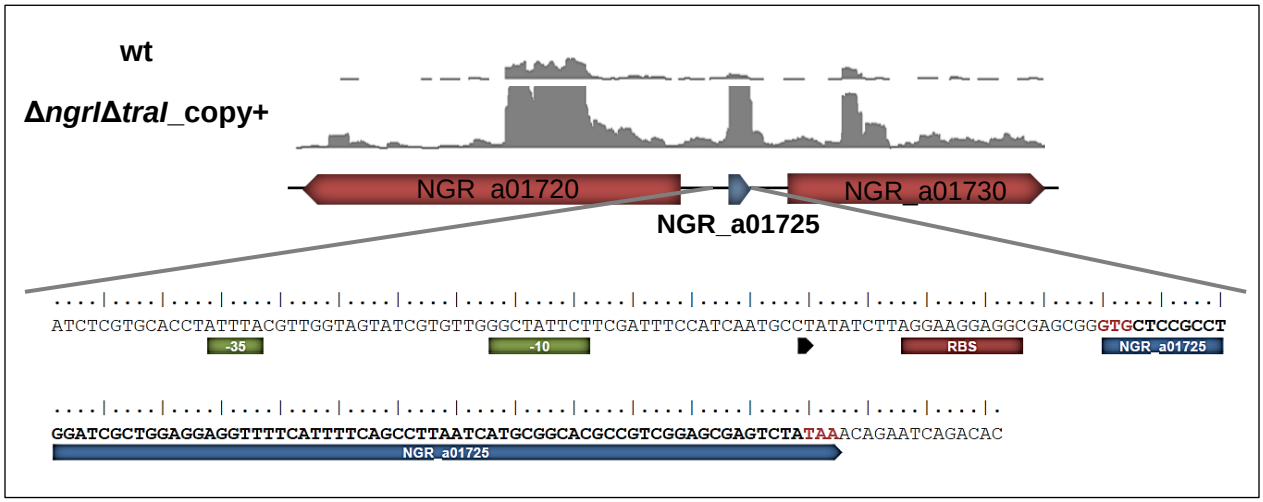

Figure 37: Genomic organization of the region around the new smORF NGR_a01725
The genomic organization of the smORF NGR_a01725 was shown with the corresponding transcript of the wt and the AI double deletion strain with the higher copy number. Below the genomic organization, the coding sequence of NGR_a01725 and the 75 bp of the leader sequence were displayed. The promoter region (green boxes), the RBS (red box) and the transcription start (black arrow) were marked in the sequence.

A closer look to the coding sequence of NGR_a01725 (Figure 37) showed that NGR_a01725 started with a GTG and ended with TAA. A possible RBS was located 8 nts upstream of the translation start GTG. The general RNA-polymerase binding sites (-35 and -10) were predicted by an online tool and were marked with green rectangles. The transcription start site (black arrow) was calculated by the transcriptomic data.

Translations fusion to verify the gene expression

For the confirmation of the gene expression to a small protein, translation fusions were constructed. Therefore, the gene of interest including promoter region was fused to the green fluorescent protein eGFP. First a translations fusion with the translation start site GTG of NGR_a01725 was constructed. The second translation fusion was constructed without the translation start site GTG. Both constructs of the translation fusions were electroporated in NGR234 wt cells and grew for 48 h at 28 °C. The fluorescence and the optical density were measured with the plate reader Synergy™ HT and the provided software Gen5 (Figure 38; B).

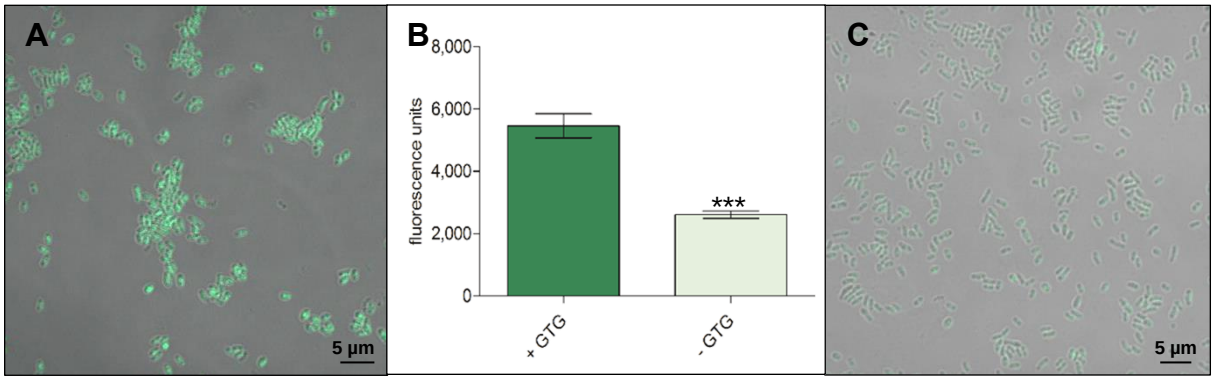

Figure 38: Translation fusion of NGR_a01725 with the green fluorescence protein eGFP with and without GTG as translation start.
For the translation fusion the promoter region and the gene NGR_a01725 were fused to the green fluorescence protein eGFP and these fragments were ligated into the broad host range vector pBBR1MCS-2. The translation fusion was electroporated into NGR234 cells. Two translations fusions were created: one harboring the translation start GTG and the second lacking the translations start. Both fusions were compared to each other. **A:** Microscope imaging of the translation fusion of A1725::eGFP harboring the translation start GTG. This image was taken with the confocal microscope LSM airyscan (Zeiss). The settings for the image are listed in Table 8. **B:** Fluorescence measurements of both translation fusions. Fluorescence was measured with the Synergy™ HT (Biotek Instruments GmbH) and the provided software Gen5. The excitation filter 485/20 and the emission filter 516/20 were used for measurements. **C:** Microscope imaging of the A1725::eGFP translation fusion lacking the translations start site GTG. The settings were listed Table 8.

To verify that the fluorescence, microscope imaging was used (Figure 38). The NGR234 cells containing the translation fusion with translation start GTG showed a higher fluorescence (5,463 ± 389.3 fluorescence units; n= 12) compared to the cells lacking the translation start GTG (2,610 ± 406.6; n= 12). Both translation fusions showed significantly different fluorescence intensities (p-value= 0.0004). By visualization the cells with the confocal microscope LSM800 airyscan, different fluorescence intensities could be detected in the bacterial cells (Figure 38; A) while the cells harboring fusion without the translation start site showed no distinct fluorescence. Only the green background fluorescence was visible (Figure 38; C). The cells containing the translation fusion with translation start showed high green

fluorescence in the bacteria cells. These experiments suggested that the gene NGR_a01725 encoded for a small protein.

Recombinant protein expression in *E. coli*

The small protein A1725 shall be expressed recombinantly by *E. coli*. Therefore, the gene was amplified and cloned into the pMALc2x expression vector as it is described in material and methods (p. 29). The expression construct was transferred into *E. coli* BL21 (DE3) and the protein was expressed in a 300 ml culture with 0.1 mM IPTG induction. The protein was purified according to manufactures protocol.

All steps of the purification were loaded on a 12 % (w/v) Tris-Tricine SDS gel to check the effectiveness of the protein purification. Of each sample 20 µg protein solution were used for SDS PAGE (Figure 39). The ordered peptide A1725 (Table 11) was used a control. This control protein band had a size of approx. 3.4 kDa. For the protein purification of MBP-A1725, a clear protein band was observed in each elution step (elution 1 - 4) corresponded to the estimated protein size of approx. 45 kDa for MBP-A1725. Some other smaller protein bands were also observed in the elution steps.

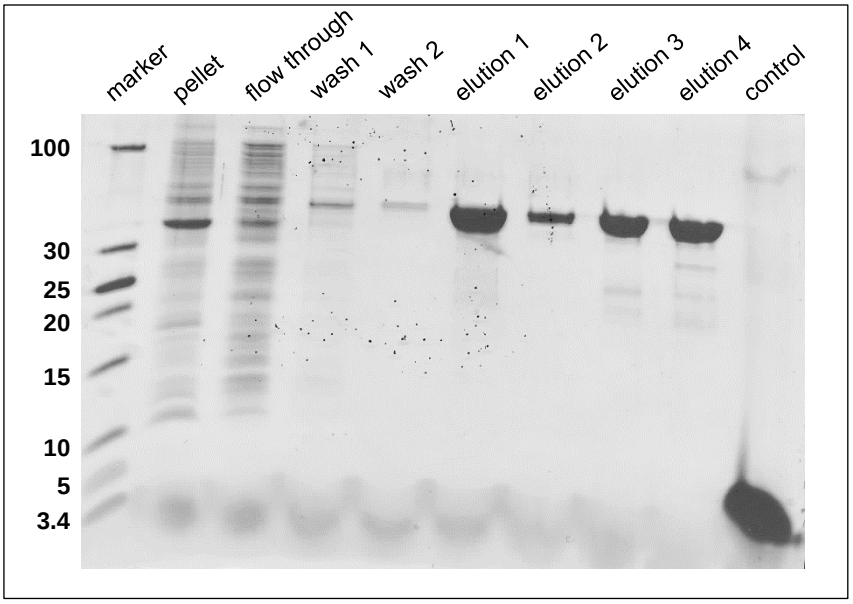

Figure 39: Recombinant overexpression of A1725 in *E. coli*
All steps of the A1725 protein purification are loaded on a 16 % (w/v) Tris-Tricine SDS gel. The ordered peptide A1725 was used as positive control. The PageRuler unstained Low Range Protein Ladder was used as protein standard. 20 µg protein of each sample was loaded into the corresponding well

Expression of A1725 in NGR234 cells

Specific anti-A1725 antibody should verified the expression of the small protein in NGR234 cells. Therefore, NGR234 Δ*ngrI*Δ*traI*_copy+ cells grew in TY broth, were harvested and then fixed with paraformaldehyde. The fixed cells were embedded in resin and cut in ultra-thin sections. For detecting the small protein these sections were incubated first with the specific antibody, which were then detected with gold-labeled secondary antibody to make then visible for TEM LEO 906 (Zeiss, Oberkochen, Germany).

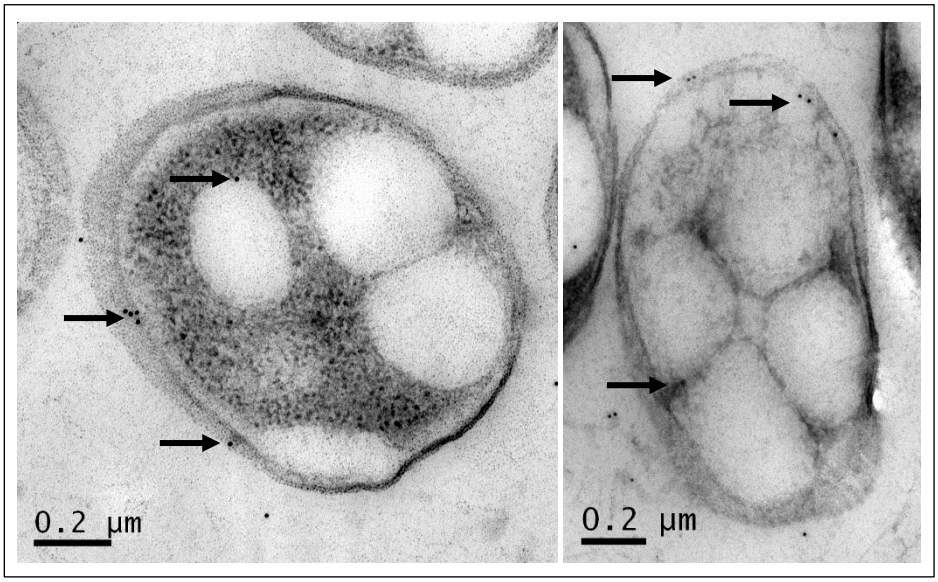

Figure 40: TEM imaging of NGR234 Δ*ngrI*Δ*traI*_copy+ cells incubated with specific anti-A1725 antibody
Cross and longitudinal sections of NGR234 cells treated with specific anti-A1725 antibodies. The specific antibodies were visualized with gold-tagged secondary antibodies, so that they are visible with TEM.

Both TEM images showed a thin section of a NGR234 Δ*ngrI*Δ*traI*_copy+ bacteria cell (Figure 40) The gold labeled proteins A1725 were visualized by black dots. In both images six black dots were observed within the cells. Most of the dots were seen closed to the membrane, but some were also in bacterial cytosol. The presence of these black dots indicated the expression of the small protein A1725 in NGR234 Δ*ngrI*Δ*traI*_copy+.

Deletion mutant of NGR_a01725 via homologous recombination

A knockout mutant was constructed for identify the effect of lacking gene NGR_a01725 to the organism. Therefore, the flanking regions of NGR_a1725 were amplified and ligated to the suicide vector pNPTS138-R6KT. This deletion construct was integrated to the NGR234 genome by conjugation. The event of the second crossing over was tested by PCR. Here two different primer pairs were applied. One primer pair bound at the flanking regions, which were

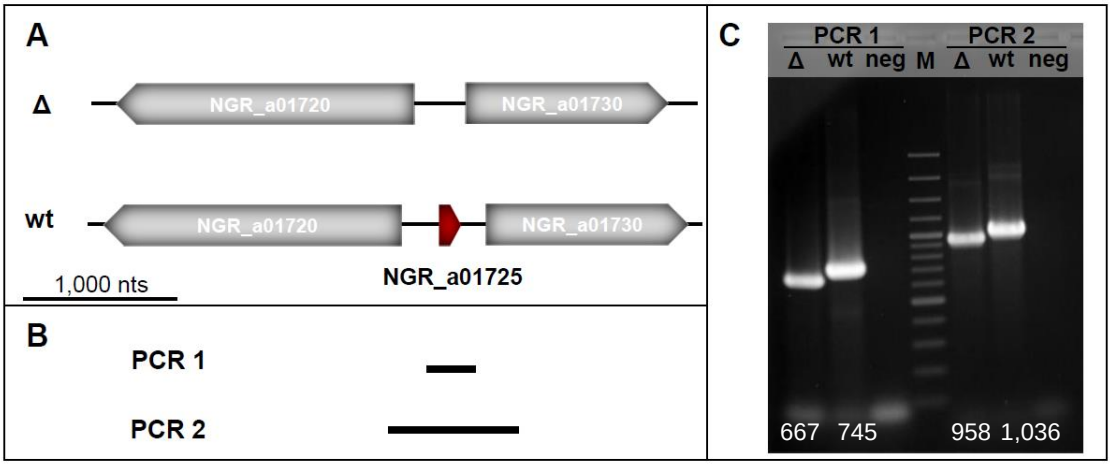

Figure 41: Construction of NGR234 Δa1725 mutant
A: Genomic organization of the gene NGR_a01725 in the newly constructed deletion mutant NGR234 Δa1725 and in the wt **B:** For verification of knockout mutant two PCR reactions with different primers were performed. PCR 1 was done with primers binding to flanking region of NGR_a01725 (upstream and downstream fragments). PCR 2 run with primers binding outside of construction region to control the correct recombination. **C:** The 2.0 % agarose gels for PCR 1 and PCR. Wt gDNA was used as positive control. The GeneRuler™ 100 bp + DNA ladder was loaded to estimate the product sizes.

used for construction (Figure 41, B). The second primer pair could amplify a longer region cover the whole area. With both primer pairs the deletion of NGR_a01725 could be verified. Both PCR products of the knockout mutant showed smaller bands than the parental control strain (Figure 41, C). The knockout PCR product resulted with outer primer had a size of 958 nts in (PCR 2) and 667 nts by using the flanking primers (PCR 1). The PCRs in the wildtype strain resulted in 78 nts bigger amplicons for both PCR reactions. The third verification of the mutant was done with sequencing approach (Table 24, appendix).

To identify the effect of this smORF-104 on organism a transcriptomic profile of the mutant compared to the parent strain was established. Characteristically for this transcriptomic profile was, that nearly all genes (%) on the symbiotic plasmid were downregulated compared to the wt (Figure 42, A). To determine whether this effect was achieved by the copy number, the copy number of pNGR234*a* was measured. Therefore, the quantitative PCR was used. With qPCR, it was recognized that the number of copies was decreased to less than half in comparison to the wild type (Figure 42, B). This led to the assumption that the symbiotic plasmid might not be present in every cell. So, a dilution streak was done, and some single colonies were tested by colony PCR and two specific symbiotic plasmid primer pairs as it was performed for NGR234Δ*repX* (Table 21). Primer pair A bound to the deletion region and amplified a 1,038 nts PCR product in the wt and a 960 nts product in the deletion strain (Figure 42, C). In the parental strain the PCR product had a size of 1,038 nts (wt) whereas in

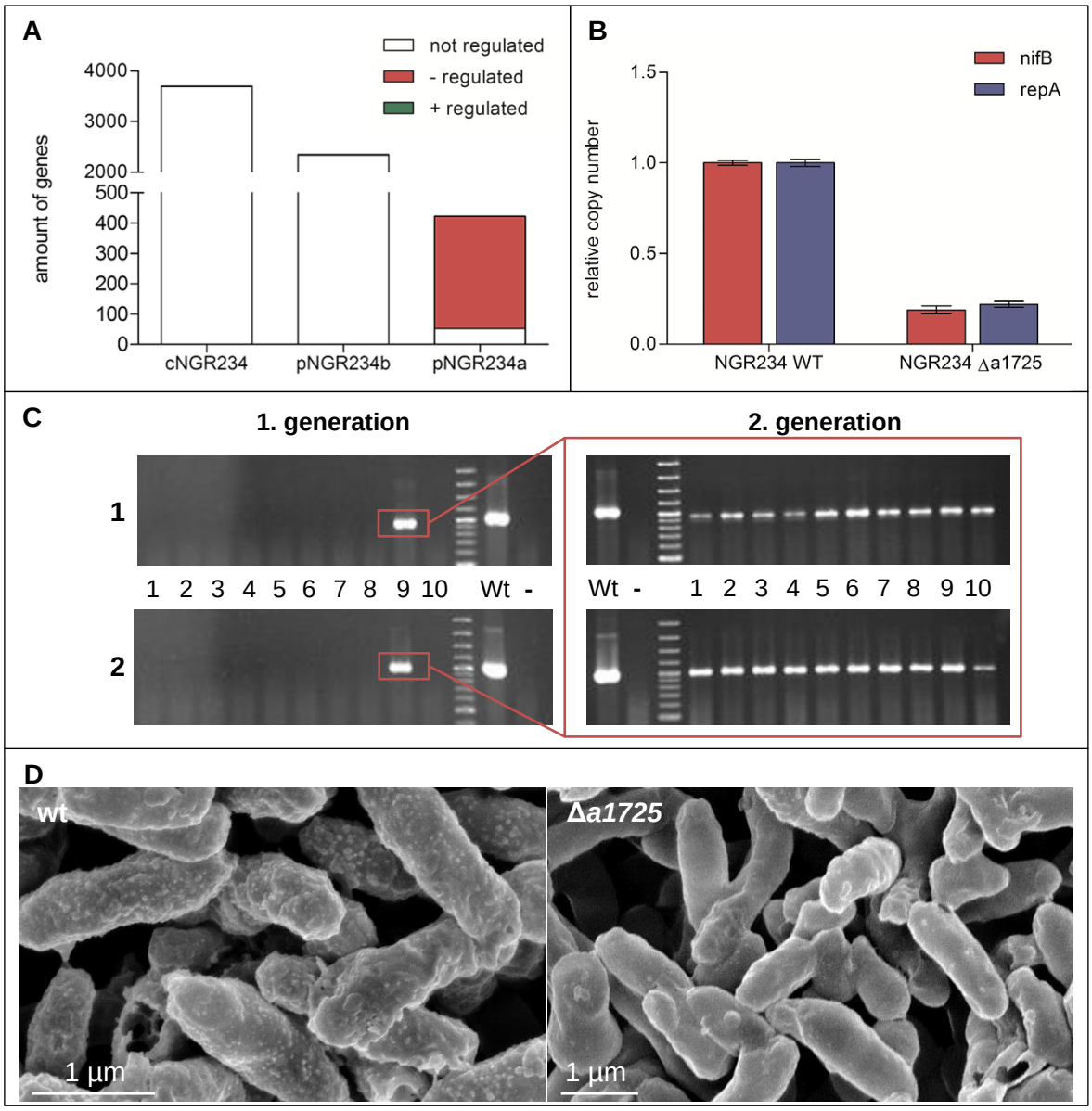

Figure 42: The effect of lacking the smORF NGR_a01725.
A: Bar diagram of NGR234 Δa1725 mutant transcriptome that was compared to the wt strain. Not regulated ORFs are presented in white, whereas downregulate ORFs are symbolized in red bars and up regulated ORFs have green colored bars. **B:** Symbiotic plasmid copy number verification of the mutant strain compared to the wt strain. The wt copy number of pNGR234a was set to 100 % (1.0) and the relative copy number in the mutant strain was related to wt number. **C:** Colony PCR with single NGR234 Δa1725 colonies. PCR 1 run with oKP149 and oKP150 and got products with 1.038 nts for the wt and 958 nts for the deletion strain. The second PCR was done with the primers oKP169 and oKP170 and resulted in a 1,000 nts product. **D:** SEM imaging of the NGR234 wt strain and the NGR234 Δa1725. The cells were coated with gold and the SEM LEO1525 run with 5 kV.

the deletion strain the product size was reduced to 960 nts (line 9). Primer pair B was completed to the *repA* region. When pNGR234a was presence a 1,000 nts PCR product was gained. This test was repeated three times with 10 colonies. Each time only one or two colonies were detectable, where the symbiotic plasmid was retained in the cell. This led to the

result that the plasmid was only stable in approx. 10 % of the cells in the NGR234 Δa1725 mutant.

To observe difference between NGR234 wt and the newly constructed mutant NGR234 Δ*a1725* in cell morphology, the NGR234 wt cells were compared with NGR234 Δ*a1725* cells by using scanning electron microscopy (SEM). Therefore, the cells were coated with gold and SEM LEO1525 run with 5 kV (Figure 42, D). The cells of NGR234 wt had structural an uneven surface, whereas mutant cells had a smooth surface. Lack of surface structure led the assumption that a genetic modification had taken place in the mutant that resulted in this phenotype.

Single nucleotide polymorphism (SNP) analysis of the NGR234 Δa1725+ mutant

To identify why the symbiotic plasmid is still present in some NGR234 Δa1725 mutants, a SNP analysis of the NGR234 Δa1725+ has been performed. Therefore, gDNA of NGR234 Δa1725+ was isolated in triplet and was sent to the Goettingen Genomics Laboratory of George-August-University in Goettingen. There an Illumina HiSeq2500 run was performed. The obtained results of the triplicates were compared with the results of the wt strain and only mutations not occurring in the wt strain were examined in detail. In total, 13 mutations were found in all three replicons of the deletion strain (Table 17). On the symbiotic plasmid only the self-induced mutation of the small ORF NGR_a01725 was observed. On the megaplasmid pNGR234*b*, three mutations in three genes were found (NGR_ab01780; NGR_b02370 and NGR_b23250, TAB). The first single mutation polymorphism was at the nucleotide position 169,219 and was in the gene that encodes for a putative Biotin carboxylase. This protein is a component of the acetyl coenzyme A carboxylase complex and catalyzes the carboxylation of the carrier protein and participates in fatty acid biosynthesis. Due this mutation the cysteine (C) was changed to adenine (A), this led the change of an amino acids (A346S) and resulted in a missense mutation. The amino acid alanine was converted to serine and since the serine side chain had a hydroxy group, the protein became more polar and the new hydroxy group could form new hydrogen bonds. But the biotin carboxylase had no influence on stability of the symbiotic plasmid. The second point mutation was at the position 232,071 and was in the gene NGR_b02370. A conserved domain search resulted that NGR_b02370 belongs to the PRK13289 super family and consisted of a potential bifunctional nitric oxide dehydrogenase/dihydropteridine reductase. At position 232,071 a base pair was deleted, and this deletion resulted in a frame shift. This frame shift had the consequence that the mRNA was translated to another amino acid succession. Thus, the protein from amino acid 45 had a different sequence with the result that the function will be lost. But also, this protein had no

influence on the symbiotic plasmid stability. The third point mutation was at position 2,393,135 and was within the gen NGR_b23250 that encodes probably for a dipeptidase. Here, a cysteine was changed to thymine. Due to the change in only the last nucleotide of this codon and the characteristics that the genetic code is degenerate, the new codon specifies the same amino acid resulting in a non-mutated protein. Such a mutation is called silent mutation. Thus, this mutation had also no effect on plasmid stability.

On the chromosome there were nine mutations that were found in all three replicates AGWSs1-AGWSs3 (Table 17). The first mutation was within the 81 nts intergenic region in-between gene coding for ATP synthase subunit C (NGR_c04480) and the gene *atpF2* (NGR_c004490) coding for FoF1 ATP synthase subunit B. Here, no influence on the symbiotic plasmid stability was recognizable by the mutation at the position 501,821. The second mutation that was observed in the SNP analysis was at position 1,228,611 on the chromosome and there the codon GTC was changed to GCC resulting in an aa exchange (V32A) in the gene *rpoB*, that encodes for the beta subunit of the bacterial RNA polymerase. At aa position 32, valine was changed to alanine. Both amino acids were characterized to the group of hydrophobic amino acids and the only difference between these amino acids was a methanediyl group. Thus, a second aa exchange was observed in the *rpoB* gene. The Aspartic acid was changed to Glutamic acid (D621E) at position 1,230,379. Both mutations had no influence on plasmid stability of pNGR234*a*. At position 1,238,154 was observed a point mutation within the gene *rpsL* coding for the 30 S ribosomal protein S12 and plays an important role in translation accuracy. But this mutation had no effect to the symbiotic plasmid stability. At the position 1,447,774, the codon AGG was changed to AAG and this mutation resulted in an amino acid exchange from arginine (R) to Lysine (K). Both amino acids belong to the group of the positive electrically charged amino acids. Interestingly, there were 3,489 nts missing on the chromosome. This missing region contained five genes coding for transposases that are also present on pNGR234*a*. A presumption that this region was meanwhile also missing in the currently working wt strain could be confirmed by PCR (Figure 55). This suggested that this lacking region had no impact on plasmid stability. Three more single nucleotide polymorphisms were detected on chromosome. The gene NGR_c32870 coding for a LuxR family transcriptional regulator had a SNP at position 3,450,413 resulted in an exchange of aspartic acid to glutamic acid but both amino acids only differentiate the length of the side chain by one methylene group. An upstream localized gene NGRc32880 also had a SNP. Here, the nucleotide exchange at position 3,452,403 led to a change in the coding amino acid from serine to glycine. But the corresponding protein an ABC transporter had no impact on plasmid maintenance. The last SNP was in the gene encoding a

transposase. The SNP at position 3,806,599 had no effect to the transposases because of being a silent mutation.

As a summary of the SNP investigation, it could be said that this analysis did not provide a clear result as to why the symbiotic plasmid was stable in NGR234Δa1725+ mutant and was lost in NGR234Δ*a1725*-.

Table 17: SNP results of the NGR234 Δa1725+ mutant in triplicates (AGWSs1-3)

#	replicon	position	mutation	AGW Ss1	AGW Ss2	AGW Ss3	annotation	gene
1	NC_000914	218,030	81 bp →CTAG	100%	100%	100%	Intergenic (-246/-192)	*NGR_a01720 ← / → NGR_a01730*
2	NC_012586	169,213	C→A	100%	100%	100%	A346S (GCC→TCC)	*NGR_b01780 ←*
3	NC_012586	232,071	Δ1 bp	100%	100%	100%	coding (135/294 nt)	*NGR_b02370 →*
4	NC_012586	2,392,135	C→T	100%	100%	100%	S384S (TCG→TCA)	*NGR_b23250 ←*
5	NC_012587	501,821	G→A	100%	100%	100%	Intergenic (+61/-21)	*NGR_c04480 → / → atpF2*
6	NC_012587	1,228,611	T→C	100%	100%	100%	V32A (GTC→GCC)	*rpoB →*
7	NC_012587	1,230,379	T→G	100%	100%	100%	D621E (GAT→GAG)	*rpoB →*
8	NC_012587	1,238,154	G→A	100%	100%	100%	R43K (AGG→AAG)	*rpsL →*
9	NC_012587	1,447,774	G→A	100%	100%	100%	R173K (AGG→AAG)	*thrA →*
10	NC_012587	1,828,597	Δ3,489 bp	100%	100%	100%		*[NGR_c17700]–NGR_c17740*
11	NC_012587	3,450,413	C→A	100%	100%	100%	D123Y (GAT→TAT)	*NGR_c32870 ←*
12	NC_012587	3,452,403	T→G	100%	100%	100%	S 419G (TGC→GGC)	*NGR_c32880 →*
13	NC_012587	3,806,599	T→G	100%	100%	100%	V21V (GTA→GTC)	*NGR_c35870 ←*

Transcriptomic analysis of the NGR234 Δa1725+ mutant

To identify what effect the small protein A1725 had on NGR234, a transcriptomic approach was conducted. Therefore, the mutant and the wt strain grew in triplicates at 28°C for 55 h. Thus, the different gene expression in the stationary growth phase could be estimated. The RNA was isolated, and the Illumina sequencing was done by our cooperation's partner of the Genomics Laboratory of George-August-University in Goettingen.

In total 185 of the 6,472 genes were differently regulated in the stationary growth phase. This led that only 2.90 % of all genes on the genome of NGR234 were differently regulated after

55 h. Interestingly, half of the regulated genes coded for hypothetical proteins whose roles in the organism were still unknown. For the chromosome it meant that 93 genes on the chromosome cNGR234 showed a different gene expression compared to NGR234 wt. Of these, 11 ORFs were up regulated and 82 genes were transcribed in a lower level (Figure 43).

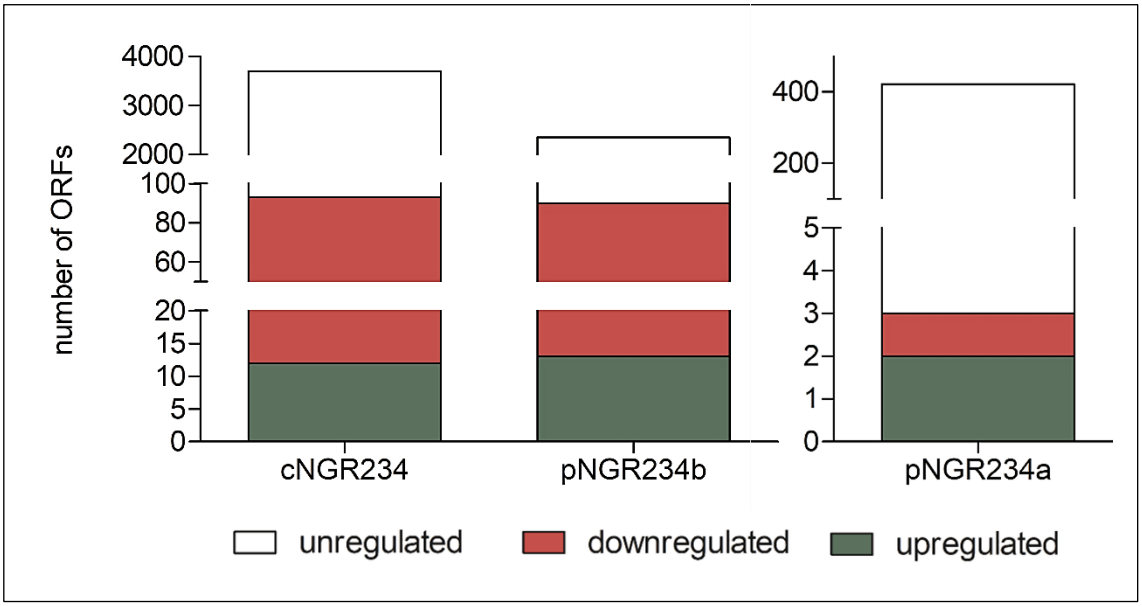

Figure 43: Transcriptomic analysis of NGR234 Δ*a1725*+ that was compared to NGR234 wt
Bar diagram containing the number of up and down regulated ORFs in and stationary (55 h) growth phase. All up regulated (+) genes were shown in green and all down regulated genes were colored in red. A colorless filling of bar indicated not regulated genes.

The mutation of NGR_01725 affected in total 91 genes on pNGR234*b*. 13 genes showed a higher transcription rate and 77 genes were down regulated. Interestingly, there were several genes, which were responsible for the formation of ABC transporter and were down regulated. Noticeable, the mutation affected only three genes on pNGR234*a* itself. Two genes showed a higher transcription level coding once for alanine ligase and second for a transposase. The third gene that showed a lower transcription rate encoded also for transposases.

In summary, a direct correlation between the deletion of NGR_a1725 and the differently regulated genes could not be proposed. The founded SNPs in the genome of NGR234 Δ*a1725*+ led mutations in the genes that could influence the transcription of different genes.

II. Structure analysis and characterization of GqqA- a Quorum Quenching hydrolase

Quorum Quenching acts as an inhibitor of bacterial cell-to-cell communication common as Quorum Sensing. The most characterized mechanism for this QS interference is the degradation of the AI molecules by different enzymes. Due to the fact, that QS regulates virulence factor production and biofilm formation, these enzyme interfering QS signal molecules play an important role as potential new drug target. In 2016, GqqA was first identified in the bacterium *Komagataeibacter europaeus* CECT 8546 and showed no similarity to known QQ enzymes and should investigate in more detail.

For the analysis of the GqqA enzyme activity the protein was overproduce and purified with the *E. coli* pET-vector system (Novagen's®, Merck KGaA, Darmstadt, Germany) like published before with some modifications (Valera et al., 2016). The structure analysis was performed in cooperation with Nadine Werner (working group Prof Betzel, University of Hamburg) at the DESY Hamburg facility. The ESI-MS/MS analysis was done in Madrid by our cooperation partner Dr. Manolo Ferrer.

Classification of GqqA to known QQ enzymes

Various enzymatic degradation pathways were possible for the inactivation of AI signal molecules. These different enzymatic pathways are classified in three enzyme classes: AHL lactonases, AHL acylases and AHL oxidoreductases. All these different enzymes classes were divided into different enzyme subfamilies. GqqA should be compared with known QQ enzymes based on their aa sequences. Additionally, to the alignment of GqqA to the known enzymes of different enzyme families, a phylogenetic classification was performed with the maximum likelihood method and JTT matrix-based model (Figure 44). The Bootstrap consensus tree was derived from 500 replicas. Branches corresponding to partitions that were rendered in less than 50 % of the bootstrap replicas were compressed. The percentage of the replication tree in which the associated taxa summarized in the bootstrap test were displayed next to the branches. All 22 known and described QQ enzymes were represented in the phylogenetic tree and clustered in the different enzyme classes. All AHL acylases were highlighted in green. The AHL lactonases were shown in red and the AHL oxidoreductases were displayed in blue. The subfamilies within the classes were highlighted with the different nuances of the colors and the family name was written next to the classes. GqqA was written in red and clustered next to the Ntn-hydrolases within the AHL acylases class. GqqA showed the closed maximum likelihood to the Amidase enzyme that was also clustered to AHL acylases but the mechanism behind was not yet clear.

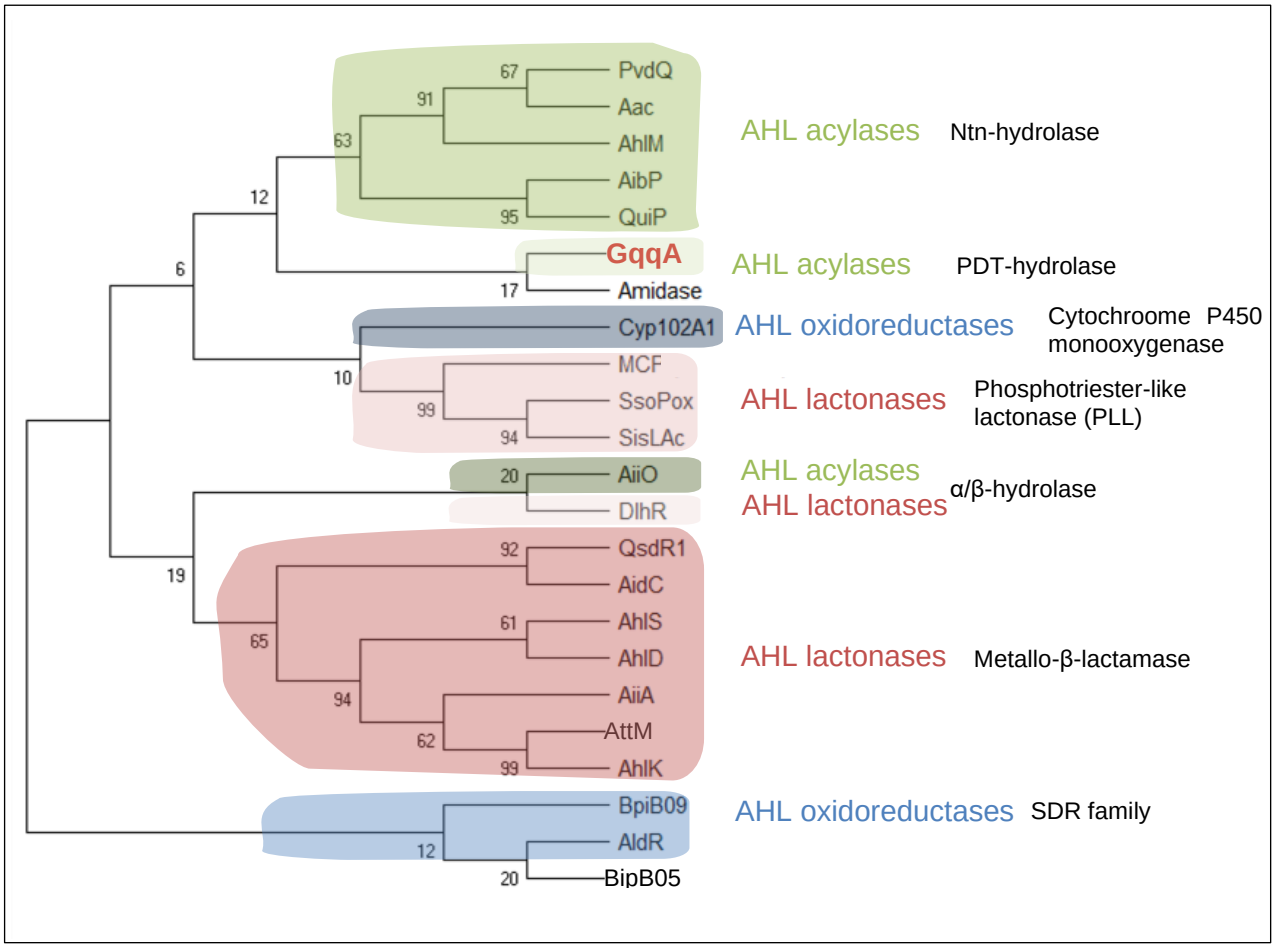

Figure 44: Phylogenetic construction of protein sequence of known QQ degrading enzymes
The phylogenetic tree was calculated by using Mega X and the Maximum Likelihood method with JTT matrix-based model. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test were shown next to the branches. This analysis involved 23 aa sequences: PvdQ (Q9I194), Aac (Q8XWC7), AhlM (AAT68473.1), AibP (Q8YDG4), QuiP (Q9I4U2), GqqA (KOEU_05990), Amidase (Q396E7), Cyp102A1 (P14779), MCP (MAP_3668c), SsoPox (Q97VT7), SisLAc (C4KKZ9), AiiO (D8X182), DlhR (ACP22150.1), QsdR1 (ACP23138.1), AidC (I7HB71), AhlS (SCV20999.1), AhlD (AAP57766.1), AiiA (P0CJ63), AttM (Q7D3U0.3), AhlK (A0A098CCK3), BpiB09 (ABU51100.1), AldR (ACP26074.1) and BpiB05 (ABU51109.1). The different enzyme classes are highlighted in different colors. The corresponding enzyme families were shown. The enzymes were chosen from Fetzner et al 2014.Corresponding accession numbers were given in brackets after enzyme name.

Structure similarity to other PDT enzymes

By using the Cofactor structure-based function prediction tool (Zhang et al., 2017). Proteins with similar structure were searched based on the PDB formats. Therefore, a TM-alignment was performed using protein structures. The TM Score scaled the structural similarity between different proteins. A value of 1 indicated a perfect match between the structures. RMSD meaning root-mean-square deviation and showed the average distance between the

aligned atoms. A value of 0.00 Å indicated an identical structure. All values between 1.00 and 3.00 Å showed similarity in structure and a value more than 3 Å meant distant structural similarity. Also, the sequence similarity between the proteins was analyzed. The coverage was the number of structural aligned residues divided by length of the query protein GqqA. The results of the structural similarity were shown in Table 18. The highest structural similarity to GqqA with a value of 0.86 was found with the putative chorismate mutase (3luy) of *Bifidobacterium adolescentis*. The distance between the aligned atoms was 2.49 Å and indicated a similarity based on structure. But the sequence of 3luy showed a lower similarity with 27.7 % similarity. The prephenate dehydratases from *Arthobacter aurescens* (3mwb), *Chlorobiumtepidum TLS* (2qmX) and from *Staphylococcus aureus* (2qmw) had high structure similarities (0.84 or 0.80 identity) to the GqqA structure. Also, the RMSD values of these proteins indicated structural similarities. The sequences of these proteins showed only a low similarity to GqqA sequence with values between 0.207 and 0.311. The prephenate dehydratase of *Streptococcous mutants* (4lub) showed only 56 % similarity in the protein structure compared to the above-mentioned PDT enzymes. There were not found any structurally similarities to known QQ enzymes by using this online tool.

Overall, there were major differences between structural and sequential similarities between the resulting proteins and the query protein. While all proteins showed a high structural similarity, the sequences of these proteins had hardly coincidences.

Table 18: Structural similarity of GqqA to other PDT proteins

No.	PDB hit	Organism	Structure similarity	RMSD*1	Sequence similarity	Coverage*2
1	3luy	*Bifidobacterium adolescentis*	0.86	2.49	0.277	0.964
2	3mwb	*Arthrobacter aurescens*	0.84	2.89	0.288	0.975
3	2qmx	*Chlorobium tepidum TLS*	0.84	2.54	0.331	0.950
4	2qmw	*Staphylococcus aureus*	0.80	2.64	0.207	0.921
5	4lub	*Streptococcus mutants*	0.56	2.77	0.237	0.633

*1RMSD Root-mean-square deviation [Å]
*2Coverage Number of structural aligned residues divided by length of query protein

Structure analysis of GqqA

(The structure and data analysis were performed in cooperation with Nadine Werner of the department of Prof. Dr. Betzel at the DESY facility (University of Hamburg).

The protein was overproduced and purified similar to previously published protocols (Valera et al., 2016) with some minor modifications. The purified protein was verified by SDS-PAGE. The purification resulted in a single protein band on protein gel and had a MW of 30.5 kDa (appendix, Figure 57) that was fitted with the calculated protein mass. The group of Prof. Dr. Betzel especially Nadine Werner did some modification at protein expression by using selenomethionine (se-methionine) instead of methionine. This served to solve the phase problems during crystallization that frequently occurred with GqqA. With selenomethionine a protein crystal was reaching 0.98 Å. GqqA without selenomethionine resulted in crystals leading crystals diffraction with a wavelength of 1.03 Å.

In the crystallization and the data collection was seen, that GqqA acted as a dimer consisting of two homologous monomers that were displayed in red and blue (Figure 45, C) Each monomer had the predicted PDT enzymes that showed high similarity to other PDT enzymes like 3luy or 2qmx (Table 18). Such as 3luy or 2qmx, GqqA shared the same folding and the same division into two domains. The periplasmatic domain was located at the N-terminal and consisted of two structurally similar subdomains (Figure 45 and Figure 46). Both subdomains had a β-sheet harboring three or four parallel and one anti-parallel strand, which were surrounded by three α-helices. The ACT domain typical for PDT enzymes (green structure, Figure 45, A) was located on the C-terminal and had two α-helices, which formed a base for the entire protein. An anti-parallel β-sheet with two strands was also seen in the ACT domain. It was known that the ACT Domain could bind allosteric effectors such as phenylalanine, also the GqqA structure showed a binding site for phenylalanine in each monomer.

Compared to the GqqA structure, the structure of the Se-Methionine labeled crystal (Figure 45; B) was lacking the middle domain (black region, Figure 45; A) from amino acid Glu86 to Thr175. It seemed that the primary structure of Se-GqqA is digested at the peptide bond between Val 85 and Glu86 and at the peptide bond between Thr175 and His 176 in a post-translational processing step. The first 85 aa (Met1 to Val85) belonged to the α-peptide (red). The green domain of the structure started at His170 and ended with the last amino acid and represented the β-peptide.

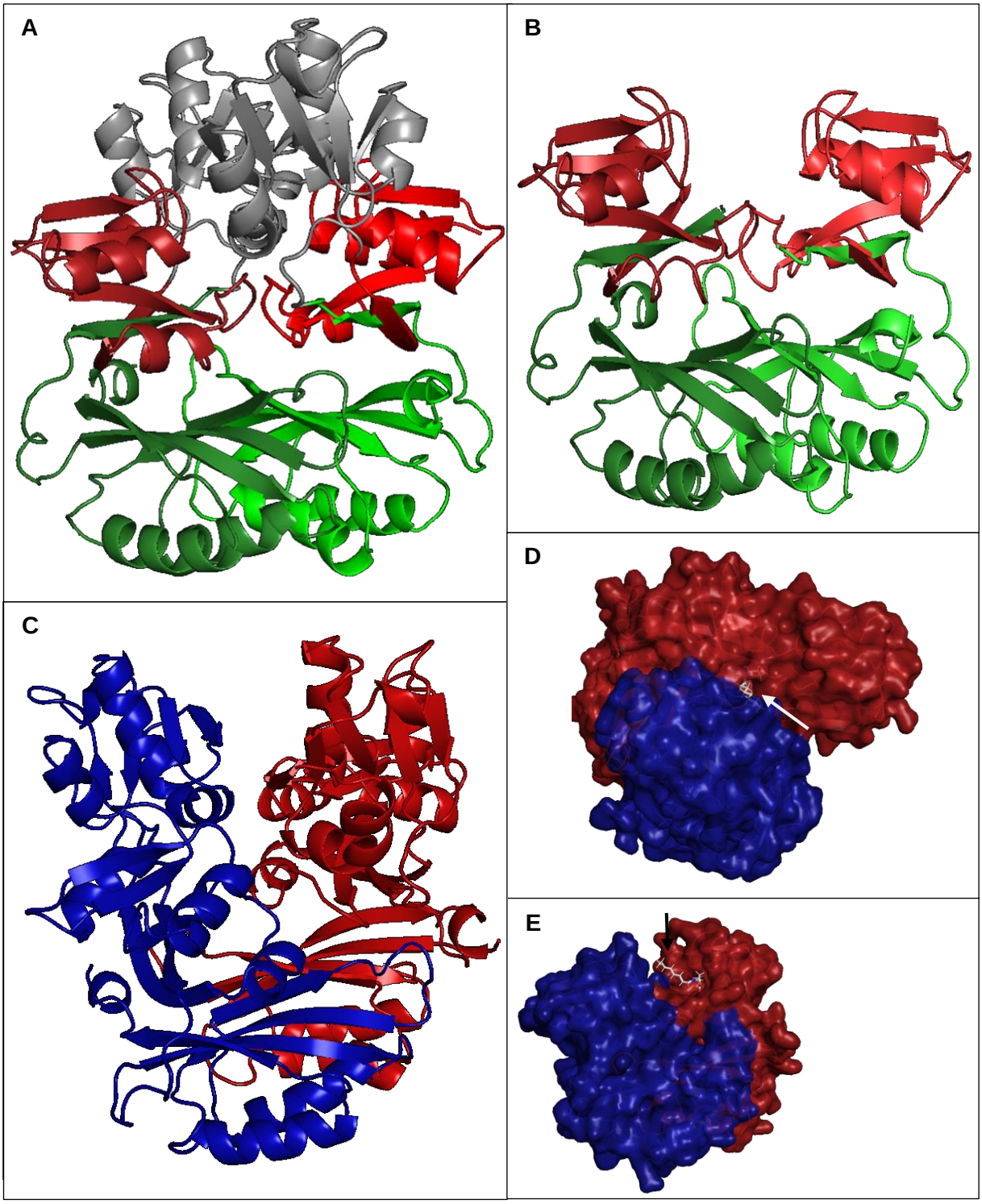

Figure 45: GqqA crystallization and 3D data collection
A: 3-D structure of GqqA using dataset collection at a wavelength of 1.03 Å from a single protein crystal. All three domains of GqqA were shown: α-peptide (red), lacking region (black) and β-peptide (green). **B**: Crystal structure of Se-GqqA that data set was collected at a wavelength of 0.98 Å from a single SeMet- labeled protein crystal. The middle part of the protein sequence was missing in the structure. Only the α-peptides (N-terminal, red) and both β-peptide (C-terminal, green) were seen.
C: Crystal structure of GqqA showing that GqqA was a dimer in solution with two homologues monomers.

D: 3-D structure of GqqA with corresponding protein surface. The middle part of the protein peptides covered up the active site of the enzyme. The substrate 3-oxo-C8-HSL was loaded to the structure (white cartoon, see white arrow). The arrow also indicates the viewing direction from above. **E:** 3-D structure of Se-GqqA with corresponding protein surface. The substrate 3-oxo-C8-HSL was docked to protein structure. The active site of GqqA is freely accessible through the missing middle domain (black domain in A). The black arrow indicates the top view of the protein for orientation.

By using docking analysis, a possible substrate binding sites could be verified within the protein sequence. The alleged substrate pocket was located in the area of the α-peptide (red, Figure 45, D, E). The GqqA crystal structure (Figure 45, D) showed how the middle domain covers the substrate pocket. By lacking the middle domain in the Se-GqqA structure, the substrate pocket was freely accessible for the substrate. The binding of the lactone ring to the enzyme was clearly recognizable, while the nonpolar acyl chain was directed to the outside of the protein (Figure 45, D).

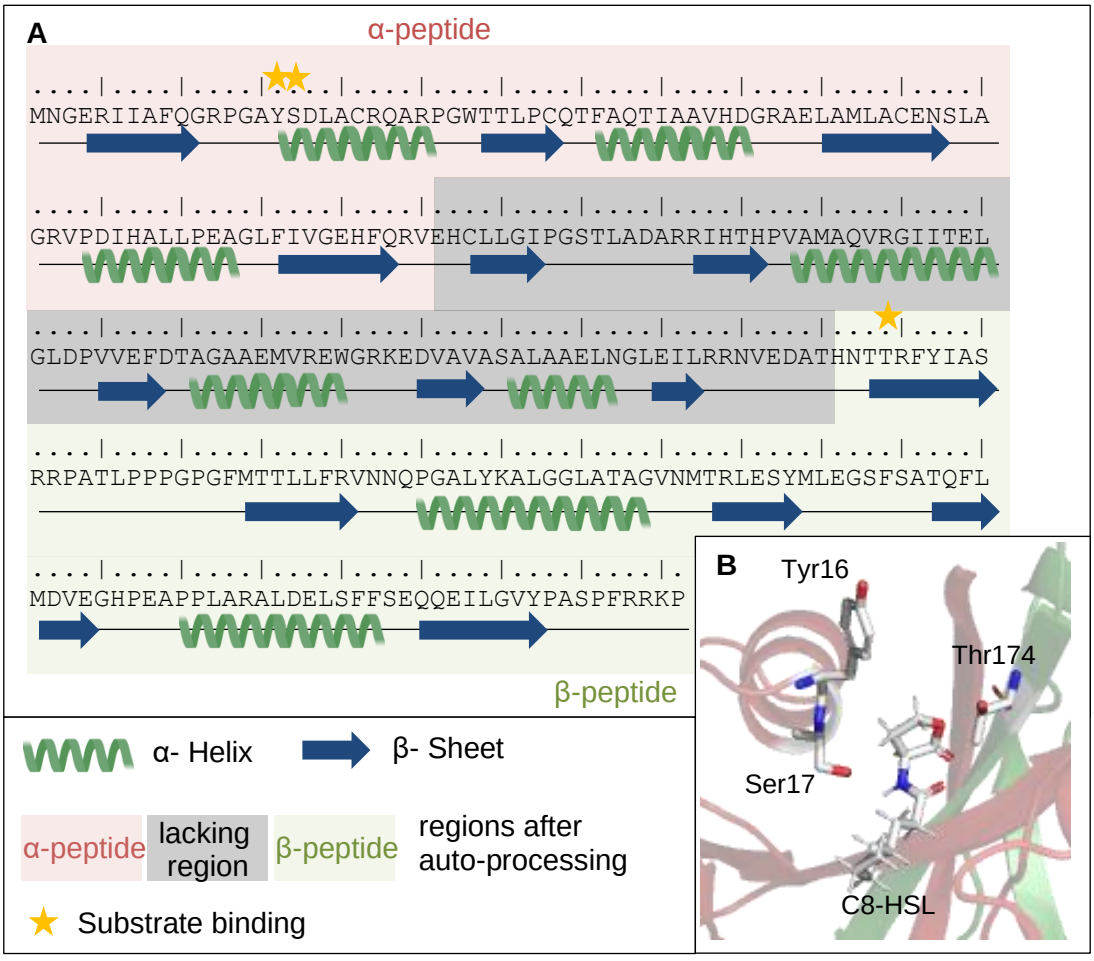

Figure 46: Secondary structure and post-translational processing of GqqA
A: The aa- sequence of GqqA with the associated secondary structure is shown. The different regions modified by post-translational auto-processing are highlighted in red (α-peptide), black (lacking region) and green (β-peptide). The possible substrate binding aa Tyr16; Ser17 and Thr174 are flagged with a star. **B:** Substrate pocket of GqqA showing the substrate octanoyl-homoserine lactone (C8-HSL) and the putative substrate binding residues Tyr16, Ser17, and Thr174. Nitrogen atoms are blue and oxygen atoms are red.

Two of the three putative substrate binding residues were located on the α-peptide (Figure 46. Here, the polar residues of Tyr16 and Ser17 consisted of hydroxy groups that could interact with the oxygens of lactone ring. The third possible residue that could interact with the lactone ring was the hydroxy group of Thr174 that was located at the beginning of the β-peptide. All these residues were highlighted with a star in the secondary structure of GqqA (Figure 46, A). That all the three residues were directed to substrate lactone ring was observed within the docking analysis of the substrate octanoyl homoserine lactone and the three-dimensional crystal structure of Se-GqqA. The docking analysis could be displayed with PyMOL (Figure 46, B).

Mutagenesis of GqqA and *in vivo* activity assay

In order to elucidate possible interaction and dimerization regions in the structure of GqqA, several different mutations in the gene *gqqA* were constructed. For this purpose, five different mutations within the *gqqA* gene (Table 12) were generated using quick exchange mutation approach and the wt version of the overexpression construct pET-21a::*gqqA* (Table 2) as template. The resulted PCR products were digested with *Dpn*I and a phosphate group was

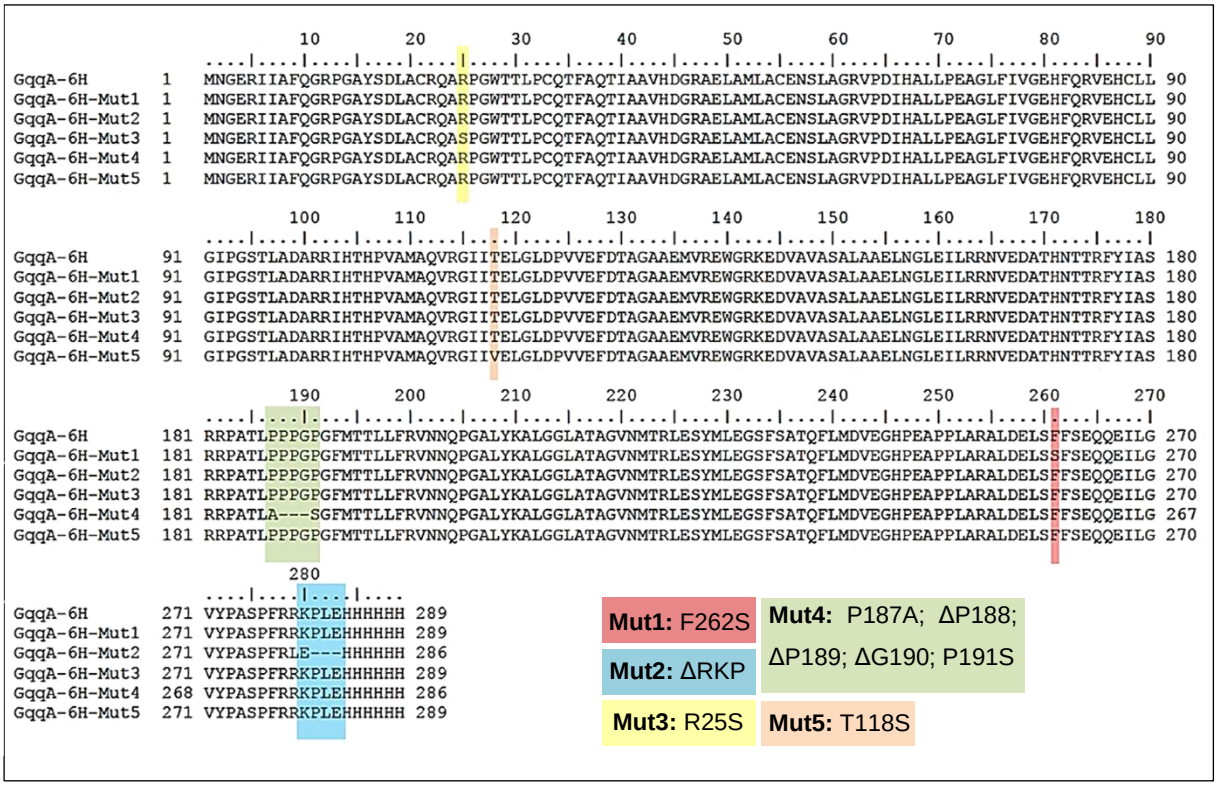

Figure 47: Overview of GqqA mutagenesis
The aa sequence of GqqA-6H wt and of all five mutants are shown. The different mutations are highlighted in different colors. Mutation 1 (F262S) was shown in red, mutation 2 (ΔRKP) in blue and mutation 3 (R25S) in yellow. The modification at the loop (Mut4) was highlighted in green and the aa exchange T118S (mutation 5) in orange. The one- and three letter code for amino acids is attached (Table 19)

added to the amplificants with a Polynucleotide kinase. The linearized amplificants were ligated and the final mutation constructs were transformed into *E. coli* DH5α. The correctness of the mutations was checked by sequencing and the correct mutations constructs were transformed into the overexpression strain *E. coli* BL21 (DE3). An overview of the mutants showed all modifications in the aa- sequence (Figure 47).

In GqqA-Mut1 (red), the amino acid Phenylalanine at position 262 was changed into Serine. This mutation could lead to the loss of hydrophobic interaction with symmetry-equivalent monomer. The second mutation (blue, GqqA-Mut2) concerned the end of the protein. The last three amino acids Arginine, Lysine and Proline was removed, and these deletions were supposed to prevent interaction with neighboring molecules. An amino acid exchange was performed with GqqA-Mut3 (yellow). Here, the amino acid Arginine was changed into Serine. The mutations could lead like mutation 1 to the loss of interaction. The surface loop (aa: 187 - 191) shall be shortened and mutated in GqqA-Mut4 (green). This mutation was intended to identify the function of this loop. The last mutation was an amino acid exchange, where the Threonine at position 118 was changed into Serine. Here too, the hydrophobic interaction to symmetry-equivalent monomer would be interrupted.

All GqqA variants were recombinant overproduced and purified using the fused His-tag. The purified proteins were loaded onto a native gel to detect the first effects of mutations (Figure 48, A). The protein GqqA-wt, GqqA-Mut1 (M1), GqqA-Mut2 (M2) and GqqA-Mut5 (M5) could be overexpressed and purified. Clear protein bands were seen on the native gel. But these bands did not run on the same level so that difference in dimerization could be suggested. GqqA-wt and GqqA-Mut1 run on the same level. It seemed that both had the same quaternary structure containing of more homologue peptide chains. GqqA-Mut2 and GqqA-Mut5 run on a lower level than GqqA-wt or GqqA-Mut1. It seemed that these native proteins consisted of less peptide chains than GqqA-wt or GqqA-Mut1. A second protein band was observed in the well of GqqA-Mut2. This protein band could be a degraded GqqA, because this band was on a lower level than the other bands. The running lines of GqqA-Mut3 and GqqA-Mut4 showed no clear protein band as it was seen in the other running lines. A small protein band was observed at the lower part of the native gel. This could be degraded peptide chains of GqqA.

To verify if one of these mutations had an influence on the functionality of the enzyme, a *in vivo* QQ activity test was performed. Therefore, the *E. coli* strains expressing the wt or the different mutated GqqA protein grew in LB medium supplemented with oxo-C8-HSL. After six hours the supernatants were collected and spotted on a filter paper that was placed on LB

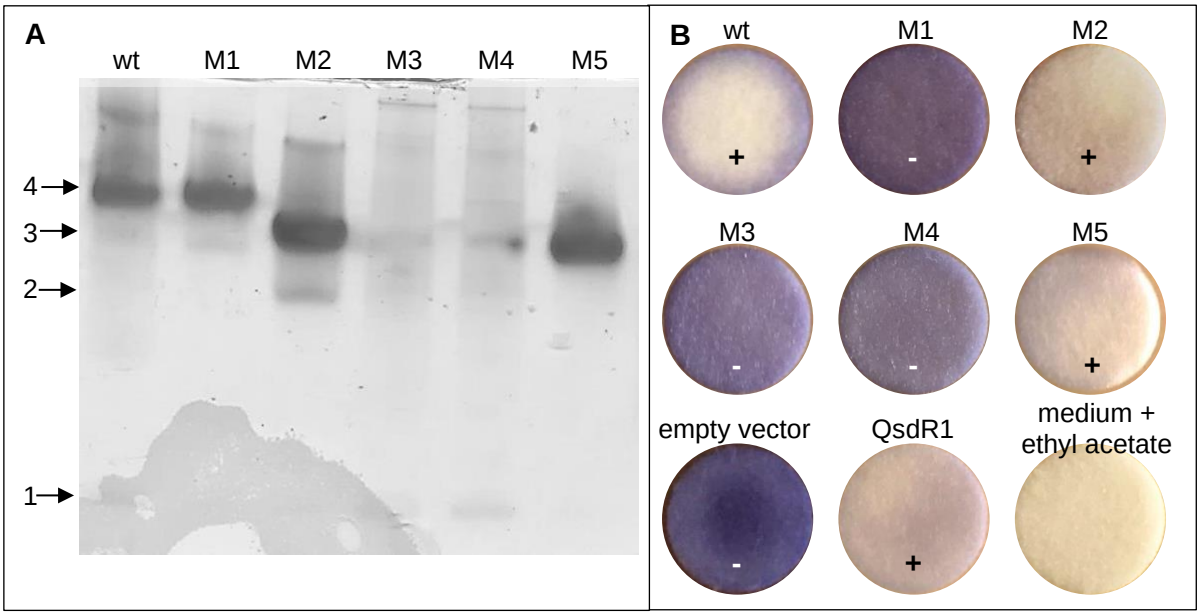

Figure 48: Native protein gel and in vivo assay of GqqA proteins
A: A precast TGX protein gel (Bio-Rad) were loaded with 20 µg of GqqA-wt, GqqA-Mut1 (M1), GqqA-Mut2 (M2), GqqA-Mut3 (M3), GqqA-Mut4 (M4) and GqqA-Mut5 (M5) and was run under native conditions. The gel was stained with Coomassie brilliant blue. **B:** Bioindicator assay of an i*n vivo* enzyme assay of GqqA wt and various mutants in *E. coli*. The reporter strain CV026 was used for indicating not degraded AI molecules (violet color). An active enzyme is shown by a colorless plate.

plates inoculated with the QS reporter strain CV026. A purple filter paper indicated no GqqA activity whereas a white filter paper showed an active expressed GqqA protein (Figure 48; B). An. *E. coli* Strain harboring the empty expression vector pET-21a was used as negative control. The HSL solvent ethyl acetate had no impact of CV026 violacein production.

A white filter plate with light violet color at the side was seen for GqqA wt. The same effect was observed for the positive control QsdR1. Also, by using the GqqA mutants GqqA-Mut2 and GqqA-Mut5, white filter paper lying on plates inoculated with CV026 were detectable. In contrast, GqqA-Mut1, GqqA-Mut3 and GqqA-Mut4 showed violet colored filter paper indicating that these mutants were not enzymatically active.

ESI-MS/MS analysis of GqqA

For identifying the mechanism behind the QQ activity of the enzyme GqqA, GqqA was shipped to our cooperation partner Dr. Manuel Ferrer (Department of Applied Biocatalysis, Institute of Catalysis, Madrid, Spain). His team performed electronic ionization coupled with Gas Chromatography and ESI-MS. The substrate 3-oxo-C8-HSL was solved in DMSO and was used for *in vitro* enzyme assay in 5 mM EPPs pH 8.0 and at 30°C. After three hours the reaction products were analyzed by Mass Spectrometry (MS) experiments using a hybrid quadrupole-time of flight mass spectrometer (QTOF, QSTAR pulsar, Applied Biosystems, California, USA) equipped with a micro electrospray ion source. The products were dissolved

in methanol and the samples were introduced into the spectrometer. The resulted products were shown in the mass spectrums (Figure 49). In addition to the assay with GqqA (+GqqA), a control assay without enzyme was also carried out (-GqqA).

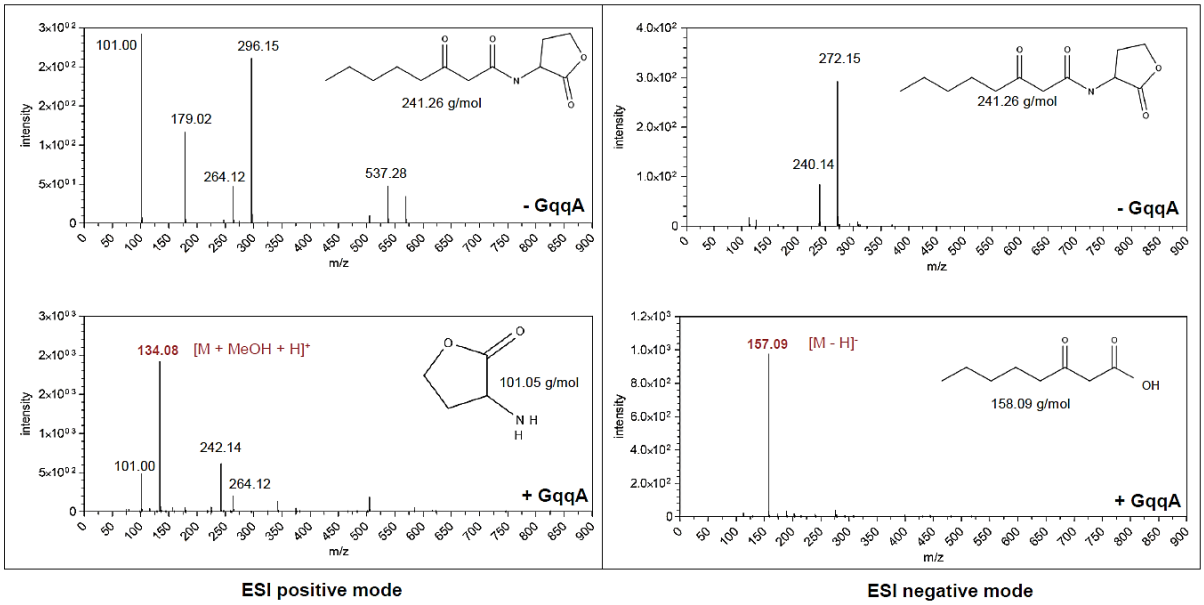

Figure 49: Mass fragmentation spectrum of the GqqA enzyme assay in positive and negative mode

The upper spectra indicated the control that took place without enzyme (-GqqA) in the positive (left) and in the negative (right) mode. MeOH was used as solvent during MS analysis, it appears an esterified substrate or product. The control spectrum in the positive mode shows an $[M + MeOH + Na]^+$ ion at an *m/z* of 264.14 and an $[2M + 2MeOH + Na]^+$ ion at m/z of 537.28 that indicate the substrate 3-oxo-octanoyl homoserine lactone. The structural formula is drawn in the spectrum. The substrate is also observed at an m/z of 240.14 $[M – H]^-$ in the negative mode. By adding GqqA to the assay (lower spectra) the substrate is only observed in lower intensity. An $[M + MeOH + H]^+$ ion is observed at an *m/z* of 134.08 in the positive mode and indicates the product homoserine lactone. In the negative mode an $[M – H]^-$ ion is detectable at an *m/z* of 157.09 and displays the 3- oxo- octanoyl acid. The homoserine lactone and the 3-oxo- octanoyl acid are displayed in the spectra. The $[M + Na]^+$ ion at an m/z 101.00 and the [2M + Na]+ at an m/z at 179.02 display.

The lower spectra represented the MS fragmentation of the reaction in positive and negative mode. The resulted products of the enzyme assay with GqqA were shown as an $[M + MeOH + H]^+$ ion in the positive mode indicating a homoserine lactone and an $[M - H]^-$ ion in the negative mode that displayed the product 3-oxo-octanoyl acid. These ions were not detectable in the control spectra without GqqA in the enzyme assay. Clear peaks were visible in the control spectra (-GqqA) representing ions for the substrate 3-oxo-octanoyl homoserine lactone. Thus, the $[M - H]^-$, $[M + Na]^+$ and $[2M + 2MeOH + Na]^+$ ions at an *m/z* of 240.14, 264.12 and 537.28 showed the substrate ion and an esterified substrate ion with the MS solvent MeOH. The solvent DMSO, in which the AI molecules were dissolved, a peak was recognizable with the $[M + H]^+$ and $[2M + H]^+$ ions at an *m/z* of 101.00 and 179.02.

These enzyme assays with the corresponding MS analysis suggested that GqqA acts as an HSL acylases, which cleaves the amide bond and releases a fatty acid and a homoserine lactone ring.

5. Discussion

Bacterial communication such as quorum sensing bases on the amount of chemical molecules called AI and is an established mechanism for recording the population density. When a threshold population density is reached, the bacterial cells collective coordinate the gene expression of certain genes. With this communications system, biofilm formation, virulence factor production or bacterial- host interactions were synchronized under differing environmental conditions (Ng and Bassler, 2009). In the past, several strategies of competing organisms have been discovered to interrupt these signaling pathways. One of the interfering strategies is the enzymatic degradation of signal molecules. The main enzymatic class of these quorum quenching enzymes are AHL lactonases catalyzing hydrolytic ring opening of lactones and AHL acylases cleaving the amide bond (Fast and Tipton, 2012; Fetzner, 2014). The third group of interfering enzymes is the enzyme class of oxidoreductases. The resulting products can also act as AI molecules but with less binding ability to regulator protein (Chowdhary et al., 2007).

The genome of NGR234 codes for two QS systems belonging to the class producing AHL as AI molecules (He et al., 2003). For example, NGR234 uses QS to coordinate bacteria-plant interaction (Hartmann et al., 2014). All genes, which are necessary for this symbiotically interaction, are located on a symbiotic plasmid that has high similarity to Ti-plasmid of the plant pathogen bacterium *A. tumefaciens* (Perret et al., 1999; Schmeisser et al., 2009).
The fact that the plant symbiont *S. fredii* NGR234 has evolved a diverse and unique tool to infect more than 120 legume and one non-legume plants makes NGR234 to a unique model organism for the investigation of plant-bacteria interaction in conjunction with bacteria communication and small proteins.

The aim of my research was first to identify the role of small proteins in a quorum sensing dependent expression and second the structural and functional analysis of a novel quorum quenching hydrolase designate as GqqA.

NGR234 Δ*ngrI*Δ*traI*_copy+ induces root hair curling in absence of apigenin

The symbiosis is initiated by a signal exchange between the plant and the bacterium. The plant releases secondary metabolites like flavonoids detected by the bacterial transcription regulator NodD1. This complex activates the synthesis of strain-specific Nod factors triggering root hair curling on susceptible legume plants (Barnett and Long, 2015) The mutant NGR234 Δ*ngrI*Δ*traI*_copy+ is not able to produce any AI molecules (Krysciak et al., 2014) and it was shown that this mutant can trigger root hair curling in a flavonoid-independent manner (Figure 6). In absence of the AI molecules the pNGR234*a* copy number increases to a 4 - 6-fold higher level compared to the parent strain and this overexpression of nearly all genes on the symbiotic plasmid can induce the formation of the symbiosis in absence of plant released flavonoids (Grote et al., 2016). This experiment shows that the mutant not only upregulates the transcription of DNA into RNA, but also translated the mRNA to functional proteins. Without Nod factor producing proteins that are encoded by *nodABC* genes, the legumes were not stimulated to curl the root hairs as it was seen for NGR234 Δ*ngrI*Δ*traI*_copy+ (Freiberg et al., 1997; Esseling et al., 2003; Parniske and Downie, 2003; Jones et al., 2007; Grote et al., 2016).

Initially, it appears that the AI controls the symbiotic plasmid copy number in cooperation to the non-coding RNA NGR_a00040 and NGR_a00050. Similar reports were already made for *Rhizobium leguminosarium.* The transcription of *repA* was increased in presence of native AI molecules (McAnulla et al., 2007).

I. Quorum sensing dependent expression of small proteins

Small proteins are defined as synthesized directly by expression of smORFs rather than by proteolytic processing of larger proteins (Storz et al., 2014). With the first annotation of genomes an ORF cutoff of 300 nts were set and smaller ORFs of less than 50 aa were largely underestimated. In recent years focused functional genomics efforts have led to the identification of various numbers of new small proteins that were encoded in the genome of Gram-positive and Gram-negative bacteria. It became clear that these small proteins have influence to several cellular functions involving in regulatory networks, virulence, transporters or sporulation (Storz et al., 2014). In NGR234 QS regulates all symbiotic genes located on symbiotic plasmid pNGR234*a* (He et al., 2003).

Since nearly all genes on pNGR234*a* were upregulated in NGR234 Δ*ngrI*Δ*traI*_copy+, this strain can be used as a tool to analyze expression of small proteins on pNGR234*a*.

The symbiotic plasmid pNGR234*a* encodes for additional 251 smORFs in-between the 422 features

The replicon pNGR234*a* encodes for 422 ORFs and in-between these ORFs are a lot of intergenic regions, which are not annotated (Schmeisser et al., 2009). Thus, these obvious, intergenic regions encode for a high number of unknown genes.

In the last years, the number of identified small proteins increased due to the possibilities for deeper sequencing methods (Boekhorst et al., 2011). These investigations in small proteins showed the high functional importance of small proteins (Wang et al., 2008). A smORF search with subsequent mapping of the receiving smORFs to transcriptomic data resulted in 251 additional entries on symbiotic plasmid of NGR234. Due all new smORFs had transcripts in NGR234 wt and/or NGR234 Δ*ngrI*Δ*traI*_copy+ suggests that RNA was produced for corresponding gene regions. Previous published analysis demonstrated that not only RNA was transcribed in mutant strain but also the function protein was synthesized, hence. NGR234 Δ*ngrI*Δ*traI*_copy+ was able to trigger root hair curling in absence of flavonoids because all genes on pNGR234*a* had a higher transcriptional level (Figure 6, Grote et al., 2016).

Proteins encoded by ORFs start with a translation start (ATG, GTG or TTG) and end with a translation stop (TAA, TAG or TGA). Between start and stop codon, there is a sequence of nucleotide triplets lying in a reading frame and coding for a specific amino acid (Turanov et al., 2009). All smORFs identified in this thesis have these described characteristics and lead to suggest that these ORFs were expressed to small proteins. However, there are non-coding RNAs (ncRNA) that have protein-like characters or sequences that have errors caused by lining up the contigs of sequenced DNA fragments (Morozova and Marra, 2008). Furthermore, internal ATG codons (coding for methionine) could be mistaken for start codons and the received smORF could be a part of larger coding region (Sieber et al., 2018).

Using novel approaches such as the high-resolution Ribo-seq, the expression of the obtained smORFs into functional small proteins can be identified. This Ribo-seq analysis was started for NGR234 Δ*ngrI*Δ*traI*_copy+. Since NGR234 cells are often challenging to handle, the results of Ribo-seq analysis were expected soon. In the meantime, the mRNA protected by ribosomes during translation was successfully isolated and sequencing approach was started. When the results are available, a global overview of the expression of all smORFs to small proteins on pNGR234*a* will be shown.

The intergenic region in-between *traI* and *repABC* encodes for an operon consisting three small proteins were essential for plasmid maintenance

An Operon is a functional unit of DNA that contains a cluster of genes under the control of a single promoter. All the genes belonging to the operon are co-transcribed into a single mRNA (Jacob and Monod, 1961). A first indication of genes that are organized in an operon resulted from the gene order and orientation, which are often conserved in two or more bacterial genomes (Ermolaeva, 2001). In the intergenic region in-between the autoinducer synthase *traI* and the plasmid replication operon *repABC* three smORFs were identified and ordered closely together and only a few bases separated the single smORFs (Figure 35). In order to identify whether these three ORFs were co-transcribed into a single mRNA, the total RNA was isolated, and revers transcribed to cDNA for qPCR analysis. The results of the qPCR proved that all three smORFs were co-transcribed into a single mRNA because all three smORFs were detected on one mRNA fragment (Figure 36).

Genes belonging to an operon were co-regulated at transcriptional level (Hodgman, 2000).Thus, a second evidence that these three genes belonged to an operon showed the transcriptional data of NGR234Δ*ngrI*Δ*traI*_copy+ (Table 25; appendix). The smORFs *repX*, *repY* and r*epA0* were all highly upregulated in mutant strain and thus manifested same regulation of transcription. It can be assumed that the expression of this operon was under the control of a promoter upstream of *repX*. Since the translation start of *repX* was located 153 nts downstream of the calculated transcription start showed that the operon consisted of a leader sequence (also known as 5'UTR) with a size of 153 nts (Figure 14). This untranslated region could contain a possible riboswitch regulating the own translation in response to effector molecules (Winkler, 2005).

Small proteins often need an interaction partner to gain a structural and functional conformation (Storz et al., 2014). The binding assay with purified RepX and cell lysate of NGR234 gave first indications that RepX and RepA0 interact to each other and this interaction indicated the assumption of a common functional role in bacteria cell that characterized the definition of operons (Hodgman, 2000). Specific antibodies of RepX and of RepA0 were bound to proteins with same molecular weight of approx. slightly less than 30 kDa (Figure 21). All molecular weights of these three proteins RepX (5.6 kDa), RepY (6.2 kDa) and RepA0 (16.2 kDa) gave a molecular weight of slightly less than 30 kDa. A new pulldown assay with detection of protein bands with a specific antibody for RepY should be confirming this suggestion.

RepABC type plasmids were divided into four classes of conjugations systems (Wetzel et al., 2015). The symbiotic plasmid of NGR234 belonged to class I RepABC type plasmids that

were composed of *traAFBH* and *traCDG* operons and an IncP-like type IV secretion system (Li et al., 1998).

The three smORFs *repXYA0* were also identified in related bacteria species harboring RepABC type plasmids (Table 16) and a calculation of a phylogenetic three comparing the genetic region of *traI-repA*, indicated a classification into two clusters. Cluster II species contained only a shorter intergenic region whereas strains of cluster I had an intergenic region of 803 nts. These clusters formation could have been achieved by horizontal gene transfer with a diverging evolution. The classification of RepABC type plasmids subdivided this type of plasmids into further subclades as previously known. Many genomes of RepABC type plasmids were not yet fully sequenced. When new sequences become available over time, further clades will emerge in future with a high complexity (Wetzel et al., 2015). In a previously published article, it was shown that *A. tumefaciens* did not have such a long intergenic region as NGR234 (Grote et al., 2016). *A. tumefaciens* did not belong to Clade II group based on the organization of the conjugation system like NGR234 or other *Sinorhizobium* species. On *A. tumefaciens* Ti-plasmid, the *tra*-operon is separated from *rep* and *trb* operon (Wetzel et al., 2015). Thus, it seems that this long intergenic region in-between *traI* and *repABC* is a unique characteristic for genus *Sinorhizobium* and could be a key to broad host range.

Since many genomes had not yet been completely sequenced, it was not possible to make a clear statement as to whether RepXYA0 only occurs in other closely related *Sinorhizobium* / *Ensifer* species or whether they are more widely distributed.

The functional role of RepXYA0 was not identified, but with the first results of characterizing these three smORFs, different hypothesis could be formulated.

When the expression of *repXYA0* was not under control of *repXYA0* promoter and no AI molecules were present (NGR234Δ*ngrI*Δ*traI*_copy+ mutant), the symbiotic plasmid copy number increased to a 6-8-fold amount compared to wt copy number (Figure 24). This indicated that the smORFs *repXYA0* were essential together with both non-coding RNA in-between *repB* and *repC* for plasmid copy number regulation (Schmeisser et al., 2009). This phenomenon was not observed in the newly constructed AI double deletion mutant, where the promoter region of *repXYA0* was reserved. Also, when *repX* was genetically added on a high copy plasmid in NGR234Δ*ngrI*Δ*traI*_copy+ cells, the effect of high symbiotic plasmid copy number was quenched to wt level. These observations promoted the hypothesis that *repXYA0* was involved in copy number regulation.

A deletion of *repX* had the effect that non symbiotic plasmid was observed in NGR234 cells (Figure 24). Without symbiotic plasmid the cells were not able to initiate symbiotic interaction. No nodules were observed in NGR234 Δ*repX* (Figure 25). Either the replication of symbiotic plasmid failed without *repX* (Figure 50, B) or neither the energy budget for maintaining pNGR234a was not acceptable for the cells and resulting in the loss of symbiotic plasmid (Figure 50, C). Because the replication of DNA is a process that consumes ATP at different steps of replication and partitioning (Jesu et al., 1998; Havey et al., 2012). All genes that were essential for bacterial survival are located on chromosome and NGR234 is viable without pNGR234*a*, hence (Morrison et al., 1983; Schmeisser et al., 2009). As a third hypothesis was that without RepXYA0 the partitioning of symbiotic plasmid failed during cell division (Figure 50, A). RepA and RepB were the responsible proteins for plasmid partitioning similar to ParA and ParB that mediated the partitioning of chromosomes to nascent daughter cells (Mazur and Koper, 2012) and RepXYA0 could be involved in stabilization of RepA-RepB-*parS* complex. Especially the mutation of *repY* indicated that distribution of pNGR234*a* did not occur correctly during cell division because the symbiotic plasmid could not be detected in

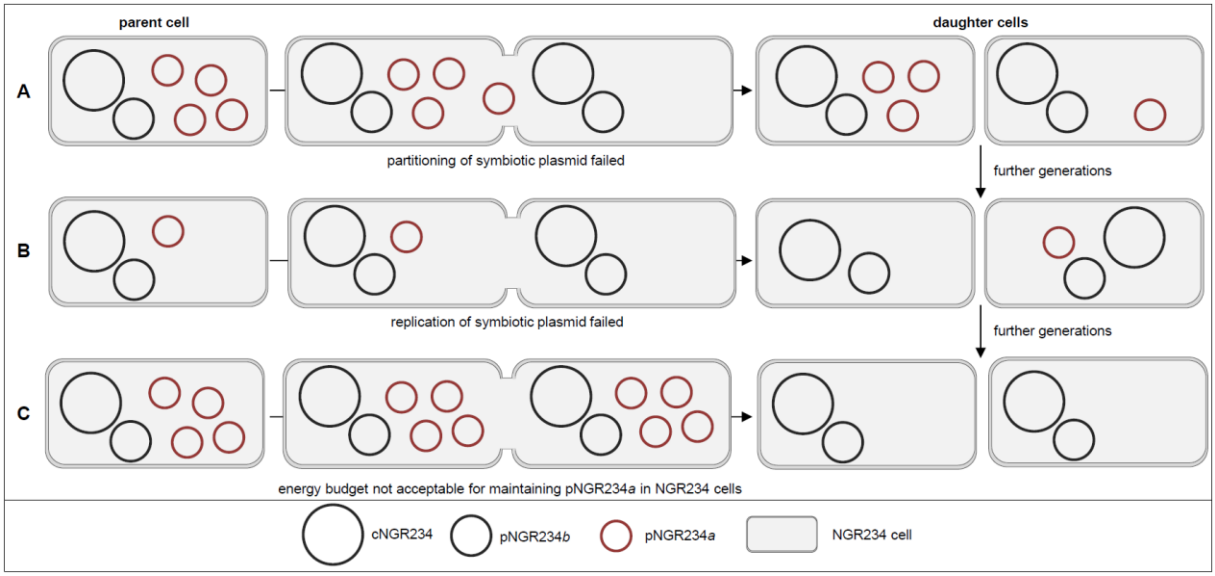

Figure 50: Hypothesis on RepXYA0 function
During cell division, the replicons must replicate and segregate to nascent daughter cells. In Rhizobia the plasmid encoded at least one *repABC* operon, which products that were involved in plasmid replication and partitioning to daughter cells in a process that was analogue to mitosis. The results of this work indicated that repXYA0 operon was involved in this system. **A:** A parent cell with high plasmid copy number started to divide but the even partitioning failed, and the plasmid were not distributed equal to both daughter cells. This led that in further generations existing free-plasmid cells. **B:** RepXYA0 were involved in replication and without *repXYA0* the replication could be failed thus that one of the nascent daughter cells did not carry a plasmid. **C:** A third hypothesis for loosing pNGR234a was that the maintenance of high plasmid copy number in bacterial cells require a lot of energy for plasmid replication. This required energy could be too high for the cells and the only opportunity was to forfeit the plasmid.

each colony (Figure 31). The faulty partitioning led to the fact that there were plasmid-free cells or even the whole culture did not carry pNGR234*a* at all since this was the fact in NGR234 Δ*repX*. All these findings suggested that *repXYA0* was essential for plasmid replication and partitioning. Hypothetically, it could be that RepX and RepY stabilized the partitioning complex (RepA-RepB-*parS*) and RepA0 stabilize the RepC-*oriV* complex recruiting other components for plasmid replication (Wetzel et al., 2015).

A high copy number of pNGR234*a* allowed expression of all infection-related genes in absence of host-specific signals and when the cell density was not detectable via QS (Figure 6). These findings that were made here implied that NGR had evolved an alternative system to ensure a successful symbiosis even in absence of a compatible host or in absence of other NGR234 cells by increasing the symbiotic plasmid copy number and expression of all relevant proteins for symbiosis (Grote et al., 2016). Such similar behavior had already discovered in human pathogenic bacterium *Yersinia* sp. The small secreted protein was involved in regulation of copy number to increase sufficient of establishing virulence. The colleagues showed that a single copy of pathogenic plasmid was insufficient in infecting in mice (Wang et al., 2016).
All in all, it could be suggested that the small proteins RepXYA0 could be the major key to broad host range because *repXYA0* was only identified in strains that were able to infect more than one host (Pueppke and Broughton, 1999; Grote et al., 2016; Temprano-Vera et al., 2018).

QS-dependent expression of *repXYA0*

The opposite direction of *repXYA0* and *traI*, and the fact that the corresponding promoter regions were located within mutually coding regions indicated a QS dependent expression of *repXYA0* (Figure 51). When the AI level reached a threshold concentration, AI bounds to transcriptional regulator TraR and activated the expression of *traI* (He et al., 2003). The TraR-AI-DNA complex initiated the binding of RNA polymerase for *traI* transcription. In this case the expression of *repXYA0* was inhibited (Figure 51, A) because The transcription components of *traI* obstructed the bonding between RNA polymerase and -10 and -35 regions of *repXYA0*. But when the TraI/R system of NGR234 was inhibited by TraM or the absence of AI or TraR, the RNA polymerase could transcribed *repXYA0* and the expression of AI synthase was blocked (Figure 51, B).

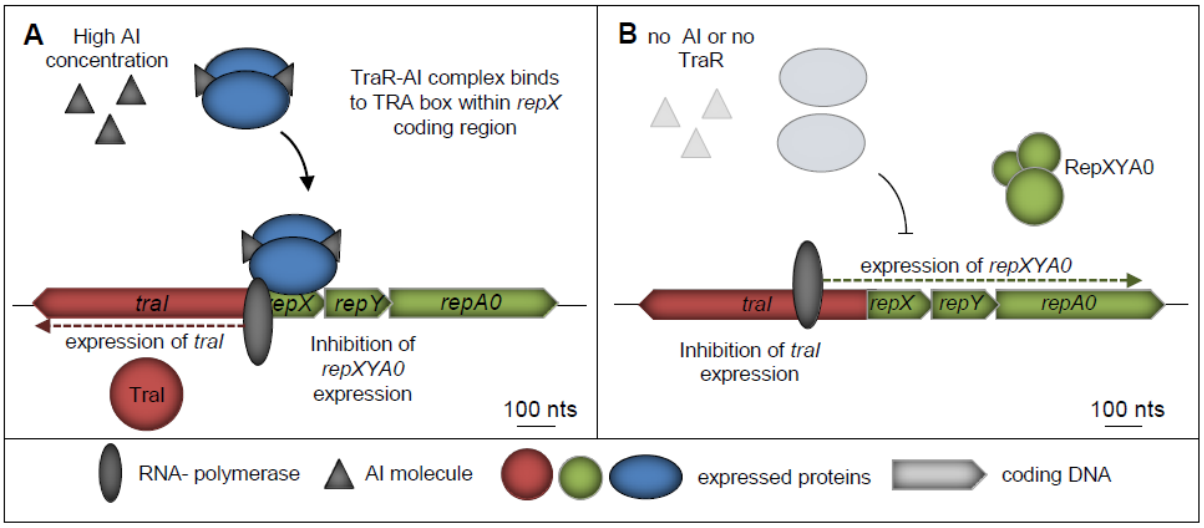

Figure 51: The model of competing expression of *traI* and *repXYA0*
A: At high AI concentration, the TraR-AI complex binds to DNA (TRA box) withing *repX* coding region and activates the transcription of *traI* gene. As long as TraR and RNA polymerase bind to *traI* promoter region, *repXYA0* cannot be transcribed. The expression of repXYA0 is inhibited by TraR. **B:** When no AI and thus no TraR is present in NGR234, RNA polymerase binds to *repXYA0* promoter region and the expression of *repXYA0* starts causing pleiotrophic effects.

That the expression was dependent on QS was also shown with genomic translations fusion and with promoter fusion in various NGR234 cells. Whereas, a red fluorescence that was observed in NGR234 Δ*traR* harboring the genomic translation fusion (*S. fredii* NGR234 Δ*traR* ::gTF*repX*::mCherry) verified the expression of *repX* in absence of TraR (Figure 19), a genomic translation fusion of *repX* in NGR234 wt showed no red fluorescence and only the blue stained DNA was observed (Figure 56, appendix). Also, the higher fluorescence that was observed in NGR234Δ*ngrI*Δ*traI* cells harboring *repX*-promoter fusion (Figure 17) indicated the QS dependent expression of the small proteins RepX, RepY and RepA0.

The expression of four proteins encoded by smORFs was verified with a diverse set of tools

The investigation of small proteins is still in the beginning and there are only few approaches for verifying the expression of small proteins (Hobbs et al., 2011; Ramamurthi and Storz, 2014; Storz et al., 2014). The best methods to verify the translation of smORFs to functional proteins are Ribo-seq analysis and protein mass spectrometry but these analysis are often not accessible (Cassidy et al., 2016; Ingolia, 2016). Thus, some other approaches were used for verification of small proteins.

The four characterized smORFs were chosen because of the high transcript in NGR234 Δ*ngrI*Δ*traI*_copy+ mutant. Here, the expression of these smORFs should be verified with a diverse set of genomics and biochemical tools.

First, all smORFs could be expressed *in vivo* in *E. coli* expression strains to give hints if these ORF would be expressed to proteins. All smORFs were cloned into expression vectors and then overexpressed in *E. coli*. After some modifications it was possible for all smORFs to express them in *E. coli*. The biggest protein (RepA0, 143 aa) could be overexpressed with traditionally pET-vector systems harboring His-tag for purification (Mierendorf et al., 1998). But the expression of small proteins in *E. coli* could be challenging because small proteins could be a part of toxin antitoxin systems and contain a single hydrophobic helix that promotes membrane insertion and the loss of membrane potential that leads cell death (Hemm et al., 2008). Here, the use of larger tags helped to solve the problems with toxicity. The larger tags improved also a better visualization onto protein gels because the proteins bands at the lower part of gels were often grainy (Hobbs et al., 2011). Thus, the small protein RepX, RepY and A1725 were recombinant overexpressed with pMALc2x-vector system. The *E. coli* protein expression could be verified with specific antibodies for RepX and RepA0. The expression of A1725 and RepY could not be confirmed with specific antibodies. But this did not mean that these ORFs were not expressed to small proteins. The expression strains of *E. coli* can produce a wide range of different type of proteins, but often these recombinant expressed proteins need some post translational modification or interaction partner that are missing in *E. coli.* Thus, it could be possible that *E. coli* was not able to overproduce the proteins and the overexpression resulted in inclusion bodies.

Here, the expression of all four small proteins was determined in NGR234 cells by using specific antibodies and transmission microscopy. Each small protein was verified in NGR234 Δ*ngrI*Δ*traI*_copy+ mutant (Figure 22, Figure 29, Figure 34 and Figure 40). Black dots corresponding to the proteins were observed in each thin section. For determination of RepX in NGR234 cells *E. coli* overexpressing RepX and an *E. coli* harboring an empty expression vector were used as controls. These controls indicated the correctness of RepX-antibody specificity. That these proteins could be detected with specific antibodies demonstrated the expression of *repX*, *repY*, *repA0* and NGR_a01725 to functional proteins.

The best evidence that RepX and A1725 were expressed to functional proteins was given with two types of translation fusion. Translation fusions harboring translation start sites (ATG or GTG) were compared to translation fusion lacking these start sites. Without ATG or GTG no fluorescence was observed whereas with translation start strong red or green fluorescence was seen (Figure 18 and Figure 38). These fusions indicated that the upstream located Shine Dalgarno sequences (SD) were recognized by ribosomes and initiate translation by aligning ribosome with start codon (Malys, 2012). In many published studies,

fluorescent proteins have been used to detect gene expression at single cell level. Because of a high signal intensity needed for detection, high copy number plasmid such as pBBR1MCS were used. However, the use of high copy number plasmids should be applied with regard to regulation of genes through the corresponding regulator, which is normally encoded on bacterial chromosome in a single copy. For *repX* also a genomic translation fusion was conducted and integrated into pNGR234*a*. A red fluorescence could be observed indicating the protein biosynthesis. But normal confocal imaging reaches the limits for detection. Only with using airyscan detection methods an observation of red fluorescence was possible (Huff, 2015) but did not allow the localization of RepX. For this approach, a super resolution microscope is necessary.

The translation fusion of NGR_a01725 and the detection with specific anti-a1725-antibodies indicated the expression of NGR_a01725 to a small protein consisting 25 aa. The handling of protein with this small size was challenging. The most tools were not adapted for small proteins (Wang et al., 2008; Storz et al., 2014). A deletion of NGR_a01725 gave first hints on the function, as the symbiotic plasmid seemed not to be stable in mutant strain. Thus, it seemed that NGR_a01725 was also involved in plasmid stability. But since pNGR234*a* was stable in a NGR234 Δa1725+ mutant and the SNP analysis could not give a hint for sudden stability in multiple generations, the investigation for the function starts again and only hypothesis could be named. First, the transcriptomic data must analyze in more detail and an interaction partner must determine. As a far-fetched hypothesis, NGR_01725 might be involved in toxin anti-toxin systems (TA-system). A second smORF was identified 87 nts upstream of NGR_a01725 that points in the opposite direction. Both smORF could be a part of a toxin-anti-toxin system similar type VI TA systems, in which two proteins were toxin or antitoxin involving in DNA replication (Aakre et al., 2013; Page and Peti, 2016). Thus, A1725 could be the toxin, which deletion has not directly influence to bacteria cell. But the first reaction of lacking the gene resulted in the loss of symbiotic plasmid in few colonies. TA systems are involved in several important cellular processes, including plasmid stability, bacterial persistence or stress response (Gerdes and Maisonneuve, 2012; Page and Peti, 2016; Yamaguchi and Inouye, 2016).

In summary, there is a lot of evidence that these smORF are expressed to functional proteins. The Ribo-seq results will further provide proof of this assumption. The function of these small proteins was not identified directly but there were a lot of evidences that these small proteins were involved in plasmid maintenance and copy number regulation. Until yet, there was

nothing reported that small proteins are involved directly in plasmid maintenance and copy number regulation in *repABC* plasmid.

II. Structure analysis and characterization of GqqA- a Quorum Quenching hydrolase

Until now several bacterial enzymes are known that interfere with quorum sensing. These are quorum quenching enzymes. The best studied class of QQ enzymes is the class of AHL lactonases, which catalyzes the opening of AHL lactone ring (Dong et al., 2000). Enzyme that cleaved the amid bond in-between lactone ring and acyl side chain were categorized to AHL acylases. The most common AHL acylase family is the Ntn-hydrolases with the best characterized enzyme PvdQ (Huang et al., 2003).

In 2016, a protein designated as GqqA was identified in the cellulose overproducer and biofilm-forming bacterium *K. europaeus* CECT 8546 interfering with AHL QS signals from gram-negative bacteria (Valera et al., 2016). The mechanism behind GqqA was not yet known and was to identify.

GqqA quenches bacterial communication by acylase activity

The aa sequence of GqqA was aligned to 22 known and functionally verified QQ enzymes and then visualized with a phylogenetic three (Figure 44). All enzyme classes were highlighted in the different colors red, green and blue. The most prominent enzymes were clustered to AHL lactonases especially in the subclasses of Metallo-β-lactamases with the specific HXHXDH motif (Dong et al., 2000). The second largest representative group is the enzyme family of AHL acylases with the best characterized class of Ntn-hydrolases (Utari et al., 2017). GqqA was clustered close to Ntn-hydrolases, however forming a new subclass of acylases and showed the closed maximum likelihood to amidase enzyme [EC 3.5.1.4] from *Burkholderia lata* acting on carbon-nitrogen bonds (Figure 52, B). The similarity to AHL acylases suggested that GqqA cleaved the amide bond between homoserine lactone-ring and acyl chain (Fetzner, 2014).

For verifying this suggestion, an *in vitro* assay was performed, and the resulted products were measured with ESI-MS/MS. By comparing the *in vitro* enzyme assay with a negative sample not supplemented with GqqA, the mechanism behind GqqA activities was identified. The spectra showed that the enzyme assay released a lactone ring (ESI positive mode, Figure 49) and the corresponding 3-oxo-octanoyl acid (negative mode, Figure 49) by using 3-oxo-C8 homoserine lactone as substrate. This proved the affiliation of GqqA to AHL acylases cleaving the amide bond and releasing lactone ring and a carboxylate (Figure 52, B).

Also, the crystallization with associated data processing of GqqA and especially of Se-GqqA gave evidence for an acylase activity. It seemed that GqqA was post-translationally modified that is typically for Ntn-hydrolases (Fast and Tipton, 2012; Utari et al., 2017). A middle part of enzyme was lacking in calculated protein structure (Figure 45). GqqA consisted of three domains. The α-peptide and the middle domain belonged to N-terminal periplasmatic domain. Both subdomains consisted of the same folding. The β-peptide belonged to PDT typical ACT domain. With lacking the middle domain, the putative substrate domain was freely accessible for substrate. Although the structure itself was not very similar to Ntn- hydrolases, but there were characteristics in the structure that showed similarity to Ntn- hydrolases such as PvdQ from *P. aeruginosa* (Huang et al., 2003). Typical for Ntn-hydrolases is the maturation process of the pro-enzyme, in which the N-terminal end of β-peptide is auto-proteolyzed and separates from α-peptide. This exposes the N-terminal single nucleophilic catalytic residue giving the name of this enzyme class (Yoon et al., 2004; Fast and Tipton, 2012). At first appearance, this N-terminal nucleophilic residue is missing in GqqA structure. But there was also a THR170 that was located at C-terminal of the middle domain. Often the electron density at the ends of peptides is usually rather unclear so that it is possible that Thr170 is part of β-peptide. This Thrβ1 (Thr170) could assume the function of Serβ1 such as in PvdQ. Members of Ntn- hydrolases use the nucleophilic oxygen or sulfur of serine, threonine or cysteine residues for initiating the hydrolysis mechanism. Ntn-Hydrolase is a widespread superfamily that exists in different subfamilies with different structures and different types of enzymes that share the same mechanism (Oinonen and Rouvinen, 2000).

So far, only a prediction of enzyme and substrate was performed using C8-HSL as substrate. Here, three putative residues appeared which were identified as possible interaction sites. But to gain insights into the substrate-binding mechanism of GqqA, a co-crystallization of GqqA with AHL with various side chain length and modifications has to determine.

Five mutated variants of GqqA were created and these mutants were tested for their functionality. The mutant GqqA-Mut2 and GqqA-Mut5 were still able to degrade 3-oxo-C8 HSL whereas the mutant GqqA-Mut1, GqqA-Mut3 and GqqA-Mut4 could not catalyze the reaction and the reporter cells staid purple (Figure 48, B). In GqqA-Mut1 an amino acid exchange was conducted (Phe261Ser). This exchange could lead the loss of hydrophobic interaction between the monomers. But the native gel suggested that GqqA-Mut1 had the same formation as GqqA wt. Thus, Phe261 had no impact to hydrophobic interaction of the monomers and was probably involved in reaction or other functional role which was inhibit due to the substitution Phe262Ser. Also, the mutations of GqqA-Mut3 and GqqA-Mut5

should have the same effect of monomer interactions. On native protein gel and also on denaturing SDS protein gel the expression of GqqA-Mut3 and GqqA-Mut4 could not be observed (Figure 48 and Figure 57, appendix). Furthermore, a second overexpression of GqqA-Mut3 and GqqA-Mut4 did not work. It seems that these residues were important for protein formation and the misfolded protein were degraded. GqqA-Mut2 was modified at the end of aa-sequence. The last residues of a protein have often no important functional role. The native formations showed difference to wt formation but did not influence QQ activity. Possibly an interaction to neighboring molecule could be interrupt, which disturbed the quaternary structure of GqqA but not the single functional unit. In GqqA-Mut5, a threonine was substituted to serine (Thr118Ser). Both amino acids had hydroxylic characteristics and can adopt the functional role. Only a formation change was observed in native protein gel that had no impact of the degradation of AHLs (Figure 48). With further mutagenesis, the residues that were involved in protein formation and in catalytic reaction should be identified in future. A main approach will be to identify more similarities to Ntn-hydrolases (Fast and Tipton, 2012; Fetzner, 2014; Utari et al., 2017) or to PDT enzymes (Tan et al., 2008).

GqqA and the functional role as PDT enzyme

Interestingly, GqqA did not show any structurally similarities to known QQ enzymes but clearly affected QS dependent gene expression such as inhibition of biofilm formation in *K. europaeus* CECT 8546. GqqA was also predicted to belong to the PDT enzymes having all characteristic domains (Valera et al., 2016). Prephenate dehydratase [EC:4.2.1.51] catalyzes prephenate to phenylpyruvate through dehydration and decarboxylation reaction (Figure 52, A) and are involved in biosynthesis of L-phe in organism that use shikimate pathway (Bentley, 1990). Also, the cofactor structure-based function prediction tool indicated the structural similarity to PDT enzymes (Table 18). High structural similarities to five described PDT enzymes were identified, while sequence similarities were not observed with high similarity. PDT enzymes often form tetramers consisting of two dimers. These dimers have two asymmetric PDT monomers (Tan et al., 2008). A prediction of the GqqA structure showed also the asymmetric accumulation of two monomers to a dimer that forms a tetramer (dimer of an asymmetric dimer) that could have four catalytic sites and could bind four L-Phe as allosteric regulator as it is described for PDT enzymes (Tan et al., 2008). These L-Phe binding site were also seen in crystal structure. With a prediction of substrate binding site, only one substrate pocket was observed per dimer and differs from PDT enzymes, where each monomer consists of a substrate pocket (Tan et al., 2008). In bacteria using shikimate pathway PDT enzymes were important to synthesize the essential amino acids phenylalanine

and tyrosine. Also, in *K. europaeus* CECT8546 GqqA was predicted to be involved in phenylalanine synthesis. Previous studies implied that GqqA was not able to complement *E. coli* PDT deleted mutant (Valera et al., 2016).

This analysis indicated that GqqA did not act as PDT enzyme and had other function in *K. europaeus* CECT8546. Thus, further work will have addressed this issue and clarify the native function in *K. europaeus* CECT8564.

Thereby, it is very well possible that GqqA acts as a bifunctional or even multifunctional and promiscuous enzyme that relies on different substrates in different environmental conditions.

Another possibility in natural role of GqqA is that GqqA acts as a bi- or multifunctional enzyme that changes the substrate in different environmental conditions. One suggestion is that more PDT enzymes could interfere with QS signals.

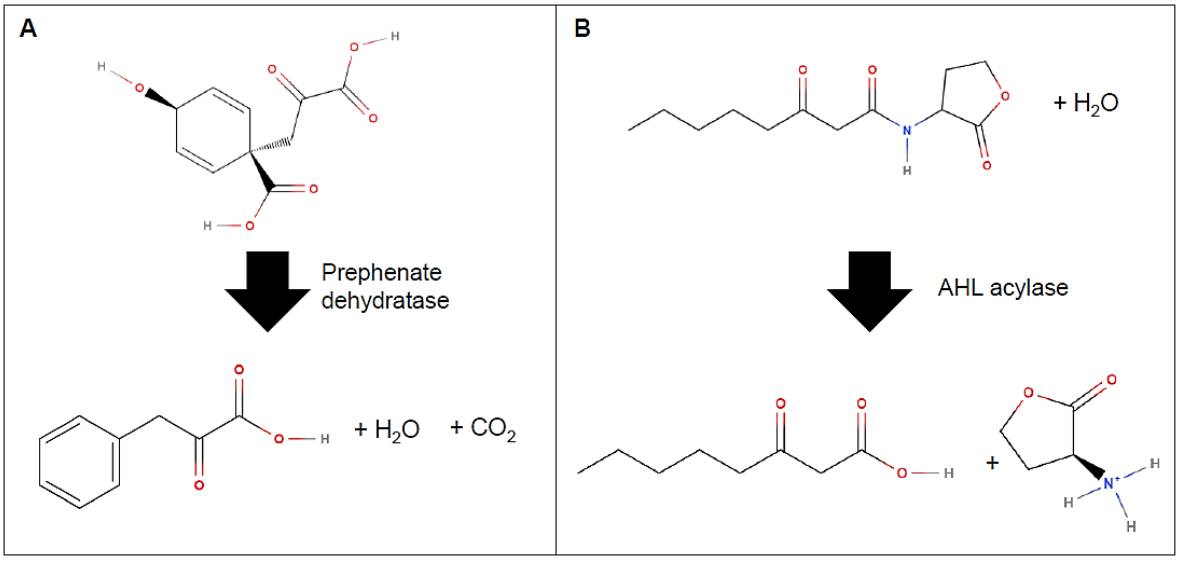

Figure 52: Prephenate dehydratase vs. AHL acylase
A: PDT dehydratase catalyzed the chemical reaction of prephenate to phenylpyruvic acid releasing H_20 and CO_2 as co-products. PDT enzymes [EC 4.2.1.51] were involved in phenylalanine, tyrosine and tryptophan biosynthesis. B: AHL acylases [EC 3.5.1.97] catalyzed the cleavage of amid bond of acyl homoserine lactone in presence of H2O. The lactone ring and a carboxylic acid were released.

In summary, the work above was identified a completely new type of QQ enzyme, that has previously not been observed. Thereby the structural data together with the analytical data clearly support that hypothesis that GqqA is a hydrolase acting on the amide bond as outline in Figure 52, B.

6. Outlook

The bacterial communication system QS allows bacteria to synchronize collective behavior in cell density-dependent manner. Only slowly we understand the complexity of these processes on population wide manner and on the level of single cells. Therefore, in this thesis two different projects were followed and that need to be continued as outlined below.

I. Quorum sensing dependent expression of small proteins

The focus of small protein was started 10 years ago and is still in the beginning. All applications that are used to identify protein functions must be adapted for small protein investigation and cannot be applied one to one. In the past it became clear, that the most genomes encode for a much higher number of genes than previously assumed. The plant symbiont NGR234 encodes for additional 251 smORFs. The global expression of these smORFs to proteins must be verified with Ribo-seq. The results of Ribo-seq will be available soon and will then have to be evaluated.

Further there were many smORFs that showed a different transcription levels in the newly constructed AI double deletion mutant that should analyzed in detailed. The detection of putative newly TRA boxes upstream of the smORFs will be one approach.

The investigation of *repXYA0* operon must be continued by detecting the NGR234 cell lysate with specific anti-RepY-antibodies. With this analysis, the same protein band probably appears at approx. 30 kDa similar to observation with anti-RepX and anti-RepA0 antibodies. For functional analysis, the role of RepXYA0 on plasmid replication must be determined. Therefore, a reporter strain should be constructed so that an effect can be seen by adding RepX, RepY or RepA0 to the reporter strain.

II. Structure analysis and characterization of GqqA- a Quorum Quenching hydrolase

In general, the native function of QQ enzymes is not known until now. The identification of the natural substrate can give hints for the origin of QS degrading enzymes. One of the main questions is the elucidation of the natural substrate of GqqA in *K. europaeus* CECT8546. GqqA shows high similarity to known PDT enzymes but is functionally different. Probably, GqqA act as PDT enzyme or the genome of *K. europaeus* encodes for another PDT enzyme, which is unknown so far and should be identified.

Another important investigation is the elucidation of the mechanism behind AI degradation of GqqA. The fact that GqqA has an acylase activity gives an important indication of the mechanism behind it. For characterization of enzymatic mechanism, the next step is to

mutate single amino acids that could be involved in enzyme activity. A co-crystallization of enzyme with substrate helps to identify residues that are involved in substrate binding. The first residues that shall be mutated are Thr170 and Thr174. One of these residues could be the catalytic nucleophilic residue. All residues that are involved in substrate binding can give hints for substrate specificity and the nature roll of GqqA. Thus, all substrate interacting residues should be identified.

References

Aakre CD, Phung NN, Huand D, Laub MT (2013) A bacterial toxin inhibits DNA replication elongation a direct interaction with the β sliding clamp. Molecular cell **52**: 617–628

Alix E, Blanc-Potard AB (2008) Peptide-assisted degradation of the Salmonella MgtC virulence factor. EMBO Journal **27**: 546–557

Allen EK, Allen ON (1950) Biochemical and Symbiotic Properties Of The Rhizobia. Bacteriological Reviews **14**: 273–330

Anders S, Pyl PT, Huber W (2014) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics **31**: 166–169

Bachmann BJ (1983) Linkage map of *Escherichia coli* K-12, edition 7. Microbiological Reviews **47**: 180–230

Barber CE, Tang JL, Feng JX, Pan MQ (1997) A novel regulatory system required for pathogenicity of Xanthomonas campestris is mediated by a small diffusible signal molecule. **24**: 555–566

Barbulova A, Chiurazzi M (2005) A procedure for in vitro nodulation studies. *In* AJ Márquez, ed, Lotus japonicus Handbook. Springer Netherlands, Dordrecht, pp 83–86

Barnett MJ, Long SR (2015) The sinorhizobium meliloti SyrM Regulon: Effects on global gene expression are mediated by syrA and nodD3. Journal of Bacteriology **197**: 1792–1806

Belin P, Moutiez M, Lautru S, Seguin J, Pernodet J-L, Gondry M (2012) The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. Natural product reports **29**: 961–979

Bentley R (1990) The shikimate pathway--a metabolic tree with many branches. Critical reviews in biochemistry and molecular biology **25**: 307–384

Bertani G (1951) Studies on Lysogenesis I. The mode of phageliberation by lysogenic *Eschericia coli* . Journal of Bacteriology **62**: 293–300

Bijtenhoorn P (2012) Characterization of novel metagenomic quorum quenching enzymes attenuating Pseudomonas aeruginosa motility , biofilm formation and virulence.

Bijtenhoorn P, Schipper C, Hornung C, Quitschau M, Grond S, Weiland N, Streit WR (2011) BpiB05, a novel metagenome-derived hydrolase acting on N-acylhomoserine lactones. Journal of Biotechnology **155**: 86–94

Boekhorst J, Wilson G, Siezen RJ (2011) Searching in microbial genomes for encoded small proteins. Microbial Biotechnology **4**: 308–313

Byers JT, Lucas C, Salmond GPC, Welch M (2002) Nonenzymatic turnover of an Erwinia carotovora quorum-sensing signaling molecule. Journal of bacteriology **184**: 1163–1171

Cao JG, Meighen EA (1989) Purification and structural identification of an autoinducer for the luminescence system of Vibrio harveyi. Journal of Biological Chemistry **264**: 21670–21676

de Carvalho MP, Abraham W-R (2012) Antimicrobial and biofilm inhibiting diketopiperazines.

Current medicinal chemistry **19**: 3564–3577

Carver T, Berriman M, Tivey A, Patel C, Bohme U, Barrell BG, Parkhill J, Rajandream M-A (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics (Oxford, England) **24**: 2672–2676

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics **28**: 464–469

Cassidy L, Prasse D, Linke D, Schmitz RA, Tholey A (2016) Combination of Bottom-up 2D-LC-MS and Semi-top-down GelFree-LC-MS Enhances Coverage of Proteome and Low Molecular Weight Short Open Reading Frame Encoded Peptides of the Archaeon Methanosarcina mazei. Journal of proteome research **15**: 3773–3783

Chan K-G, Atkinson S, Mathee K, Sam C-K, Chhabra SR, Cámara M, Koh C-L, Williams P (2011) Characterization of N-acylhomoserine lactone-degrading bacteria associated with the Zingiber officinale (ginger) rhizosphere: Co-existence of quorum quenching and quorum sensing in Acinetobacter and Burkholderia. BMC Microbiology **11**: 51

Chen X, Schauder S, Potier N, Van Dorsselaer A, Pelczer I, Bassler BL, Hughson FM (2002) Structural identification of a bacterial quorum-sensing signal containing boron. Nature **415**: 545–549

Cheng F, Ma A (2017) ORIGINAL RESEARCH N - - regulation of genes mediating motility and pathogenicity in Pseudomonas syringae pathovar tabaci 11528. 1–14

Choi E, Lee K-Y, Shin D (2012) The MgtR regulatory peptide negatively controls expression of the MgtA Mg2+ transporter in Salmonella enterica serovar Typhimurium. Biochemical and biophysical research communications **417**: 318–323

Chowdhary PK, Keshavan N, Nguyen HQ, Peterson JA, Gonzalez JE, Haines DC (2007) Bacillus megaterium CYP102A1 oxidation of acyl homoserine lactones and acyl homoserines. Biochemistry **46**: 14429–14437

Cutting S, Anderson M, Lysenko E, Page A, Tomoyasu T, Tatematsu K, Tatsuta T, Kroos L, Ogura T (1997) SpoVM, a small protein essential to development in Bacillus subtilis, interacts with the ATP-dependent protease FtsH. Journal of Bacteriology **179**: 5534–5542

D'Haeze W, Holsters M (2002) Nod factor structures, responses, and perception during initiation of nodule development. Glycobiology **12**: 79R-105R

Darch SE, West SA, Winzer K, Diggle SP (2012) Density-dependent fitness benefits in quorum-sensing bacterial populations. Proceedings of the National Academy of Sciences **109**: 8259–8263

Davies DG, Parsek MR, Pearson JP, Iglewski BH, Costerton JW, Greenberg EP (1998) Davies DG, Parsek MR, Pearson JP, Iglewski BH, Costerton JW, Greenberg EP. 19The involvement of cell-to-cell signals in the development of a bacterial biofilm. **280**: 295–299

Denarie J, Debelle F, Rosenberg C (1992) Signaling and host range variation in nodulation. Annual review of microbiology **46**: 497–531

Dong Y-H, Xu J-L, Li X-Z, Zhang L-H (2000) AiiA, an enzyme that inactivates the acylhomoserine lactone quorum-sensing signal and attenuates the virulence of

Erwinia carotovora Proceedings of the National Academy of Sciences **97**: 3526 LP – 3531

Engebrecht J, Silverman M (1984) Identification of genes and gene products necessary for bacterial bioluminescence. Proceedings of the National Academy of Sciences **81**: 4154–4158

Ermolaeva MD (2001) Prediction of operons in microbial genomes. Nucleic Acids Research **29**: 1216–1221

Esseling JJ, Lhuissier FGP, Emons AMC (2003) Nod Factor-Induced Root Hair Curling : Continuous Polar Growth towards the Point of Nod Factor Application 1. **132**: 1982–1988

Fast W, Tipton PA (2012) The enzymes of bacterial census and censorship. Trends in Biochemical Sciences **37**: 7–14

Fetzner S (2014) Quorum quenching enzymes. Journal of Biotechnology. doi: 10.1016/j.jbiotec.2014.09.001

Firmin JL, Wilson KE, Rossen L, Johnston AWB (1986) Flavonoid activation of nodulation genes in Rhizobium reversed by other compounds present in plants. Nature **324**: 90–92

Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X (1997) Molecular basis of symbiosis between Rhizobium and legumes.

Fuqua C, Greenberg EP (1998) Self perception in bacteria: quorum sensing with acylated homoserine lactones. Current opinion in microbiology **1**: 183–189

Fuqua C, Winans SC (1996) Conserved cis-acting promoter elements are required for density-dependent transcription of *Agrobacterium tumefaciens* conjugal transfer genes. Journal of Bacteriology **178**: 435–440

Gannoun-Zaki L, Alibaud L, Carrère-Kremer S, Kremer L, Blanc-Potard AB (2013) Overexpression of the KdpF Membrane Peptide in Mycobacterium bovis BCG Results in Reduced Intramacrophage Growth and Altered Cording Morphology. PLoS ONE. doi: 10.1371/journal.pone.0060379

Gannoun-Zaki L, Belon C, Dupont C, Hilbert F, Kremer L, Blanc-Potard AB (2014) Overexpression of the Salmonella KdpF membrane peptide modulates expression of kdp genes and intramacrophage growth. FEMS Microbiology Letters **359**: 34–41

Gargioli C, Muraglia E, Cortese R, Francesco R De, Neddermann P, Marco S Di (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. EMBO Journal **21**: 4393–4401

Gaßel M, Moellenkamp T, Puppe W, Altendorf K (1999) The KdpF subunit is part of the K+-translocating Kdp complex of Escherichia coli and is responsible for stabilization of the complex in vitro. Journal of Biological Chemistry **274**: 37901–37907

Gerdes K, Bech FW, Jorgensen ST, Lobner-olesen A, Rasmussen PB, Atlung T, Boe L, Karlstrom O, Molin S, Von K (1986) Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid Rl and its homology with the relF gene product of the E. coli relB operon. The EMBO Journal **5**: 2023–2029

Gerdes K, Jacobsen JS, Franch T (1997) Plasmid Stabilization by Post-Segregational Killing.

Genetic Engineering **19**: 49–50

Gerdes K, Maisonneuve E (2012) Bacterial persistence and toxin-antitoxin loci. Annual review of microbiology **66**: 103–123

Geurts R, Bisseling T (2002) Rhizobium Nod Factor Perception and Signalling. 239–249

Gonza JE, Marketon MM (2003) Quorum Sensing in Nitrogen-Fixing Rhizobia. **67**: 574–592

Grandclément C, Tannières M, Moréra S, Dessaux Y, Faure D (2015) Quorum quenching: Role in nature and applied developments. FEMS Microbiology Reviews **40**: 86–116

Grote J, Krysciak D, Petersen K, Güllert S, Schmeisser C, Förstner KU, Krishnan HB, Schwalbe H, Kubatova N, Streit WR (2016) The Absence of the N-acyl-homoserine-lactone Autoinducer Synthase Genes traI and ngrI Increases the Copy Number of the Symbiotic Plasmid in Sinorhizobium fredii NGR234. Frontiers in Microbiology. doi: 10.3389/fmicb.2016.01858

Grote J, Krysciak D, Schorn A, Dahlke RI, Soonvald L, Müller J, Hense BA, Schwarzfischer M, Sauter M, Schmeisser C, et al (2014) Evidence of autoinducer-dependent and -independent heterogeneous gene expression in Sinorhizobium fredii NGR234. Applied and Environmental Microbiology. doi: 10.1128/AEM.01689-14

Gurich N, González JE, Gurich N, Gonza JE (2009) Role of Quorum Sensing in Sinorhizobium meliloti-Alfalfa Symbiosis Role of Quorum Sensing in Sinorhizobium meliloti -Alfalfa Symbiosis □ †. doi: 10.1128/JB.00376-09

Hanahan D, Harbor CS (1983) Studies on Transformation of Escherichia coli with Plasmids Department of Biochemistry and Molecular Biology. **0**: 557–580

Hartmann A, Rothballer M, Hense BA, Schröder P (2014) Bacterial quorum sensing compounds are important modulators of microbe-plant interactions. Frontiers in Plant Science **5**: 131

Havey JC, Vecchiarelli AG, Funnell BE (2012) ATP-regulated interactions between P1 ParA, ParB and non-specific DNA that are stabilized by the plasmid partition site, parS. Nucleic Acids Research **40**: 801–812

He X, Chang W, Pierce DL, Seib LO, Wagner J, Fuqua C (2003) Quorum Sensing in Rhizobium sp . Strain NGR234 Regulates Conjugal Transfer (tra) Gene Expression and Influences Growth Rate. **185**: 809–822

Hemm MR, Paul BJ, Schneider TD, Storz G, Rudd KE (2008) Small membrane proteins found by comparative genomics and ribosome binding site models. Molecular Microbiology **70**: 1487–1501

Higashii S (1967) Transfer of clover infectivity of Rhizobium trifolii to Rhizobium phaseoli as mediated by an episomic factor. J Gen Appl Microbio **13**: 391–403

Hobbs EC, Fontaine F, Yin X, Storz G (2011) An expanding universe of small proteins. Current Opinion in Microbiology **14**: 167–173

Hobbs EC, Yin X, Paul BJ, Astarita JL, Storz G (2012) Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. Proceedings of the National Academy of Sciences **109**: 16696–16701

Hodgman TC (2000) A historical perspective on gene/protein functional assignment. Bioinformatics **16**: 10–15

Huang JJ, Han J-I, Zhang L-H, Leadbetter JR (2003) Utilization of Acyl-Homoserine Lactone Quorum Signals for Growth by a Soil Pseudomonad and Pseudomonas aeruginosa PAO1. Applied and Environmental Microbiology **69**: 5941 LP – 5949

Huff J (2015) The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. Nature Methods **12**: i–ii

Hwang I, Li PL, Zhang L, Piper KR, Cook DM, Tate ME, Farrand SK (1994) TraI, a LuxI homologue, is responsible for production of conjugation factor, the Ti plasmid N-acylhomoserine lactone autoinducer. Proceedings of the National Academy of Sciences of the United States of America **91**: 4639–4643

Iida A, Ohnishi Y, Horinouchi S (2008) Control of acetic acid fermentation by quorum sensing via N-acylhomoserine lactones in Gluconacetobacter intermedius. Journal of bacteriology **190**: 2546–2555

Iida A, Ohnishi Y, Horinouchi S (2009) Identification and characterization of target genes of the GinI/GinR quorum-sensing system in Gluconacetobacter intermedius. Microbiology (Reading, England) **155**: 3021–3032

Ingolia NT (2016) Ribosome Footprint Profiling of Translation throughout the Genome. Cell **165**: 22–33

Irmer S, Podzun N, Langel D, Heidemann F, Kaltenegger E, Schemmerling B, Geilfus C-M, Zörb C, Ober D (2015) New aspect of plant–rhizobia interaction: Alkaloid biosynthesis in *Crotalaria* depends on nodulation. Proceedings of the National Academy of Sciences **112**: 4164–4169

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology **3**: 318–356

Jeffrey PD, Chen G, Chen L, Fuqua C, Shi Y (2007) Structural basis for antiactivation in bacterial quorum sensing. Proceedings of the National Academy of Sciences **104**: 16474–16479

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Henrik HN, Alves BJR, Morrison MJ (2012) Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. Agronomy for Sustainable Development. doi: 10.1007/s13593-011-0056-7

Jesu A, Giraldo R, Di N (1998) Replication and Control of Circular Bacterial Plasmids. **62**: 434–464

Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: The Sinorhizobium - Medicago model. Nature Reviews Microbiology **5**: 619–633

Kabsch W (2010) XDS. Acta Crystallographica Section D **66**: 125–132

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols **10**: 845

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Research **32**: 526–531

Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, Peterson KM (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS , carrying different antibiotic-resistance cassettes. **166**: 175–176

Krysciak D, Grote J, Orbegoso MR, Utpatel C, Foerstner KU, Li L, Schmeisser C, Krishnan HB, Streit WR (2014) RNA sequencing analysis of the broad-host-range strain sinorhizobium fredii NGR234 Identifies a large set of genes linked to quorum sensing-dependent regulation in the background of a trai and ngri deletion mutant. Applied and Environmental Microbiology **80**: 5655–5671

Krysciak D, Orbegoso MR, Schmeisser C, Streit WR (2015) Molecular Keys to Broad Host Range in Sinorhizobium fredii NGR234, USDA257, and HH103. Biological Nitrogen Fixation **1**–**2**: 325–336

Krysciak D, Schmeisser C, Preuß S, Riethausen J, Quitschau M, Grond S, Streit WR (2011) Involvement of Multiple Loci in Quorum Quenching of Autoinducer I Molecules in the Nitrogen-Fixing Symbiont Rhizobium (Sinorhizobium). **77**: 5089–5099

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: An information aesthetic for comparative genomics. Genome Research **19**: 1639–1645

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution **35**: 1547–1549

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

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods **9**: 357–359

Lassak J, Henche A, Binnenkade L, Thormann KM (2010) ArcS , the Cognate Sensor Kinase in an Atypical Arc System of Shewanella oneidensis MR-1 □ †. **76**: 3263–3274

Levin PA, Fan N, Ricca E, Driks A, Losick R, Cutting S (1993) An unusually small gene required for sporulation by Bacillus subtilis. Molecular Microbiology **9**: 761–771

Li PL, Everhart DM, Farrand SK (1998) Genetic and sequence analysis of the pTiC58 trb locus, encoding a mating-pair formation system related to members of the type IV secretion family. Journal of bacteriology **180**: 6164–6172

Li Y-H, Tian X (2012) Quorum Sensing and Bacterial Social Interactions in Biofilms. Sensors (Basel, Switzerland) **12**: 2519–2538

Lin Y-H, Xu J-L, Hu J, Wang L-H, Ong SL, Leadbetter JR, Zhang L-H (2003) Acyl-homoserine lactone acylase from Ralstonia strain XJ12B represents a novel and potent class of quorum-quenching enzymes. Molecular microbiology **47**: 849–860

Lopez-lara IM, Berg JDJ Van Den, Thomas-oates JE, Glushka J, Ben JJ, Spaink HP (1995) Structural identification of the iipo-chitin oligosaccharide nodulation signals of Rhizobium loti. 15:

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology **15**: 550

Malys N (2012) Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes. Molecular biology reports **39**: 33–39

Marie C, Deakin WJ, Ojanen-Reuhs T, Diallo E, Reuhs B, Broughton WJ, Perret X (2004) TtsI, a key regulator of Rhizobium species NGR234 is required for type III-dependent protein secretion and synthesis of rhamnose-rich polysaccharides. Molecular plant-microbe interactions : MPMI **17**: 958–966

Mazur A, Koper P (2012) Rhizobial plasmids - replication, structure and biological role. Central European Journal of Biology **7**: 571–586

McAnulla C, Edwards A, Sanchez-Contreras M, Sawers RG, Downie JA (2007) Quorum-sensing-regulated transcriptional initiation of plasmid transfer and replication genes in Rhizobium leguminosarum biovar viciae. Microbiology **153**: 2074–2082

McClean KH, Winson MK, Fish L, Taylor A, Chhabra SR, Camara M, Daykin M, Lamb JH, Swift S, Bycroft BW, et al (1997) Quorum sensing and Chromobacterium violaceum: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. Microbiology (Reading, England) **143 (Pt 1**: 3703–3711

Mierendorf RC, Morris BB, Hammer B, Novy RE (1998) Expression and Purification of Recombinant Proteins Using the pET System. Methods in molecular medicine **13**: 257–292

Modell JW, Hopkins AC, Laub MT (2011) A DNA damage checkpoint in Caulobacter crescentus inhibits cell division through a direct interaction with FtsW.[Erratum appears in Genes Dev. 2011 Aug 1;25(15):1662]. Genes Dev **25**: 1328–1343

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics **92**: 255–264

Morrison NA, Hau CY, Trinick MJ, Shine J, Rolfe BG (1983) Heat Curing of a Sym Plasmid in a Fast-Growing Rhizobium sp. That Is Able to Nodulate Legumes and the Nonlegume Parasponia sp. Journal of Bacteriology **153**: 527–531

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

Nealson KH (1977) Autoinduction of bacterial luciferase. Occurrence, mechanism and significance. Archives of microbiology **112**: 73–79

Nealson KH, Hastings JW (1979) Bacterial bioluminescence: its control and ecological significance. Microbiological reviews **43**: 496–518

Nelson MS, Sadowsky MJ (2015) Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. Frontiers in Plant Science **6**: 1–11

Ng W-L, Bassler BL (2009) Bacterial Quorum-Sensing Network Architectures. Annual Review of Genetics **43**: 197–222

Notredame C, Higgins DG, Heringa J (2000) T-coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology **302**: 205–217

Oinonen C, Rouvinen J (2000) Structural comparison of Ntn-hydrolases. Protein science : a publication of the Protein Society **9**: 2329–2337

Page R, Peti W (2016) Toxin-antitoxin systems in bacterial growth arrest and persistence. Nature Chemical Biology **12**: 208–214

Papenfort K, Bassler BL (2016) Quorum sensing signal-response systems in Gram-negative bacteria. Nature reviews Microbiology. doi: 10.1038/nrmicro.2016.89

Papenfort K, Said N, Welsink T, Lucchini S, Hinton JCD, Vogel J (2009) Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Molecular Microbiology **74**: 139–158

Parniske M, Downie JA (2003) Plant biology Locks , keys and symbioses. **425**: 3–4

Perret X, Broughton WJ (1999) Monitoring symbiotic gene expression in *Rhizobium sp .* NGR234. Plant-Microbe Interactions

Perret X, Freiberg C, Rosenthal a, Broughton WJ, Fellay R (1999) High-resolution transcriptional analysis of the symbiotic plasmid of Rhizobium sp. NGR234. Molecular microbiology **32**: 415–25

Perret X, Staehelin C, Broughton WJ (2003) Molecular Basis of Symbiotic Promiscuity. Microbiology and Molecular Biology Reviews **64**: 180–201

Pesci EC, Milbank JBJ, Pearson JP, McKnight S, Kende AS, Greenberg EP, Iglewski BH (1999) Quinolone signaling in the cell-to-cell communication system of Pseudomonas aeruginosa Proceedings of the National Academy of Sciences **96**: 11229 LP – 11234

Peter J, Young W, Haukka KE (1996) Diversity and phylogeny of rhizobia. New Phytologist **133**: 87–94

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry **25**: 1605–1612

Piper KR, Beck von Bodman S, Farrand SK (1993) Conjugation factor of Agrobacterium tumefaciens regulates Ti plasmid transfer by autoinduction. Nature **362**: 448–450

Pueppke SG, Broughton WJ (1999) *Rhizobium* sp. Strain NGR234 and *R. fredii* USDA257 Share Exceptionally Broad, Nested Host Ranges. Molecular Plant-Microbe Interactions **12**: 293–318

Ramamurthi KS, Storz G (2014) The small protein floodgates are opening; now the functional analysis begins. BMC Biology. doi: 10.1186/s12915-014-0096-y

Rasmussen PB, Gerdes K, Molin S (1987) Genetic analysis of the parB+ locus of plasmid R1. MGG Molecular & General Genetics **209**: 122–128

Ren D, Bedzyk LA, Ye RW, Thomas SM, Wood TK (2004) Stationary-Phase Quorum-Sensing Signals Affect Autoinducer-2 and Gene Expression in Escherichia coli. Applied and Environmental Microbiology **70**: 2038–2043

Rosas Olvera M, Vivès E, Molle V, Blanc-Potard A-B, Gannoun-Zaki L (2017) Endogenous

and Exogenous KdpF Peptide Increases Susceptibility of Mycobacterium bovis BCG to Nitrosative Stress and Reduces Intramacrophage Replication . Frontiers in Cellular and Infection Microbiology **7**: 115

Sambrook JF, Russell DW (2001) Molecular Cloning: A Laboratory Manual, 3rd ed., Vols 1,2 and 3, 3rd ed. Cold Spring Harbor Laboratory Press, New York, NY

Schipper C, Hornung C, Bijtenhoorn P, Quitschau M, Grond S, Streit WR (2009) Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in pseudomonas aefuginosa. Applied and Environmental Microbiology **75**: 224–233

Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quéré A, Wollherr A, Heinemeyer I, Morgenstern B, Pommerening-Röser A, Flores M, et al (2009) Rhizobium sp. strain NGR234 possesses a remarkable number of secretion systems. Applied and Environmental Microbiology **75**: 4035–4045

Sessitsch A, Howieson JG, Perret X, Antoun H, Martínez-Romero E (2002) Advances in Rhizobium research. Critical Reviews in Plant Sciences **21**: 323–378

Shaw PD, Ping G, Daly SL, Cha C, Cronan JEJ, Rinehart KL, Farrand SK (1997) Detecting and characterizing N-acyl-homoserine lactone signal molecules by thin-layer chromatography. Proceedings of the National Academy of Sciences of the United States of America **94**: 6036–6041

Shell SS, Wang J, Lapierre P, Mir M, Chase MR, Pyle MM, Gawande R, Ahmad R, Sarracino DA, Ioerger TR, et al (2015) Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. PLoS Genetics **11**: 1–31

Sieber P, Platzer M, Schuster S (2018) The Definition of Open Reading Frame Revisited. Trends in Genetics **34**: 167–170

Simon R, Priefer U, Pühler A (1983) A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. Nature Publishing Group

Solovyev V, Salamov A (2010) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. Nova Science Publisher

Storz G, Wolf YI, Ramamurthi KS (2014) Small Proteins Can No Longer Be Ignored. Annual Review of Biochemistry **83**: 753–777

Streit WR, Schmitz R a, Perret X, Staehelin C, Deakin WJ, Raasch C, Liesegang H, Broughton WJ (2004) An Evolutionary Hot Spot: the pNGR234 b Replicon of Rhizobium sp. Strain NGR234. Journal of Bacteriology **186**: 535–542

Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of molecular biology **189**: 113–130

Swiderska A, Berndtson AK, Cha MR, Li L, Beaudoin GM 3rd, Zhu J, Fuqua C (2001) Inhibition of the Agrobacterium tumefaciens TraR quorum-sensing regulator. Interactions with the TraM anti-activator. The Journal of biological chemistry **276**: 49449–49458

Tan K, Li H, Zhang R, Gu M, Clancy ST, Joachimiak A (2008) Structures of open (R) and

close (T) states of prephenate dehydratase (PDT)-Implication of allosteric regulation by l-phenylalanine. Journal of Structural Biology **162**: 94–107

Tempe J, Petit A, Holsters M, Montagu M, Schell J (1977) Thermosensitive step associated with transfer of the Ti plasmid during conjugation: Possible relation to transformation in crown gall. Proceedings of the National Academy of Sciences of the United States of America **74**: 2848–2849

Temprano-Vera F, Rodríguez-Navarro DN, Acosta-Jurado S, Perret X, Fossou RK, Navarro-Gómez P, Zhen T, Yu D, An Q, Buendía-Clavería AM, et al (2018) Sinorhizobium frediiStrains HH103 and NGR234 form nitrogen fixing nodules with diverse wild soybeans (Glycine soja) from Central China but Are Ineffective on Northern China Accessions. Frontiers in Microbiology. doi: 10.3389/fmicb.2018.02843

Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences of the United States of America **76**: 4350–4354

Trinick MJ (1980) Relationships Amongst the Fast- growing Rhizobia of Lablab purpureus, Leucaena leucocephala, Mimosa spp., Acacia farnesiana and Sesbania grandiflora and their Affinities with Other Rhizobial Groups. Journal of Applied Bacteriology **49**: 39–53

Turanov AA, Lobanov A V, Fomenko DE, Morrison HG, Sogin ML, Klobutcher LA, Hatfield DL, Gladyshev VN (2009) Genetic code supports targeted insertion of two amino acids by one codon. Science (New York, NY) **323**: 259–261

Uroz S, Chhabra SR, Cámara M, Williams P, Oger P, Dessaux Y (2005) N-acylhomoserine lactone quorum-sensing molecules are modified and degraded by Rhodococcus erythropolis W2 by both amidolytic and novel oxidoreductase activities. Microbiology **151**: 3313–3322

Utari PD, Vogel J, Quax WJ (2017) Deciphering physiological functions of AHL quorum quenching acylases. Frontiers in Microbiology **8**: 1–13

Valera MJ, Mas A, Streit WR, Mateo E (2016) GqqA, a novel protein in Komagataeibacter europaeus involved in bacterial quorum quenching and cellulose formation. Microbial Cell Factories **15**: 88

Valera MJ, Poehlein A, Torija MJ, Haack FS, Daniel R, Streit WR, Mateo E, Mas A (2015) Draft Genome Sequence of Komagataeibacter europaeus CECT 8546, a Cellulose-Producing Strain of Vinegar Elaborated by the Traditional Method. Genome announcements **3**: e01231-15

Viprey V, Rosenthal A, Broughton WJ, Perret X (2000) Genetic snapshots of the Rhizobium species NGR234 genome. Genome biology **1**: RESEARCH0014

Wang F, Xiao J, Pan L, Yang M, Zhang G, Jin S, Yu J (2008) A systematic survey of mini-proteins in bacteria and archaea. PLoS ONE. doi: 10.1371/journal.pone.0004027

Wang H, Avican K, Fahlgren A, Erttmann SF, Nuss AM, Dersch P, Fallman M, Edgren T, Wolf-Watz H (2016) Increased plasmid copy number is essential for Yersinia T3SS function and virulence. Science. doi: 10.1126/science.aaf7501

Wang H, Yin X, Wu Orr M, Dambach M, Curtis R, Storz G (2017) Increasing intracellular

magnesium levels with the 31-amino acid MgtS protein. Proceedings of the National Academy of Sciences **114**: 5689 LP – 5694

Wetzel ME, Olsen GJ, Chakravartty V, Farrand SK (2015) The repABC plasmids with quorum-regulated transfer systems in members of the rhizobiales divide into two structurally and separately evolving groups. Genome Biology and Evolution **7**: 3337–3357

Williams P, Cámara M (2009) Quorum sensing and environmental adaptation in Pseudomonas aeruginosa: a tale of regulatory networks and multifunctional signal molecules. Current Opinion in Microbiology **12**: 182–191

Winkler WC (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Current Opinion in Chemical Biology **9**: 594–602

Xavier KB, Bassler BL (2003) LuxS quorum sensing: more than just a numbers game. Current opinion in microbiology **6**: 191–197

Yamaguchi Y, Inouye M (2016) Toxin-Antitoxin Systems in Bacteria and Archaea. Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria **1**: 97–107

Yin X, Wu Orr M, Wang H, Hobbs EC, Shabalina SA, Storz G (2019) The small protein MgtS and small RNA MgrR modulate the PitA phosphate symporter to boost intracellular magnesium levels. Molecular Microbiology. doi: 10.1111/mmi.14143

Yoon J, Oh B, Kim K, Park J, Han D, Kim KK, Cha SS, Lee D, Kim Y (2004) A Bound Water Molecule Is Crucial in Initiating Autocatalytic Precursor Activation in an N-terminal Hydrolase. Journal of Biological Chemistry **279**: 341–347

Zhang C, Freddolino PL, Zhang Y (2017) COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic acids research **45**: W291–W299

Zhu J, Miller MB, Vance RE, Dziejman M, Bassler BL, Mekalanos JJ (2002) Quorum-sensing regulators control virulence gene expression in Vibrio cholerae. Proceedings of the National Academy of Sciences **99**: 3129–3134

Zhu J, Winans SC (1998) Activity of the quorum-sensing regulator TraR of Agrobacterium tumefaciens is inhibited by a truncated, dominant defective TraR-like protein. Molecular Microbiology **27**: 289–297

Abbreviations

3-oxo-C12-HSL *3-oxo-dodecanoyl homoserine lactone*
3-oxo-C8-HSL *3-oxo-octanoyl homoserine lactone*

A. Agrobacterium
aa *amino acid*
APS *ammonium persulfate*
AHL *N-acyl-homoserine lactone*
AI *autoinducer*
AP *alkaline phosphatase*
AT *Agrobacterium tumefaciens*

$_{bidest}$ *bi-distilled*
BLAST *Basic Local Alignment Search*

C. Chromobacterium
CV026 *Chromobacterium violaceum*

Da *Dalton*
DAPI *4',6-diamidino-2-phenylindole*
DKP *Diketopiperzine*
DMSO *Dimethyl sulfoxide*
DNA *Deoxyribonucleic acid*
dpi *days past infection*
DSF *diffusible signal factor*

E. Escherichia
EPPS *4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid*
ESI *Electrospray ionization*
EtOH *ethanol*

FU *fluorescence unit*

x g *gravitational force equivalent*
gDNA *genomic DNA*
gen^{R} *gentamycin resistance*

HH103 *Sinorhizobium fredii HH103*
HRP *horse radish peroxidase*
HSL *homoserine lactone*

IPTG *Isopropyl ß-d-1-thiogalactopyranoside*
IS *ion spray voltage*

LB *lysogeny broth*
LC-MS/MS *Liquid chromatography–mass spectrometry*
LCO *lipo-chitin oligosaccharides*

Mbp *mega base pair*
MBP *maltose binding protein*
MEGA *Molecular Evolutionary Genetics Analysis*

NCBI *National Center for Biotechnology Information*
NFR *nod factor receptor*
NGR234 *Sinorhizobium fredii NGR234*
Nod *nodulation*
NTA *Nitrilotriacetic acid*
nts *nucleotides*

OD *optical density*
ORF *open reading frame*

P. Pseudomonas
PAGE *polyacrylamide gel electrophorese*
PBS *Phosphate-buffered saline*
PCR *polymerase chain reaction*

PQS *Pseudomonas quinolone signal*

qPCR *quantitative polymerase chain reaction*

QQ *quorum quenching*

QS *quorum sensing*

R. Rhizobium

RBS *Ribosome binding site*

Ribo-seq *Ribosome profiling*

RNA *Ribonucleic acid*

rpm *rounds per minute*

S. Sinorhizobium

SD *Shine Dalgarno*

SDS *Sodium dodecyl sulfate*

se- *seleno-*

SEM *scanning electron microscopy*

seq *sequencing*

smORF *small open reading frame*

SNP *Single nucleotide polymorphism*

$Spec^R$ *spectinomycin resistance*

TBS *Tris-buffered saline*

TBST *Tris-buffered saline with tween*

TEM *transmission electrone microscopy*

Tet^R *tetracyclin resistance*

TLC *Thin-layer chromatography*

TM *template modeling*

TY *tryptone-yeast*

USDA257 *Sinorhizobium fredii USDA257*

V. Vibrio

vol *volume*

wt *wild type*

wt. *weight*

YEM *yeat-extract-mannitol*

Δ *deletion of*

Table 19: Twenty-one amino acids with corresponding three and one letter codes

Alanine	**Valine**	**Isoleucine**	**Leucine**	**Methionine**	**Phenylalanine**
Ala	Val	Ile	Leu	Met	Phe
A	V	I	L	M	F
Tyrosine	**Tryptophan**	**Serine**	**Threonine**	**Asparagine**	**Glutamine**
Tyr	Trp	Ser	Thr	Asn	Gln
Y	W	S	T	N	Q
Cystein	**Glycine**	**Proline**	**Arginine**	**Histidine**	**Lysine**
Cys	Gly	Pro	Arg	His	Lys
C	G	P	R	H	K
Aspartic acid	**Glutamic acid**				
Asp	Glu				
D	E				

Table 20: Nucleobase- functional units of the genetic code

Adenine	Guanine	Cytosine	Thymine	Uracil
A	G	C	T	U

Figures

Figures

Tables

Acknowledgements

First, I would like to thank Prof. Dr. Wolfgang Streit to give me the opportunity to accomplish my work in his group and on this project. I would like to thank him for his support throughout the last four years.

Then, I would like to thank Prof. Dr. Christian Betzel for reviewing this work as the second evaluator of my PhD thesis.

Furthermore, I would like to acknowledge the whole department of microbiology and biotechnology at Biocentre Klein Flottbek. We had a lot of fun working together. Thanks for all the nice discussions about research or all other topics. It was a great pleasure to work with this fantastic Mibi gang. Many thanks also to our “PowerLab 3.114” and its researchers: Jennifer Chow, Dominik Danso, Pablo Perez Garcia, Hilke Duin, Eva Oellingrath, Hongli Zhang, Birhanu Kinfu and our best-technician assistant Angela Jordan. This lab will never again have as much power as it has in recent years with these guys. I would also thank Dr. Christel Vollstedt. Since my bachelor thesis, she has been giving me ideas and answers to a lot of my scientific questions.

I would also like to thank Prof. Dr. Betzel and Nadine Werner, Dr. Manolo Ferrer, Anja Poehlein, Prof. Dr. Cynthia Sharma and Elke Woelken for their constructive cooperation.

Special thank goes to Hilke Duin. She is a great partner in Team Rhizobia and over time she has become a friend you won’t want to miss. I hope that despite separate ways, we keep in contact.

Finally, I would like to thank my family, friends and Mathis Michel for the support and trust they have given me throughout my studies and finishing my PhD.

Appendix

Table 21: Primers used in present work

Primer	Sequence (5'-3' direction)	Tm [°C]	Reference
oKP42	GCGTCGACATGAATTTCTCCGTCGTTGTTG	61	This work
oKP43	GCAAGCTTCTACAAGACGCAAGGGGGCCTTC	69	This work
oKP68	GCGTCGACATGAATTTCTCCGTCGTTGTTG	61	This work
oKP76	GCGAATTCATGGTGAAAGCAAGTCGAGCAGAAC	67	This work
oKP77	GCAAGCTTTTGCAGGGGAACTCTCGAATAACG	66	This work
oKP82	GCAAGCTTCTACAAGACGCAAGGGGGCCTTC	69	This work
oKP112b	GGGAATTCGAGCGTCGCCTCATGAATGG	65	This work
oKP116	GGGCTAGCGATTCACAGGAGAACAATTCGTGTC	64	This work
oKP117	GCGTCGACTACTGTCCGTTACAGGCCATGC	66	This work
oKP130	ATGAATTCGTGCTCCGCCTGGATCG	64	This work
oKP131	GCAAGCTTATAGACTCGCTCCGACG	60	This work
oKP135	AAGAATTCATGGTGAGCAAGGGCGAGGAG	68	This work
oKP136	GCAAGCTTACTTGTACAGCTCGTCCATGC	67	This work
oKP141	TTACGCGACATGCCGCTCCTCCC	73	This work
oKP142	CTCGCCGGCAAAACATGCGTGTTTC	71	This work
oKP146	CTCCTTACGCTGATTGG	56	This work
oKP147	TCATTGCAGGGGAACTC	58	This work
oKP153	GCGGATCCATCCAACCACTCGATCGTTGC	65	This work
oKP154	CAGAATTCGCAGATTCTCGCGATCTCAAAGCC	68	This work
oKP155	AAGCTAGCCAAGACGCAAGGGGGCCTTC	67	This work
oKP156	GCCTCGAGATCCAACCACTCGATCGTTGC	65	This work
oKP157	GCCTCGAGCGTAAGGAGAAATCAGTGGAACAG	59	This work
oKP158	GCGAATTCCATCACACCGGAGCGGACATAG	66	This work
oKP164	TTGAGAGCGGTAGTTGG	57	This work
oKP165	ACCGCTACACCGAAATC	58	This work
oKP166	GCGCTAGCCTAGGAACCTTGTAGGATCCTACAACTG	64	This work
oKP178	GCGAATTCTAGAGGAATGAGCGTTGAGGTGAC	65	This work
oKP179	GCCTCGAGGAATTGTTCTCCTGTGAATCCTAC	61	This work
oKP180	GCCTCGAGCCCATTGTCAATCCCAATCTTCTG	63	This work

Primer	Sequence (5'-3' direction)	Tm [°C]	Reference
oKP181	AAGTCGACGCGTGTTTCTCCTTCATTCCTAGGC	67	This work
oKP199	AATGTCGACGTGTCCATAGACGGCCGTGATC	66	This work
oKP200	GGCAAGCTTTTACAGGCCATGCCCCGCAAAAATAG	69	This work
oKP221	CGAACTTTCGTCCTTTTCGGAACAGCAGGAGATC	71	This work
oKP222	TCCAGCGCACGCGCCAGT	70	This work
oKP223	CTCGAGCACCACCACCACC	65	This work
oKP224	CCGGAAGGGCGATGCGGG	68	This work
oKP225	CATGCAAGGTCCGAATAGGCCCCGGCAGGCCTCGCC CGGCTGGA	85	This work
oKP226	CATGCAAGGTCCGAATAGGCCC	66	This work
oKP229	CGGCATCATTGTCGAACTGGGACTGGACCCAG	74	This work
oKP230	CGCACCTGCGCCATGGCG	70	This work
oKP235	ACAATCTCCATGCGGCCGGAAGGAGAC	71	This work
oKP236	TTCATGATGATCCAGCAGCGCAACGG	72	This work
oKP237	CGCGCCTCAGCAGAACAATTCGG	68	This work
oKP238	GCAGCTCAACAAAGCCGTCGATGACCTC	70	This work
pET-for	ATATAGGCGCCAGCAACC	59	Merck KGaA
pET-rev	TCCGGATATAGTTCCTC	52	Merck KGaA
RT_recA-f	CGGCTCGTAGAGGACAAATCG	62	J. Grote
RT_recA-r	CAATGATGCGCCCTTTCGG	61	J. Grote
RT_c03800-f	GCGAGATGAAGGGCTATCTGG	62	J. Grote
RT_c03800-r	GCGCGACGTCCTTGATATG	60	J. Grote
RT_repA-f	GCAGCAGTTCCCACCGAATG	63	J. Grote
RT_repA-r	GCACGTAGTTCCTGGCTTCC	62	J. Grote
RT_nifB-f	GCTTGAAAGCCTGACCAACAC	62	J. Grote
RT_nifB-r	GGCGCAGTCATATTTGCGATTG	62	J. Grote

Restriction sites are underlined, Annealing temperature was calculated with the Tm calculator of Thermo Scientific (https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html

Table 22: P-values of the various NGR234 strains containing *repX* promoter fusion

p-value analysis	Wt	Δ*ngrI* Δ*traI*-copy+	Δ*ngrI* Δ*traI*	Δ1725+	Δ*traR*	Δ*traI*	Δ*ngrI*	ANU265	ANU265 +*repX*	Anu265 +a1725
Wt	1.0000	0.2390	0.0000	0.1803	0.3085	0.1360	0.0075	0.0000	0.0000	0.0000
Δngrl Δtral copy+	0.2390		0.0000	0.7549	0.7437	0.7303	0.1145	0.0000	0.0000	0.0000
Δ*ngrI* Δ*traI*	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Δ1725+	0.1803	0.7549	0.0000		0.5348	0.9984	0.2740	0.0000	0.0000	0.0000
Δ*traR*	0.3085	0.7437	0.0000	0.5348		0.4803	0.0374	0.0000	0.0000	0.0000
Δ*traI*	0.1360	0.7303	0.0000	0.9984	0.4803		0.2238	0.0000	0.0000	0.0000
Δ*ngrI*	0.0075	0.1145	0.0000	0.2740	0.0374	0.2238		0.0000	0.0000	0.0000
ANU265	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.0041	0.0287
ANU265 +*repX*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0041		0.1719
Anu265 +a1725	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0287	0.1719	

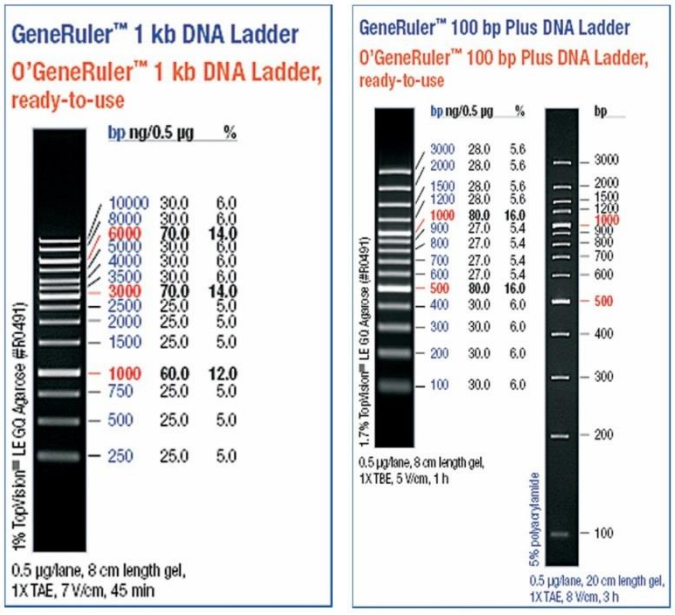

Figure 53: GeneRuler DNA ladders that were used for agarose gel electrophorese.
The Ladders were ordered from Thermo Scientific (Darmstadt, Germany).
https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/thermo-scientific-nucleic-acid-electrophoresis-purification/dna-electrophoresis-thermo-scientific/dna-ladders-thermo-scientific/generuler-dna-ladders.html

Table 23: Gene fragment for construction of GqqA-Mut4

(5'→ 3')
CGGGCTTGAAATCCTGCGCCGCAATGTGGAGGACGCCACGCATAACACCACGCGCTTTTA CATCGCATCCCGCAGGCCCGCGACCCTGGCGTCGGGCTTCATGACCACGCTGCTGTTCCG CGTCAACAACCAGCCCGGCGCGCTGTACAAGGCGCTGGGTGGCCTTGCCACGGCGGGGG TGAACATGACGC

Gene fragment was ordered at Eurofins genomics GmbH (Ebersberg, Germany)

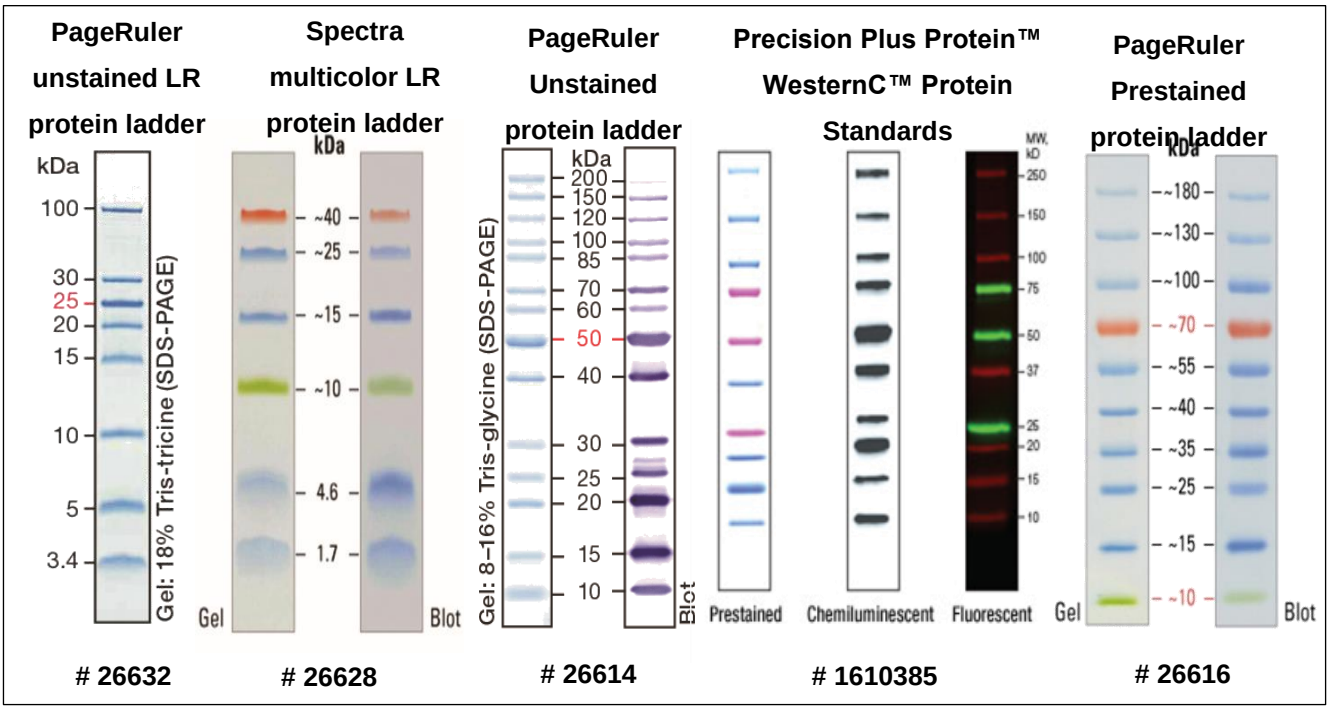

Figure 54: PageRuler Protein Ladders that were used for protein gel and western blot
The Ladders were ordered from Thermo Scientific (Darmstadt, Germany).
https://www.thermofisher.com/de/de/home/life-science/protein-biology/protein-gel-electrophoresis/protein-standards-ladders.html.
The Precision Plus ProteinTM WesternCTM Protein Standard was ordered from Bio-Rad (Munich, Germany), https://www.bio-rad.com/de-de/category/protein-ladders-standards-markers?gclid=EAIaIQobChMIhfm14b_i5AIVB-N3Ch1y1wtvEAAYASAAEgJm6PD_BwE&WT.knsh_id=315733ef-91c2-494f-98a8-7e38b38b7176&WT.srch=1&ID=09507551-2848-4bd1-a3c1-650b4d41aa48&WT.mc_id=170125005874

Table 24: Sequencing results of the produced constructs and mutants

(all sequences are shown in 5'→ 3' direction)
>NGR234 Δ*ngrI*Δ*traI* TCGGCCGTTGGCGAAGCGCTCGACCTGGATGATGTCGCGCACACGCCGGCCTCGCGGCG TACGCTCGATGGAAACGATGAGGTCGACAACCTCCCCGATCACCTCGTGCATCGGCTGTT GGCTTGCTTCGGCCGTCAGCTGCTCCAGGCGCCGGAGCGCGGAAGTCGCCGTGTTCGAG TGGATTGTTGCCACGCCGCCTGGATGTCCGGTGTTCCATGCCTTGAGCAAGGTGAGGGCA

GCACCATCGCGCACCTCGCCGACGACTATTCGGTCGGGACGCAGGCGCATCGTGCTCTTG
AGCAGCCGCGCCATGTCAACCGTATCGCTGGTACGAAGAAGAACGGCGTTGTCGGCCGCA
CACTGGATTTCGGCGGTGTCTTCGAGGATGACGAGCCGGTCTTCCGGCGCGCTCCTCACG
ATCTCATGGATTACTGCATTTGCGAGCGTCGTTTTGCCGGAGGCAGTTCCACCGGAAATGA
TGATGTTGAGGCGTGAGTCAATGGCGCTGCGGATCGTCGCAGCCTGGGCCTCGGTCATCA
CACCGGAGCGGACATAGTCGTCGAGTGGGATCAGGCGAGATGCCCGACGACGGATTGTG
AAGGAGGGCTTCGCGACCACTGGCGGCAATAGACCCTCGAAACGATGTCCGCCTATCGGA
AGCTCACCGGAAATGATGGGTTGTTCGGTGTCCACTTCGGATTGGAGGGCATGAGCAACC
GTCCCAATCACCATTTCCGCCGCAGCCGACGACATCTCGCCAGCGGGCGCGACGCCGTG
CCCGAGCCGTTCTATGAACAGCTTACCATCGGGATTCAGCATTATCTCTACTACGTTGCTAT
CGTCCAAAGCAACGCAGAGCTGGTCGCCAAGCGCCTCCTGAAGCTTGCGGACGAGGCGA
GGATGAGAGCGAAGCTGTTCCACTGATTTCTCCTTACGCTCGAGATCCAACCACTCGATCG
TTGCTCGATACGGCTAGGATGTAGGTCGGTTGAAGATCGTCGAAAGTATCTGACTCACAGC
CTCCGACGACATTGACTTCCCAGCCCAAACGATCGGAAAAGACGCGGGCGCGCAGTTCGT
GATGGCTGCGGAGAAGTTGGGCTTCTTCGATATTTCGCGGCTTTGAGATCGCGAGAATCTG
CATGAATTTCTCCGTCGTTGTTGGACCGCGGGGAAATCAGCACAAATCGGAATCGGGTGG
CAGTTGTAGGATCCTACAAGGTTCCCTGGAACGCGTTGGGCGTTGCGTCGCGTCACGTCT
GCCTTTGCAGACCCGAAGGCCCCCTTGCGTCTTGTAGGATTCACAGGAGAACAATTCGTGT
CCATAGACGGCCGTGATCCGCGTCTAACCCTCAGCGGACAAATACAAATTTGCGAAAGTAA
TCCTTCGGCTCACTTGAGCCAAGCAGGGTCTTTTCAACCCATTGTCAATCCCAATCTTCTGG
ACGGTCAGGAAGAATCACCTTCTATTTTTGCGGGGCATGGCCTGTAACGGACAGTACGCCG
CAGAGAGGGACAGATGGTGAAAGCAAGTCGAGCAGAACCAATTGCTGGGCCCAATGAAGA
AGTTGGCAGCAAAGTTATGTGCCCCGGCTCTCTTTCCGGGCAACCGCAGCAACTCAGAACT
CGAATATATTCACCAAGGCTGAGTCGGGTGAGTCGCGGGGAGCAATACGACCGCTATTTC
GCGGTGACGTTTGGTAGTCAGGCGAACGGCCGTATTTCTCCCCTGGAAAAGTGCTGGCGC
CNNGCGGAGCGCAAG

>NGR234 gTF*repX*::mCherry

GATTGGCCAAAAGTGCTGATGATTTGAGAGCGGTAGTTGGCGGGGCGGAGCCTCATGAAT
GTGTCTGCCTTTGCCGGCAGAGTGCCCGCCACGGCCACGGTCGCTCCGATCGGCCTTGG
CTCGCCAATGCGCTGAAGCGGCCAACCGACGCGGGCGAGGATCCGTTCGAACCGAAGAT
CCGTCACCGTGACGATTTCGGTGTAGCGGTTAGCCACCGACCATTCGATGATCCCCGCGA
ACATTGTGAGCGTCGCCTCATGAATGGAGCCGTCTCCCCTGCTCTCAGCAAGGAAGGTGT
CCACGCAAAAGCGCGAGCTCTCGACCATAGAGGAATGAGCGTTGAGGTGACCGGCAGAAA
GGAGCGAGGGGAAAACGTTGGCCACCATCGTGGGGCCGAGCGCAGGCAGTAGGCGAGCA
CATCCAACCACTCGATCGTTGCTCGATACGGCTAGGATGTAGGTCGGTTGAAGATCGTCGA
AAGTATCTGACTCACAGCCTCCGACGACATTGACTTCCCAGCCCAAACGATCGGAAAAGAC
GCGGGCGCGCAGTTCGTGATGGCTGCGGAGAAGTTGGGCTTCTTCGATATTTCGCGGCTT
TGAGATCGCGAGAATCTGCATGAATTTCTCCGTCGTTGTTGGACCGCGGGGAAATCAGCAC
AAATCGGAATCGGGTGGCAGTTGTAGGATCCTACAAGGTTCCCTGGAACGCGTTGGGCGT
TGCGTCGCGTCACGTCTGCCTTTGCAGACCCGAAGGCCCCCTTGCGTCTTGTCTAGAATG
GTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTG
CACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCG
CCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCT
TCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACC
CCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCG

TGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACNCCTCCTTGCAGGAC
GGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTA
ATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGG
CGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACG
CTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACG
TCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGA
ACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAGCTAGCG
ATTCACAGGAGAACAATTCGTGTCCATAGACGGCCGTGATCCGCGTCTAACCCTCAGCGGA
CAAATACAAATTTGCGAAAGTAATCCTTCGGCTCACTTGAGCCAAGCAGGGTCTTTTCAACC
CATTGTCAATCCCAATCTTCTGGACGGTCAGGAAGAATCACCTTCTATTTTTGCGGGGCATG
GCCTGTAACGGACAGTACGCCGCAGAGAGGGACAGATGGTGAAAGCAAGTCGAGCAGAAC
CAATTGCTGGGCCCAATGAAGAAGTTGGCAGCAAAGTTATGTGCCCCGGCTCTCTTTCCGG
GCAACCGCAGCAACTCAGAACTCGAATATATTCACCAAGGCTGAGTCGGGTGAGTCGCGG
GGAGCAATACGACCGCTATTTCGCGGTGACGTTTGGTAGTCAGGCGAACGGCCGTATTTCT
CCCCTGGAAAAGTGCTGGCGCCTGGCGGAGCGCAAGCGTCCATCTAAGTGCCTAGGAATG
AAGGAGAAACACGCATGTTTTGCCGGCGAGCAGGGGCTGCTCGGGACGCGGTATAAACCC
TTTGTTAACTCTATATTCCTTGTCAGTCGGCGAGAATCGCACATAA

>NGR234 Δ*repY*

AAACGTTGGCCACCATCGTGGGGCCGAGCGCAGGCAGTAGGCGAGCACATCCAACCACTC
GATCGTTGCTCGATACGGCTAGGATGTAGGTCGGTTGAAGATCGTCGAAAGTATCTGACTC
ACAGCCTCCGACGACATTGACTTCCCAGCCCAAACGATCGGAAAAGACGCGGGCGCGCAG
TTCGTGATGGCTGCGGAGAAGTTGGGCTTCTTCGATATTTCGCGGCTTTGAGATCGCGAGA
ATCTGCATGAATTTCTCCGTCGTTGTTGGACCGCGGGGAAATCAGCACAAATCGGAATCGG
GTGGCAGTTGTAGGATCCTACAAGGTTCCCTGGAACGCGTTGGGCGTTGCGTCGCGTCAC
GTCTGCCTTTGCAGACCCGAAGGCCCCCTTGCGTCTTGTAGGATTCACAGGAGAACAATTC
CTCGAGCCCATTGTCAATCCCAATCTTCTGGACGGTCAGGAAGAATCACCTTCTATTTTTGC
GGGGCATGGCCTGTAACGGACAGTACGCCGCAGAGAGGGACAGATGGTGAAAGCAAGTC
GAGCAGAACCAATTGCTGGGCCCAATGAAGAAGTTGGCAGCAAAGTTATGTGCCCCGGCT
CTCTTTCCGGGCAACCGCAGCAACTCAGAACTCGAATATATTCACCAAGGCTGAGTCGGGT
GAGTCGCGGGGAGCAATACGACCGCTATTTCGCGGTGACGTTTGGTAGTCAGGCGAACGG
CCGTATTTCTCCCCTGGAAAAGTGCTGGCGCCTGGCGGAGCGCAAGCGT

>NGR234 Δa1725

GCTAGATCTATAAGGCGACCAGAGCCCGGGGTCGGCATGATACGTATCACGAGCTCCATG
CGCTTCATGCGGTTGTTGAGAATACGTTTAACGTCGCCGATACCATCTAATGCGAACACGA
TCGTTTTTTGTACCCCCGCCTGTGCCAATTGAGAGACCAAGCGGGCAAGAATTGGCATGCC
GCCGACCGAGCTCAGCGGCGCTGGACAGCGATCTTTCCAAGTCGAGATTATCAGTCGCGG
GTCAGTGACAATAATCGCTAAACTCTGGACGTTGGGCATTCTTCAACTCGCATTGCGGGGG
CTAACATTACCAGATCACGTGCAAACTTTGATCATGCTGACCGAGAAGATAAGCATGTGGTC
GGCACCGCCACTAACGTGAACAAAGGCATAATCGCACTGAGCTTCGTGGAGAGTCTCAGC
ATAGAGCGCCATCTCGTGCACCTATTTACGTTGGTAGTATCGTGTTGGGCTATTCTTCGATT
TCCATCAATGCCTATATCTTAGGAAGGAGGCGAGCGGGTCTAGATCAGACACTCGACTTGT
TAGACTGGCGACACGGCCAAAAGCAGTGCCGATTGTCTGATGCGGGAAGGTGTCGGCGG
CTACGGATACGCTGCTCTTGGACTTGTCGTGTGCTCGAAAGGACCCGGCCCCCCAGTAAC
ACGTCCGGATGTCTCGTAAACTGCAACCAACATAGCGCCCGTGTGCAGGTCACCATGGAA
CGACAAATTTCATACAGAAGCCTTTCGAAGATACCTGGCCTATTCTCCTTGCTTCTTGCTGC

AACCTTGTCCCGCCTGGCCGGACGCATGTTCGTCCTCACATTGGTGCTGTTCGCGCTGGC
GCGTTTTTCGTCACCAGTTCTGGCTGGGTGGCTTA

>pMALc2x::*repX*

ATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTC
TCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATC
TTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCC
CGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGG
CAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGC
TGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGA
AAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGC
TGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGC
GTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAAC
ACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGC
GATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGT
GTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGC
GCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATC
TGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGC
TGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGC
CCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGT
ACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCG
CAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGG
GAAGGATTTCAGAATTCGGATCCTCTAGAGTCGACATGAATTTCTCCGTCGTTGTTGGACC
GCGGGGAAATCAGCACAAATCGGAATCGGGTGGCAGTTGTAGGATCCTACAAGGTTCCCT
GGAACGCGTTGGGCGTTGCGTCGCGTCACGTCTGCCTTTGCAGACCCGAAGGCCCCCTTG
CGTCTTGTAG

>pMALc2x::*repY*

ATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTC
TCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCC
GGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATC
TTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCC
CGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGG
CAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGC
TGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGA
AAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGC
TGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGC
GTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAAC
ACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGC
GATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGT
GTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGC
GCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATC
TGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGC
TGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGC
CCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGT
ACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCG

CAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGATCGAGG
GAAGGATTTCAGAATTCGTGTCCATAGACGGCCGTGATCCGCGTCTAACCCTCAGCGGACA
AATACAAATTTGCGAAAGTAATCCTTCGGCTCACTTGAGCCAAGCAGGGTCTTTTCAACCCA
TTGTCAATCCCAATCTTCTGGACGGTCAGGAAGAATCACCTTCTATTTTTGCGGGGCATGG
CCTGTAA

>pET-21a::repA0

TTAAGAAGGAGATATACATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCC
GAATTCATGGTGAAAGCAAGTCGAGCAGAACCAATTGCTGGGCCCAATGAAGAAGTTGGCA
GCAAAGTTATGTGCCCCGGCTCTCTTTCCGGGCAACCGCAGCAACTCAGAACTCGAATATA
TTCACCAAGGCTGAGTCGGGTGAGTCGCGGGGAACAATACGACCGCTATTTCGCGGTGAC
GTTTGGTAGTCAGGCGAACGGCCGTATTTCTCCCCTGGAAAAGTGCTGGCGCCTGGCGGA
GCGCAAGCGTCCATCTAAGTGCCTAGGAATGAAGGAGAAACACGCATGTTTTGCCGGCGA
GCAGGGGCTGCTCGGGACGCGGTATAAACCCTTTGTTAACTCTATATTCCTTGTCAGTCGG
CGAGAATCGCACATAATAACGATTATGTCGGTTCTCGGGCCGAAAAACCGTTATTCGAGAG
TTCCCCTGCAAAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGG
CTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATA

>pMALc2x::a1725

TTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGG
GCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCA
AGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTC
CGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGA
AGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTT
GGCGAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCC
GAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGC
CAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAA
CAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGATTTCAGAATTCGTG
CTCCGCCTGGATCGCTGGAGGAGGTTTTCATTTTCAGCCTTAATCATGCGGCACGCCGTCG
GAGCGAGTCTATAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT
GGCGTT

>pBBR1MCS-2::P*repX*::mCherry

TCTAGAACTAGTGGATCCATCCAACCACTCGATCGTTGCTCGATACGGCTAGGATGTAGGT
CGGTTGAAGATCGTCGAAAGTATCTGACTCACAGCCTCCGACGACATTGACTTCCCAGCCC
AAACGATCGGAAAAGACGCGGGCGCGCAGTTCGTGATGGCTGCGGAGAAGTTGGGCTTCT
TCGATATTTCGCGGCTTTGAGATCGCGAGAATCTGCGAATTCATGGTGAGCAAGGGCGAG
GAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCC
GTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCA
CCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATC
CTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCC
GACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAG
GACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCTTGCAGGACGGCGAGTTCATCTAC
AAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC
ATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGA
GATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCA
CCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGG
ACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCC

GCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAGCTTATCGATACCGTCGACCTC
G

>pBBR1MCS-2::TF*repX*::mCherry

TCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCCATCCAACCACTCGATCGTTGCTCG
ATACGGCTAGGATGTAGGTCGGTTGAAGATCGTCGAAAGTATCTGACTCACAGCCTCCGAC
GACATTGACTTCCCAGCCCAAACGATCGGAAAAGACGCGGGCGCGCAGTTCGTGATGGCT
GCGGAGAAGTTGGGCTTCTTCGATATTTCGCGGCTTTGAGATCGCGAGAATCTGCATGAAT
TTCTCCGTCGTTGTTGGACCGCGGGGAAATCAGCACAAATCGGAATCGGGTGGCAGTTGT
AGGATCCTACAAGGTTCCCTGGAACGCGTTGGGCGTTGCGTCGCGTCACGTCTGCCTTTG
CAGACCCGAAGGCCCCCTTGCGTCTTGTCTAGAATGGTGAGCAAGGGCGAGGAGGATAAC
ATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGC
CACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCG
CCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTC
AGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAA
GCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCG
TGGTGACCGTGACCCAGGACTCCTCCTTGCAGGACGGCGAGTTCATCTACAAGGTGAAGC
TGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGG
AGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAG
AGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGC
CAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTC
CCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCAC
CGGCGGCATGGACGAGCTGTACAAGTAAGCTTGATATCGAATTCCTGCAGCCCGGGGGAT
CCACTAGTTCTAGA

>pBBR1MCS-2::TFneg-*repX*::mCherry

GAATTCCATCCAACCACTCGATCGTTGCTCGATACGGCTAGGATGTAGGTCGGTTGAAGAT
CGTCGAAAGTATCTGACTCACAGCCTCCGACGACATTGACTTCCCAGCCCAAACGATCGGA
AAAGACGCGGGCGCGCAGTTCGTGATGGCTGCGGAGAAGTTGGGCTTCTTCGATATTTCG
CGGCTTTGAGATCGCGAGAATCCTCGAGAATTTCTCCGTCGTTGTTGGACCGCGGGGAAAT
CAGCACAAATCGGAATCGGGTGGCAGTTGTAGGATCCTACAAGGTTCCCTGGAACGCGTT
GGGCGTTGCGTCGCGTCACGTCTGCCTTTGCAGACCCGAAGGCCCCCTTGCGTCTTGTCT
AGAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTC
AAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGA
GGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCC
TGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAA
GCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGA
GCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCTTGCA
GGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCC
CGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGG
ACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTAC
GACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTA
CAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAG
TACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAAGC
TT

>pBBR1MCS-2::TF*a1725*::eGFP

TCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCCAAACTTTGATCATGCTGACCGAGA

AGATAAGCATGTGGTCGGCACCGCCACTAACGTGAACAAAGGCATAATCGCACTGAGCTTC
GTGGAGAGTCTCAGCATAGAGCGCCATCTCGTGCACCTATTTACGTTGGTAGTATCGTGTT
GGGCTATTCTTCGATTTCCATCAATGCCTATATCTTAGGAAGGAGGCGAGCGGGTGCTCCG
CCTGGATCGCTGGAGGAGGTTTTCATTTTCAGCCTTAATCATGCGACACGCCGTCGGAGCG
AGTCTATCTAGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTG
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGG
CGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT
GCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGA
GGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGA
CAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG
CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC
TGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACG
AGCTGTACAAGTAAAAGCTT

>pBBR1MCS-2::TFneg-*a1725*::eGFP

TCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCCAAACTTTGATCATGCTGACCGAGA
AGATAAGCATGTGGTCGGCACCGCCACTAACGTGAACAAAGGCATAATCGCACTGAGCTTC
GTGGAGAGTCTCAGCATAGAGCGCCATCTCGTGCACCTATTTACGTTGGTAGTATCGTGTT
GGGCTATTCTTCGATTTCCATCAATGCCTATATCTTAGGAAGGAGGCGAGCGGCTCGAGCT
CCGCCTGGATCGCTGGAGGAGGTTTTCATTTTCAGCCTTAATCATGCGACACGCCGTCGGA
GCGAGTCTATCTAGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATC
CTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA
GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCC
CGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTA
CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT
CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGC
CGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGG
CAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGC
TGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGA
AGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGG
ACGAGCTGTACAAGTAAAAGCTT

>pBBR1MCS-2::*repX*

CAGGAAACAGCTATGACCATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACA
AAAGCTGGGTACCGGGCCCCCCCTCGAGGTCGACATGAATTTCTCCGTCGTTGTTGGACC
GCGGGGAAATCAGCACAAATCGGAATCGGGTGGCAGTTGTAGGATCCTACAAGGTTCCCT
GGAACGCGTTGGGCGTTGCGTCGCGTCACGTCTGCCTTTGCAGACCCGAAGGCCCCCTTG
CGTCTTGTAGAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTTCTAGA

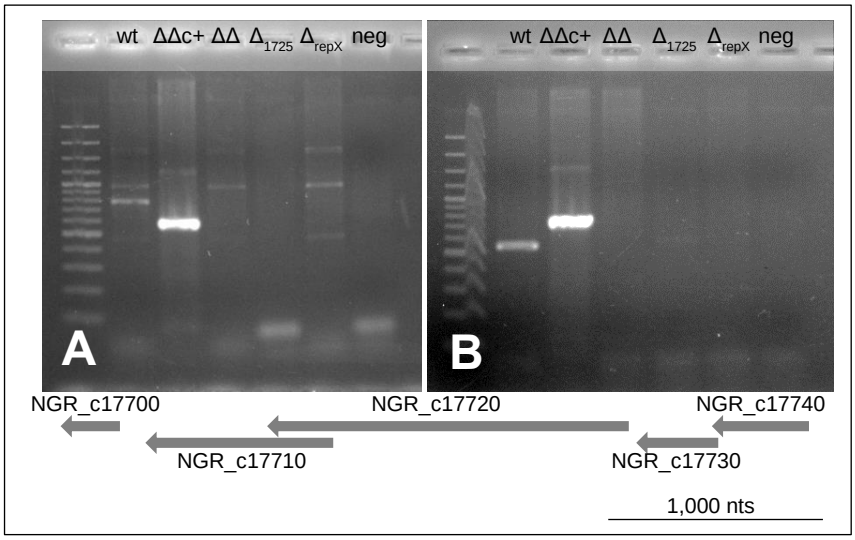

Figure 55: Colony PCR for detecting transposase fragment in various NGR234 strains
colony PCRs were performed with oKP235/oKP236 and oKP237/oKP238 that only can amplified these region on cNGR234

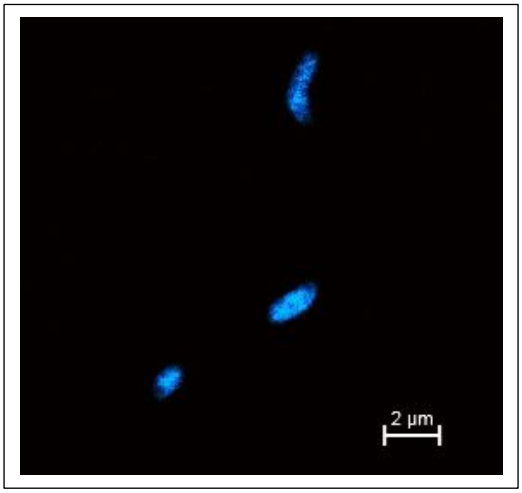

Figure 56: Confocal microscope imaging of the RepX::mCherry genomic translation fusion in NGR234 wt
The gene repX was fused to the red fluorescence protein mCherry and integrated into the genome of NGR234 wt. For microscopy the cells grew for 48 h and were prepared for fluorescence imaging. The settings for the microscope LSM800 airyscan are listed in material and methods (Table 7).

Appendix

Table 25: Transcriptomic data of new small ORFs on pNGR234*a*

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
3,933	4,070		*smORF-1*	45	27.36	12.60	19.98	-1.03	0.020	0.101	196.89	82.51	139.70	-1.25	0.000	0.000
4,441	4,575		*smORF-2*	44	28.29	5.82	17.06	-2.12	0.000	0.001	119.27	117.14	118.21	-0.02	0.916	0.959
4,868	4,948		*smORF-3*	26	10.22	1.73	5.97	-2.33	0.001	0.008	66.14	61.15	63.64	-0.12	0.639	0.810
5,560	5,640		*smORF-4*	26	51.27	9.16	30.22	-2.38	0.000	0.000	21.34	17.93	19.64	-0.25	0.562	0.756
7,649	7,816		*smORF-5*	55	23.48	17.88	20.68	-0.40	0.354	0.607	14.88	14.14	14.51	-0.07	0.879	0.934
8,015	8,080		*smORF-6*	21	16.54	11.23	13.89	-0.47	0.377	0.634	21.73	17.47	19.60	-0.32	0.448	0.667
30,092	30,241		*smORF-7*	49	31.86	18.42	25.14	-0.82	0.049	0.182	201.35	106.16	153.76	-0.92	0.000	0.000
31,333	31,488		*smORF-8*	51	71.12	40.49	55.81	-0.80	0.033	0.147	24.14	35.58	29.86	0.57	0.103	0.259
37,314	37,231	c	*smORF-9*	27	40.56	9.00	24.78	-2.09	0.000	0.000	50.50	105.08	77.79	1.06	0.000	0.000
37,875	37,726	c	*smORF-10*	49	16.89	8.07	12.48	-1.07	0.033	0.146	57.18	177.15	117.16	1.64	0.000	0.000
38,147	37,974	c	*smORF-11*	57	469.63	136.51	303.07	-1.73	0.000	0.000	223.50	445.90	334.70	1.00	0.000	0.000
40,016	39,882	c	*smORF-12*	44	37.58	31.74	34.66	-0.24	0.524	0.750	3277.88	2710.65	2994.27	-0.27	0.032	0.099
40,193	40,294		*smORF-13*	33	51.62	32.79	42.20	-0.66	0.053	0.190	9.59	11.58	10.59	0.26	0.645	0.810
41,037	40,915	c	*smORF-14*	40	469.04	369.77	419.40	-0.34	0.138	0.352	196.18	210.55	203.36	0.11	0.536	0.734
41,203	41,114	c	*smORF-15*	29	470.47	359.83	415.15	-0.37	0.266	0.526	104.50	112.20	108.35	0.11	0.610	0.787
43,081	43,179		*smORF-16*	32	14.49	5.93	10.21	-1.09	0.050	0.184	26.28	20.80	23.54	-0.33	0.382	0.612
43,347	43,457		*smORF-17*	36	26.39	9.57	17.98	-1.51	0.001	0.012	148.49	70.14	109.32	-1.08	0.000	0.000
45,166	45,128	c	*smORF-18*	12	110.52	82.13	96.32	-0.41	0.151	0.377	374.53	581.01	477.77	0.63	0.000	0.000
48,411	48,467		*smORF-19*	18	141.20	95.25	118.22	-0.53	0.112	0.306	5.29	37.42	21.35	2.83	0.000	0.000
49,507	49,656		*smORF-20*	49	88.47	88.68	88.57	0.01	0.980	0.995	44.10	258.95	151.53	2.57	0.000	0.000
51,972	52,028		*smORF-21*	18	31.62	30.86	31.24	-0.13	0.772	0.884	37.25	76.68	56.96	1.05	0.000	0.001
52,402	52,515		*smORF-22*	37	9.12	36.65	22.88	1.84	0.004	0.032	0.49	3.40	1.95	2.85	0.065	NA
52,739	52,879		*smORF-23*	46	1.04	15.81	8.43	2.98	0.001	0.009	2.62	14.83	8.72	2.50	0.001	0.006
58,458	58,363	c	*smORF-24*	31	2.59	19.98	11.29	2.38	0.001	0.008	34.95	58.50	46.73	0.74	0.011	0.041
58,605	58,769		*smORF-25*	54	26.96	25.44	26.20	-0.19	0.653	0.810	100.40	69.50	84.95	-0.53	0.017	0.058
58,847	58,803	c	*smORF-26*	14	10.43	52.17	31.30	2.07	0.000	0.000	242.91	213.55	228.23	-0.19	0.260	0.494
59,328	59,089	c	*smORF-27*	79	126.96	538.74	332.85	2.02	0.000	0.000	39.27	36.63	37.95	-0.09	0.770	0.889
60,693	60,568	c	*smORF-28*	23	3.04	11.71	7.38	1.88	0.014	0.078	20.52	32.80	26.66	0.67	0.070	0.194
60,939	60,895	c	*smORF-29*	14	12.35	31.42	21.89	1.22	0.014	0.078	3.98	7.34	5.66	0.89	0.243	0.477

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
61,013	61,096		*smORF-30*	27	42.89	32.39	37.64	-0.41	0.303	0.568	8.23	7.54	7.89	-0.13	0.835	0.923
63,132	63,031	c	*smORF-31*	33	14.71	24.44	19.57	0.75	0.109	0.302	87.19	122.18	104.69	0.49	0.035	0.107
65,351	65,250	c	*smORF-32*	33	22.33	39.53	30.93	0.78	0.075	0.236	230.86	90.17	160.52	-1.35	0.000	0.000
65,903	66,037		*smORF-33*	44	10.26	14.85	12.56	0.48	0.368	0.622	57.26	49.83	53.54	-0.20	0.447	0.667
67,198	67,103	c	*smORF-34*	31	31.28	100.01	65.64	1.50	0.001	0.006	19.75	15.34	17.54	-0.36	0.397	0.622
68,211	68,092	c	*smORF-35*	39	99.59	72.49	86.04	-0.46	0.091	0.273	37.64	35.93	36.79	-0.08	0.803	0.906
68,550	68,603		*smORF-36*	17	29.07	16.10	22.58	-0.72	0.124	0.328	6.42	10.20	8.31	0.65	0.309	0.535
75,019	75,084		*smORF-37*	21	6.39	12.71	9.55	1.25	0.058	0.201	8.18	6.84	7.51	-0.27	0.679	0.841
75,736	75,587	c	*smORF-38*	49	24.00	21.58	22.79	-0.18	0.679	0.825	20.46	23.27	21.87	0.19	0.646	0.810
84,215	84,271		*smORF-39*	18	12.57	26.03	19.30	1.12	0.018	0.092	0.40	1.75	1.08	1.88	0.298	NA
85,236	85,093	c	*smORF-40*	47	5.66	14.83	10.24	1.24	0.043	0.171	8.80	5.13	6.97	-0.78	0.239	0.471
87,172	87,321		*smORF-41*	49	7.45	16.77	12.11	1.31	0.037	0.159	29.76	74.52	52.14	1.32	0.000	0.000
94,595	94,443	c	*smORF-42*	50	1.60	11.68	6.64	2.23	0.008	0.048	41.84	34.57	38.20	-0.27	0.368	0.594
99,004	98,840	c	*smORF-43*	25	5.16	21.10	13.13	1.81	0.013	0.074	8.16	8.47	8.32	0.03	0.963	0.978
99,437	99,532		*smORF-44*	31	33.15	36.05	34.60	0.20	0.734	0.856	19.30	18.75	19.02	-0.01	0.979	0.989
99,743	99,814		*smORF-45*	23	6.37	165.65	86.01	4.18	0.000	0.000	32.73	43.53	38.13	0.43	0.183	0.403
100,450	100,274	c	*smORF-46*	58	1.08	10.47	5.77	2.46	0.007	0.043	8.77	10.56	9.67	0.26	0.643	0.810
101,494	101,598		*smORF-47*	34	12.53	45.00	28.76	1.86	0.000	0.004	15.59	49.63	32.61	1.66	0.000	0.000
103,108	103,170		*smORF-48*	20	51.98	98.84	75.41	0.86	0.093	0.276	2.10	1.24	1.67	-0.80	0.557	NA
103,310	103,462		*smORF-49*	50	7.24	25.49	16.36	1.61	0.002	0.017	18.12	44.30	31.21	1.27	0.001	0.004
103,796	103,972		*smORF-50*	58	5.62	13.94	9.78	1.15	0.074	0.236	15.73	34.72	25.22	1.16	0.003	0.015
104,936	105,010		*smORF-51*	24	35.92	36.79	36.35	0.09	0.831	0.926	5.75	2.41	4.08	-1.24	0.183	0.403
109,225	109,389		*smORF-52*	54	9.24	14.17	11.71	0.90	0.195	0.445	5.64	6.24	5.94	0.13	0.868	0.934
110,044	110,163		*smORF-53*	39	8.06	20.94	14.50	1.28	0.052	0.188	27.94	41.50	34.72	0.56	0.092	0.237
110,808	110,918		*smORF-54*	36	11.68	8.98	10.33	-0.31	0.568	0.781	13.63	16.31	14.97	0.26	0.575	0.762
111,321	111,395		*smORF-55*	24	4.56	24.89	14.72	2.25	0.000	0.003	15.44	11.61	13.53	-0.41	0.410	0.635
114,381	114,467		*smORF-56*	28	31.66	98.80	65.23	1.66	0.000	0.001	7.04	8.28	7.66	0.24	0.708	0.855
114,548	114,670		*smORF-57*	40	10.60	16.08	13.34	0.62	0.255	0.512	27.01	32.72	29.87	0.28	0.411	0.635
116,003	115,869	c	*smORF-58*	44	22.16	24.45	23.30	0.09	0.837	0.929	3.84	6.35	5.10	0.70	0.399	0.622
116,619	116,717		*smORF-59*	33	194.40	238.40	216.40	0.30	0.320	0.582	138.74	127.27	133.00	-0.12	0.542	0.739
117,137	117,081	c	*smORF-60*	18	14.28	63.37	38.82	2.05	0.000	0.000	0.92	1.24	1.08	0.46	0.784	NA

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	**end**	**Orien-tation ***	**smORF- No**	**aa**	**WT**	**Δ*ngrl* Δ*tral* _copy+**	**baseMean**	**log2Fold Change**	**pvalue**	**padj**	**WT**	**Δ*ngrl*Δ*tral***	**baseMean**	**log2Fold Change**	**pvalue**	**padj**
119,224	119,325		*smORF-61*	33	13.13	23.52	18.32	0.85	0.124	0.328	34.87	36.04	35.45	0.05	0.875	0.934
119,880	119,824	c	*smORF-62*	18	16.05	63.50	39.78	2.00	0.000	0.005	8.35	41.64	25.00	2.27	0.000	0.000
119,889	119,975		*smORF-63*	28	9.72	18.64	14.18	0.87	0.088	0.267	41.85	82.28	62.07	0.98	0.000	0.002
120,602	120,432	c	*smORF-64*	56	296.99	186.02	241.51	-0.68	0.004	0.029	53.11	86.36	69.73	0.69	0.009	0.035
122,427	122,308	c	*smORF-65*	39	26.35	18.40	22.37	-0.46	0.343	0.597	30.47	28.39	29.43	-0.10	0.770	0.889
122,683	122,558	c	*smORF-66*	41	40.15	26.30	33.22	-0.53	0.336	0.593	85.64	66.55	76.09	-0.37	0.163	0.373
124,044	123,967	c	*smORF-67*	25	27.02	11.03	19.03	-1.21	0.052	0.188	54.13	24.90	39.51	-1.12	0.004	0.018
124,153	124,049	c	*smORF-68*	34	77.71	33.33	55.52	-1.19	0.014	0.078	100.59	24.08	62.34	-2.07	0.000	0.000
125,563	125,492	c	*smORF-69*	23	879.64	127.59	503.62	-2.62	0.000	0.000	748.08	52.69	400.38	-3.83	0.000	0.000
125,817	125,963		*smORF-70*	48	59.74	70.34	65.04	0.16	0.662	0.816	55.65	29.19	42.42	-0.93	0.001	0.006
127,021	126,935	c	*smORF-71*	28	2.10	10.48	6.29	1.79	0.019	0.099	74.80	63.60	69.20	-0.24	0.329	0.552
130,145	130,050	c	*smORF-72*	31	10.11	7.12	8.61	-0.41	0.472	0.715	747.75	53.96	400.86	-3.79	0.000	0.000
130,244	130,209	c	*smORF-73*	11	70.72	14.11	42.41	-2.30	0.000	0.000	871.95	47.81	459.88	-4.19	0.000	0.000
130,275	130,391		*smORF-74*	38	15.57	14.26	14.92	-0.05	0.932	0.968	8.32	4.47	6.39	-0.91	0.197	0.420
130,521	130,411	c	*smORF-75*	36	471.06	39.70	255.38	-3.30	0.000	0.000	511.39	27.23	269.31	-4.24	0.000	0.000
131,010	131,180		*smORF-76*	56	552.95	68.69	310.82	-2.83	0.000	0.000	20709.13	525.82	10617.48	-5.30	0.000	0.000
132,161	132,039	c	*smORF-77*	40	24.54	35.84	30.19	0.46	0.327	0.586	294.78	411.97	353.38	0.49	0.004	0.016
132,654	132,607	c	*smORF-78*	15	9.70	12.88	11.29	0.35	0.503	0.739	7.19	11.13	9.16	0.60	0.336	0.563
142,712	142,813		*smORF-79*	33	5.68	10.14	7.91	0.72	0.256	0.512	44.03	81.06	62.55	0.88	0.001	0.003
148,759	148,923		*smORF-80*	54	11.32	9.84	10.58	-0.27	0.613	0.797	7.41	7.22	7.31	-0.05	0.939	0.975
149,972	149,877	c	*smORF-81*	31	6.59	21.21	13.90	1.67	0.004	0.032	53.71	54.14	53.93	0.01	0.969	0.982
150,122	150,205		*smORF-82*	27	33.24	136.51	84.88	1.95	0.000	0.000	50.46	53.20	51.83	0.08	0.759	0.884
150,997	150,830	c	*smORF-83*	55	13.13	25.72	19.42	1.02	0.028	0.132	24.47	23.61	24.04	-0.05	0.905	0.954
151,132	151,188		*smORF-84*	18	18.10	44.24	31.17	1.27	0.010	0.058	26.09	20.20	23.15	-0.37	0.340	0.567
159,388	159,212	c	*smORF-85*	58	9.20	13.10	11.15	0.45	0.446	0.694	141.09	219.42	180.25	0.64	0.000	0.002
166,486	166,424	c	*smORF-86*	20	18.42	5.37	11.90	-1.72	0.002	0.020	3231.00	734.76	1982.88	-2.14	0.000	0.000
188,328	188,281	c	*smORF-87*	15	16.87	10.73	13.80	-0.63	0.230	0.479	152.12	51.17	101.64	-1.57	0.000	0.000
189,688	189,575	c	*smORF-88*	37	14.06	13.09	13.57	-0.25	0.631	0.799	13.44	8.86	11.15	-0.59	0.268	0.497
195,626	195,730		*smORF-89*	34	2.64	14.58	8.61	1.96	0.006	0.039	91.27	107.02	99.15	0.24	0.315	0.539
195,978	195,814	c	*smORF-90*	54	7.75	10.54	9.15	0.28	0.642	0.803	67.80	43.35	55.57	-0.64	0.015	0.053

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
196,726	196,643	c	*smORF-91*	27	36.18	31.40	33.79	-0.18	0.618	0.797	12.58	15.59	14.08	0.30	0.529	0.729
196,727	196,819		*smORF-92*	30	20.56	43.14	31.85	1.00	0.027	0.127	3.05	1.27	2.16	-1.29	0.341	NA
198,438	198,304	c	*smORF-93*	44	13.11	10.71	11.91	-0.43	0.446	0.694	25.31	21.35	23.33	-0.24	0.527	0.728
202,335	202,373		*smORF-94*	12	6.63	14.23	10.43	1.10	0.105	0.294	1.82	2.99	2.41	0.74	0.526	NA
204,996	204,853	c	*smORF-95*	47	7.71	16.19	11.95	0.88	0.117	0.316	141.85	138.04	139.95	-0.04	0.839	0.923
205,162	205,043	c	*smORF-96*	39	26.63	48.78	37.70	0.80	0.040	0.164	16.57	17.69	17.13	0.09	0.843	0.923
207,153	207,049	c	*smORF-97*	34	114.17	172.20	143.18	0.56	0.198	0.445	37.68	18.50	28.09	-1.01	0.009	0.034
208,327	208,280	c	*smORF-98*	15	12.44	122.38	67.41	2.98	0.000	0.000	12.57	2.95	7.76	-2.08	0.002	0.011
208,886	208,752	c	*smORF-99*	44	15.07	15.36	15.22	0.13	0.792	0.895	19.19	11.01	15.10	-0.79	0.089	0.234
209,245	209,355		*smORF-100*	36	14.56	17.60	16.08	0.17	0.718	0.848	73.97	46.28	60.12	-0.67	0.010	0.037
212,536	212,601		*smORF-101*	21	11.49	5.95	8.72	-0.74	0.216	0.462	26.34	44.93	35.64	0.77	0.016	0.056
215,385	215,215	c	*smORF-102*	56	5.77	47.57	26.67	2.58	0.000	0.001	11.71	26.60	19.15	1.17	0.008	0.032
217,940	217,821	c	*smORF-103*	39	3.13	50.65	26.89	3.40	0.000	0.000	19.33	11.81	15.57	-0.70	0.133	0.322
218,028	218,105		*smORF-104*	25	30.37	162.28	96.32	2.31	0.000	0.000	31.11	24.40	27.75	-0.36	0.309	0.535
220,079	219,936	c	*smORF-105*	47	53.71	11.93	32.82	-2.12	0.000	0.001	111.90	132.31	122.11	0.24	0.290	0.516
221,703	221,671	c	*smORF-106*	10	23.74	21.39	22.56	-0.21	0.604	0.795	0.83	3.43	2.13	1.98	0.154	NA
227,580	227,470	c	*smORF-107*	36	23.46	15.95	19.70	-0.56	0.183	0.433	51.62	31.59	41.61	-0.70	0.021	0.070
227,599	227,769		*smORF-108*	56	3.13	14.63	8.88	1.73	0.016	0.086	131.34	34.25	82.80	-1.94	0.000	0.000
239,136	239,056	c	*smORF-109*	26	36.50	42.51	39.50	0.16	0.689	0.830	115.56	64.12	89.84	-0.85	0.000	0.001
239,302	239,225	c	*smORF-110*	25	6.93	17.42	12.18	1.52	0.014	0.078	3.49	2.69	3.09	-0.38	0.695	0.846
241,354	241,452		*smORF-111*	32	175.70	1029.68	602.69	2.49	0.000	0.000	572.29	80.86	326.57	-2.83	0.000	0.000
241,479	241,592		*smORF-112*	37	0.54	25.53	13.04	4.14	0.000	0.000	628.58	79.61	354.09	-2.98	0.000	0.000
243,020	242,931	c	*smORF-113*	29	7.17	23.70	15.43	1.67	0.008	0.052	6.94	10.79	8.87	0.62	0.304	0.532
246,962	247,129		*smORF-114*	55	6.03	41.06	23.54	2.70	0.000	0.000	77.98	69.52	73.75	-0.17	0.482	0.691
250,010	250,162		*smORF-115*	50	53.15	50.19	51.67	-0.06	0.872	0.951	113.76	122.82	118.29	0.12	0.578	0.762
250,310	250,459		*smORF-116*	49	90.61	86.34	88.47	-0.07	0.845	0.934	254.95	138.85	196.90	-0.88	0.000	0.000
250,633	250,785		*smORF-117*	50	10.67	22.67	16.67	1.10	0.037	0.159	12.59	24.98	18.79	0.96	0.051	0.148
251,244	251,173	c	*smORF-118*	23	17.86	16.08	16.97	-0.15	0.731	0.856	88.75	72.28	80.52	-0.29	0.264	0.496
251,505	251,392	c	*smORF-119*	37	113.43	132.91	123.17	0.22	0.510	0.742	32.62	33.93	33.28	0.06	0.854	0.929
251,508	251,621		*smORF-120*	37	16.82	16.87	16.85	0.00	0.993	0.997	84.21	38.78	61.49	-1.11	0.000	0.000
255,795	255,715	c	*smORF-121*	26	4.25	11.55	7.90	1.07	0.099	0.284	9.26	6.88	8.07	-0.43	0.485	0.692

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
259,158	259,223		*smORF-122*	21	38.36	40.92	39.64	0.05	0.923	0.962	56.00	67.92	61.96	0.27	0.285	0.514
259,496	259,434	c	*smORF-123*	20	61.00	55.25	58.12	-0.18	0.601	0.794	12.42	21.91	17.17	0.83	0.066	0.182
261,585	261,505	c	*smORF-124*	26	10.02	18.18	14.10	0.58	0.267	0.527	11.61	11.31	11.46	-0.03	0.958	0.978
262,130	261,996	c	*smORF-125*	44	2.85	12.43	7.64	2.59	0.003	0.021	8.61	10.85	9.73	0.31	0.593	0.775
262,856	262,725	c	*smORF-126*	43	8.88	10.87	9.87	0.57	0.338	0.593	5.65	17.39	11.52	1.61	0.005	0.023
263,032	263,094		*smORF-127*	20	170.38	251.69	211.04	0.56	0.067	0.229	12.70	13.31	13.00	0.06	0.907	0.954
264,262	264,110	c	*smORF-128*	50	262.69	83.78	173.24	-1.63	0.000	0.000	171.56	133.88	152.72	-0.36	0.052	0.150
265,063	265,158		*smORF-129*	31	13.76	5.97	9.87	-1.18	0.029	0.135	16.81	28.96	22.88	0.79	0.048	0.141
265,269	265,382		*smORF-130*	37	31.75	10.39	21.07	-1.57	0.000	0.003	60.32	66.71	63.52	0.15	0.549	0.745
273,894	273,992		*smORF-131*	32	9.85	18.16	14.00	0.69	0.180	0.430	28.67	17.34	23.00	-0.72	0.055	0.155
274,115	274,062	c	*smORF-132*	17	5.66	14.68	10.17	1.22	0.039	0.162	7.94	5.51	6.72	-0.51	0.469	0.684
274,139	274,231		*smORF-133*	30	11.84	15.74	13.79	0.59	0.315	0.581	35.43	9.56	22.50	-1.90	0.000	0.000
274,527	274,408	c	*smORF-134*	39	21.62	39.49	30.55	0.79	0.046	0.178	81.42	49.85	65.64	-0.71	0.004	0.016
276,917	276,804	c	*smORF-135*	37	396.73	352.23	374.48	-0.18	0.543	0.764	233.77	105.84	169.81	-1.14	0.000	0.000
290,323	290,222	c	*smORF-136*	33	14.08	7.44	10.76	-0.84	0.169	0.414	1.35	0.99	1.17	-0.44	0.784	NA
292,269	292,358		*smORF-137*		148.03	74.20	111.11	-0.94	0.003	0.021	82.20	80.22	81.21	-0.04	0.862	0.933
292,448	292,513		*smORF-138*	21	47.06	36.20	41.63	-0.42	0.350	0.607	331.88	578.55	455.22	0.80	0.000	0.000
293,185	293,328		*smORF-139*	47	12.44	13.68	13.06	0.07	0.914	0.962	37.03	26.31	31.67	-0.50	0.132	0.322
293,906	293,944		*smORF-140*	12	396.77	276.20	336.48	-0.52	0.073	0.236	14.84	15.94	15.39	0.10	0.831	0.923
294,436	294,570		*smORF-141*	44	3415.49	2362.28	2888.88	-0.52	0.073	0.236	1768.63	2186.88	1977.75	0.31	0.011	0.040
296,668	296,808		*smORF-142*	46	617.64	2207.20	1412.42	1.81	0.000	0.000	3248.59	1817.76	2533.17	-0.84	0.000	0.000
298,934	299,089		*smORF-143*	51	12.46	27.63	20.05	0.95	0.057	0.201	87.24	82.51	84.87	-0.08	0.736	0.874
299,423	299,346	c	*smORF-144*	25	6.52	16.50	11.51	1.34	0.023	0.114	8.77	14.91	11.84	0.80	0.159	0.367
309,244	309,357		*smORF-145*	37	19.70	13.16	16.43	-0.48	0.300	0.568	19.93	11.86	15.90	-0.74	0.107	0.267
309,428	309,463		*smORF-146*	11	23.20	15.03	19.11	-0.55	0.219	0.464	15.80	9.46	12.63	-0.73	0.167	0.381
312,697	312,650	c	*smORF-147*	15	28.58	115.74	72.16	1.88	0.001	0.006	28.49	41.45	34.97	0.54	0.093	0.240
312,815	312,768	c	*smORF-148*	15	15.46	54.64	35.05	1.86	0.001	0.011	15.30	46.17	30.73	1.59	0.000	0.000
313,094	312,984	c	*smORF-149*	36	4.54	44.43	24.48	3.01	0.000	0.000	6.60	27.59	17.10	2.06	0.000	0.000
313,252	313,356		*smORF-150*	34	9.01	15.98	12.49	0.90	0.101	0.284	13.41	11.10	12.25	-0.26	0.604	0.786
313,854	313,738	c	*smORF-151*	38	11.90	10.91	11.40	-0.19	0.711	0.844	113.75	75.37	94.56	-0.59	0.012	0.044
314,165	314,016	c	*smORF-152*	49	15.01	24.61	19.81	0.58	0.201	0.445	118.68	13.36	66.02	-3.15	0.000	0.000

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
331,989	332,111		*smORF-153*	40	5.68	15.40	10.54	1.27	0.035	0.149	5.39	9.52	7.45	0.78	0.273	0.499
332,180	332,317		*smORF-154*	45	12.92	8.27	10.59	-0.44	0.453	0.699	1357.05	678.03	1017.54	-1.00	0.000	0.000
334,288	334,332		*smORF-155*	14	27.04	19.52	23.28	-0.48	0.234	0.485	148.98	95.91	122.44	-0.64	0.001	0.005
334,429	334,575		*smORF-156*	48	72.46	58.96	65.71	-0.31	0.383	0.640	249.06	230.94	240.00	-0.11	0.516	0.717
336,385	336,302	c	*smORF-157*	27	249.42	110.95	180.19	-1.10	0.071	0.236	4947.43	489.57	2718.50	-3.34	0.000	0.000
336,799	336,734	c	*smORF-158*	21	123.27	26.99	75.13	-2.09	0.000	0.001	15854.20	606.53	8230.36	-4.71	0.000	0.000
338,268	338,435		*smORF-159*	55	15.07	22.75	18.91	0.69	0.152	0.377	120.90	27.46	74.18	-2.15	0.000	0.000
338,808	338,894		*smORF-160*	28	3.56	10.74	7.15	1.51	0.034	0.147	7.37	3.94	5.65	-0.92	0.220	0.448
339,466	339,603		*smORF-161*	45	14.19	16.75	15.47	0.29	0.552	0.773	121.48	61.49	91.48	-0.98	0.000	0.000
339,854	339,949		*smORF-162*	31	42.83	33.07	37.95	-0.36	0.303	0.568	259.68	108.08	183.88	-1.27	0.000	0.000
340,210	340,160	c	*smORF-163*	16	15.08	10.52	12.80	-0.63	0.254	0.512	511.53	117.32	314.43	-2.12	0.000	0.000
340,475	340,320	c	*smORF-164*	51	174.62	61.86	118.24	-1.50	0.000	0.000	1647.12	509.76	1078.44	-1.69	0.000	0.000
340,647	340,582	c	*smORF-165*	21	71.77	38.42	55.09	-0.84	0.025	0.123	135.80	156.40	146.10	0.21	0.268	0.497
341,663	341,770		*smORF-166*	35	25.94	14.69	20.31	-0.78	0.077	0.240	5.43	2.46	3.95	-1.13	0.212	0.438
343,825	343,980		*smORF-167*	51	26.35	14.17	20.26	-0.89	0.101	0.284	33.43	39.64	36.53	0.23	0.486	0.692
344,483	344,542		*smORF-168*	19	70.17	48.11	59.14	-0.53	0.283	0.550	73.89	46.88	60.38	-0.65	0.009	0.034
344,771	344,896		*smORF-169*	41	25.08	70.69	47.88	1.47	0.008	0.048	75.39	55.24	65.32	-0.46	0.081	0.218
344,913	345,092		*smORF-170*	59	35.23	69.98	52.60	1.00	0.039	0.163	112.55	114.34	113.45	0.02	0.918	0.959
346,113	346,018	c	*smORF-171*	31	74.71	49.48	62.09	-0.61	0.081	0.251	265.32	226.07	245.69	-0.23	0.177	0.396
346,276	346,190	c	*smORF-172*	28	42.01	24.47	33.24	-0.62	0.198	0.445	201.05	189.33	195.19	-0.09	0.606	0.786
346,351	346,307	c	*smORF-173*	14	16.95	15.53	16.24	-0.12	0.824	0.922	18.02	14.34	16.18	-0.34	0.476	0.690
346,429	346,494		*smORF-174*	21	27.17	34.24	30.70	0.32	0.420	0.671	3.55	5.91	4.73	0.74	0.368	0.594
346,595	346,660		*smORF-175*	21	9.85	13.81	11.83	0.32	0.576	0.786	26.67	33.08	29.87	0.30	0.387	0.613
346,908	346,861	c	*smORF-176*	15	16.91	7.00	11.95	-1.26	0.015	0.082	4.07	4.90	4.49	0.29	0.726	0.868
351,249	351,326		*smORF-177*	25	14.58	6.59	10.59	-1.24	0.029	0.135	2.10	3.25	2.67	0.58	0.604	0.786
357,429	357,560		*smORF-178*	43	7.82	18.03	12.92	1.02	0.091	0.273	13.03	16.14	14.59	0.30	0.522	0.723
358,061	358,171		*smORF-179*	36	6.85	18.05	12.45	1.06	0.064	0.222	25.84	21.49	23.66	-0.27	0.481	0.691
368,557	368,384	c	*smORF-180*	57	58.14	24.47	41.31	-1.24	0.000	0.003	292.83	773.26	533.04	1.40	0.000	0.000
369,419	369,279	c	*smORF-181*	46	12.92	10.89	11.90	-0.32	0.526	0.750	40.41	39.36	39.89	-0.05	0.868	0.934
369,570	369,451	c	*smORF-182*	39	11.68	7.47	9.58	-0.58	0.287	0.554	83.49	61.04	72.27	-0.44	0.080	0.216

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
370,718	370,608	c	*smORF-183*	36	12.76	15.68	14.22	0.28	0.582	0.789	27.59	30.69	29.14	0.16	0.645	0.810
380,414	380,542		*smORF-184*	42	26.18	19.23	22.70	-0.45	0.298	0.568	372.33	494.49	433.41	0.41	0.003	0.015
381,984	381,868	c	*smORF-185*	38	73.13	44.47	58.80	-0.73	0.041	0.166	43.33	75.49	59.41	0.80	0.005	0.020
384,389	384,255	c	*smORF-186*	44	59.55	51.20	55.37	-0.17	0.632	0.799	116.06	163.30	139.68	0.50	0.015	0.053
384,540	384,379	c	*smORF-187*	53	86.26	59.63	72.95	-0.51	0.080	0.247	262.25	193.26	227.75	-0.44	0.010	0.037
389,930	389,844	c	*smORF-188*	28	14.51	9.33	11.92	-0.48	0.393	0.643	16.68	10.81	13.74	-0.65	0.196	0.419
392,354	392,235	c	*smORF-189*	39	28.44	18.97	23.71	-0.57	0.165	0.408	317.84	167.21	242.52	-0.92	0.000	0.000
392,606	392,532	c	*smORF-190*	24	25.53	25.48	25.51	0.01	0.986	0.997	46.75	22.65	34.70	-1.05	0.001	0.006
395,125	395,244		*smORF-191*	39	11.75	9.06	10.41	-0.33	0.567	0.781	1.35	1.95	1.65	0.55	0.690	NA
398,815	398,868		*smORF-192*	17	17.11	10.34	13.73	-0.76	0.218	0.464	3.12	4.88	4.00	0.66	0.462	0.680
398,926	399,102		*smORF-193*	58	940.81	857.31	899.06	-0.13	0.727	0.854	602.16	412.70	507.43	-0.54	0.000	0.001
399,375	399,298	c	*smORF-194*	25	30.26	20.12	25.19	-0.61	0.127	0.329	72.59	73.88	73.23	0.02	0.922	0.962
399,423	399,518		*smORF-195*	31	366.84	179.85	273.35	-1.01	0.001	0.005	1236.58	1029.95	1133.27	-0.26	0.037	0.115
407,808	407,710	c	*smORF-196*	32	10.93	26.39	18.66	1.07	0.026	0.127	10.52	11.53	11.02	0.13	0.812	0.911
408,130	407,987	c	*smORF-197*	47	14.92	21.79	18.36	0.49	0.302	0.568	4.10	3.45	3.78	-0.23	0.803	0.906
416,317	416,222	c	*smORF-198*	31	113.00	83.68	98.34	-0.43	0.194	0.445	1186.17	754.95	970.56	-0.65	0.000	0.000
417,172	417,107	c	*smORF-199*	21	33.91	34.30	34.11	-0.04	0.915	0.962	36.89	32.90	34.89	-0.17	0.581	0.762
417,761	417,808		*smORF-200*	15	2253.89	207.37	1230.63	-3.22	0.000	0.000	2196.55	91.60	1144.08	-4.59	0.000	0.000
418,962	418,997		*smORF-201*	11	1625.29	256.19	940.74	-2.56	0.000	0.000	7300.75	384.07	3842.41	-4.25	0.000	0.000
419,020	419,169		*smORF-202*	49	228.45	101.18	164.81	-1.15	0.004	0.027	38962.9	2880.81	20921.9	-3.76	0.000	0.000
419,342	419,497		*smORF-203*	51	1224.71	271.46	748.08	-2.09	0.000	0.000	3517.07	418.80	1967.93	-3.07	0.000	0.000
420,456	420,494		*smORF-204*	12	17.71	32.27	24.99	0.86	0.097	0.283	27.21	7.52	17.37	-1.85	0.000	0.000
423,217	423,125	c	*smORF-205*	30	14.97	8.64	11.80	-0.57	0.317	0.582	3.31	3.41	3.36	0.00	0.999	0.999
423,899	424,075		*smORF-206*	58	19.72	41.02	30.37	1.04	0.021	0.105	131.26	178.22	154.74	0.45	0.016	0.056
428,628	428,551	c	*smORF-207*	25	24.64	17.45	21.05	-0.39	0.469	0.715	16.43	27.61	22.02	0.75	0.065	0.180
428,765	428,667	c	*smORF-208*	32	28.38	27.29	27.83	0.00	0.997	0.997	58.39	96.66	77.52	0.74	0.002	0.010
430,298	430,164	c	*smORF-209*	44	39.20	49.60	44.40	0.42	0.291	0.560	58.92	70.30	64.61	0.26	0.304	0.532
430,687	430,550	c	*smORF-210*	45	39.16	50.67	44.92	0.32	0.361	0.614	52.72	64.21	58.47	0.29	0.262	0.494
431,031	430,942	c	*smORF-211*	29	48.57	60.47	54.52	0.36	0.269	0.527	171.73	94.95	133.34	-0.85	0.000	0.000
431,152	431,087	c	*smORF-212*	21	42.27	100.49	71.38	1.22	0.009	0.052	3.11	2.41	2.76	-0.35	0.741	0.876
432,332	432,466		*smORF-213*	44	11.36	14.94	13.15	0.32	0.526	0.750	12.71	11.77	12.24	-0.11	0.828	0.922

Appendix

smORFs					wt vs Δ*ngrl*Δ*tral*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrl*Δ*tral* (transcriptom Göttingen, this work)					
start	end	Orien-tation *	smORF- No	aa	WT	Δ*ngrl* Δ*tral* _copy+	baseMean	log2Fold Change	pvalue	padj	WT	Δ*ngrl*Δ*tral*	baseMean	log2Fold Change	pvalue	padj
433,016	433,066		*smORF-214*	16	12.12	10.98	11.55	-0.13	0.822	0.921	1.33	1.73	1.53	0.38	0.785	NA
434,297	434,470		*smORF-215*	57	35.16	71.78	53.47	0.96	0.051	0.184	11.67	33.62	22.64	1.51	0.000	0.002
442,070	442,210		*smORF-216*	46	23.91	36.54	30.23	0.57	0.205	0.446	61.66	79.53	70.59	0.37	0.128	0.313
444,353	444,397		*smORF-217*	14	5.55	32.86	19.21	2.57	0.000	0.000	73.07	145.79	109.43	0.99	0.000	0.000
449,691	449,524	c	*smORF-218*	55	5.33	13.96	9.65	1.04	0.104	0.293	7.28	15.11	11.19	1.03	0.071	0.194
449,809	449,720	c	*smORF-219*	29	15.42	37.99	26.70	1.45	0.001	0.012	67.15	45.45	56.30	-0.56	0.041	0.123
451,289	451,188	c	*smORF-220*	33	7.86	23.02	15.44	1.24	0.048	0.181	7.47	10.57	9.02	0.51	0.414	0.635
451,880	451,749	c	*smORF-221*	43	5.33	22.07	13.70	1.67	0.005	0.035	24.99	37.19	31.09	0.56	0.122	0.302
451,955	451,893	c	*smORF-222*	20	6.20	11.01	8.60	0.70	0.245	0.501	2.19	1.42	1.81	-0.60	0.641	NA
452,298	452,399		*smORF-223*	33	3.69	10.62	7.16	1.12	0.099	0.284	12.63	23.33	17.98	0.88	0.047	0.139
452,960	453,109		*smORF-224*	49	17.95	24.01	20.98	0.35	0.465	0.713	28.84	54.43	41.63	0.91	0.003	0.013
461,020	461,148		*smORF-225*	42	5.72	13.03	9.38	1.02	0.137	0.352	26.74	20.56	23.65	-0.38	0.305	0.532
462,970	462,869	c	*smORF-226*	33	26.11	47.24	36.68	0.79	0.068	0.229	59.71	107.82	83.76	0.86	0.000	0.001
463,923	464,057		*smORF-227*	44	17.06	13.77	15.42	-0.34	0.510	0.742	4.01	4.36	4.19	0.14	0.873	0.934
465,605	465,706		*smORF-228*	33	13.41	9.93	11.67	-0.26	0.635	0.800	15.19	9.52	12.36	-0.66	0.190	0.410
478,801	478,920		*smORF-229*	39	6.78	10.67	8.73	0.52	0.400	0.650	23.12	18.38	20.75	-0.32	0.413	0.635
479,236	479,292		*smORF-230*	18	17.47	14.39	15.93	-0.33	0.462	0.711	0.86	2.19	1.52	1.33	0.398	NA
479,994	479,827	c	*smORF-231*	55	9.76	28.26	19.01	1.43	0.005	0.034	72.89	76.00	74.45	0.07	0.786	0.896
483,313	483,405		*smORF-232*	30	73.35	40.20	56.77	-0.82	0.045	0.175	12.76	16.28	14.52	0.36	0.452	0.669
484,053	483,973	c	*smORF-233*	26	98.25	58.23	78.24	-0.72	0.069	0.230	782.29	581.81	682.05	-0.43	0.004	0.015
490,485	490,565		*smORF-234*	26	11.38	9.75	10.57	-0.26	0.657	0.814	73.05	72.03	72.54	-0.02	0.946	0.975
491,239	491,340		*smORF-235*	33	27.36	22.14	24.75	-0.34	0.425	0.672	100.72	99.46	100.09	-0.01	0.952	0.978
491,418	491,594		*smORF-236*	58	34.36	32.59	33.48	-0.06	0.886	0.951	117.85	102.21	110.03	-0.21	0.345	0.574
492,093	492,046	c	*smORF-237*	15	15.53	20.17	17.85	0.25	0.573	0.785	7.72	10.90	9.31	0.53	0.381	0.612
492,447	492,328	c	*smORF-238*	39	9.78	21.68	15.73	1.06	0.042	0.169	165.78	133.58	149.68	-0.31	0.103	0.260
495,593	495,685		*smORF-239*	30	36.48	76.85	56.66	1.08	0.006	0.038	7.01	7.93	7.47	0.18	0.781	0.896
497,767	497,862		*smORF-240*	31	11.47	13.24	12.35	0.05	0.926	0.963	18.24	42.89	30.56	1.25	0.001	0.003
498,200	498,120	c	*smORF-241*	26	0.54	12.25	6.39	3.18	0.001	0.012	14.03	20.77	17.40	0.56	0.200	0.422
499,735	499,652	c	*smORF-242*	27	37.19	39.37	38.28	0.11	0.792	0.895	160.18	113.45	136.82	-0.50	0.013	0.046
502,837	502,685	c	*smORF-243*	50	10.84	7.87	9.36	-0.17	0.790	0.895	8.27	7.31	7.79	-0.18	0.774	0.890
503,061	503,129		*smORF-244*	22	169.41	152.37	160.89	-0.14	0.587	0.789	752.58	444.07	598.32	-0.76	0.000	0.000

Appendix

smORFs					wt vs Δ*ngrI*Δ*traI*_copy+ (transcriptom würzburg, Grote et al. 2016)						wt vs Δ*ngrI*Δ*traI* (transcriptom Göttingen, this work)					
start	**end**	**Orien-tation ***	**smORF- No**	**aa**	**WT**	**Δ*ngrI* Δ*traI* _copy+**	**baseMean**	**log2Fold Change**	**pvalue**	**padj**	**WT**	**Δ*ngrI*Δ*traI***	**baseMean**	**log2Fold Change**	**pvalue**	**padj**
508,248	508,159	c	*smORF-245*	29	30.06	29.11	29.59	-0.02	0.963	0.987	54.81	58.35	56.58	0.08	0.745	0.877
523,516	523,439	c	*smORF-246*	25	112.31	94.45	103.38	-0.26	0.389	0.643	283.83	192.04	237.93	-0.56	0.000	0.003
525,154	525,092	c	*smORF-247*	20	24.43	25.51	24.97	0.07	0.861	0.946	9.54	12.46	11.00	0.41	0.463	0.680
525,172	525,237		*smORF-248*	21	18.88	13.55	16.21	-0.49	0.279	0.544	74.89	56.35	65.62	-0.41	0.089	0.234
535,461	535,616		*smORF-249*	51	8.92	799.15	404.03	6.50	0.000	0.000	44.71	19.45	32.08	-1.20	0.000	0.002
535,637	535,810		*smORF-250*	57	3.20	1298.75	650.97	7.86	0.000	0.000	6.40	2.87	4.64	-1.12	0.209	0.433
535,838	104		*smORF-251*	143	4.08	555.27	279.68	6.66	0.000	0.000	58.03	28.88	43.45	-1.00	0.002	0.010

*orientation: C indicates that the ORF is complement and is located on the antisense strang

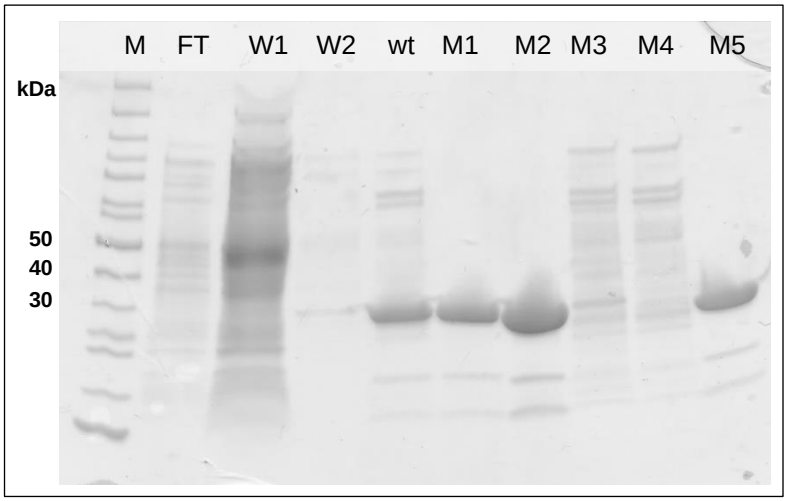

Figure 57: SDS PAGE of GqqA wt and the constructed GqqA mutants

12 % Mini-PROTEAN® TBE Precast Gel was loaded with protein solution of GqqA protein purification. The steps flow through (FT), washing step 1 (W1) and 2 (W2) and the pooled elution step of GqqA wt (wt) were loaded on the gel. Only the pooled elution step of the five mutants was used for SDS PAGE (M1- M5). The PageRuler Unstained protein ladder was used as protein standard (BIO-RAD)