

- revised version -

**A novel function-based screen for detecting RubisCO
active clones from metagenomic libraries:
elucidating the role of RubisCO associated enzymes.**

Dissertation

with the aim of achieving the degree of
Doctor rerum naturalium (Dr. rer. nat.)

at the Department of Biology
Subdivision at the Faculty of Mathematics, Informatics and Natural Sciences
of the Universität Hamburg

submitted by
Stefanie Böhnke
from Neubrandenburg

Hamburg 2014

Genehmigt vom

Fachbereich Biologie der Universität Hamburg

auf Antrag von Frau Jun. Prof. Dr. Mirjam Perner

Weiterer Gutachter der Dissertation: Herr Prof. Dr. Wolfgang Streit

Tag der Disputation: 14.11.2014

Jun. Prof. Dr. Mirjam Perner

Molecular Biology of Microbial Consortia

Microbiology and Biotechnology

Department of Biology

Ohnhorststr. 18

D-22609 Hamburg

Tel. +49 (0)40 42816 -444

Fax +49 (0)40 42816 -459

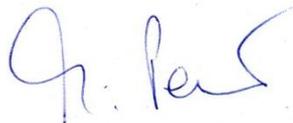
mirjam.perner@uni-hamburg.de

—
14.08.2014

English language declaration

—
As a native English speaker, I hereby declare that I have checked the thesis "A novel function-based screen for detecting RubisCO active clones from metagenomic libraries: elucidating the role of RubisCO associated enzymes." by Stefanie Böhnke for grammatically correct English and the scientific accuracy of the language.

Sincerely,



Mirjam Perner

Publications

The results that have arisen from this study were submitted, at the time of print, as follows:

Böhnke, S. & Perner, M. (2014) Transposon mutagenesis of a hydrothermal vent metagenomic fragment reveals cues for RubisCO regulation and activation, *Environmental Microbiology*, in preparation

Böhnke, S. & Perner, M. (2014) A function-based screen for seeking RubisCO active clones from metagenomes: novel enzymes influencing RubisCO activity, *The ISME Journal*, accepted.

Rabausch, U., Jürgensen, J., Ilmberger, N., **Böhnke, S.**, Fischer, S., Schubach, B., Schulte, M.M. & Streit, W.R. (2013) Functional screening of metagenome and genome libraries for detection of novel flavonoid modifying enzymes, *Applied and Environmental Microbiology* 79 (15) 4551-4563

Perner, M., Gonnella, G., Hourdez, S., **Böhnke, S.**, Kurtz, S. & Girguis, P. (2013) *In-situ* chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field, *Environmental Microbiology* 15 (5) 1551-1560

Table of Contents

Table of Contents	I
List of Figures	III
List of main Tables	IV
List of appended Tables	V
Abstract	VI
Zusammenfassung	VII
1 Introduction	1
1.1 Autotrophic carbon fixation	1
1.2 Autotrophic carbon fixation at hydrothermal deep-sea vent habitats.....	3
1.3 The Calvin Benson cycle	5
1.4 RubisCO and RubisCO associated enzymes	7
1.5 Metagenomics.....	12
1.6 Intention of this work	13
2 Material and Methods	14
2.1 Bacterial strains and respective cultivation techniques	14
2.1.1 Bacterial strains.....	14
2.1.2 Cultivating <i>T. crunogena</i> TH-55.....	14
2.1.3 Cultivating different <i>E. coli</i> strains.....	14
2.2 Vectors and constructs.....	15
2.3 Sample collection	17
2.4 (Meta)-genomic fosmid libraries	18
2.4.1 Meta-(genomic) DNA isolation.....	18
2.4.2 Multiple displacement amplification (MDA)	19
2.4.3 Construction of (meta)-genomic libraries	19
2.5 Establishing a functional screen to seek recombinant RubisCOs from metagenomes	20
2.5.1 <i>T. crunogena</i> TH-55's RubisCO activity.....	20
2.5.2 <i>T. crunogena</i> TH-55's RubisCO recombinantly expressed in <i>E. coli</i>	21
2.5.3 RubisCO activities from a TH-55 genomic fosmid clone	22
2.6 Seeking RubisCOs from metagenomic libraries	24
2.7 Working with RubisCOs from metagenomes	25
2.7.1 Primer walk and sequence editing.....	26
2.7.2 Subcloning of RubisCO gene clusters from metagenomic fosmid inserts.....	26

2.7.3	Transposon mutagenesis	27
2.7.4	Transcription experiments with TH-55	27
2.7.5	Experiments pairing transcription with RubisCO activity	28
2.7.6	Complementation experiments	29
3	Results	30
3.1	(Meta)-genomic fosmid libraries	30
3.2	Establishing a functional screen to seek RubisCOs from metagenomes	32
3.2.1	RubisCO activity of TH-55 cells and recombinant RubisCO versions.....	32
3.2.2	Up-scaling	35
3.3	Seeking RubisCOs from metagenomic libraries	37
3.4	Working with RubisCOs from metagenomes	41
3.4.1	Transposon insertion libraries.....	41
3.4.2	Transcription experiments with TH-55	47
3.4.3	Experiments pairing transcription with RubisCO activity	48
3.4.4	Complementation experiments	51
4	Discussion	52
4.1	Evaluating the newly established RubisCO screen by comparing hit rates of investigated metagenomic libraries	53
4.2	Investigating the novel recombinant RubisCO from the metagenome of 'Drachenschlund'	55
4.2.1	Subcloning of the RubisCO gene cluster unexpectedly causes a dramatic loss of RubisCO activity	55
4.2.2	The consequence of transposon insertions outside of the RubisCO gene cluster relative to RubisCO activity	56
4.2.3	The consequence of transposon insertions within the RubisCO gene cluster relative to the RubisCO activity	59
5	Conclusion.....	64
	References.....	65
	Appendix A: Abbreviations and accession numbers	71
	Appendix B: Primers used in this study	74
	Appendix C: Measured RubisCO activities	79

List of Figures

Figure 1: Heterotrophic life at hydrothermally influenced habitats.	3
Figure 2: Light conditions within the water column.	4
Figure 3: Autotrophic carbon fixation at hydrothermally influenced deep-sea habitats.	5
Figure 4: The schematic representation of the Calvin Benson cycle including the oxygen site reaction of RubisCO.	6
Figure 5: Phylogenetic relationship of <i>cbbL</i> and <i>cbbM</i> structural genes.	8
Figure 6: Comparison of RubisCO gene cluster arrangements.	9
Figure 7: Location map of sampled hydrothermal sites.	17
Figure 8: Treatment of chimney material prior to DNA isolation.	18
Figure 9: Schematic view of major points successively processed to verify that the function-based screen works on a single scale.	20
Figure 10: Overview of the experimental setup to upscale the function-based screen and apply it to the metagenomic scale.	24
Figure 11: Illustration of conducted experiments to characterize the metagenome derived RubisCOs and flanking gene regions.	25
Figure 12: Classification of sequences derived from insert end-sequencing of selected fosmids from the four metagenomic libraries constructed within this study.	31
Figure 13: Specific RubisCO activity of TH-55 dependent on growth.	32
Figure 14: Specific RubisCO activities of TH-55 native and recombinant TH-55 RubisCOs visualized together with gene arrangements of recombinant versions.	34
Figure 15: HPLC analyses of differently sized metagenomic pools.	36
Figure 16: Specific RubisCO activities of TH 55's native and recombinant RubisCOs compared with metagenome derived recombinant RubisCOs as well as respective gene arrangements.	38
Figure 17: Gene arrangement of ORFs encoded on the metagenomic fragment.	40
Figure 18: Specific RubisCO activities and insertion positions of tested transposon clones.	42
Figure 19: Phylogenetic relationship of <i>orf06</i>	43
Figure 20: Specific RubisCO activity of transposon clone 7II ($\Delta cbbO-m$ to <i>lysR2</i>) and respective gene arrangement.	47
Figure 21: Transcription of RubisCO encoding genes and flanking gene regions from TH-55.	47
Figure 22: (Co)-transcription of the RubisCO gene cluster.	48
Figure 23: Transcription of <i>cbbL</i> and <i>cbbM</i> in transposon clones with deletions in <i>orf06</i> , <i>lysR1</i> , <i>lysR2</i> and <i>cbbO-m</i>	49
Figure 24: Transcription of <i>cbbL</i> and <i>cbbM</i> in transposon clone 7II.	50
Figure 25: Complementation experiments.	51

List of main Tables

Table 1: Properties of the six currently known autotrophic carbon fixation pathways.	2
Table 2: Properties of different RubisCO forms.	7
Table 3: Bacterial strains used in this study.	14
Table 4: Antibiotics and supplements used in this study.	15
Table 5: Vectors used in this study.	15
Table 6: Constructs created in this study.	16
Table 7: Concentrations and purities of metagenomic DNA before and after DNA.	30
Table 8: Meta-(genomic) libraries constructed within this study.	30
Table 9: Comparison of specific RubisCO activities from different publications.	33
Table 10: Numbers of screened clones and corresponding numbers of positive tested ones.	37
Table 11: Open reading frames (ORFs) identified on the metagenomic fragment.	39
Table 12: Insertion positions of tested transposon clones with 35.2 kb inserts and respective RubisCO activities.	44
Table 13: Insertion positions of tested transposon clones with 13 kb inserts and respective RubisCO activities.	46

List of appended Tables

Appendix A: Abbreviations and Accession numbers

Appendix Table A 1: Abbreviations and GenBank accession numbers of strains used for Figure 5 and Figure 6.	71
Appendix Table A 2: Gene abbreviations used in this study.	72
Appendix Table A 3: Abbreviations and GenBank accession numbers of RubisCO encoding genes used for Figure 19.	73

Appendix B: Primers used in this study

Appendix Table B 1: Primers used for primer walking.	74
Appendix Table B 2: Primers used for cloning and validation procedure.	76
Appendix Table B 3: Primer pairs used to analyzed (co)-transcription.	77
Appendix Table B 4: Primer pairs used to analyze relative transcript abundance of RubisCO form I and II.	78
Appendix Table B 5: Primer pairs used for complementation experiments.	78

Appendix C: Measured RubisCO activities

Appendix Table C 1: Specific RubisCO activities of TH-55 and recombinant RubisCOs expressed in <i>E. coli</i>	79
---	----

Abstract

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO; EC 4.1.1.39) catalyzes the key and primary carbon fixation reaction of the Calvin Benson cycle (CB cycle). This unique enzyme is present in all plants, cyanobacteria and most autotrophic bacteria (phototrophs and chemolithoautotrophs) and is involved in the assimilation of most of the inorganic carbon fixed by all primary producers on Earth (>99.5% of 105×10^9 tons/year). Most studies that have enhanced our understanding of RubisCO functioning have been conducted with plants or cultured bacteria. The functionality of environmental RubisCOs as well as the importance of gene products encoded on flanking DNA regions is however still enigmatic. Currently no sequence independent approach is available, that enables seeking of RubisCOs directly from the environment by functionality alone. Therefore, the vast majority of functional RubisCOs and RubisCO associated genes from uncultured organisms (>99%) remains inaccessible. This study describes a novel, function-based approach suited to seek RubisCO active enzymes directly from metagenomic libraries. Within this study twelve environmental, recombinant RubisCOs were identified through this screen and resembled genes from *Thiomicrospira crunogena*. The 35.2 kb comprising metagenomic fragments consist of a RubisCO gene cluster and flanking DNA regions. The relevance of potential RubisCO associated genes for expressing a fully functional RubisCO was further investigated for one clone exemplarily, by making single genes inoperative due to transposon insertions. This approach uncovered one gene (*orf06*) whose gene product has never been associated with RubisCO activity before, but is directly or indirectly involved in positively regulating the transcription of *cbbM* and *cbbL*. Significantly changed RubisCO activities were furthermore found for clones with insertions in eleven other genes, whose gene products were assigned to functions of putative transcriptional regulators or those believed to be vital for RubisCO activation. This screen opens the door to detect up until now unexplored RubisCOs from the otherwise inaccessible uncultured majority and enables us to better understand the functioning of prokaryotic RubisCOs.

Zusammenfassung

Das Enzym Ribulose-1,5-bisphosphat Carboxylase/Oxygenase (RubisCO; EC 4.1.1.39) katalysiert die Schlüsselreaktion des Calvin Benson (CB) Zyklus, eine Reaktion bei dem erstmalig im Ablauf des Zyklus Kohlenstoff fixiert wird. Dieses einzigartige Enzym kommt in allen Pflanzen, Cyanobakterien und vielen autotrophen Bakterien vor (phototrophe und chemotrophe Prokaryoten) und fast der gesamte, weltweit durch Primärproduzenten assimilierte anorganische Kohlenstoff kann auf die Aktivität der RubisCO zurückgeführt werden (>99.5% von insgesamt 105×10^9 Tonnen pro Jahr). Die meisten Studien die das Wissen um die Funktionsweise der RubisCO erweitert haben, wurden mit Pflanzen oder kultivierbaren Mikroorganismen durchgeführt. Die Funktionalität von RubisCO Enzymen aus der Umwelt sowie die Bedeutung flankierender Genregionen ist jedoch noch immer ungeklärt. Derzeit gibt es keinen funktionsbasierten Ansatz, der es ermöglicht RubisCO aktive Enzyme direkt aus der Umwelt zu isolieren. Die Mehrheit funktionsfähiger RubisCOs und RubisCO assoziierter Gene unkultivierter Mikroorganismen (>99%) bleibt folglich unzugänglich. In dieser Studie wird erstmals ein funktionsbasierter Ansatz vorgestellt, der es ermöglicht Metagenombanken nach Klonen mit RubisCO-Aktivität zu durchsuchen und dementsprechend RubisCOs direkt in Umweltproben zu detektieren. Unter Anwendung des neu etablierten Durchmusterungsverfahrens wurden insgesamt zwölf RubisCO aktive Klone identifiziert, die sequenzielle Übereinstimmungen zu *Thiomicrospira crunogena* aufwiesen. Die 35.2 kb umfassenden metagenomischen DNA Fragmente bestehen aus einem RubisCO Gen Cluster und flankierender DNA. Des Weiteren wurde die Bedeutung der auf dem metagenomischen DNA Fragment kodierten Genen in Bezug auf die Expression einer voll funktionsfähigen RubisCO analysiert, indem einzelne Gene durch Transposon Insertionen ausgeschaltet wurden. Hier wurde unter anderem ein Gen detektiert (*orf06*), das für ein hypothetisches Protein kodiert das niemals zuvor mit der Funktionsfähigkeit des RubisCO Enzyms in Verbindung gebracht wurde, jedoch direkt oder indirekt die Transkription sowohl von *cbbL* als auch von *cbbM* aktivierend reguliert. Signifikant veränderte RubisCO Aktivitäten wurden darüber hinaus für elf weitere ausgeschaltene Gene gemessen, die entweder für Proteine kodieren die als transkriptionelle Regulatoren annotiert sind oder für Proteine von den angenommen wird, dass sie an der RubisCO Aktivierung beteiligt sind. Mit diesem Screening Verfahren ist es nun möglich RubisCOs von bisher unkultivierten Mikroorganismen zu detektieren und deren Funktionsweise zu analysieren.

1 Introduction

1.1 Autotrophic carbon fixation

The ground for all heterotrophic life on earth is provided by autotrophic organisms, which have the facility to synthesize primary biomass solely from inorganic carbon compounds. Annually 7×10^{16} grams of carbon are fixed by terrestrial and aquatic autotrophs, whereby 2.8×10^{18} kilo joule of energy are conserved (Berg 2011). Thus, autotrophic carbon fixation is an essential process of the global carbon cycle, representing the only connection between the inorganic and the organic world. Next to plants and algae various microorganisms have the ability to fix carbon autotrophically (Hügler and Sievert 2011). However, the spectra of utilizable carbon fixation pathways for autotrophic microorganisms are, as currently known, broader because they can make use of six different pathways while eukaryotes are limited to one cycle, namely the Calvin Benson (CB) cycle (Minic and Thongbam 2011). The other five so far known alternative carbon fixation pathways additionally used by autotrophic microorganisms are: the reverse tricarboxylic acid (rTCA) cycle, the reductive acetyl-coenzyme A (rACA) pathway, the 3-hydroxypropionate (3-HP) bicycle, the 3-hydroxypropionate / 4-hydroxybutyrate (3-HP/4-HB) cycle and the dicarboxylate / 4-hydroxybutyrate (DC/4-HB) cycle (see Table 1). Due to phylogenetic relationships it has been proposed that the CB cycle evolved quite late in evolution of the bacterial branch while the rTCA cycle and the rACA pathway developed much earlier (Pereto *et al.* 1999). Moreover it has been suggested that the rACA pathway may be closer to the ancestral CO₂ fixation route, which is reasoned by several unique characteristics like e.g. (i) the marginal amounts of needed energy, (ii) the capability to utilize CO, a highly reduced common volcanic gas or (iii) the excessive utilization of coenzymes, metals (Fe, Co, Ni, Mo or W) and Fe-S centers (Fuchs 2011). The six autotrophic CO₂ fixation pathways furthermore differ with respect to the energy request, which means the number of ATP equivalents required to synthesize 1 pyruvate (Bar-Even *et al.* 2011). The temperature requirements represent another important characteristic (Hügler and Sievert 2011) just as the oxygen sensitivity of correlated enzymes (Fuchs 2011) (for details on properties of each pathway see Table 1). Based on these different characteristics it has been recapitulated that prevailing environmental conditions strongly influence which autotrophic pathway is predominantly represented in a habitat like e.g. in hydrothermally influenced environments (Berg *et al.* 2010b).

Table 1: Properties of the six currently known autotrophic carbon fixation pathways (modified from Berg *et al.* 2010b).

pathway	ATP equivalents*	distribution	CO ₂ -fixing enzymes [oxygen tolerance]	CO ₂ species**	key enzymes [oxygen tolerances]
Calvin Benson (CB) cycle	7	plants, algae and aerobic or facultative aerobic bacteria; genes encoding for RubisCO and/or Phosphoribulokinase are also present in some archaea	RubisCO [+]	CO ₂	RubisCO [+] Phosphoribulokinase [+]
reverse tricarboxylic acid (rTCA) cycle	2	microaerophil or obligate anaerobic bacteria	2-Oxoglutarate synthase [-] ¹ Isocitrate dehydrogenase [+] Pyruvate synthase [-] ² PEP carboxylase [+]	CO ₂ CO ₂ CO ₂ HCO ₃ ⁻	2-Oxoglutarate synthase [-] ATP-citrate lyase [+]
reductive acetyl coenzyme A (rACA) pathway	<1	obligate anaerobic Bacteria & archaea	CO-dehydrogenase-acetyl-CoA synthase [-] Formylmethanofuran dehydrogenase (in methanogenes) [+] Pyruvate synthase [-]	CO ₂ CO ₂ CO ₂	CO-dehydrogenase-acetyl-CoA synthase [-] enzymes reducing CO ₂ to methyltetrahydrofolate [+/-]
3-hydroxypropionate (3-HP) bicycle	7	aerobic green non-sulfur bacteria	Acetyl-CoA–propionyl CoA carboxylase [+]	HCO ₃ ⁻	Malonyl-CoA-reductase [+] Propionyl-CoA synthase [+] Malyl CoA lyase [+]
3-hydroxypropionate–4-hydroxybutyrate (3-HP/4-HB) cycle	9	(micro)-aerobic <i>Sulfolobales</i> ; genes encoding for characteristic enzymes are also present in the aerobic “marine group I”	Acetyl-CoA–propionyl CoA carboxylase [+]	HCO ₃ ⁻	Acetyl-CoA–propionyl CoA carboxylase [+] enzymes reducing malonyl-CoA to propionyl-CoA [+] Methylmalonyl-CoA mutase [+] 4-Hydroxybutyryl-CoA dehydratase [+]
dicarboxylate–4-hydroxybutyrate (DC/4-HB) cycle	5	mostly present in anaerobic autotrophic representatives of <i>Thermoproteales</i> and <i>Desulfurococcales</i>	Pyruvate synthase [-] PEP carboxylase [-]	CO ₂ HCO ₃ ⁻	4-Hydroxybutyryl-CoA dehydratase [+] ³

* ATP equivalents needed for the synthesis of one pyruvate; RubisCO – Ribulose-1,5-bisphosphate carboxylase/oxygenase; ** refers to the CO₂ fixing enzymes and describes the active CO₂ species which is incorporated; ATP – adenosine-5'-triphosphate; ¹with the exception of one of the two 2-Oxoglutarate synthases of *Hydrogenobacter thermophilus* which seems to be relatively oxygen stable (Yamamoto *et al.* 2006); ²Pyruvate synthase of the strictly anaerobic bacterium *Desulfovibrio africanus* is an exception because it is highly stable against oxygen (Pieulle *et al.* 1997); ³4-Hydroxybutyryl-CoA dehydratase is inactivated by oxygen in *Clostridia* (Scherf *et al.* 1994) but it has been suggested that it may be sufficiently stable in *Crenarchaeota* at low oxygen concentrations to maintain active (Berg *et al.* 2010a). [+]
oxygen tolerant enzymes; [-]
oxygen intolerant enzymes.

1.2 Autotrophic carbon fixation at hydrothermal deep-sea vent habitats

Hydrothermal deep-sea systems form when high temperature, reduced fluids ascend from inner earth and mix with the ambient oxygenated, cold sea water (Perner *et al.* 2013a). They represent oasis of life at the otherwise inhabitable, nutrient scarce deep seafloor. Hardly any primary produced organic matter from the surface of the oceans arrives at hydrothermal deep-sea habitats (Dick *et al.* 2013). However, a flourishing heterotrophic community consisting of shrimps, mussels or tubeworms, evolved (see Figure 1).

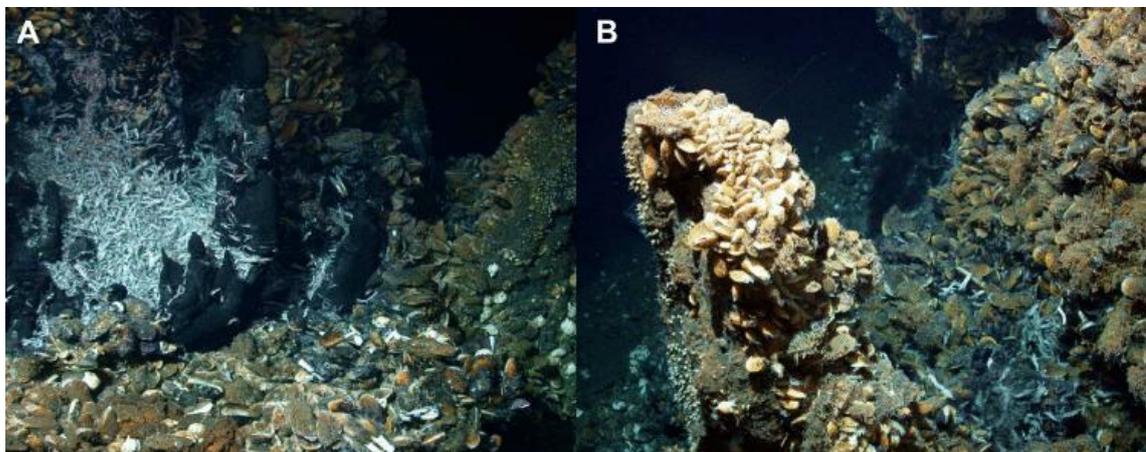


Figure 1: Heterotrophic life at hydrothermally influenced habitats. Two pictures of the Irina II chimney complex located in the Logatchev hydrothermal vent field are shown: A) the shrimp gab (dense shrimps-aggregations of *Rimicaris cf. exoculatus*) and parts of the northern chimney; B) the northern chimney, where one half is covered by *Bathymodiolus puteoserpentis*, while the other half is overcast by clutches of gastropods. Photos © ROV QUEST Bremen, MARUM.

Thus, primary biomass needs to be produced directly on site. Since CO_2 concentrations of emanating hydrothermal fluids can be immensely high (up to 215 nmol per kg, Kelley *et al.* 2002) the ground for diverse microorganisms associated with autotrophic growth is provided. However, to fix this CO_2 autotrophic microorganisms need energy and since no visible light reaches the deep-sea the local primary production cannot be based on sunlight (Kelley *et al.* 2002 and references therein) (see Figure 2). Indeed chemolithotrophic microbes can yield energy by oxidizing reduced solutes such as hydrogen or sulfide (hydrogen or sulfide oxidation), which are readily available as part of emanating hydrothermal fluids (see Figure 3) (Perner *et al.* 2013b).

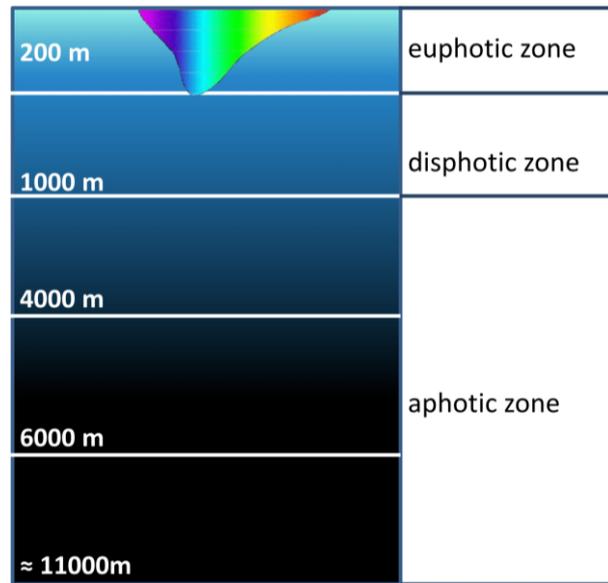


Figure 2: Light conditions within the water column. The three zones of the ocean are depicted with: (i) the upper 200 m called euphotic or “sunlight” zone, where sunlight is able to enter the water, (ii) the disphotic zone, also known as “twilight” zone where the intensity of light decreases as depth increases and (iii) the aphotic or “midnight” zone starting at 1000 m where visible light does not penetrate, and it is pitch dark.

Hydrothermal biotopes can provide a wide range of various thermal and chemical conditions. Thus, hydrothermal fluids can be highly enriched e.g. in sulfide or hydrogen, if circulating through basaltic rocks or ultramafic-hosted systems, respectively (Fouquet *et al.* 2010). It has been calculated, that the primary biomass production is driven by chemosynthetic energy gained from sulfide oxidation in sulfide-rich vents and from hydrogen oxidation in hydrogen-rich systems (McCollom and Shock 1997, Mccollom 2007). Furthermore it has been suggested that theoretically more primary biomass can be produced by hydrogen oxidizers compared to sulfide-oxidizers (Heijnen and Vandijken 1992) which might be reasoned by the fact that hydrogen-rich ultramafic systems can supply more chemical energy than basalt-hosted sulfide-rich systems (McCollom and Shock 1997, Mccollom 2007). This however might be reflected in the distribution of key carbon fixation pathway enzymes, where e.g. key enzymes of pathways with high energy requests are predominantly present in hydrogen-rich ultramafic systems while key enzymes of pathways with low energy requests are more represented in sulfide-rich basalt-hosted vents (for details on ATP requirements of each autotrophic carbon fixation pathway see Table 1). In addition to the energy demands the prevailing temperature as well as the concentration of oxygen differ significantly, which results from the mixing of the emanating hot, highly reduced hydrothermal fluids with the cold, ambient seawater (Perner *et al.* 2013a).

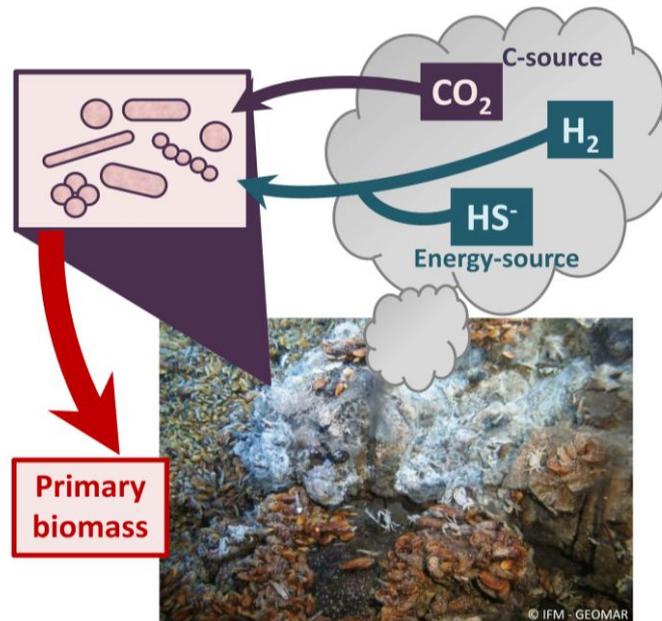


Figure 3: Autotrophic carbon fixation at hydrothermally influenced deep-sea habitats.

Here steep thermal and chemical gradients develop and form multiple microhabitats (Perner *et al.* 2009). Since various enzymes involved in proceeding autotrophic carbon fixation pathways are restricted by temperature maxima (Hügler and Sievert 2011) one may conclude that the occurrence of key carbon fixation enzymes may reflect the thermal limitations of the investigated vent habitat. Same argumentation may hold true for the prevailing oxygen concentration, since key enzymes of some autotrophic carbon fixation pathways are highly sensitive to oxygen while others can tolerate fully oxic conditions (Bar-Even *et al.* 2011, Fuchs 2011, see also Table 1). Thus, it is highly likely that the prevailing oxygen concentration in the environment is an additional decisive factor in determining which autotrophic pathway is predominantly used for primary production.

1.3 The Calvin Benson cycle

A comparison of all autotrophic CO₂ fixation pathways reveals that the Calvin Benson (CB) cycle accounts for most of Earth's net primary production (>99.5% of 105 x 10⁹ tons/year) (Field *et al.* 1998, Raven 2013). Two key enzymes are characteristic for this cycle, namely the Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO, EC 4.1.1.39), which catalyzes the attachment of CO₂ to ribulose-1,5-bisphosphate (Rulp) to produce 2 molecules of 3-D-phosphoglycerate (3-PGA) and the Phosphoribulokinase (PRuk, EC 2.7.1.19), which completes the cycle by regenerating Rulp (see Figure 4). The classical CB cycle can function under fully oxic conditions, but a 'wasteful' oxygenase side reaction results from this (see Figure 4), where only one

molecule 3-PGA, but one molecule 2-D-phosphoglycolate are formed (Tabita *et al.* 2008). This reaction is 'wasteful' because 2-D-phosphoglycolate is toxic in higher amounts and has to be removed via photorespiration, while 3-PGA is used to fuel an additional cycle and finally no product for biosynthesis is gained (Berg 2011). Since the CB cycle requires seven ATP equivalents and five NAD(P)Hs for the synthesis of one pyruvate it is obvious, that this mechanism is one of the most energy intensive ones among all autotrophic carbon fixation pathways (for comparison with the other five autotrophic carbon fixation pathways see Table 1) (Bar-Even *et al.* 2011). The CB cycle is the only one existent in eukaryotes (all higher plants and algae). However, it is also present in most autotrophic bacteria (Berg 2011) including phototrophs and chemolithoautotrophs and its key enzyme RubisCO is still believed to be the most abundant protein on Earth (Ellis 1979, Raven 2009). It was furthermore recapitulated that the CB cycle is quantitatively the most important autotrophic carbon fixation pathway throughout the world (Berg 2011) reasoning that this cycle is of fundamental importance for global primary production. Hence, it is essential to better understand the functioning, assembling and regulation of its central key enzyme RubisCO.

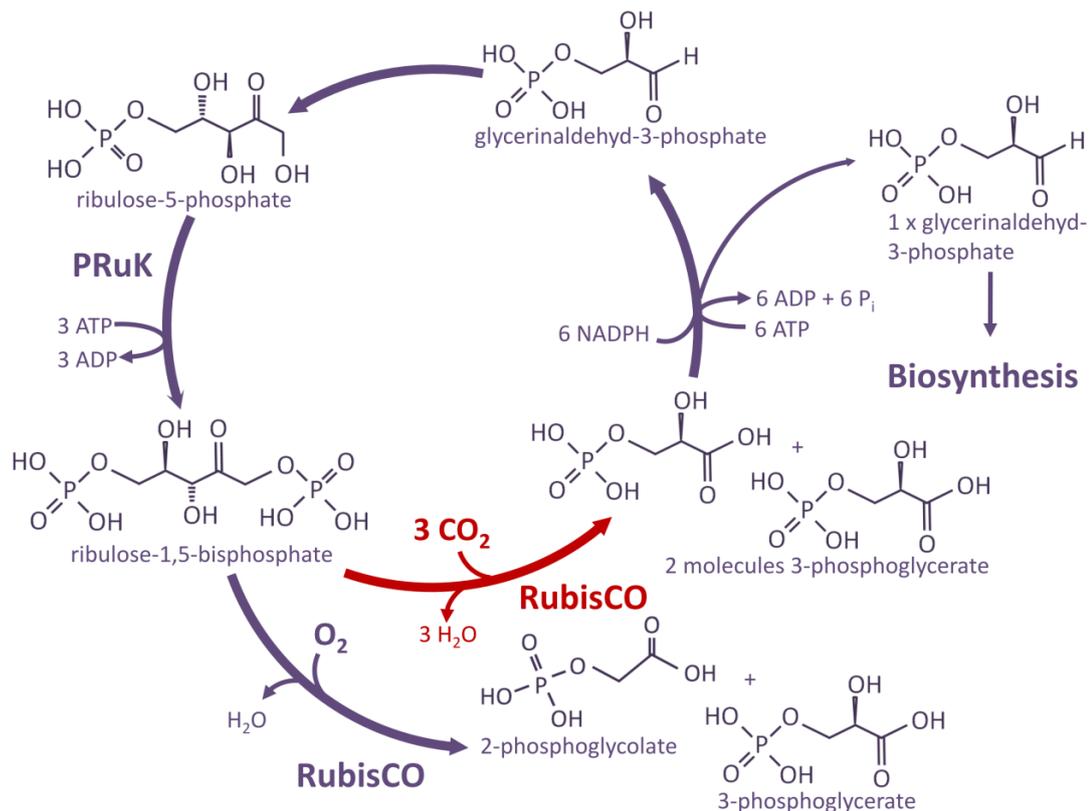
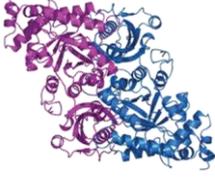
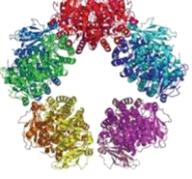
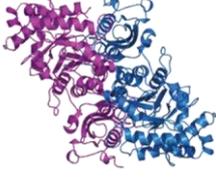


Figure 4: The schematic representation of the Calvin Benson cycle including the oxygen site reaction of RubisCO. Carboxylation reaction catalyzed by Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) is indicated in red. Abbreviations are as follows: PRuK – Phosphoribulokinase; ATP – adenosine-5'-triphosphate; ADP – adenosine-5'diphosphate; NADPH - nicotinamide adenine dinucleotide phosphate.

1.4 RubisCO and RubisCO associated enzymes

Since RubisCO is incapable of discriminating between the fairly similar substrates CO₂ and O₂ the ‘wasteful’ oxygenase site reaction is unavoidable (Lorimer and Andrews 1973), but even if operating under anaerobic conditions RubisCO’s low catalytic efficiency remains unaffected (Berg 2011). However, four different forms of RubisCO (form I to IV) evolved (see Table 2), probably as a consequence of evolutionary adaptation to changing environmental conditions like e.g. the decreasing atmospheric CO₂ concentration paired with an increasing O₂ content (Badger and Bek 2008). RubisCO form I has for example a higher specificity for CO₂ vs. O₂ than form II, but form III has the least specificity to CO₂ and thus a poor capability to discriminating between CO₂ and O₂ (Berg 2011). Only form I, II and III are capable of catalyzing the Rubp-dependent carboxylation reaction, while form IV (RubisCO-like protein) is known to be involved in the methionine salvage pathway (Tabita *et al.* 2007). As of today only form I and form II, which are both present in proteobacteria, have been shown to operate in the classical autotrophic CB cycle (Badger and Bek 2008).

Table 2: Properties of different RubisCO forms.

	Form I RubisCO	Form II RubisCO	Form III RubisCO	Form IV RubisCO-like
metabolism	CB cycle	CB cycle	not known to participate in autotrophy	methionine salvage pathway
sources	plants, algae, cyanobacteria & γ -proteobacteria (<i>Thiomicrospira</i>)	photo- and chemoautotrophic α -, β -, & γ -proteobacteria (<i>Thiomicrospira</i>)	archaea like <i>Aciduliprofundum boonei</i>	<i>Bacillus subtilis</i>
specificity to CO₂^[2]	$\Omega = 25$ to 75	$\Omega = 10$ to 15	$\Omega = 4$	-----
subunit composition	large and small subunits (L ₈ S ₈)	large subunit (ranging from L ₂₋₈)	large subunit (L ₂ or (L ₂) ₅)	large subunit (L ₂)
quaternary structures^[1]				

^[1] Quaternary structures are deduced from Tabita *et al.* (2008).

^[2] Specificity factors to CO₂ are deduced from Berg (2011).

The encoding genes, *cbbLS* (form I) and *cbbM* (form II), have been detected in ubiquitous marine environments such as the photic (Pichard *et al.* 1997) or aphotic

zone (Swan *et al.* 2011) of the water column, hydrothermal vents (Perner *et al.* 2007b), cold seeps (Elsaied and Naganuma 2001) or intertidal sediments (Nigro and King 2007). Here they appear to be highly abundant and are therefore likely responsible for major amounts of carbon assimilation in marine habitats (46.2% of $>104.4 \times 10^9$ tons/year $\cong >48.2 \times 10^9$ tons/year (Field *et al.* 1998)). The role of RubisCO form III (limited to archaea) for carbon fixation is on the contrary still enigmatic (for details on properties of all four forms of RubisCO see Table 2). The form I RubisCO can be further classified in green-like type IA and type IB RubisCOs and red-like type IC and type ID RubisCOs (Delwiche and Palmer 1996, Elsaied and Naganuma 2001, Tabita *et al.* 2007) (see Figure 5).

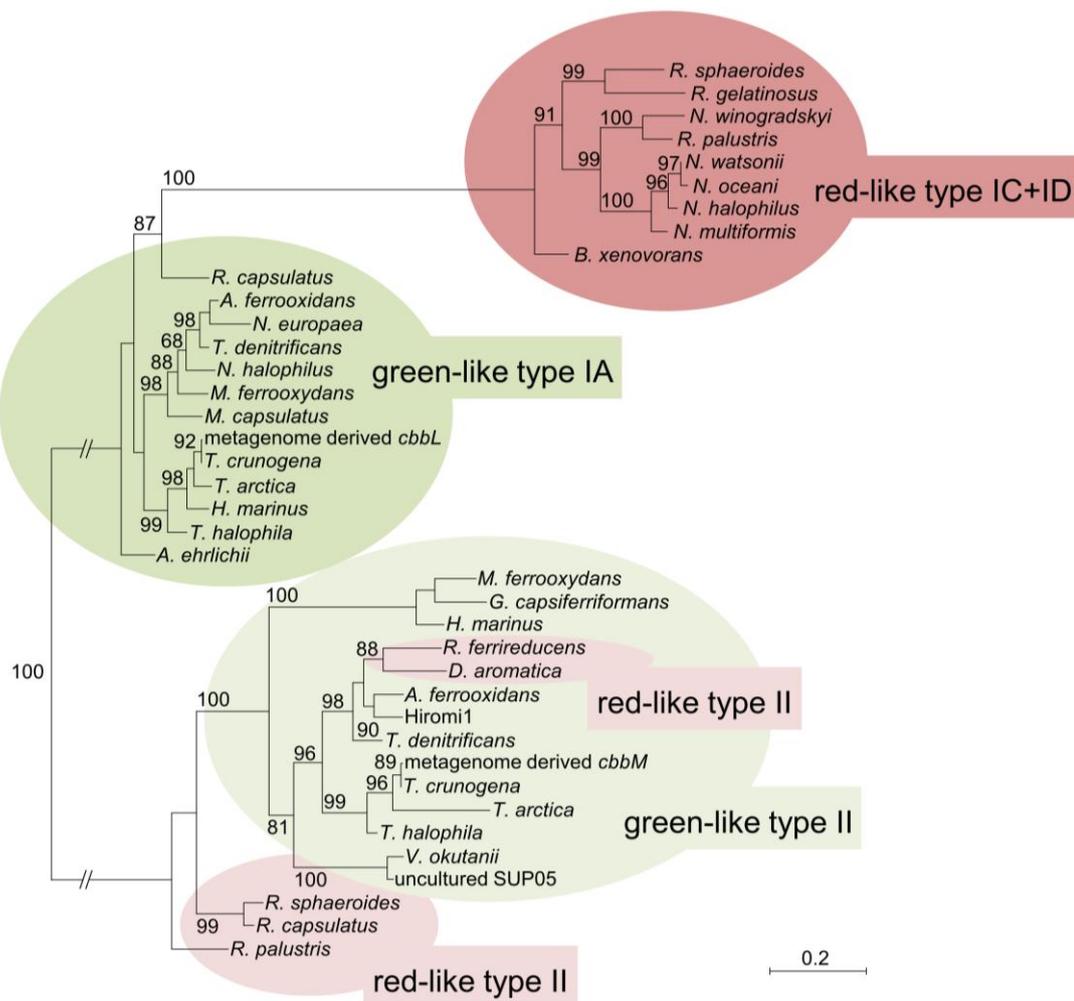


Figure 5: Phylogenetic relationship of *cbbL* and *cbbM* structural genes. The phylogenetic tree calculated for the amino-acid sequence of *cbbL* and *cbbM* of representative microorganism using Maximum-Likelihood analyses. Bootstrap values, calculated for 100 replicates, are presented as percentages at the node and are indicated only when above 80%. Abbreviations and accession numbers of shown species are listed in Appendix Table A 1. The scale bar represents the expected number of changes per amino acid position.

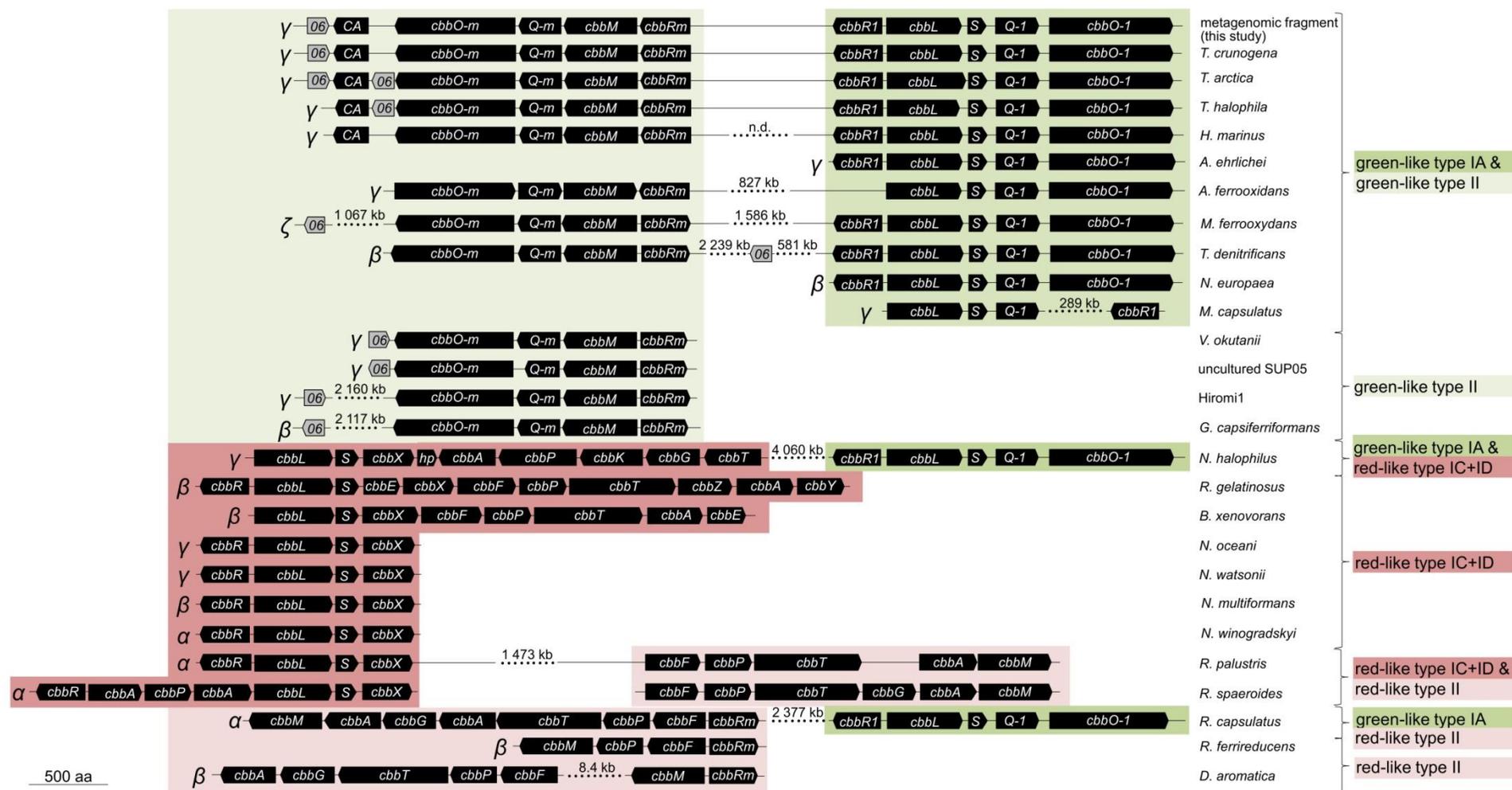


Figure 6: Comparison of RubisCO gene cluster arrangements. The RubisCO gene cluster arrangements of different bacteria with green-like and red-like RubisCO form I and form II are shown. Cyanobacterial RubisCOs and RubisCO genes encoded as part of a carboxysome operon were not included in the overview. Further genes of the CB cycle are not shown if they are scattered across the genome. Open reading frames (ORF) are indicated as arrows in the direction of transcription. Abbreviations and accession numbers of species are listed in Appendix Table A 1. Standard gene abbreviations were used (for details see Appendix Table A 2).

Gene arrangements and the presence of specific RubisCO associated genes differ on various genomes: only in plants, red algae and proteobacteria with red-like type I RubisCO, genes encoding a RubisCO activase (*cbbX* or *rca*) are present (see Figure 6) which appear to be responsible for RubisCO activation (Mueller-Cajar *et al.* 2011, Portis 2003 and references therein). In contrast, microorganisms with a green-like type I RubisCO have *cbbQ* and/or *cbbO* genes on their genomes instead (see Figure 6). Phylogenetic relationships of RubisCO form I structural genes (*cbbL*) support this classification (see Figure 5). If comparing gene arrangements of *cbbM* structural genes and flanking regions a similar classification is conjecturable, since comparison of gene arrangements of different representatives also reveal two different types (see Figure 6): (i) one where the structural gene *cbbM* is surrounded by genes encoding enzymes associated with the classical CB cycle (comparable to red-like form I RubisCO) and (ii) one with *cbbQ* and/or *cbbO* genes adjacent to the *cbbM* structural gene (comparable to green-like form I RubisCO). However, phylogenetic analysis did not confirm this classification approach (see Figure 5).

Over multiple years RubisCO genes and enzymes were studied (Li *et al.* 1993, Mueller-Cajar *et al.* 2011, Portis 2003) but despite many open questions remain unanswered. For plant RubisCO it has been shown that the presence of catalytically active form I RubisCO depends on a carbamylation reaction, where CO₂ reacts at the active site lysine with Mg²⁺ as co-factor (Portis 2003). However, uncarbamylated form I RubisCO tends to bind its substrate ribulose-1,5-bisphosphate (Rbp) prematurely, forming an inactive complex (Mueller-Cajar *et al.* 2011). In order to restore RubisCO activity, Rbp needs to be released from the active site of competitively inhibited RubisCO, enabling the essential carbamylation step. In green algae and plants this activation reaction is catalyzed by an enzyme named RubisCO activase (Portis 1990). The functioning of RubisCO activase is also affected by several other aspects like e.g.: the concentration of Rbp or prevailing stromal ATP/ADP ratio, which as a result contributes to the level of higher-plant RubisCO activity as well (Portis 1990). Furthermore it is known, that the activity of plant RubisCO is influenced by light intensity, which also correlates with the presence of RubisCO activase (Zhang *et al.* 2002). Thus it is obvious that RubisCO activation in plants and green-algae is a highly regulated, complex system.

Beyond that a red-type RubisCO activase has recently been discovered, being responsible for activating RubisCO of red algae and proteobacteria with red-type form I RubisCO, respectively (Mueller-Cajar *et al.* 2011). However, it seems like this RubisCO activase system is not applicable uniformly for all types of form I RubisCO. A RubisCO activase encoding gene (*rca*) has for instance been detected in cyanobacterial *Anabena* sp., but not in *Synechosystis* sp. (Portis 2003), indicating that activation of cyanobacterial RubisCO differs from that of higher-plant enzyme (Marcus and Gurevitz 2000). Furthermore, evidences for

the presence of enzymes, catalyzing posttranslational RubisCO activation in bacteria, namely CbbQ (AAA+ ATPase domain) and CbbO (von Willebrand factor, type A), have been suggested for *Pseudomonas hydrogothermophila* and *Hydrogenovibrio marinus* (Hayashi *et al.* 1997, Hayashi *et al.* 1999). However, this could not be confirmed for the *Solemya* symbiont, where RubisCO form I activity did not differ significantly regardless of whether *cbbQ* and *cbbO* were co-expressed (Schwedock *et al.* 2004). *Solemya* symbiont's *cbbO* and *cbbQ* show sequential similarities to genes encoded in the nitric oxide reductase gene cluster (de Boer *et al.* 1996), namely *norQ* (77% to *norQ* of *Nitrosomonas* sp.) and *norD* (73% to *norD* of *Thioflavicoccus mobilis*), respectively, suggesting that these genes likely operate in a generalized function (Schwedock *et al.* 2004) or possess different roles obligatory for corresponding organisms. However, it is still questionable whether enzymatic activation, comparable to those described for plant RubisCO, exist for the prokaryotic “green-like” form I RubisCO and thus, the activation mechanism of prokaryotic “green-like” form I RubisCO is still enigmatic. The same holds true for activation of form II RubisCO and it is moreover even unknown whether form II RubisCO generally needs to be activated or not.

Little is furthermore known about regulatory mechanisms behind prokaryotic RubisCO expression. CbbR genes, which were classified as transcriptional regulators of the LysR family, have been found in many genomes adjacent to RubisCO structural genes (Kusian and Bowien 1997, Scott *et al.* 2006) (see Figure 6). The *H. marinus* chromosome encodes for instance for two of these regulatory proteins, namely CbbR1 and CbbRm, which were located upstream of the RubisCO structural genes *cbbLS-1* and *cbbM*, respectively (Toyoda *et al.* 2005). Experiments aiming at the physiological role of these *H. marinus* CbbRs suggested that they regulate the expression of the adjacent RubisCO genes (Toyoda *et al.* 2005), a presumption which is further supported by studies on *Rhodopseudomonas palustris* and *Rhodobacter sphaeroides* CbbRs (Dubbs *et al.* 2000, Joshi *et al.* 2009). Since it has been shown that expression of both forms of RubisCO depends on the CO₂ concentration (Yoshizawa *et al.* 2004), it is furthermore assumed that the expression of the correlated CbbRs were governed by CO₂ concentrations as well (Toyoda *et al.* 2005). However the regulation mechanism behind RubisCO expression is not completely covered yet and it is still not known whether additional proteins, others than the structural genes, are involved in RubisCO assembling or activation. One promising approach to further investigate the role of potential RubisCO associated genes encoded close to RubisCO structural genes (not further afar than 30 to 40 kb) represents the research area of ‘Metagenomics’.

1.5 Metagenomics

Three decades ago Staley and Konopka (1985) encapsulate what dawned upon other scientist years before (Winterberg 1898) by coining the term “great plate anomaly”, which describes the fact, that the number of cells seen under a microscope in any environmental sample (e.g.: soil, water or marine sediments) differ significantly to the number of cultivable ones. Later it has been estimated that less than 1% of all microorganisms can be brought in culture (Amann *et al.* 1995), a phenomenon many scientist still have a focus on (Epstein 2013). One approach to avoid this cultivation bottleneck is ‘Metagenomics’, which is a culture-independent method of direct cloning, in principle firstly implemented for 16S ribosomal RNA sequences by Lane and colleagues in 1985. Nowadays the whole metagenomic DNA of one sample (e.g.: soil, water or marine sediments) is isolated and large metagenomic DNA fragments can directly be cloned into suited vector systems (e.g.: cosmids, BACs or fosmids) (Handelsman *et al.* 1998, Streit and Schmitz 2004). Then vector-DNA constructs are transferred in an easily cultivable host organism, which is in most cases *E. coli* (Handelsman *et al.* 1998). The metagenomic library can now be sought for genes of interest performing either sequence- or function-based screening approaches. However, inherent limitations of sequence-based screening exist because only sequences with significant similarities to known genes can be detected. Furthermore it remains largely unanswered whether the detected environmental gene is functional or not and it is generally not known how this gene is regulated and activated. By contrast function based screening approaches really open the door to tap the tremendous potential of the otherwise inaccessible uncultured majority since novel biocatalyst (Chow *et al.* 2012) and drugs (Rabausch *et al.* 2013) can be explored or ecological issues can be addressed whereby e.g. the occurrence and functionality of metabolic pathways or respective key enzymes can be elucidated (Böhnke and Perner 2014).

1.6 Intention of this work

The aim of this study was to establish a solely activity-based approach for identifying RubisCO active fosmid clones from metagenomic libraries originating from hydrothermal deep-sea habitats. Therefore a suitable, functional screening procedure is expected to be established that allows seeking recombinantly expressed RubisCOs directly from environmental DNA (metagenomic libraries). In parallel four metagenomic fosmid libraries are intended to be constructed with metagenomic DNA isolated from thermally and chemically distinct hydrothermal deep-sea vent samples. These four libraries together with two already existing libraries are finally in vision to be screened for clones with recombinant RubisCO activity by using the newly established RubisCO screen. Fosmids of clones exhibiting RubisCO activity are supposed to be analyzed to elucidate the role of flanking genes and resulting gene products, which at the end may contribute to better understand RubisCO regulation and activation mechanisms.

2 Material and Methods

2.1 Bacterial strains and respective cultivation techniques

2.1.1 Bacterial strains

Bacterial strains used in this study as well as respective characteristics are listed in Table 3.

Table 3: Bacterial strains used in this study.

strain	characteristics ^[1]	reference/source
<i>E. coli</i> EPI300	host strain for fosmid libraries; F ⁻ , <i>mcrA</i> $\Delta(mrr-hsdRMS-mcrBC)$, $\Phi80dlacZ\Delta M15$, $\Delta lacX74$, <i>recA1</i> , <i>endA1</i> , <i>araD139</i> , $\Delta(ara, leu)7697$, <i>galU</i> , <i>galK</i> , λ^- , <i>rpsL</i> , <i>nupG</i> , <i>trfA</i> , <i>tonA</i> , <i>dhfr</i>	epicentre [®] (Madison, WI, USA)
<i>T. crunogena</i> TH-55	wild-type strain (DSMZ no. 12353)	DSMZ (Braunschweig, Germany)

^[1] Abbreviations describing geno- and phenotypes were made according to Bachmann (1983).

2.1.2 Cultivating *T. crunogena* TH-55

T. crunogena TH-55 (DSMZ No. 12353) was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany) and was cultivated in artificial seawater medium supplemented with 40 mM thiosulfate and 10 mM HEPES, pH 8.0 (T-ASW) at 28°C and 130 rpm as described previously (Dobrinski *et al.* 2005, Jannasch *et al.* 1985).

2.1.3 Cultivating different *E. coli* strains

E. coli cultures were routinely grown on lysogeny broth (LB) medium (Bertani 1951) at 37°C with the exception of clones cultivated in order to measure recombinant RubisCO activities, where the growth temperature was lowered to 28°C. Required supplements were added after LB media was autoclaved but not before a temperature below to 55°C was reached. Used concentrations for stock solutions as well as final concentrations in the media are summarized in Table 4.

Table 4: Antibiotics and supplements used in this study.

substance	stock solution	final concentration	solvent	treatment
<u>antibiotics</u>				
ampicillin	100 mg/ml	100 µg/ml	H ₂ O	filtered sterile
chloramphenicol	50 mg/ml	12.5 µg/ml	96% EtOH	filtered sterile
kanamycin	50 mg/ml	50 µg/ml	H ₂ O	filtered sterile
tetracycline	10 mg/ml	10 µg/ml	70% EtOH	filtered sterile
<u>other supplements</u>				
IPTG	100 mg/ml	100 µg/ml	H ₂ O	filtered sterile
maltose	20%	0.2%	H ₂ O	filtered sterile
MgSO ₄	1 M	10 mM	H ₂ O	autoclaved
x-gal	50 mg/ml	50 µg/ml	DMF	filtered sterile

2.2 Vectors and constructs

Vectors used in this study and respective characteristics are summarized in Table 5.

Table 5: Vectors used in this study.

vector	features ^[1]	size [kb]	reference/source
pUC19	cloning vector, Amp ^R , pMB1-type ColE1 ori, <i>lacZ</i> , LacZ-promoter	3.0	Gibco/BRL [®] (Life Technologies, Darmstadt, Germany)
pCC1FOS TM	fosmid vector, Chl ^R , <i>redF</i> , oriV, ori2, <i>repE</i> , <i>parA</i> , <i>parB</i> , <i>parC</i> , <i>cos</i> , <i>loxP</i> , <i>lacZ</i> , T7 promoter	8.139	epicentre [®] , (Madison, WI, USA)

^[1] Abbreviations describing vector features were made according to manufactures' protocols.

Constructs created in this study and respective characteristics are listed in Table 6.

Table 6: Constructs created in this study.

construct	vector	insert size [kb]	characteristics
pCC1FOS:: <i>cbbLS</i>	pCC1FOS	2.6	<i>cbbLS</i> , insert cloned from TH-55 genomic DNA
pCC1FOS:: <i>cbbM</i>	pCC1FOS	2.3	<i>cbbM</i> , insert cloned from TH-55 genomic DNA
Tc6F3	pCC1FOS	38.1	genomic fosmid vector containing TH-55's RubisCO gene cluster (13 kb) and approximately 25.1 kb flanking DNA
71C2			
74E1			
74C10			
77H1			
77F4			
77D9			
78G5	pCC1FOS	35.2	metagenomic fosmid vectors containing the RubisCO gene cluster (13 kb) and 22.2 kb flanking DNA
78E10			
80G3			
81G7			
81E1			
84G4			
71C2II	pCC1FOS	13.0	fosmid vector containing the RubisCO gene cluster (13 kb) subcloned from 71C2
84G4II	pCC1FOS	13.0	fosmid vector containing the RubisCO gene cluster (13 kb) subcloned from 84G4
pUC19:: <i>orf06</i>	pUC19	1.0	<i>orf06</i> , insert cloned from 71C2
pUC19:: <i>lysR2-1</i>	pUC19	2.5	<i>lysR1</i> and <i>lysR2</i> , insert cloned from 71C2
pUC19:: <i>cbbO-m</i>	pUC19	3.1	<i>cbbO-m</i> , insert cloned from 71C2

2.3 Sample collection



Figure 7: Location map of sampled hydrothermal sites. Names of sampled chimneys are put in parentheses behind the associated hydrothermal vent field. The figure was created by using Google earth software.

Four marine, hydrothermally influenced samples originating from geographically distinct vent fields along the Mid-Atlantic Ridge (see Figure 7) were used for the construction of metagenomic libraries. These samples were collected within the DFG-SPP 1144 priority program “From Mantle to Ocean: Energy-, Material-, and Life-cycles at Spreading Axes”. Sampling was done by a remote operated vehicle (ROV 6000, GEOMAR, Kiel) during the MSM 10-3 cruise (January/February 2009) with the RV Maria S. Merian and the MAR-SUED V cruise (March/April 2009) with the RV Meteor. Two different kinds of samples were investigated, namely chimney samples and hydrothermal fluid samples. Chimney samples were collected from (i) ‘Sisters Peak’ at 4°48’S/12°22’W (see Figure 7) at a water depth of 2,982 m (part of the Comfortless Cove vent field) (Haase *et al.* 2007), from (ii) ‘Mephisto’ at 4°47’S/12°22’W (see Figure 7) at a water depth of 3,042 m (part of the Red Lion vent field) (Haase *et al.* 2007) and from (iii) ‘Site B’ at 14°45’N/44°58’W (Figure 7) at a water depth of 3,047 m (part of the Logatchev vent field) (Perner *et al.* 2007a and references therein). Chimney samples were stored immediately at -70°C until further investigations. Hydrothermal fluids were collected from the Nibelungen vent field at 8°18’S/13°30’W (see Figure 7) at a water depth of 2,915 m (Melchert *et al.* 2008) from the interface zone between hot fluids emanating from the crater ‘Drachenschlund’ and ambient seawater. Around 200 ml of hydrothermal fluids were concentrated onboard on a 0.2 µm polycarbonate filter and kept at -20°C until further analyses. Detailed sampling procedures and further information on the

sampling sites are described elsewhere (Perner *et al.* 2007a, Perner *et al.* 2007b, Perner *et al.* 2009, Perner *et al.* 2013b).

2.4 (Meta)-genomic fosmid libraries

Five fosmid libraries were constructed: (i) one genomic library with DNA material from *T. crunogena* TH-55 and (ii) four metagenomic libraries with metagenomic DNA isolated from chimney material of 'Sisters Peak', 'Mephisto' and 'Site B' or from hydrothermal fluid material of 'Drachenschlund'.

2.4.1 Meta-(genomic) DNA isolation

In order to isolate genomic DNA *T. crunogena* TH-55 was cultivated as described above (see 2.1.2, respectively) and at least 1 liter of the culture was harvested at the end of exponential growth phase through centrifugation (11,300 x g, 20 minutes and 4°C). The cell pellet was washed once in 5 ml TE-buffer [10 mM Tris-HCl (pH = 8.0) and 1 mM EDTA] and stored at -20°C until proceeding with DNA isolation. Prior to isolating metagenomic DNA from chimney samples the massive chimney material was stepwise broken up like it is visualized in Figure 8. Ground chimney material was kept at -20°C until DNA isolation was continued. The fluids of 'Drachenschlund' concentrated on a polycarbonate filter were directly utilized for DNA isolation without any pre-treatment.



Figure 8: Treatment of chimney material prior to DNA isolation. The chimney samples were firstly crushed with hammer and chisel and then ground using mortar and pestil.

Genomic DNA from *T. crunogena* TH-55 as well as metagenomic DNA from ground chimney samples of 'Sisters Peak' and 'Mephisto' and metagenomic DNA from the fluid sample of 'Drachenschlund' were isolated by using a common phenol-chloroform extraction method with TE-sucrose buffer [10 mM Tris-HCl (pH 8.0), 1 mM EDTA and 20% sucrose (w/v)], lysozyme solution [2 mg/ml TE-buffer (pH 8.0) and 1 mg/ml RNaseA], proteinase K [1 mg/ml final concentration] and sarkosyl [5% final concentration] (Streit *et al.* 1993). DNA was precipitated over night at -20°C by adding sodium acetat (0.3 M final concentration) and isopropanol (0.6 volumes). After centrifugation (16,100 x g, 20 minutes and 4°C) metagenomic DNA pellets were washed twice with ethanol (70%), dried at 37°C and resuspended in 30 µl nuclease-free water. The metagenomic DNA from the ground chimney

material of 'Site B' was isolated by using the UltraClean[®] Microbial DNA Isolation Kit (MO BIO Laboratories, Inc.) following manufacturer's protocol.

2.4.2 Multiple displacement amplification (MDA)

With the objective of generating highly concentrated, purified DNA, as required for the construction of metagenomic libraries, multiple displacement amplification (MDA) was performed for each hydrothermally influenced DNA sample with phi29 DNA polymerase (REPLI-g Kit, Qiagen, Hilden, Germany) according to the manufacturer's instruction. For this purpose, three parallel samples with 2.5 µl starting material were subjected to MDA for each of the four samples. The three parallels of one sample were finally pooled and purified using phenol/chloroform extraction. Therefore the volume of the pooled MDA was firstly increased with nuclease free water from 120 µl to 600 µl in order to minimize the loss of DNA during each purification step. After that pooled MDA samples were washed with equal volume of phenol-chloroform followed by washing with equal volume of chloroform. Precipitation was done overnight by adding sodium acetate (0.3 M final concentration) and isopropanol (1 v/v). MDA treated DNA was pelleted by centrifugation (16,100 x g, 20 minutes and 4°C), followed by two washing steps with ethanol (70%). Dried metagenomic DNA was finally resuspended in 30 µl nuclease-free water. Phylogenetic analyses of 16S rRNA genes of bacteria using 27F and 1492R (Lane 1991), cloning, sequencing (Eurofins MWG Operon, Ebersberg, Germany) and evaluation of data as described (Perner *et al.* 2009) were performed before and after MDA to validate the quality of the metagenomic DNA and to randomly monitor the MDA based bias (data not shown).

2.4.3 Construction of (meta)-genomic libraries

Genomic- and MDA treated metagenomic DNA were used to construct (meta)-genomic fosmid libraries. Here, the CopyControl[™] Fosmid Library Production Kit (epicentre[®], Madison, WI, USA) was used according to manufacturer's instructions. Clones containing fosmids with DNA inserts were selected on LB agar plates by blue-white screening [therefore IPTG (100 µg*ml⁻¹) and x-gal (50 µg*ml⁻¹) were added] and chloramphenicol (12.5 µg*ml⁻¹) addition. Quality control of the fosmid inserts was randomly performed by exemplarily testing at least 20 fosmid clones of each library. Insert sizes were checked through restriction analyses with three different enzymes (*Bam*H1, *Eco*R1 and *Hind*III, Thermo Scientific, Waltham, MA, USA). Additionally, inserts from 20 fosmid clones of each library were sequenced from both insert ends (Eurofins MWG Operon) using T7 promoter primer and pCC1FOS reverse sequencing primers (see Appendix Table B 1). Once the quality of a library was validated, clones were transferred to microtiterplates [120 µl LB medium per well

supplemented with chloramphenicol ($12.5 \mu\text{g}\cdot\text{ml}^{-1}$), incubated overnight at 37°C , mixed with glycerin (35% final concentration in each well) and stored at -70°C until further usage.

2.5 Establishing a functional screen to seek recombinant RubisCOs from metagenomes

Within this study a functional screen for seeking recombinant RubisCOs from metagenomic libraries has been established. Steps and associated methods, which were vital to ensure that the envisaged HPLC based activity assay in principle works, are visualized in Figure 9 and are described in detail in the following:

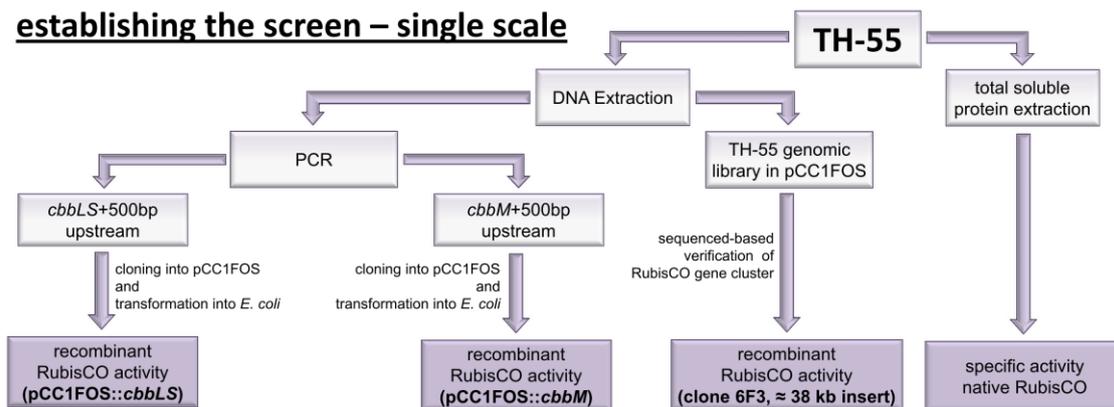


Figure 9: Schematic view of major points successively processed to verify that the function-based screen works on a single scale (partially from Böhnke and Perner 2014).

2.5.1 *T. crunogena* TH-55's RubisCO activity

2.5.1.1 Preparation of crude extracts

In order to verify that the function-based screen works for prokaryotic RubisCOs *T. crunogena* TH 55's RubisCO was firstly investigated. Pure cultures of *T. crunogena* TH-55 were harvested at three different points of time during growth and RubisCO activities were measured. For this purpose cell pellets were washed twice with buffer A [100 mM Tris-HCl (pH 7.8), 10 mM MgCl_2 , 1 mM EDTA, 25 mM NaHCO_3 , and 1 mM DTT] and resuspended in 2 ml of the same buffer. Cells were disrupted using the french pressure cell press method, passing each sample three times through the system. Lysates were centrifuged ($19,580 \times g$, 20 minutes and 4°C) and supernatants were used for analyzing the specific RubisCO activity. The concentration of total protein in the crude extracts was measured as described previously (Bradford and Williams 1976) with bovine serum albumin as a standard.

2.5.1.2 RubisCO activity assay

The enzyme assay was performed in 1.5 ml reaction tubes at 25°C . The assay mixture (150 μl final volume) contained buffer A, 0.2 mg unpurified, total protein, and 5 mM

ribulose-1.5-bisphosphate (Rubp), where the Rubp addition is the initiation step. If RubisCO enzymes are in the sample Rubp is converted to 3-phosphoglycerate (3-PGA) over time. To monitor this conversion, subsamples (50 μ l) were taken at different time points (0, 10, and 30 minutes) from the assay mixture. The reactions in the subsamples were stopped by exposure to 95°C for 3 minutes, whereby proteins were denatured and then removed by centrifugation (16,100 x g, 20 minutes and 4°C). The supernatant was then used to quantify the concentration of Rubp and 3-PGA with High Pressure Liquid Chromatography (HPLC). HPLC procedures were performed as previously described (Jakob and Saenger 1985) with some modifications. The HPLC LaChrom Elite[®] system from Hitachi (Tokyo, Japan) was used with a Lichrospher[®] 100 RP-18e column (VWR International GmbH, Darmstadt, Germany), consisting of particles with 5 μ m diameter. Detection was performed at 200 nm instead of 220 nm, because the absorption of Rubp and 3-PGA is higher at this wavelength range. Jakob and Saenger avoided wavelengths below 220 nm because the absorption of the eluent they used, namely methanol, increased significantly in this range. For that reason acetonitrile was used as eluent, which was best qualified for high sensitivity analysis at short UV wavelengths (Williams 2004). Furthermore the flow was adapted from 1.2 ml per minute to 0.6 ml per minute to reach best separation performance.

2.5.2 *T. crunogena* TH-55's RubisCO recombinantly expressed in *E. coli*

Since *T. crunogena* RubisCO genes have not been expressed in *E. coli* so far, it was tested whether *T. crunogena* TH-55's RubisCO can be recombinantly expressed in *E. coli* or not. Necessary methodological steps are summarized in Figure 9 and are described in the following.

2.5.2.1 Cloning of *T. crunogena* TH-55's RubisCO structural genes

T. crunogena TH-55 cells, cultivated as has been described above (see 2.1.2), were harvested after 24 hours of growth through centrifugation (11,300 x g, 20 minutes and 4°C). Genomic DNA was extracted using the UltraClean[®] Microbial DNA Isolation Kit (MO BIO Laboratories, Inc.) according to manufacturer's instructions. The DNA fragment encoding genes for the large (*cbbL*) and small (*cbbS*) subunit of RubisCO form I from *T. crunogena* TH-55 were PCR amplified using Pfu DNA Polymerase (Thermo Scientific) and the newly designed primer pairs CbbL_ncr690_F and CbbL_ncr139_R (see Appendix Table B 2) targeting 690 bp upstream of the *cbbL* gene and 139 bp downstream of the *cbbS* gene, respectively. The conditions were: 95°C for 30 seconds, 52°C for 30 seconds and 72°C for 6 minutes (32 cycles). The *cbbM* gene fragment, encoding the large subunit of RubisCO form II from TH-55 was PCR amplified also by using the Pfu DNA Polymerase (Thermo Scientific) and the self-designed primer pair cbbM_ncr332_F and cbbM_ncr557_R, targeting the

non-coding regions located 557 bp upstream and 332 bp downstream of *cbbM*. The conditions were 95°C for 30 seconds, 49°C for 30 seconds and 72°C for 5 minutes (32 cycles). The resulting PCR products of *cbbLS* (2,623 bp) and *cbbM* (2,269 bp) were cloned using the pCC1FOS vector system (epicentre®) and transformed into competent *E. coli* Epi 300 cells (epicentre®) by heat shock. Clones containing fosmids with DNA inserts were selected by blue-white screening and chloramphenicol (12.5 µg*ml⁻¹) addition. Fosmid inserts of generated clones were sequenced from both insert ends (Eurofins MWG Operon) using pCC1FOS forward and reverse sequencing primers (see manual for the CopyControl™ Fosmid Library Production Kit, epicentre®) to verify the presence of the *cbbLS* and *cbbM* genes on the fosmids, respectively. Validated clones were tested for RubisCO activity by following the HPLC-based activity assay as described in the following.

2.5.2.2 *T. crunogena* TH-55's recombinant RubisCO activity

In order to measure recombinant RubisCO activities of structural RubisCO genes, (*cbbLS* and *cbbM*) *E. coli* clones with pCC1FOS::*cbbLS* and pCC1FOS::*cbbM* were grown at 28°C on 200 ml pre-heated LB medium supplemented with chloramphenicol (12.5 µg*ml⁻¹) in 1 l flasks (130 rpm) and harvested after 18 hours by centrifugation (9,800 x g, 5 minutes and 4°C). Crude extracts were prepared for *T. crunogena* TH-55 as above-mentioned (see 2.5.1.1) and RubisCO activities were measured by following the HPLC-based activity assay also in the same way as it has been described for *T. crunogena* TH-55 before (see 2.5.1.2).

2.5.3 RubisCO activities from a TH-55 genomic fosmid clone

Since all metagenomic libraries are constructed with the low-copy pCC1FOS™ vector system (epicentre®), it has to be verified that recombinant RubisCO activities are measurable with the HPLC based approach described above (see 2.5.1.2), if such a fosmid clone was used. Therefore the genomic *T. crunogena* TH-55 fosmid library (see 2.4) were screened on a sequence based level for clones harboring the RubisCO form I encoding structural genes (*cbbLS*). Initially PCR-enabled pools of 96 genomic fosmid clones were used. These pools were prepared beginning with copying a microtiterplate of the *T. crunogena* TH-55 library, where each well of the new microtiterplate was filled with 200 µl of LB medium supplemented with chloramphenicol (12.5 µg*ml⁻¹). The copy was inoculated with the cells of the original plate via stamp technique and incubated overnight at 37°C. The 96 grown *E. coli* cultures were then transferred to a single reaction tube and centrifuged (16,100 x g, 45 seconds and 4°C). Cell pellets of pooled cultures were subsequently resuspended in at least 200 µl TE_{DNA} buffer [10 mM Tris-HCl (pH = 8.0) and 0.1 mM EDTA], heated up to 90°C for 10 minutes to disrupt the cell walls and centrifuged (16,100 x g, 10 minutes and 4°C) to remove cell debris. Finally, supernatants were transferred in a nuclease free tube and 1 µl of this PCR-enabled

pool was used as a template to amplify a part of the *T. crunogena* form I RubisCO encoding gene region. Here, the self-designed primer pair *rbcl+S_for* and *rbcl+S_rev* (see Appendix Table B 2) were used with following PCR conditions: denaturation at 95°C for 45 seconds, primer annealing at 57°C for 45 seconds and elongation at 72°C for 2 minutes (32 cycles). If a pool of 96 clones was tested positive for the amplification of RubisCO encoding structural genes it was broken down until the one clone harboring the targeted genes (*cbbLS*) was identified. For this purpose 12 PCR-enabled pools of the columns 1 to 12, where each pool consists of 8 clones in total and 8 PCR-enabled pools of the rows A to H, where each pool consists of 12 clones in total, were prepared and tested through PCR following the same conditions as used before for pools of 96 clones. By a crossover comparison of the PCR based hits for RubisCO encoding genes in all rows and columns of one microtiterplate, the one fosmid clone harboring the RubisCO encoding gene was identified. The RubisCO activity of the identified *cbbLS* comprising genomic clone was furthermore determined. Therefore the clone was grown at 28°C on 200 ml pre-heated LB medium supplemented with chloramphenicol (12.5 µg*ml⁻¹) and autoinduction solution [1x final concentration (epicentre®)] in 1 l flasks with shaking (130 rpm) and allowed to grow for 18 hours before harvested by centrifugation (9 800 x g, 5 minutes and 4°C). The crude extract was prepared as afore-mentioned (see 2.5.1.1) and RubisCO activity was measured by following the HPLC-based activity assay in the same way as described before (see 2.5.1.2).

(12.5 $\mu\text{g}\cdot\text{ml}^{-1}$ chloramphenicol) over night at 37°C (see Figure 10A). Fosmid clones were then swamped off with 10 ml LB medium (12.5 $\mu\text{g}\cdot\text{ml}^{-1}$ chloramphenicol), and 200 μl of this cell suspension was used as inoculum for a working culture (see Figure 10A). These working cultures were grown at 28°C on 200 ml pre-heated LB medium supplemented with chloramphenicol (12.5 $\mu\text{g}\cdot\text{ml}^{-1}$) and autoinduction solution [1x final concentration (epicentre®)] in 1 l flasks with shaking (130 rpm) and allowed to grow for 18 hours before harvesting by centrifugation (9,800 x g, 10 minutes, 4°C). Cell pellets were then washed twice with buffer A and cells were disrupted by the french pressure cell press method, followed by centrifugation (19,580 x g, 20 minutes, 4°C), as described above (see 2.5.1.1). Crude extracts were finally used for measuring RubisCO activity (see Figure 10A) as aforementioned (see 2.5.1.2) but with the exception that subsamples were taken at 0, 30 and 120 minutes instead of 0, 10 and 30 minutes to extend reaction times and maximize turnover rates. These experiments with the simulated pools revealed that a pool of 24 fosmid clones was best qualified for the detection of RubisCO activities, this pool-size was used for all further functional screenings with RubisCO active clones from metagenomic libraries (see Figure 10B). Here, metagenomic fosmid clones were cultivated and treated in the same way as described for the simulated-pools at the beginning of this section, but with the exception that RubisCO activity was measured at 25°C and 55°C to increase the spectra of detectable RubisCOs. If RubisCO activity was confirmed, the pool of 24 metagenomic clones was broken down in two pools of 12 clones, followed by 4 pools of 3 clones, until the one fosmid clone exhibiting the RubisCO activity was identified.

2.7 Working with RubisCOs from metagenomes

To elucidate the functions of genes recognized on metagenomic fragments recovered by the function-based screen different experiments were conducted, which are summarized in Figure 11 and fully described in the following:

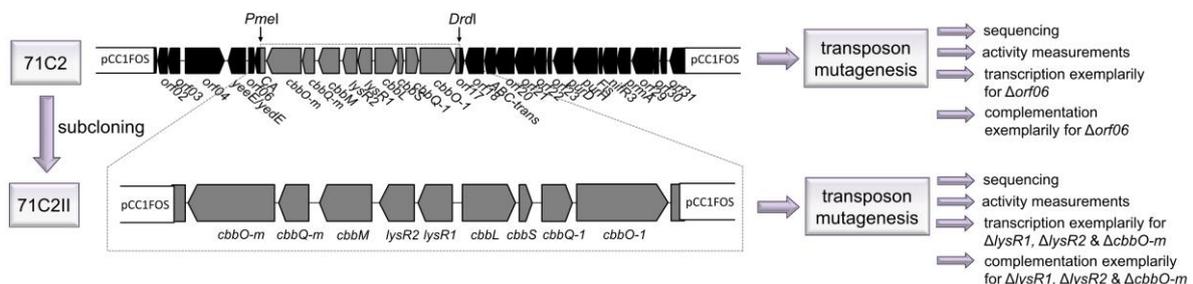


Figure 11: Illustration of conducted experiments to characterize the metagenome derived RubisCOs and flanking gene regions. The experiments performed with the recognized RubisCO active metagenomic fragment to elucidate the roles of distinct genes and respective gene products are summarized. A detailed magnification of the 35.2 kb and the 13 kb gene cluster are shown in Figure 16 and Figure 18 (partially from Böhnke and Perner 2014).

2.7.1 Primer walk and sequence editing

Fosmid inserts of RubisCO active metagenome derived clones were sequenced starting from both insert ends (Eurofins MWG Operon) using T7 promoter primer and pCC1FOS reverse sequencing primers (see manual for the CopyControl™ Fosmid Library Production Kit, epicentre®) and continued through primer walking with primers indicated in Appendix Table B 1. Since sequencing with the degenerate RubisCO primer *cbbL* R (Campbell and Cary 2004) also resulted in an evaluable sequence, a third origin for primer walk was utilizable, while the degenerate forward primer *cbbL* F, (Campbell and Cary 2004) was not applicable. All regions were sequenced at least twice. Sequences were edited with Lasergene Software SeqMan (DNA Star, Madison, WI, USA) and consensus sequences were arranged with the same software. Homologies were evaluated by comparing generated sequences with those deposited in the NCBI database, using Blast search.

2.7.2 Subcloning of RubisCO gene clusters from metagenomic fosmid inserts

In order to investigate the functionality of RubisCO structural genes, including the relevance of flanking gene regions, the RubisCO gene clusters of two metagenome derived clones, namely 84G4 and 71C2, were subcloned. Therefore the RubisCO gene cluster was cut out in each case from the whole metagenomic fragment using FastDigest *DrdI* and *MssI* restriction enzymes (Thermo Scientific) according to manufacturer's instructions. The digested fosmid DNA was separated by agarose gel electrophoresis (0.8% TAE agarose gel, 50 V, 2.5 h) and the 13 kb DNA fragment was re-extracted from the gel using the Gel/PCR DNA Fragments Extraction Kit (Geneaid, New Taipei City, Taiwan) following manufacturer's instruction. The resulting 13 kb DNA fragment was cloned into the pCC1FOS™ vector (epicentre®) and transformed into competent Epi 300 cells (epicentre®) with heat shock. Clones containing fosmids with DNA inserts were selected by blue-white screening [IPTG (100 µg*ml⁻¹) and x-gal (50 µg*ml⁻¹) were added] and chloramphenicol (12.5 µg*ml⁻¹) addition. 96 and 24 clones for subcloning of 71C2 and 84G4, respectively were used for colony PCR using KO_FOSsite primer, binding on the pCC1FOS vector and KO_Insertsite primer (see Appendix Table B 2), to proof the presence and the right orientation of the RubisCO gene cluster. The conditions were 95°C for 45 seconds (denaturation), 54°C for 45 seconds (primer annealing) and 72°C for 45 seconds (elongation) (32 cycles). One clone each was then sequenced from both insert ends (Eurofins MWG Operon) using T7 promoter primer and pCC1FOS reverse sequencing primer (see Appendix Table B 1). The subclones, named 71C2II and 84G4II, were then tested for RubisCO activity by following the HPLC-based activity assay described above for *T. crunogena* TH-55 (see 2.5.1).

2.7.3 Transposon mutagenesis

To unravel which genes and respective products in and outside of the RubisCO gene cluster contribute to a fully functional RubisCO in the metagenomic clones two transposon mutant libraries were constructed using the EZ-Tn5TM <KAN-2> Tnp TransposomeTM Kit (epicentre[®]) according to manufacturer's instructions. The first transposon library was constructed with the fosmid of the RubisCO active metagenome derived fosmid clone 71C2, harboring the whole metagenomic DNA fragment (35.2 kb) and the second transposon library is based on the fosmid of the subclone 71C2II, comprising only the metagenome derived RubisCO gene cluster (13 kb). Clones containing fasmids with <KAN-2> insertions in the fosmid were selected on LB agar plates due to antibiotic addition: (i) chloramphenicol (12.5 µg*ml⁻¹) for the selection of the fosmid vector and (ii) kanamycin (50 µg*ml⁻¹) to verify the successful transposon insertion. Resulting colonies were transferred to microtiterplates [120 µl LB medium per well supplemented with chloramphenicol (12.5 µg*ml⁻¹) and kanamycin (50 µg*ml⁻¹)], incubated overnight at 37°C and stored at -70°C, after glycerin (35% final concentration) was added to each culture, until further usage. Fasmids with transposon insertions of single transposon clones were then isolated from autoinduced cultures (for detailed information on autoinduction procedure see the manual for the CopyControlTM Fosmid Library Production Kit, epicentre[®]) using the High-Speed Plasmid Mini Kit (Geneaid) according to manufacturer's instruction excluding the mentioned optional washing step and with the exception that the fosmid DNA was eluted with pre-heated nuclease free water instead of the provided elution- or TE-buffer for elongated 60 minutes. Isolated fasmids were then sequenced with the KAN-2 FP-1 forward primer (see manual for the EZ-Tn5TM <KAN-2>Tnp TransposomeTM Kit, epicentre[®]) to identify the exact insertion position. Selected clones were further tested for their RubisCO activities. Therefore transposon clones were cultivated at 28°C on 200 ml pre-heated LB medium supplemented with chloramphenicol (12.5 µg*ml⁻¹), kanamycin (50 µg*ml⁻¹) and autoinduction solution [1x final concentration (epicentre[®])] in 1 l flasks with shaking (130 rpm) and harvested after 18 hours by centrifugation (9 800 x g, 10 minutes and 4°C). Subsequently crude extracts were prepared as it has been mentioned before (see 2.5.1.1) and used as template to operate the HPLC-based activity assay described above (see 2.5.1.2).

2.7.4 Transcription experiments with TH-55

Total RNA from TH-55 was isolated with the PrestoTM Mini RNA Bacteria Kit (Geneaid) according to manufacturer's instructions. For this purpose, TH-55 was cultivated for 22 hours at 28°C (see 2.1.2). During growth the pH was re-adjusted to 7.8 with sterile NaHCO₃ [1 mM stock solution] and cells were allowed to grow for additional 2 hours at 28°C, before the culture was harvested (17,600 x g, 20 min, 4°C). Cell pellets were washed twice with buffer

A without DTT [100 mM Tris-HCl (pH 7.8), 10 mM MgCl₂, 1 mM EDTA and 25 mM NaHCO₃]. Once total RNA was extracted genomic DNA was removed using RTS DNase[™] Kit (MO BIO Laboratories, Inc.) with some modifications to the manufacturer's instructions: incubation time at 37°C was extended to an hour and an additional microliter of RTS DNase was added after half an hour. Then total cDNA was synthesized (Invitrogen's SuperScript[®] VILO[™] cDNA Synthesis Kit, Life Technologies[™], Darmstadt, Germany), serving as a template for the amplification of different gene to gene interspaces to elucidate which genes are co-transcribed in TH-55. Amplification was done using the GoTaq[®] Green Master Mix (Promega, Mannheim, Germany) following the manufacturer's protocol and the self-designed primers listed in Appendix Table B 3. PCR conditions were: denaturation at 95°C for 45 seconds, primer annealing as listed in Appendix Table B 3 for 45 seconds and elongation at 72°C for approximately 30 seconds, depending on the expected size of PCR amplicates (1 minute per 1,000 bp; corresponding sizes are also listed in Appendix Table B 3) for 32 PCR cycles overall. Controls with the template lacking reverse transcriptase were performed with PCR.

2.7.5 Experiments pairing transcription with RubisCO activity

Total RNA from the metagenomic fosmid clone 71C2, the subcloned 71C2II and the transposon clones 96Δ*orf06*, 6IIΔ*lysR1*, 149IIΔ*lysR2* and 17IIΔ*cbbO-m* were isolated. Therefore mentioned fosmid clones were autoinduced by growing them for 18 h at 37°C on LB medium supplemented with chloramphenicol (12.5 μg*ml⁻¹) and autoinduction solution (epicentre[®]). RNA was isolated using the UltraClean[®] Microbial RNA Isolation Kit (MO BIO Laboratories, Inc.) according to manufacturer's protocol, with the exception that only 1 ml cell culture was harvested instead of the recommended 2 ml. Genomic DNA was removed and total cDNA was synthesized as has been described before (see 2.7.4). cDNA served as a template for the amplification of small fragments of 330 bp and 306 bp within *cbbL* and *cbbM*, respectively, to examine the relative variations of transcript amounts of RubisCO form I and II structural genes in the above mentioned transposon clones and respective intact versions. However, cDNA was diluted (1:50) prior to performing PCR. SYBR[®] Select Master Mix, CFX (Applied Biosystems[®] by life technologies, Darmstadt, Germany) was used to avoid unspecific PCR bands. Primers used for amplifications are listed in Appendix Table B 4. PCR conditions were 95°C for 3 minutes followed by 32 cycles of 95°C for 45 seconds, 50°C or 52°C for *cbbL* and *cbbM*, respectively for 45 second, 72°C for 30 seconds and 5 minutes final elongation at 72°C. Controls with the template lacking reverse transcriptase were performed with PCR.

2.7.6 Complementation experiments

To test whether the activity of the transposon clones could be restored, transposon clones 96 Δ *orf06*, 149II Δ *lysR2*, 6II Δ *lysR1* and 17II Δ *cbbO-m* were complemented with pUC19::*orf06*, pUC19::*lysR2-1*, and pUC19::*cbbO-m*, respectively. For this purpose, chemically competent cells were prepared from all transposon clones mentioned, following standard procedures (The QIAexpressionist™ 06/2003, Qiagen, Venlo, Provincz Limburg, Netherlands). Additionally, the intact versions of the inserted genes (i) *orf06*, (ii) *lysR2* and *lysR1* as well as (iii) *cbbO-m* were PCR amplified using Pfu DNA Polymerase (Thermo Scientific) and newly designed primer pairs (i) *orf06_ncr444_F* and *orf06_ncr161_R* (ii) *lysR2-1_ncr193_F* and *lysR2-1_ncr383_R* and (iii) *CbbOm_for2* and *CbbOm-rev2*, respectively (for details on primer sequences see Appendix Table B 5). The conditions were: denaturation at 95°C for 30 seconds, primer annealing at (i) 50°C, (ii) 50°C and (iii) 53°C for 30 seconds and elongation at 72°C for (i) 2 minutes (ii) 5 minutes and (iii) 6:20 minutes (32 cycles), respectively. Purified PCR products of *orf06* (1,013 bp), *lysR2-1* (2,457 bp) and *cbbO-m* (3,137 bp) were cloned using the pCC1FOS vector system (epicentre®) and transformed into the previously prepared chemically competent *E. coli* cells of transposon clones 96 Δ *orf06*, 149II Δ *lysR2*, 6II Δ *lysR1*, 17II Δ *cbbO-m* by heat shock. Clones containing both vectors were selected on LB agar plates with chloramphenicol (12.5 $\mu\text{g}\cdot\text{ml}^{-1}$) addition for the selection on fosmids, with kanamycin (50 $\mu\text{g}\cdot\text{ml}^{-1}$) addition to select for the transposon insertion, and with ampicillin (100 $\mu\text{g}\cdot\text{ml}^{-1}$) addition for the selection on the pUC19 vector. To verify that both vectors contain inserts blue-white screening was also applied. Generated clones were then autoinduced for 18 hours at 37°C and applied for plasmid preparation (High-Speed Plasmid Mini Kit, Geneaid, New Taipei City, Taiwan). Clones were then digested with *Bam*H1 (Thermo Scientific) to evaluate the insert size of the pUC19 vector. pCC1FOS and pUC19 of verified clones were separated by agarose gel electrophoresis (0.8% TAE agarose gel, 120 V, 35 minutes) and re-extracted from the gel using the Gel/PCR DNA Fragments Extraction Kit (Geneaid). pUC19 vectors were sequenced with M13 forward and reverse sequencing primers to verify that the entire intact fragments were present in each case and pCC1FOS vectors were sequenced again with the KAN-2 FP-1 forward primer targeting the kanamycin cassette to verify that the insertion position is consistent. Verified clones were tested for RubisCO activity by following the HPLC-based activity assay as described before (see 2.5.1.2).

3 Results

3.1 (Meta)-genomic fosmid libraries

Five fosmid libraries were constructed: (i) one genomic library with DNA material from *T. crunogena* TH-55 and (ii) four metagenomic libraries with DNA isolated from chimney material of 'Sisters Peak', 'Mephisto' and 'Site B' as well as DNA isolated from hydrothermal fluid material of 'Drachenschlund'. Since DNA concentrations and purities of isolated metagenomic DNA did not meet the standards needed for the construction of fosmid libraries MDA was performed. DNA concentrations and purities of metagenomic DNA before and after MDA are summarized in Table 7.

Table 7: Concentrations and purities of metagenomic DNA before and after DNA.

sampling site	concentration before MDA [ng* μ l ⁻¹]	ratio _{260/280} before MDA	concentration after MDA [ng* μ l ⁻¹]	ratio _{260/280} after MDA
Sisters Peak, 5°S	113	1.42	840	1.89
Site B, 15°N	5	1.66	64	1.91
Mephisto, 5°S	9	1.39	774	1.64
Drachenschlund, 8°S	42	1.46	188	1.82

The number of generated clones as well as average insert sizes of the five constructed fosmid libraries are summarized in Table 8.

Table 8: Meta-(genomic) libraries constructed within this study.

sample origin	sample type	number of clones	average insert size
genomic library			
<i>T. crunogena</i> TH-55	pure culture	2 304	29 kb
metagenomic libraries			
<u>basalt-hosted vent system</u>			
Sisters Peak, 5°S	sulfide active venting chimney	14 000	39 kb
Mephisto, 5°S	sulfide active venting chimney	20 500	26 kb
<u>ultramafic-hosted vent system</u>			
Site B, 15°N	sulfide active venting chimney	17 000	31 kb
Drachenschlund, 8°S	diffuse fluids	20 200	37 kb

Blast search of sequences generated from end sequencing of all isolated fosmids from the *T. crunogena* TH-55 library showed sequential similarities to *T. crunogena* XCL-2 (Scott *et al.* 2006).

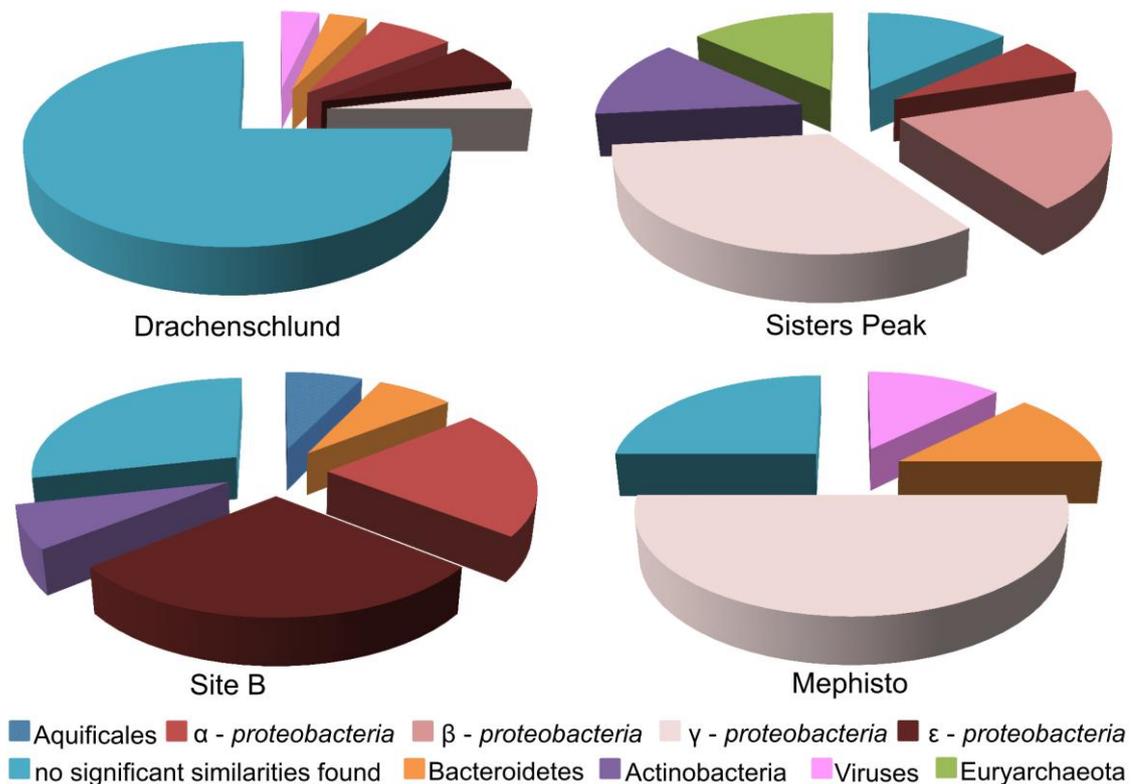


Figure 12: Classification of sequences derived from insert end-sequencing of selected fosmids from the four metagenomic libraries constructed within this study. Sequences were compared with those deposited in the NCBI database performing Blastn search.

The sequenced inserts from the metagenomic fosmid libraries resembled microorganisms commonly found in hydrothermal habitats. At 'Drachenschlund' library sequential similarities to e.g.: *T. crunogena* (Scott *et al.* 2006), *Sulfurospirillum* sp. (Campbell *et al.* 2006) or *Aquifex aeolicus* (Guiral *et al.* 2012) were found. Derived sequences from insert end-sequencing of all constructed metagenomic libraries were classified on the level of phyla, which is visualized in Figure 12. The results show that *Gamma*- and *Epsilon*proteobacteria are highly abundant at 'Sisters Peak', 'Site B' and 'Mephisto' which is typical for hydrothermal vent habitats (Perner *et al.* 2007a, Perner *et al.* 2014) (see Figure 12). However, the sequences from the fosmids of the 'Drachenschlund' library mainly show no significant similarities to sequences deposited in the NCBI database. This might be indicative for a high abundance of for now unknown microorganisms at this site.

3.2 Establishing a functional screen to seek RubisCOs from metagenomes

A HPLC-based approach was used to establish a new functional screen to seek RubisCO active fosmid clones from metagenomic libraries.

3.2.1 RubisCO activity of TH-55 cells and recombinant RubisCO versions

It was initially verified that the envisaged HPLC based approach in principle works to measure prokaryotic RubisCO activities. Here a well-known representative organism was investigated, namely *T. crunogena* TH-55. The specific RubisCO activity of *T. crunogena* TH-55 at different stages of growth (see Figure 9) was 210 ± 12 nmol 3-PGA*min⁻¹*mg⁻¹ after 15.5 hours of growth, 252 ± 19 nmol 3-PGA*min⁻¹*mg⁻¹ after 25 hours of growth and 228 ± 23 nmol 3-PGA*min⁻¹*mg⁻¹ after 47 hrs of growth (see Figure 13 and Appendix Table C 1).

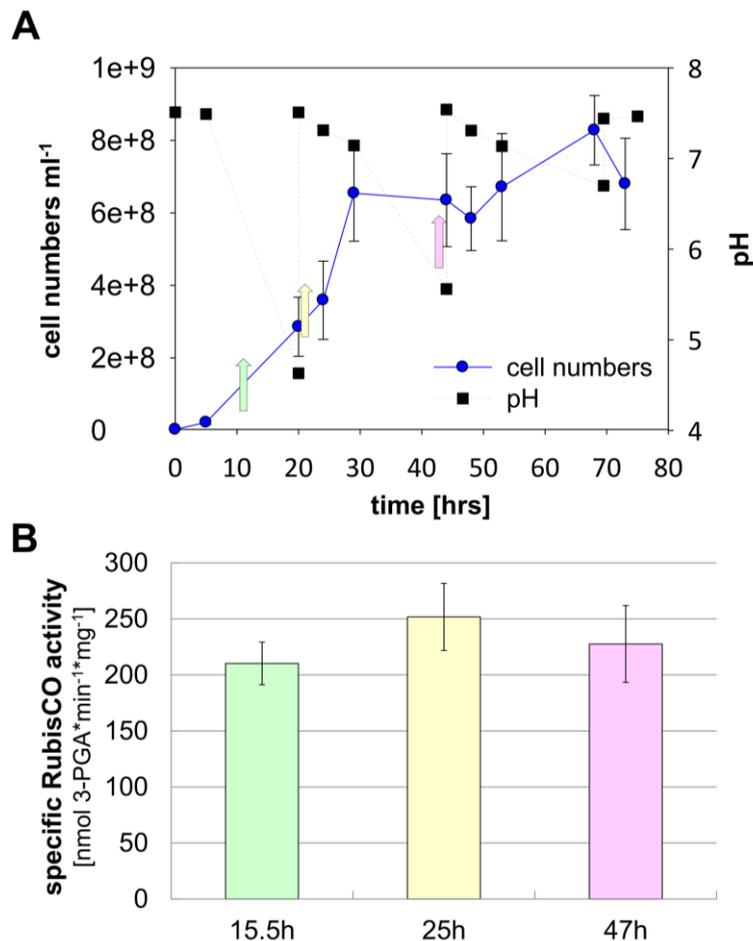


Figure 13: Specific RubisCO activity of TH-55 dependent on growth. (A) Growth curve of *T. crunogena* TH-55. Cell numbers and pH values are displayed over time. When the pH dropped it was re-adjusted to 7.8. Time points of activity measurements are indicated by colored arrows, where green is 15.5 h, yellow is 25 h and pink is 47 h. (B) Specific RubisCO activities of TH-55 crude extracts, measured after 15.5 h (green bar), 25 h (yellow bar) and 47 h (pink bar) of growth (Böhnke and Perner 2014, Supplementary Information).

These activities were in the same magnitude of what has been measured for other *Thiomicrospira* isolates (see Table 9), demonstrating that the conducted HPLC-based RubisCO assay in principle works for prokaryotic RubisCO enzymes.

Table 9: Comparison of specific RubisCO activities from different publications.

tested organism	analyses	purification state	specific activity [nmol*min ⁻¹ *mg ⁻¹]	reference
<i>T. crunogena</i>	radioactive	crude extract (not purified)	410 ± 40	(Scott <i>et al.</i> 2006)
<i>T. thermophila</i>	spectrophotometrical	purified	222 ± 40	(Takai <i>et al.</i> 2005)
<i>T. crunogena</i>	HPLC	crude extract (not purified)	252 ± 19	this study

To further validate the functioning of the RubisCO screen for recombinantly expressed versions, three different types of cloned TH-55's RubisCO genes and gene clusters were initially tested. The first one was a genomic fosmid clone (6F3) with a 38.1 kb large insert which contains the RubisCO gene cluster (*cbbQOM lysR2 lysR1 cbbLSQO*) and further 25.1 kb flanking DNA regions. The second cloned version was a fosmid clone where the small and large subunit of TH-55's RubisCO form I were inserted (pCC1FOS::*cbbLS*) and the third tested recombinant version was a fosmid clone whose insert encoded the large subunit of form II (pCC1FOS::*cbbM*). The measured specific RubisCO activities of all three versions are summarized in Figure 14A and Appendix Table C 1 and respective gene arrangements are visualized in Figure 14B.

The specific RubisCO activity of the *T. crunogena* TH-55 genomic fosmid clone 6F3 reached 455 ± 30 nmol 3-PGA*min⁻¹*mg⁻¹, which is nearly twice as much as what has been measured for the native *T. crunogena* TH-55's enzyme (252 ± 19 nmol 3-PGA*min⁻¹*mg⁻¹). This activity increase may be explained by the copy number of pCC1FOS vector in *E.coli* cells, which could range from 10 to 200 copies per cell after autoinduction (see manual for the CopyControl™ Fosmid Library Production Kit, epicentre®). Thus it is highly likely that more RubisCO is expressed in the autoinduced metagenomic fosmid clone than in the natural host *T. crunogena* TH-55.

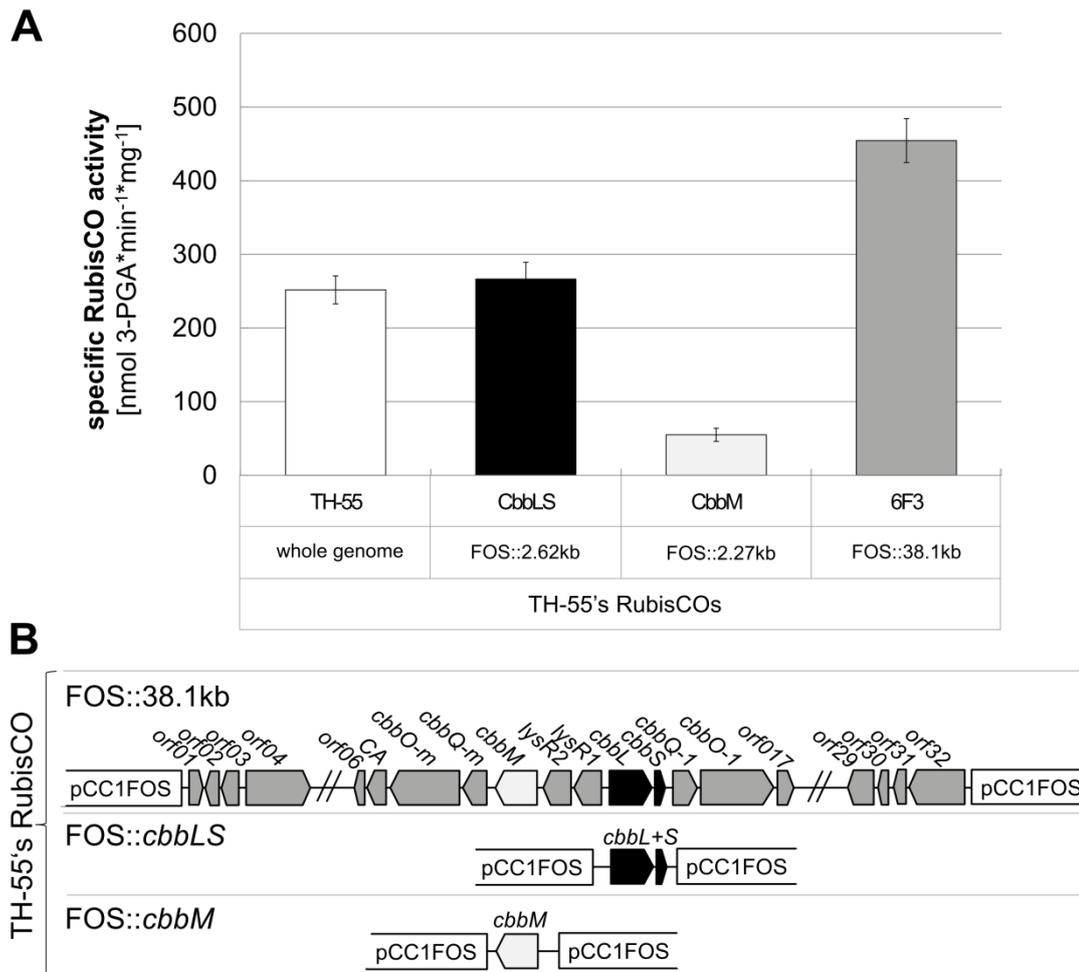


Figure 14: Specific RubisCO activities of TH-55 native and recombinant TH-55 RubisCOs visualized together with gene arrangements of recombinant versions. (A) Specific RubisCO activities measured from crude extracts of TH-55 (TH-55, grown for 25h), clone 6F3 (38.1 kb insert size) derived from a genomic fosmid library constructed with TH-55 DNA (FOS::38.1kb), a pCC1FOS clone containing TH-55's *cbbLS* structural genes (FOS::cbbLS) and a pCC1FOS clone containing TH-55's *cbbM* structural gene (FOS::cbbM). (B) Schematic representation of the gene arrangements of the DNA inserts. Identified open reading frames (ORF) are indicated as grey arrows and structural RubisCO genes are indicated in black or light grey, for *cbbLS* or *cbbM*, respectively, in the direction of transcription. ORFs were numbered serially from *orf01* to *orf32* (for details on annotations of respective ORFs see Table 11). Standard gene abbreviations were used (see Appendix Table A 2) (modified from Böhnke and Perner 2014).

RubisCO activity measurement with *T. crunogena* TH-55's form I RubisCO structural genes (pCC1FOS::cbbLS) individually expressed in *E. coli* reveals a specific RubisCO activity of 266 ± 23 nmol 3-PGA*min⁻¹*mg⁻¹, which equates to 59% of the activity measured for the genomic fosmid clone 6F3 (455 ± 30 nmol 3-PGA*min⁻¹*mg⁻¹), while the expression of *T. crunogena* TH-55's form II RubisCO structural gene (pcc1FOS::cbbM) individually in *E. coli* reaches with 55 ± 9 nmol 3-PGA*min⁻¹*mg⁻¹ only 12% of the total specific RubisCO activity of the genomic fosmid clone 6F3 ($455 \pm$

30 nmol 3-PGA*min⁻¹*mg⁻¹). RubisCO form I has a higher specificity for CO₂ vs. O₂ than form II, thus a better capability to discriminate between CO₂ and O₂ (Berg 2011). Consequently it is not surprising that the RubisCO measurements conducted under aerobic conditions lead to significantly higher RubisCO activities for recombinant RubisCO form I as compared to RubisCO form II. Nevertheless both activities of individually expressed form I (pCC1FOS::*cbbLS*) and form II (pCC1FOS::*cbbM*) RubisCOs together still only reach 71% of the total RubisCO activity measured for the genomic fosmid clone 6F3, harboring next to the RubisCO structural genes *cbbLS* and *cbbM* additional 33.2 kb flanking DNA regions. This suggested that further enzymes encoded on the flanking regions are of fundamental importance for the expression of a proper functioning RubisCO enzyme, which is capable to reach maximal activities. However, the initial question whether the HPLC-based screening approach is applicable through recombinantly expressed RubisCO versions of *T. crunogena* TH-55 was positively answered. The conducted experiments even demonstrated that the RubisCO screen is appropriated to measure RubisCO activities from fosmid clones with large DNA inserts (TH-55 genome clone 6F3) on a single scale and thus offer a positive control utilizable for all further applications.

3.2.2 Up-scaling

Within this study a high number of clones is in vision to be screened for RubisCO activity and this is hardly manageable if using the single scale screening approach. For that reason it was tested if the specific RubisCO activity of the genomic fosmid clone 6F3 is still detectable when it is grown in a pool together with 11, 23, 47 or 95 inactive clones (always 6F8) (see Figure 10). It was shown, that pools of 96 and 48 clones are unsuitable, because the decrease of the Rubp peak as well as the formation of the 3-PGA peak was not detectable anymore. However, a clear detection of RubisCO activity was observed when 12 or 24 clones were pooled (see Figure 15). Since a high number of clones are proposed to be tested for RubisCO activity, all subsequent screenings were performed with the largest possible pool and thus with the pool of 24 fosmid clones. These up-scaling experiments furthermore make obvious, that the functional RubisCO screen was successfully established and is now available as a powerful tool to seek novel RubisCOs from metagenomes, even if the native host is not cultivable.

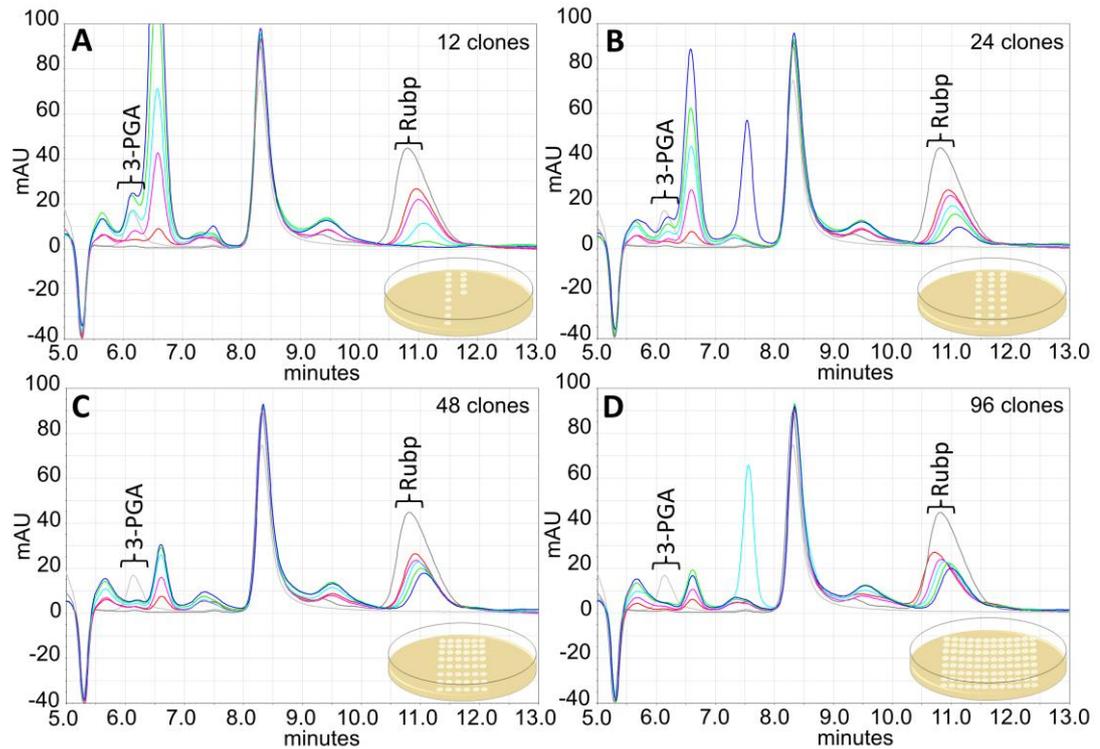


Figure 15: HPLC analyses of differently sized metagenomic pools. The crude extracts of 12 (A); 24 (B); 48 (C) and 96 (D) pooled fosmid clones were measured to determine whether the RubisCO activity of one active clone among 11, 23, 47 and 95 other clones is still detectable. Standards of the ribulose-1,5-bisphosphate (Rubp) and 3-*D*-phosphoglycerate (3-PGA) were measured before and after sample analyses to observe the correlated peak area and retention times of the substrates. Chromatogram graphs of standards are indicated in grey and observed retention times are denoted by brackets. Subsamples taken at different time points are color-coded, where $t=0$ minutes is red, $t=10$ minutes is pink, $t=30$ minutes is light blue, $t=60$ minutes is green and $t=120$ minutes is dark blue (Böhnke and Perner 2014, Supplementary Information).

3.3 Seeking RubisCOs from metagenomic libraries

Four metagenomic libraries that have been constructed with metagenomic DNA isolated from either chimney material of ‘Sisters Peak’, ‘Mephisto’ and ‘Site B’ or from hydrothermal fluid material of ‘Drachenschlund’ (see Table 8) were searched for RubisCO active fosmid clones by using the self-developed RubisCO screen. Furthermore two other already existing metagenomic libraries, i.e. (i) ‘Lilliput’ and (ii) ‘Irina II’, were investigated and also searched for RubisCO active clones using the newly established RubisCO screen. These two libraries were constructed with metagenomic DNA isolated from low temperature, diffuse hydrothermal fluids which were collected from the hydrothermal vent fields (i) Lilliput at 9°S (Perner *et al.* 2007b) and (ii) Irina II at 15°N (Perner *et al.* 2013a) located along the Mid-Atlantic Ridge (MAR).

Table 10: Numbers of screened clones and corresponding numbers of positive tested ones.

metagenomic libraries	total number of clones	number of clones tested	positive tested clones
<u>basalt-hosted vent system</u>			
Sisters Peak, 5°S	14,000	1,344	0
Mephisto, 5°S	20,500	288	0
Lilliput, 9°S	15,000	2,016	0
<u>ultramafic-hosted vent system</u>			
Site B, 15°N	17,000	1,824	0
Irina II, 15°N	18,000	1,920	0
Drachenschlund, 8°S	20,200	1,152	12

Although six different metagenomic libraries were screened for RubisCO active fosmid clones only one library revealed hits (see Table 10). From the 1,152 screened fosmid clones of the ‘Drachenschlund’ library, twelve fosmid clones were recovered, exhibiting RubisCO activities ranging from 427 ± 16 to 484 ± 44 nmol 3-PGA*min⁻¹*mg⁻¹ (see Figure 16 and Appendix Table C 1). Restriction analyses showed that these twelve clones had approximately 35 kb DNA inserts and shared the same restriction pattern (data not shown) suggesting that the fosmid clones were highly similar to each other. Sequencing supported this and even demonstrated a 100% identity on DNA level for fragment A (3.1 kb), fragment B (13 kb) and fragment C (2.9 kb) (see Figure 16). All three fragments shared a high sequential similarity to *Thiomicrospira crunogena* XCL-2 (90% to 96%) (Scott *et al.* 2006). Genes of fragment A were similar to genes of hypothetical proteins and an ATPase family associated enzyme (90% DNA sequence similarity).

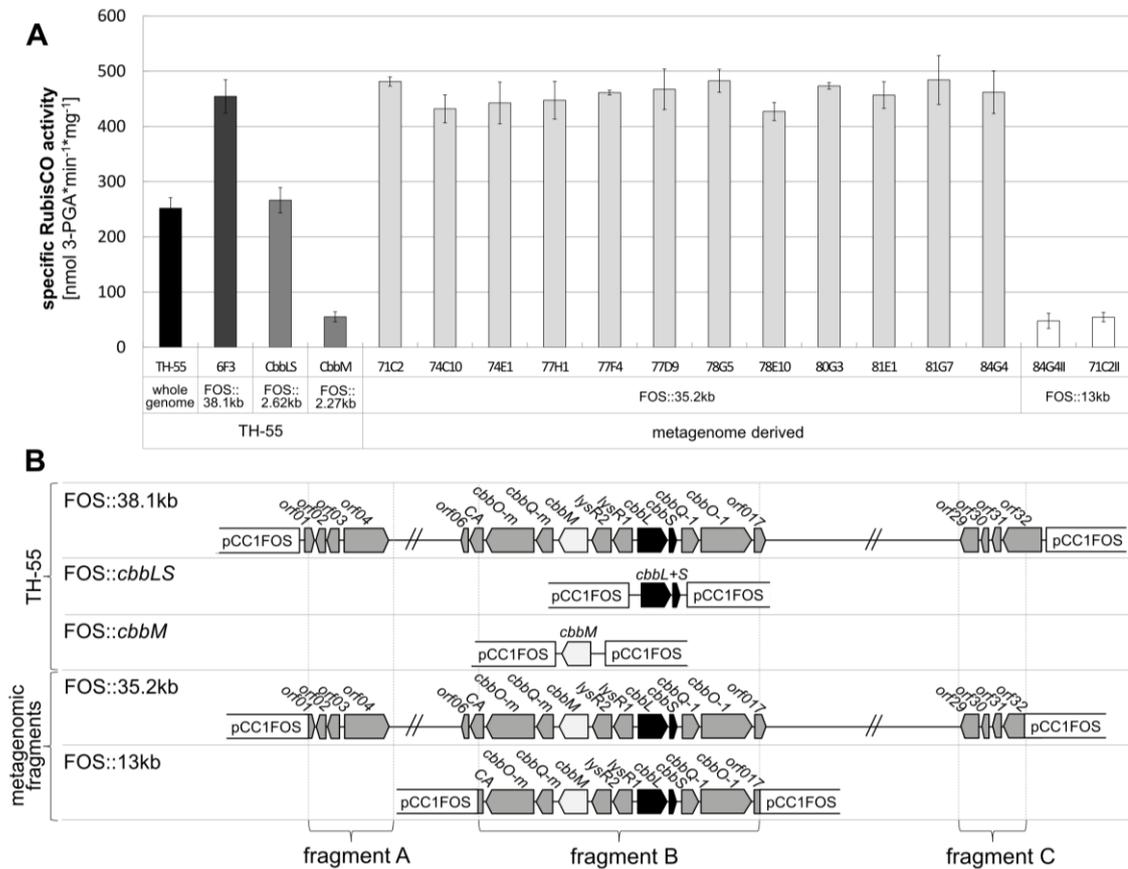


Figure 16: Specific RubisCO activities of TH 55’s native and recombinant RubisCOs compared with metagenome derived recombinant RubisCOs as well as respective gene arrangements. (A) Specific RubisCO activities measured from crude extracts of TH 55 (TH 55, grown for 25h), clone 6F3 (38.1 kb insert size) derived from a genomic fosmid library constructed with TH-55 DNA (FOS::38.1kb), a pCC1FOS clone containing TH-55’s *cbbLS* structural genes (FOS::cbbLS), a pCC1FOS clone containing TH-55’s *cbbM* structural gene (FOS::cbbM), fosmid clones derived from the ‘Drachenschlund’ metagenome library (71C2, 74C10, 74E1, 77H1, 77F4, 77D9, 78G5, 78E10, 80G3, 81E1, 81G7 and 84G4, summarized as FOS::35.2 kb) and fosmid subclones (84G4II and 71C2II) containing the RubisCO gene cluster (fragment B; 13 kb) (summarized as FOS::13 kb). (B) Schematic representation of the gene arrangements of the DNA inserts. Identified open reading frames (ORF) are indicated as grey arrows and structural RubisCO genes are indicated in black or light grey, for *cbbLS* or *cbbM*, respectively, in the direction of transcription. ORFs were numbered serially from *orf01* to *orf32* (for details on annotations of respective ORFs see Table 11. Standard gene abbreviations were used (see Appendix Table A 2) Fragments A, B and C denote regions which were sequenced for all twelve RubisCO active metagenome derived fosmid clones sought from the ‘Drachenschlund’ library through the function based RubisCO screen (Böhnke and Perner 2014).

Genes of fragment B resembled those from the RubisCO gene cluster of *T. crunogena* XCL-2 (96% DNA sequence similarity): the *cbbMQO* and *cbbLSQO* clusters were juxtaposed and were separated by two transcriptional regulators (*lysR1* and *lysR2*), which were located in the same orientation as the *cbbMQO* genes (see Figure 16). Genes of fragment C resembled genes encoding an Acetyl-CoA carboxylase, a biotin

Results

carboxyl carrier protein and a 3-Dehydroquinate dehydratase (94% DNA sequence similarity) (for all identified genes see Figure 16 and Table 11).

Table 11: Open reading frames (ORFs) identified on the metagenomic fragment (Böhnke and Perner 2014, Supplementary Information).

ORF	aa	function ^[1]	identity [%]	accession
<i>orf01</i>	partial/189	hypothetical protein Tcr0411	-	YP_390681
<i>orf02</i>	206	conserved hypothetical protein Tcr0412	89	YP_390682
<i>orf03</i>	270	hypothetical protein Tcr0413	91	YP_390683
<i>orf04</i>	854	ATPaseAAA	99	YP_390684
<i>orf05</i>	385	YeeE/YedE family protein	99	YP_390689
<i>orf06</i>	135	hypothetical protein Tcr0420	97	YP_390690
<i>orf07</i>	218	Carbonate dehydratase (CA)	100	YP_390691
<i>orf08</i>	757	von Willebrand factor, type A (<i>cbbO-m</i>)	99	YP_390692
<i>orf09</i>	266	ATPase, AAA-type (<i>cbbQ-m</i>)	99	YP_390693
<i>orf10</i>	459	RubisCO form II, large subunit (<i>cbbM</i>)	99	YP_390694
<i>orf11</i>	314	transcriptional regulator (<i>lysR2</i>)	100	YP_390695
<i>orf12</i>	308	transcriptional regulator (<i>lysR1</i>)	98	YP_390696
<i>orf13</i>	472	RubisCO form I, large subunit (<i>cbbL</i>)	99	YP_390697
<i>orf14</i>	116	RubisCO form I, small subunit (<i>cbbS</i>)	98	YP_390698
<i>orf15</i>	272	ATPase (<i>cbbQ-1</i>)	100	YP_390699
<i>orf16</i>	777	von Willebrand factor, type A (<i>cbbO-1</i>)	99	YP_390700
<i>orf17</i>	177	hypothetical protein Tcr0431	98	YP_390701
<i>orf18</i>	408	hypothetical protein Tcr0432	97	YP_390702
<i>orf19</i>	233	ABC transporter	99	YP_390703
<i>orf20</i>	425	hypothetical protein Tcr0434	98	YP_390704
<i>orf21</i>	431	hypothetical protein Tcr0435	99	YP_390705
<i>orf22</i>	260	hypothetical protein Tcr0436	98	YP_390706
<i>orf23</i>	91	hypothetical protein Tcr0437	96	YP_390707
<i>orf24</i>	430	Phosphoribosylamine-glycine ligase (<i>purD</i>)	98	YP_390708
<i>orf25</i>	520	AICARFT_IMPCHas (<i>purH</i>)	98	YP_390709
<i>orf26</i>	87	Fis family transcriptional regulator	100	YP_390710
<i>orf27</i>	320	dihydrouridine synthase TIM-barrel protein (<i>nifR3</i>)	98	YP_390711
<i>orf28</i>	294	50S ribosomal protein L11 Methyltransferase (<i>prmA</i>)	97	YP_390712
<i>orf29</i>	449	biotin carboxylase	99	YP_390713
<i>orf30</i>	150	biotin carboxyl carrier protein	99	YP_390714
<i>orf31</i>	144	3-Dehydroquinate dehydratase	100	YP_390715
<i>orf32</i>	partial/731	Protein-disulfide reductase	-	YP_390716

^[1] Functions are deduced from XCL-2 annotations exhibiting the highest similarities (Scott *et al.* 2006).

Furthermore the whole fosmid insert of one clone of all duplicative twelve clones was sequenced exemplarily, which illustrates that the genes encoded on the metagenomic DNA insert in principle follows the arrangement of the homologous genes from XCL-2 (Tcr0411 to Tcr0446) (Scott *et al.* 2006). Blast search reveal a DNA similarity of 96% but disclose a lack of a cascade of four genes, namely an ApbE-like lipoprotein (Tcr0416), a hypothetical protein with a FMN-binding domain (Tcr0417) as well as two hypothetical proteins with unknown function (Tcr0415 and Tcr0418) (see Figure 17).

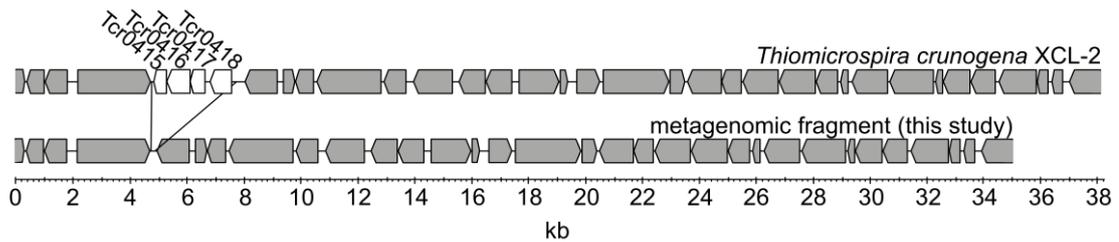


Figure 17: Gene arrangement of ORFs encoded on the metagenomic fragment. The genes encoded on the insert of the ‘Drachenschlund’ metagenomic fosmid clone were shown parallel to the respective area from the closest relative *T. crunogena* XCL-2. Genes are indicated as arrows in the direction of transcription. Grey arrows denote genes present in both gene arrangements and white arrows display genes only present in XCL-2, but not in the ‘Drachenschlund’ metagenomic fosmid clone, namely: Tcr0415: hypothetical protein; Tcr0416: ApbE like lipoprotein; Tcr0417: flavin mononucleotide (FMN) binding protein and Tcr0418: hypothetical protein (Böhnke and Perner 2014, Supplementary Information).

The nucleotide sequence of the entire 35.2 kb metagenomic DNA insert was submitted to the GenBank database under the accession number KJ639815.

3.4 Working with RubisCOs from metagenomes

3.4.1 Transposon insertion libraries

As sequencing has revealed the RubisCO gene cluster consists next to the RubisCO structural genes (*cbbM* and *cbbLS*) of genes which are believed to be responsible for RubisCO regulation, activation or assembling (*cbbO-m*, *cbbQ-m*, *lysR1*, *lysR2*, *cbbO* and *cbbQ*) (Hayashi *et al.* 1997, Toyoda *et al.* 2005). In order to simplify subsequent investigations on this RubisCO gene cluster, flanking genes which are not believed to be associated with RubisCO functioning are intended to be removed. Therefore the respective gene cluster (13 kb) was subcloned exemplarily for two metagenomic fosmid clones (71C2 and 84G4). Subsequently RubisCO activities of both subclones, designated 71C2II and 84G4II, were measured with 55 ± 8 nmol 3-PGA*min⁻¹*mg⁻¹ and 48 ± 14 nmol 3-PGA*min⁻¹*mg⁻¹, respectively (Figure 16, Appendix Table C 1). The unexpected significant activity decrease by a factor of around five suggests that there is at least one gene outside of the RubisCO gene cluster whose gene product influences RubisCO functioning. To identify responsible gene(s) a transposon mutant library was constructed with the fosmid clone 71C2, carrying the entire metagenomic DNA fragment with the RubisCO gene cluster and flanking DNA regions encoded on (35.2 kb). To further discover whether *cbbO-m*, *cbbQ-m*, *lysR1*, *lysR2*, *cbbO* and *cbbQ* have a share in maximizing recombinant RubisCO activity a second transposon mutant library was constructed with the fosmid subclone 71C2II, harboring the RubisCO gene cluster without any flanking DNA region (13 kb). Exact insertion positions were identified for 384 transposon clones over all, but RubisCO activities were measured only for 46 selected ones (for methodological details see 2.7.3 and for summarized results see Figure 18, Table 12 and Table 13).

3.4.1.1 Transposon insertions outside of the RubisCO gene cluster

Specific RubisCO activities of 25 transposon clones with insertions outside of the 13 kb RubisCO gene cluster were measured, revealing seven clones with significantly lowered RubisCO activities relative to the original, intact 71C2 clone (see Figure 18A and B). Here, transposon clone 96 exhibited the most dramatic activity loss (see Figure 18A and B as well as Table 12). The gene in which the kanamycin cassette was inserted (insertion at 6 of 135 aa) was designated *orf06*. No putative conserved domains could be detected when performing Blastx search, but similarities to several hypothetical proteins of different *Beta-*, *Gamma-* and *Zetaproteobacteria* were found (see Figure 19), with highest resemblance to gene Tcr0420 of *T. crunogena* XCL-2 (97% aa similarity) (see Table 11).

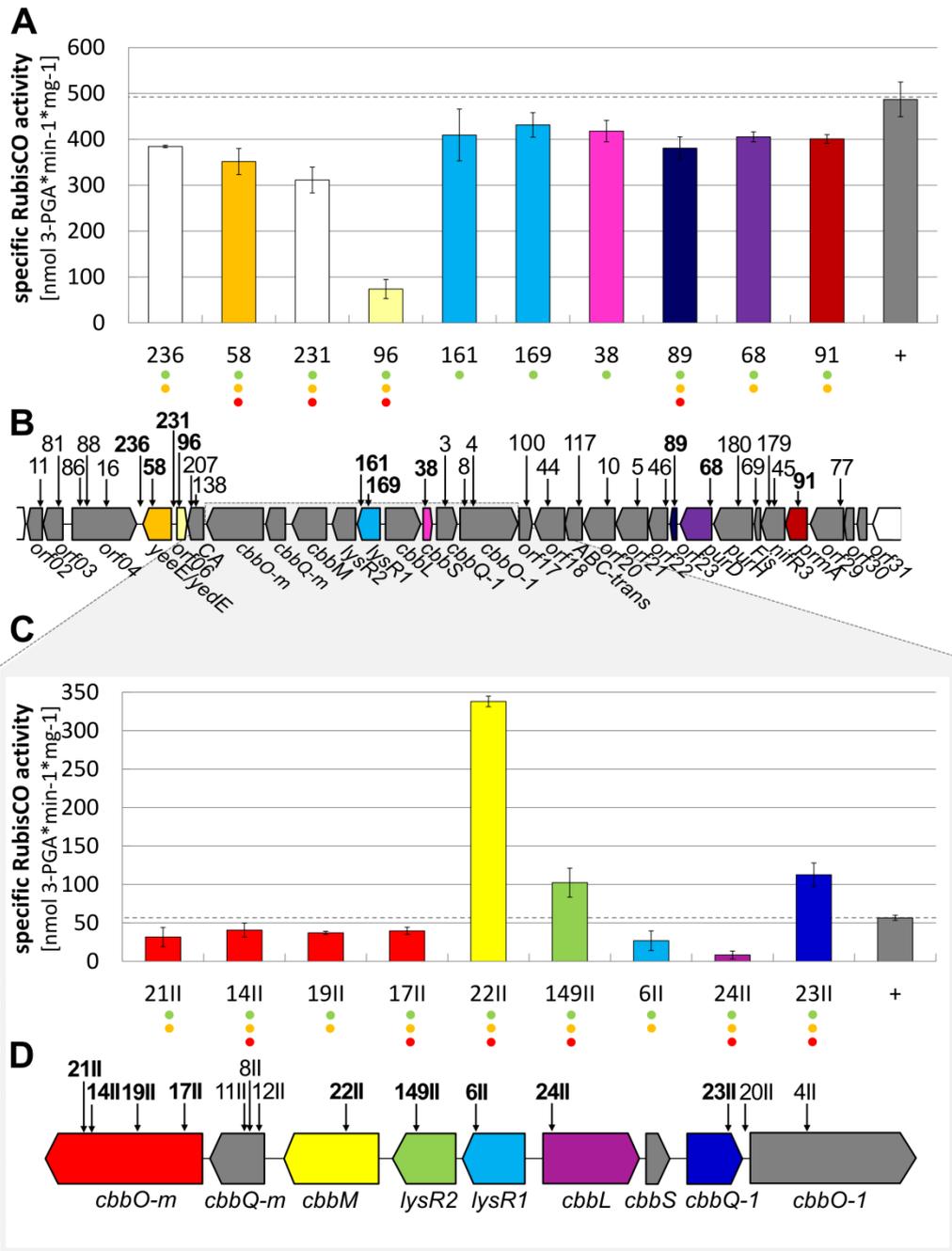


Figure 18: Specific RubisCO activities and insertion positions of tested transposon clones. Specific RubisCO activities (A) and schematic gene arrangement (B) of transposon clones carrying the entire 35.2 kb insert derived from the metagenomic fosmid clone 71C2. Specific RubisCO activities (C) and schematic gene arrangement (D) of transposon clones carrying the 13 kb insert derived from the fosmid clone 71C2II. Only specific RubisCO activities for transposon clones with significant differences to the intact 35.2 kb (A) or 13 kb (C) fosmid clones (intact fosmid clones indicated as +) are displayed. Identified ORFs were numbered consecutively, designated *orf01* to *orf32* (for detailed abbreviations see Table 11) and indicated as arrows in the direction of transcription. Insertion sites are denoted by vertical black arrows and transposon clone numbers. Clones exhibiting significant changes in RubisCO activity relative to the intact clones are highlighted in bold. Genes with significant activity change are color coded according to the bars in A or C. The level of significance is denoted by dots under the activity where green is 0.05 (95%), yellow is 0.01 (99%) and red is 0.001 (99.9%) (Böhnke and Perner 2014).

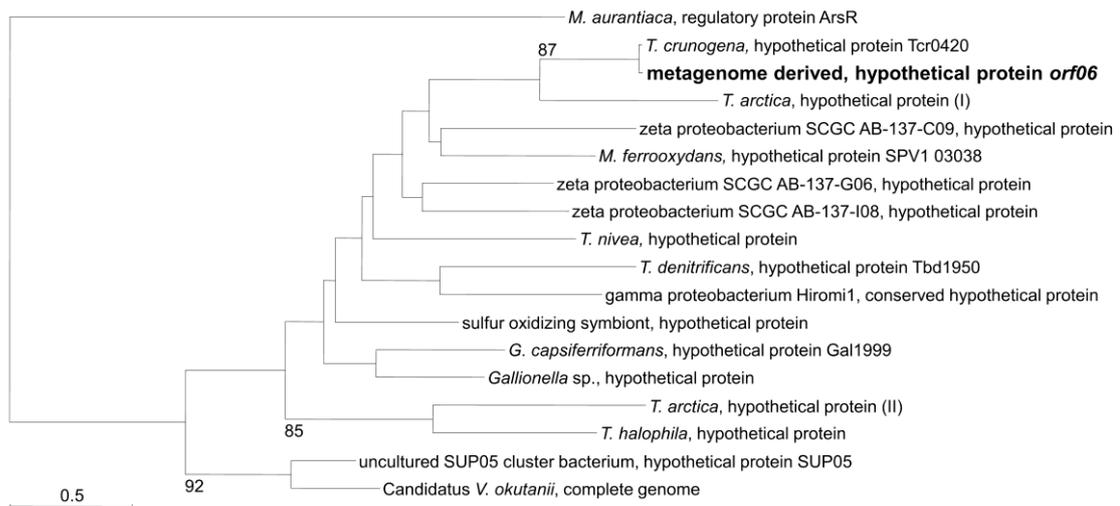


Figure 19: Phylogenetic relationship of *orf06*. The phylogenetic tree calculated for the amino-acid sequence of *orf06* and its closest relatives, using Maximum-Likelihood analyses. Bootstrap values, calculated for 100 replicates, are presented as percentages at the node and are indicated only when above 70%. Abbreviations and accession numbers of shown sequences are listed in Appendix Table A 3. The scale bar represents the expected number of changes per amino acid position (Böhnke and Perner 2014, Supplementary Information).

Further transposon clones with insertions outside of the RubisCO gene cluster and significantly decreased specific RubisCO activities were: clone 236 (insertion located in the non-coding region between *orf04* and *orf05*), clone 58 (Δ *orf05*), clone 231 (insertion located in the non-coding region upstream of *orf06*), clone 89 (Δ *orf23*), clone 68 (Δ *orf24*) and clone 91 (Δ *orf28*) (see Figure 18, for further details on similarity of ORFs to genes of *T. crunogena* XCL-2 see Table 11; for insertion positions on fosmids of transposon clones and corresponding specific RubisCO activities see Table 12).

Table 12: Insertion positions of tested transposon clones with 35.2 kb inserts and respective RubisCO activities (Böhnke and Perner 2014, Supplementary Information).

ORF	clone number	insertion position [aa]	total orf length [aa]	specific RubisCO activity [nmol 3-PGA*min ⁻¹ *mg ⁻¹]
TH-55	/	/	/	252 ± 19
pCC1FOS::35.2 kb ^[1]	/	/	/	481 ± 8
<i>orf02</i>	11	28	206	470 ± 12
<i>orf03</i>	81	52	270	433 ± 14
<i>orf04</i>	86	246	854	491 ± 28
	88	290		450 ± 30
	16	499		451 ± 0.1
<i>ncr orf04-05</i>	236	448 ^[2]	450 ^[3]	384 ± 3
<i>orf05</i>	58	237	385	351 ± 28
<i>ncr orf05-06</i>	231	34 ^[2]	200 ^[3]	311 ± 28
<i>orf06</i>	96	6	135	81 ± 20
<i>orf07</i>	138	139	218	522 ± 27
	207	212		437 ± 40
<i>orf12 (lysR1)</i>	169	220	308	432 ± 27
	161	285		409 ± 56
<i>orf14 (cbbS)</i>	38	111	116	418 ± 23
<i>orf15 (cbbQ-1)</i>	3	180	272	455 ± 19
<i>orf16 (cbbO-1)</i>	8	62	777	511 ± 21
	4	157		482 ± 7
<i>orf17</i>	100	143	177	460 ± 1
<i>orf18</i>	44	329	408	450 ± 1
<i>orf19</i>	117	98	233	447 ± 9
<i>orf20</i>	10	128	425	435 ± 13
<i>orf21</i>	5	245	431	508 ± 10
<i>orf22</i>	46	18	260	463 ± 37
<i>orf23</i>	89	52	91	380 ± 25
<i>orf24</i>	68	10	430	405 ± 11
<i>orf25</i>	180	397	520	444 ± 10
<i>orf26</i>	69	87	87	449 ± 8
<i>orf27</i>	179	22	320	459 ± 18
	45	151		438 ± 10
<i>orf28</i>	91	147	294	401 ± 9
<i>orf29</i>	77	21	449	444 ± 32

^[1] intact metagenome derived fosmid containing the RubisCO gene cluster and flanking DNA

^[2] The insertion position in non-coding region (*ncr*) is given in nucleotides.

^[3] The total length of non-coding region (*ncr*) is given in nucleotides.

3.4.1.2 Transposon insertions within the RubisCO gene cluster

The specific RubisCO activities of 21 transposon clones with insertions inside of the RubisCO gene cluster were measured resulting in nine transposon clones with significantly decreased and four transposon clones with significantly increased RubisCO activities when comparing their RubisCO activities to the activity of the respective intact metagenome derived clones 71C2 (pCC1FOS::35.2 kb) and 71C2II (pCC1FOS::13 kb) (see Figure 18, Table 12 and Table 13).

Four transposon clones among these with significantly changed RubisCO activities exhibit insertions scattered across the two LysR family transcriptional regulators *lysR2* and *lysR1*. Transposon clone 149II with an insertion at aa position 271 (of 315 aa) of *lysR2* (*orf11*) for instances display a significantly increased RubisCO activity (see Figure 18C and D), while three insertions in *orf12* (*lysR1*) at the aa positions (i) 220, (ii) 264 and (iii) 285 (of total 308 aa) led to a significant loss of activity for clones (i) 169, (ii) 6II and (iii) 161, respectively (see Figure 18, Table 12 and Table 13).

Twelve selected transposon clones holding insertions in *orf08* (*cbbO-m*), *orf09* (*cbbQ-m*), *orf16* (*cbbO-1*) and *orf15* (*cbbQ-1*) were furthermore tested. The four transposon clones 17II, 19II, 14II and 21II with insertions scattered across *orf08* (*cbbO-m*) display significantly decreased RubisCO activities (Figure 18C, D), but no changes in RubisCO activity were observed when *orf09* (*cbbQ-m*) (transposon clones 8II, 11II and 12II) or *orf16* (*cbbO-1*) (transposon clones 4, 8 and 4II) were impaired. The insertions in *orf15* (Δ *cbbQ-1*) of (i) clone 23II harboring the RubisCO gene cluster (13 kb) and (ii) clone 3 equipped with the entire metagenomic DNA fragment (35.2 kb) offer contrasting results. Thus clone 23II show significantly increased activities, while the RubisCO activity of transposon clone 3 remained unchanged (see Figure 18, Table 12 and Table 13).

Additionally, three transposon clones with insertions in the RubisCO structural genes *cbbS*, *cbbL* and *cbbM* were investigated. The specific RubisCO activity was significantly lowered when RubisCO form I structural genes, *cbbL* (clone 24II) or *cbbS* (clone 38), were impaired (see Figure 18). By contrast an insertion in the RubisCO form II structural gene *cbbM* (clone 22II) resulted in an increase of RubisCO activity by a factor of 6.2 relative to the intact version of clone 71C2II (pCC1FOS::13 kb) (see Figure 18, Table 12 and Table 13).

An insertion in the transposon clone 7II also resulted in a roughly six-fold higher RubisCO activity (301 ± 21 nmol 3-PGA*min⁻¹*mg⁻¹) compared to the intact metagenomic clone 71C2II (pCC1FOS::13 kb) (see Figure 20A). Sequencing of this clone revealed, that the kanamycin cassette was inserted in the *lysR2* gene at position

Results

185 aa (of 315 aa in total), but that the adjacent genes *cbbM* and *cbbQ-m* as well as a part of the *cbbO-m* gene were cut out (see Figure 20B).

Table 13: Insertion positions of tested transposon clones with 13 kb inserts and respective RubisCO activities (Böhnke and Perner 2014, Supplementary Information).

ORF	clone number	insertion position [aa]	total orf length [aa]	specific RubisCO activity [nmol 3-PGA*min ⁻¹ *mg ⁻¹]
TH-55	/	/	/	252 ± 19
pCC1FOS::13 kb ^[1]	/	/	/	55 ± 8
<i>orf08 (cbbO-m)</i>	17II	189	757	40 ± 5
	19II	333		41 ± 9
	14II	569		37 ± 2
	21II	578		32 ± 13
<i>orf09 (cbbQ-m)</i>	12II	46	266	55 ± 5
	8II	94		59 ± 3
	11II	109		63 ± 11
<i>orf10 (cbbM)</i>	22II	171	459	338 ± 7
<i>orf11 (lysR2)</i>	149II	271	314	102 ± 19
<i>orf12 (lysR1)</i>	6II	264	308	31 ± 12
<i>orf13 (cbbL)</i>	24II	41	472	8 ± 5
<i>orf15 (cbbQ-1)</i>	23II	229	272	113 ± 15
ncr <i>orf15-16</i>	20II	85 ^[2]	105 ^[3]	66 ± 9
<i>orf16 (cbbO-1)</i>	4II	303	777	63 ± 7

^[1] metagenome derived intact version containing only RubisCO gene cluster

^[2] insertion positions in non-coding regions (ncr) given in nucleotides

^[3] total length of non-coding region (ncr) given in nucleotides

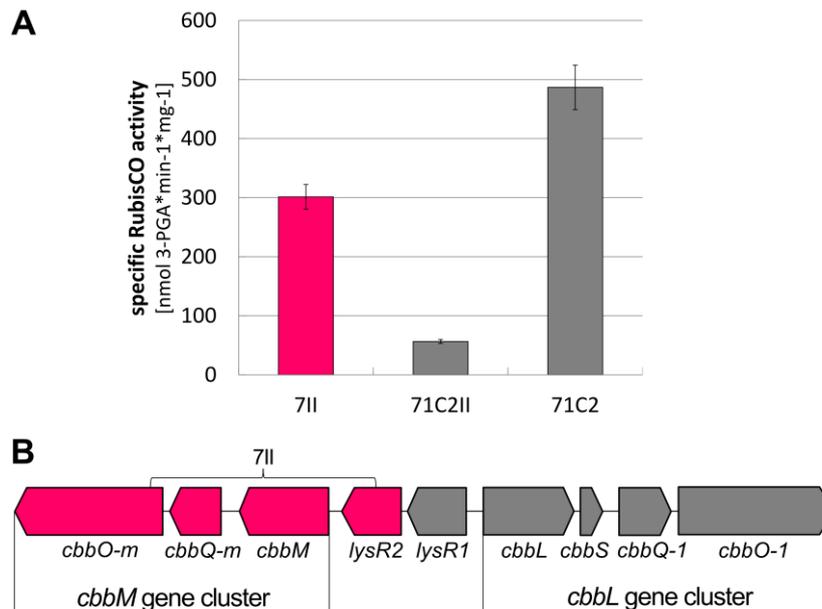


Figure 20: Specific RubisCO activity of transposon clone 7II ($\Delta cbbO-m$ to *lysR2*) and respective gene arrangement. Specific RubisCO activity (A) and schematic gene arrangement (B) of transposon clone 7II carrying the *cbbL* gene cluster derived from the fosmid clone 71C2II, but lack the *cbbM* gene cluster inclusively the *lysR2* gene. The deleted gene region is denoted by a bracket and impaired genes are color-coded in pink just as the corresponding bar displaying the specific RubisCO activity. RubisCO activities of intact 13 kb (71C2II) or 35.2 kb (71C2) fosmid clones are shown in grey (A). Identified ORFs were indicated as arrows in the direction of transcription and standard gene abbreviations were used (for details see Appendix Table A 2).

3.4.2 Transcription experiments with TH-55

TH-55's RNA was investigated in order to determine whether RubisCO associated genes were co-transcribed or not. Results are summarized in Figure 21, but were made on the basis of the agarose gels shown in Figure 22. Thus, it has been shown that in TH-55 the *orf06* homolog was transcribed separately from the Carbonic anhydrase and the YeeE/YedE family protein. Furthermore, it was observed that *cbbLS* was transcribed separately from *cbbM*, but that (i) *cbbQ-m* and *cbbO-m* as well as (ii) *cbbO-1* and *cbbQ-1* were co-transcribed, respectively. However, neither *cbbQO-m* nor *cbbQO-1* were co-transcribed with *cbbM* or *cbbLS*, respectively. Additionally this transcription experiment suggests that *lysR2*, *lysR1* and *cbbLS* were co-transcribed.

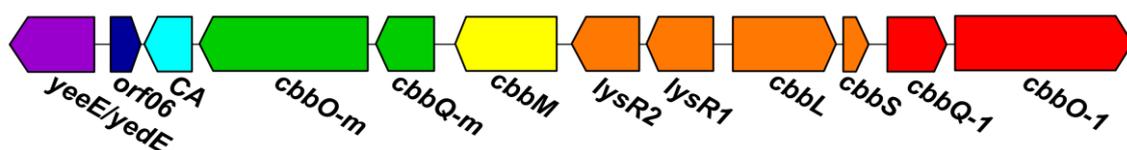


Figure 21: Transcription of RubisCO encoding genes and flanking gene regions from TH-55. Genes are indicated as arrows directed towards the way of transcription. Co-transcribed genes are displayed in the same color. Abbreviations are listed in Appendix Table A 2 (Böhnke and Perner 2014).

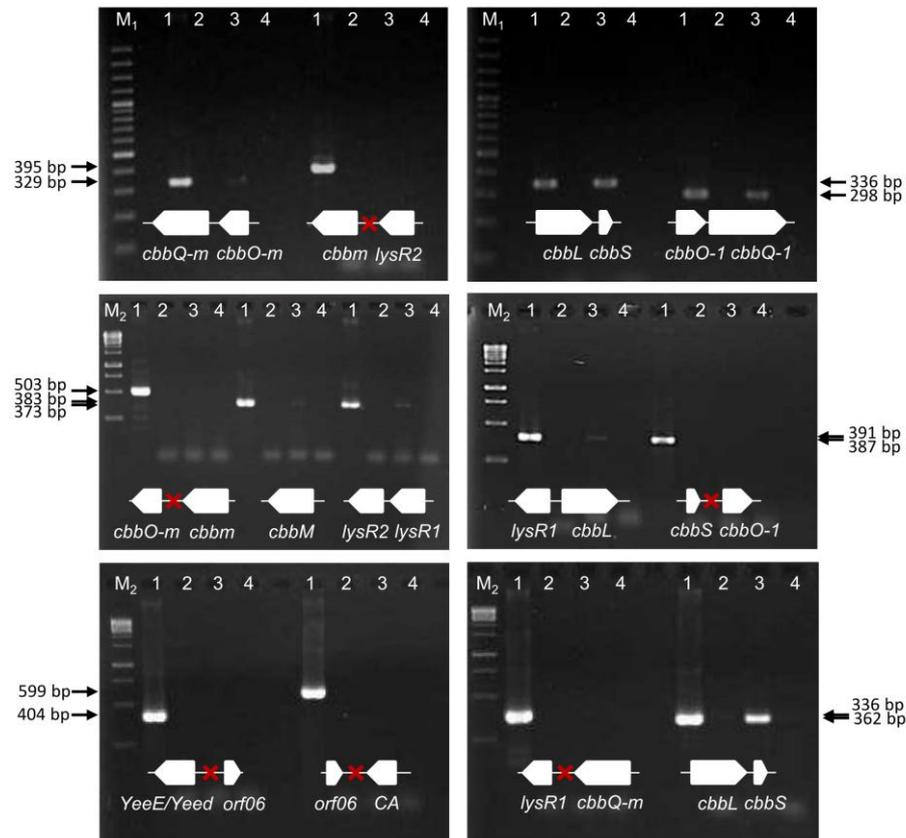


Figure 22: (Co)-transcription of the RubisCO gene cluster. Different gene to gene interspaces of the RubisCO gene cluster of TH-55 were amplified to elucidate which genes are co-transcribed. Amplificates were separated by agarose gel electrophoresis (1 x concentrated 2% TBE agarose gel at 120 V for 20 min). Abbreviations are as follows: M₁ – GeneRuler 100 bp DNA Ladder (Thermo Scientific), M₂ – GeneRuler 1 kb DNA Ladder (Thermo Scientific), 1 - positive control, where the DNA isolated from TH-55 was used as PCR template, 2 - negative control, where the corresponding template lacking reverse transcriptase was amplified, 3 - the cDNA of TH-55 was applied, 4 - negative amplification control of the Master Mix without any template (Böhnke and Perner 2014, Supplementary Information).

3.4.3 Experiments pairing transcription with RubisCO activity

The five transposon clones exhibiting the most dramatic activity changes, excluding those with insertions in RubisCO structural genes, were investigated to elucidate whether corresponding genes are involved in transcriptional processes or not. Therefore the transcription amount of RubisCO form I (*cbbL*) and II (*cbbM*) of clones 96 ($\Delta orf06$), 149II ($\Delta lysR2$), 6II ($\Delta lysR1$), 17II ($\Delta cbbO-m$) and 7II ($\Delta cbbO-m$ to *lysR2*) were compared with those of unimpaired, original fosmid clones 71C2II (pCC1FOS::13 kb) and 71C2 (pCC1FOS::35.2 kb). Corresponding agarose gels of ‘real’ transposon clones are summarized in Figure 23, while those of the accidentally generated clone 7II were shown in Figure 24. Agarose gels of the single inserted transposon clones showed, that transcription of *cbbL* and *cbbM* in transposon clone 96 ($\Delta orf06$) is down-regulated.

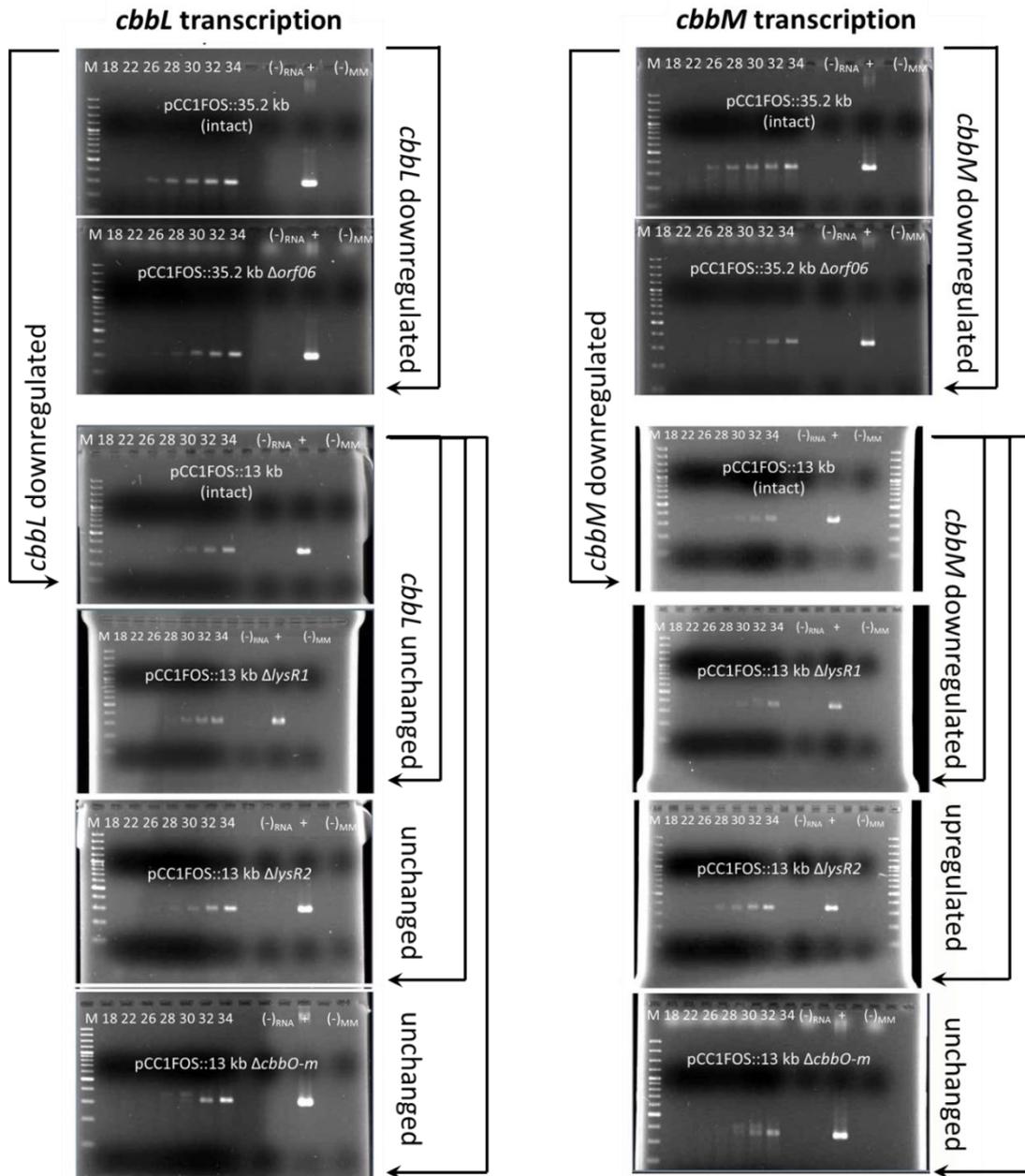


Figure 23: Transcription of *cbbL* and *cbbM* in transposon clones with deletions in *orf06*, *lysR1*, *lysR2* and *cbbO-m*. The amount of *cbbL* and *cbbM* amplified within 18, 22, 26, 28, 30, 32 and 34 PCR cycles of different RubisCO active transposon clones, namely clone 96 with pCC1FOS::*Δorf06*; clone 149II with pCC1FOS::*ΔlysR2*, clone 6II with pCC1FOS::*ΔlysR1* and clone 17II with pCC1FOS::*ΔcbbO-m* were compared to the respective intact versions of the 35.2 kb or 13 kb insert exhibiting RubisCO active clones. Amplificates were separated by agarose gel electrophoresis (1 x concentrated 2% TBE agarose gel at 120 V for 20 min). Abbreviations are as follows: M – GeneRuler 100 bp DNA Ladder (Thermo Scientific), (-)_{RNA} negative control, where the corresponding template lacking reverse transcriptase was used for PCR, (+) positive control, where the fosmid DNA isolated from clone 71C2 was used as PCR template, (-)_{MM} negative control of the Master Mix in the PCR without applying any template (Böhnke and Perner 2014, Supplementary Information).

Furthermore it seems that in transposon clone 149II ($\Delta lysR2$) the transcription level of *cbbL* is unaffected, but that *cbbM* transcription was up-regulated, just the opposite of what has been observed for clone 6II ($\Delta lysR1$), where *cbbL* transcription level also remains stable, but *cbbM* appears to be down-regulated. For clone 17II ($\Delta cbbO-m$) an unchanged transcription level for *cbbL* and *cbbM* is suggested. No transcript for *cbbM* was received for clone 7II, but the transcription level of *cbbL* is up-regulated instead (see Figure 24).

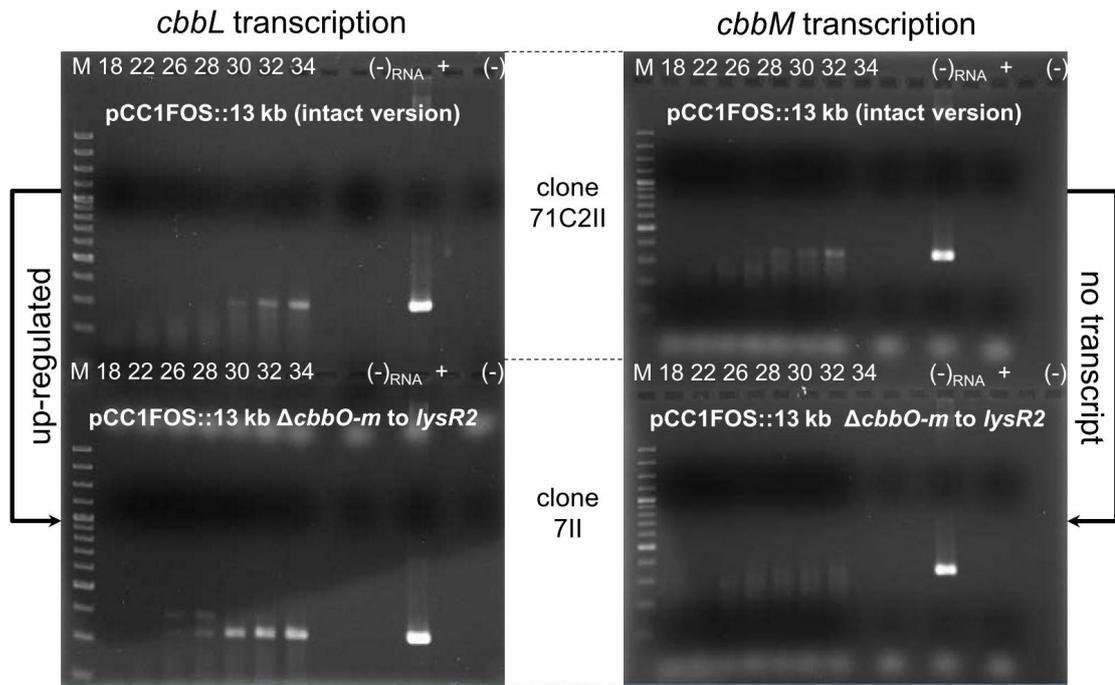


Figure 24: Transcription of *cbbL* and *cbbM* in transposon clone 7II. The amount of *cbbL* and *cbbM* amplified within 18, 22, 26, 28, 30, 32 and 34 PCR cycles of transposon clone 7II, carrying the *cbbL* gene cluster derived from the fosmid clone 71C2II, but lack the *cbbM* gene cluster inclusively the *lysR2* gene (pCC1FOS:: $\Delta cbbO-m$ to *lysR2*), were compared to the respective intact version 71C2 (pCC1FOS::13 kb). Amplificates were separated by agarose gel electrophoresis (1 x concentrated 2% TBE agarose gel at 120 V for 20 minutes). Abbreviations are as follows: M – GeneRuler 100 bp DNA Ladder (Thermo Scientific), (-)_{RNA} negative control, where the corresponding template lacking reverse transcriptase was used for PCR, (+) positive control, where the fosmid DNA isolated from clone 71C2 was used as PCR template, (-)_{MM} negative control of the Master Mix in the PCR without applying any template.

3.4.4 Complementation experiments

The four transposon clones with the most extreme activity changes, excluding those with insertions in RubisCO structural genes and those with the lack of the whole *cbbM* gene cluster (clone 7II), were complemented to demonstrate that the original RubisCO activity is restorable (see Figure 25). Transposon clone 96 (pCC1FOS::*Δorf06*) was complemented with *puc19::orf06*, resulting in the recovery of most of the original RubisCO activity. Complementation of transposon clone 149II (pCC1FOS::*ΔlysR2*) with *pUC19::lysR2-1* results in a decrease of the prior increased RubisCO activity and the full RubisCO activity could be restored if complementing transposon clone 6II (pCC1FOS::*ΔlysR1*) with *pUC19::lysR2-1* or transposon clone 17II (pCC1FOS::*ΔcbbO-m*) with *pUC19::cbbO-m*.

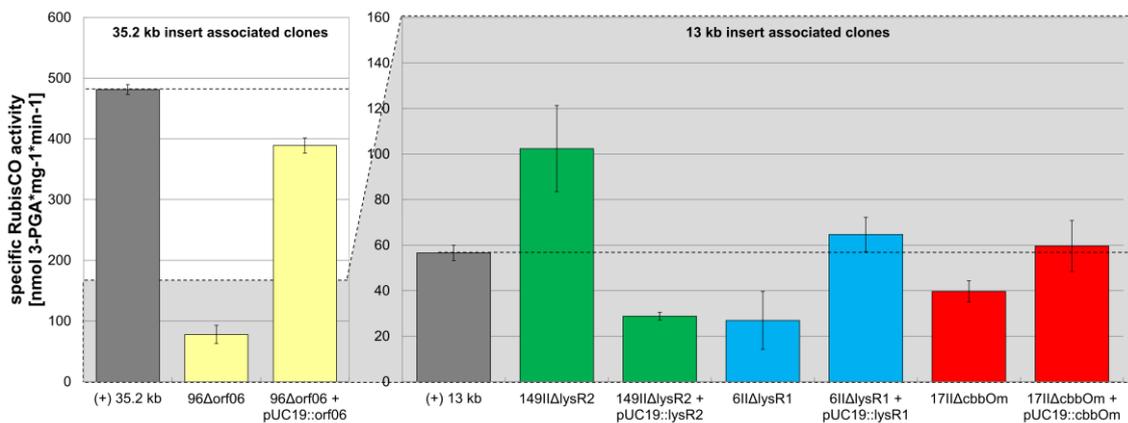


Figure 25: Complementation experiments. Specific RubisCO activities of crude extracts from the transposon clones with deletions in certain genes (*Δorf06*, *ΔlysR2*, *ΔlysR1* and *ΔcbbO-m*) and the complemented versions with additionally cloned *pUC19::orf06*, *pUC19::lysR2*, *pUC19::lysR1* and *pUC19::cbbO-m* are shown. Bars of impaired and complemented versions of one clone are displayed in the same color were 96Δ*orf06* and 96Δ*orf06* + *pUC19::orf06* are shown in yellow, 149Δ*lysR2* and *pUC19::lysR2* are shown in green, 6IIΔ*lysR1* and *pUC19::lysR1* are shown in blue and 17IIΔ*cbbO-m* and *pUC19::cbbO-m* are shown in red. RubisCO activities of intact 35.2 kb [(+) 35.2 kb] or 13 kb [(+) 13 kb] fosmid clones are shown in grey (Böhnke and Perner 2014, Supplementary Information).

4 Discussion

Most studies that have enhanced our understanding of RubisCO functioning have been conducted with cultured bacteria (Tabita *et al.* 2008 and references therein). Many of them have investigated RubisCOs of marine bacterial strains, like e.g.: *Beggiatoa* spp. (Nelson *et al.* 1989), *Thiomicrospira* spp. (Brinkhoff *et al.* 1999, Jannasch *et al.* 1985) or *Rhodospseudomonas palustris* (Joshi *et al.* 2009, and references therein), which offers first suggestions about RubisCO functioning but still only for selected cultured representatives. The functionality of environmental RubisCOs or the importance of gene products encoded on flanking DNA regions is however scarcely understood. Studies aiming environmental RubisCOs currently mainly rely on sequence searches (e.g., Xie *et al.* 2011). One sequence- and function-based combinational approach furthermore reveal a gene expression system with *Rhodobacter capsulatus* deletion strain SBI/II, applicable to investigate the functionality of environmental RubisCO genes by searching through a sequence based-screen first and determine the recombinant RubisCO activity in a surrogate host afterwards (Witte *et al.* 2010). However, sequenced-based approaches always come along with limitations, because only sequences with significant similarities to known genes are detectable and thus, one will never find novel *cbbLS* and *cbbM* encoding genes differing significantly from already known sequences of cultured representatives. Currently no sequence independent approach is available, which is suited to seek RubisCOs directly from metagenomic libraries by functionality alone. Therefore, the vast majority of functional RubisCOs and RubisCO associated genes from uncultured organisms (>99%) remains inaccessible. This study describes a novel, solely function-based approach suited to seek RubisCO active enzymes from the large majority of uncultivable organisms of any environmental sample like e.g.: soil samples, water samples, marine sediment samples or as in this study hydrothermally influenced deep-sea vent samples. It was shown that the HPLC based approach (i) in principle works to measure prokaryotic RubisCO activities from *T. crunogena* TH-55 (see 3.2.1) and that the RubisCO screen is appropriated to measure RubisCO activities from fosmid clones with large DNA inserts (TH-55 genome clone 6F3) (ii) on a single scale (see 3.2.1) and (iii) on a metagenomic scale from pools of 24 fosmid clones (see 3.2.2). Hence the functional RubisCO screen was successfully established. In the following this screen was used to seek RubisCO active fosmid clones from six metagenomic libraries with vent-origin. Fosmids of clones exhibiting RubisCO activity were furthermore analyzed and the role of flanking genes and resulting gene products for expressing a fully functional RubisCO was investigated.

4.1 Evaluating the newly established RubisCO screen by comparing hit rates of investigated metagenomic libraries

To explore the distribution and biochemical properties of RubisCOs in hydrothermally influenced habitats six metagenomic libraries constructed from thermally and chemically distinct hydrothermal deep-sea vent environments (see Table 10) were searched for RubisCO active fosmid clones by using the newly established function-based screen. The six different sampling sites were chosen to uncover possible interconnections between the distribution of RubisCO enzymes and the abiotically distinct venting environments, with respect to e.g.: (i) hydrogen-rich ultramafic systems (energy-high) versus basalt-hosted sulfide-rich systems (energy-low), (ii) O₂ - rich versus O₂ - low habitats and (iii) high temperature versus low temperature vent environments. Although only within one library, namely the 'Drachenschlund' library, clones with recombinant RubisCO activity could be discovered (for detailed results see section 3.3), it is possible to draw first conclusions from that finding and recognize first evidences for a potential connection between the distribution of RubisCO and the prevailing abiotic conditions at the Nibelungen site. The Nibelungen vent field, of which 'Drachenschlund' is a part of, is an ultramafic systems where the fluids are highly enriched in hydrogen (22 µM hydrogen measured in sampled fluids (Perner *et al.* 2013b)). Thus theoretically plenty of energy is available to fuel the energetically cost intensive CB cycle, possibly reasoning why the key CB cycle enzyme RubisCO was found at this hydrothermal site. The occurrence of functional RubisCO encoding genes furthermore might reflect that hydrogen is not only present in high concentrations, but as energy source really available for microorganisms. However if there exist a connection evident between the type of host rock, thus available energy sources, and the distribution of RubisCOs it is questionable why no RubisCO active fosmid clones were detectable within the libraries of the other energy – rich ultramafic sites ('Site B' and 'Irina II'). An explanation for that might be that one of the other alternative pathways with high energy requirements like e.g.: the 3-HP bicycle or the 3-HP / 4-HB cycle is more abundant at these sites and that a higher number of clones need to be screened in order to seek recombinantly expressed, environmental RubisCOs. Nevertheless, it also very well may be that the RubisCO screen conducted under atmospheric conditions is not qualified to cover all forms of RubisCOs, which might be reasoned by the prevailing atmospheric O₂ content. Thus, it might be possible that more RubisCO active fosmid clones are detectable if the RubisCO screen was adapted to anaerobic conditions, which however remains to be proven by performing further experiments. During sampling at 'Drachenschlund' the online-monitored fluid tempera-

ture was between 90 and 120°C (Perner *et al.* 2013b). At the first glance this seems not to be supportive for the operation of the CB cycle, which is as so far known limited to mesophilic conditions (Berg 2011). Nevertheless this might be explained by the steep thermal gradients which form when hot, highly reduced hydrothermal fluids mix with cold, ambient seawater, whereby multiple thermal microhabitats were constituted, providing grounds for mesophilic but also (hyper)-thermophilic microorganisms. Thus, the occurrence of RubisCO enzymes not directly reflects the predominant high fluid temperature at 'Drachenschlund', but may substantiate that the afore-mentioned multiple thermal microhabitats really exist at this hydrothermal site.

Overall twelve fosmid clones exhibiting RubisCO activities were identified from the 'Drachenschlund' library (see Figure 14, Appendix Table C 1). Restriction analyses suggested that these twelve clones were highly similar to each other, which was further supported by sequencing, where for all twelve clones identical sequences were generated matching each other with one-hundred percent DNA identity. This multiple occurrence of the same fragment in the 'Drachenschlund' metagenomic library (1% of all tested clones) could be affiliate to methodological procedures like e.g.: DNA extraction, MDA based bias or the construction of the metagenomic library it-self. But it could just as well be indicative for one species occurring in a high abundance at the investigated hydrothermally influenced sampling site. Since sequences of all twelve clones resemble *T. crunogena* XCL-2 (Scott *et al.* 2006) (90%, 96% and 94% for fragment A, B and C, respectively; for detailed results see 3.3) when Blast search was performed this highly abundant species at 'Drachenschlund' would be *Thiomicrospira*. However, these results may also be related to *E. coli*'s inability to express enzymes of other organisms due to weak recognition of intrinsic promoters and associated factors (Perner *et al.* 2011) and thus reflecting a limitation of the newly established screen, but which generally holds true for all function-based metagenomic screening procedures. Thus, it has been shown that 60 to 70% of foreign DNA is not expressible by *E. coli* (Gabor *et al.* 2004, Rondon *et al.* 1999), which is however the most preferred host organism used in the majority of conducted metagenomic studies (Handelsman *et al.* 1998). It might for example be possible that RubisCO of *T. crunogena* is expressible in *E. coli* due to the taxonomically close affiliation (both organisms were classified as *Gammaproteobacteria*), while RubisCOs of distantly related organisms remained unexpressed and thus still undetectable. Therefore it would be important to figure out the whole spectra of detectable RubisCOs if applying the HPLC-based RubisCO screen in subsequent studies.

4.2 Investigating the novel recombinant RubisCO from the metagenome of 'Drachenschlund'

4.2.1 Subcloning of the RubisCO gene cluster unexpectedly causes a dramatic loss of RubisCO activity

In order to simplify subsequent investigations on the metagenome derived RubisCO gene cluster, flanking genes which are not supposed to be associated with RubisCO functioning were removed by subcloning. However, subsequent RubisCO activity measurements of two subclones (71C2II and 84G4II) surprisingly reveal a decrease of the activity by a factor of about 5 compared to those measured for the original, intact clones 71C2 and 84G4 harboring the entire metagenomic DNA fragment inclusively flanking regions of additional 22.2 kb DNA (for detailed results see section 3.4.1, Figure 16 and Appendix Table C 1). This significant activity change unexpectedly suggests that there is at least one gene outside of the RubisCO gene cluster whose gene product influences RubisCO functioning.

In order to further detect the gene(s) responsible for this activity loss a transposon mutant library was constructed with the fosmid clone 71C2, carrying the entire metagenomic DNA fragment consisting of the RubisCO gene cluster and flanking DNA regions (35.2 kb). Furthermore it was examined whether *cbbO-m*, *cbbQ-m*, *lysR1*, *lysR2*, *cbbO* and *cbbQ* have a share in maximizing recombinant RubisCO activity by constructing a second transposon mutant library with the fosmid subclone 71C2II, harboring the RubisCO gene cluster without any flanking DNA region (13 kb). Additionally the four transposon clones with the most extreme activity changes, excluding those with insertions in RubisCO structural genes and those with the lack of the whole *cbbM* gene cluster (clone 7II), were complemented to demonstrate that the original RubisCO activity is restorable (see Figure 25). Furthermore the five transposon clones exhibiting the most remarkable activity changes, excluding those with insertions in RubisCO structural genes, were investigated to elucidate if corresponding genes are involved in transcriptional processes or rather in subsequent cellular processes like e.g.: regulation, assembling or activation of RubisCO.

4.2.2 The consequence of transposon insertions outside of the RubisCO gene cluster relative to RubisCO activity

4.2.2.1 Lowered RubisCO activity as a result of the impaired *orf06*

Transposon clone 96 ($\Delta orf06$) exhibited with 81 ± 20 nmol 3-PGA*min⁻¹*mg⁻¹ the most dramatic activity loss of all tested clones with insertions outside of the RubisCO gene cluster (see Figure 18A and B). Its specific RubisCO activity reaches only 17% of the total activity measured for the intact clone 71C2 (pCC1FOS::35.2 kb; 481 ± 8 nmol 3-PGA*min⁻¹*mg⁻¹). The measured activity was comparable to that of the subclones 71C2II and 84G4II (55 ± 8 and 48 ± 14 nmol 3-PGA*min⁻¹*mg⁻¹, respectively), which had the RubisCO gene cluster but lacked the flanking regions (see section 3.4.1, Figure 14 and Appendix Table C 1). If transposon clone 96 (pCC1FOS:: $\Delta orf06$) was complemented with puc19::*orf06* most of the original RubisCO activity could be restored (389 ± 12 nmol 3-PGA*min⁻¹*mg⁻¹ \cong 81%) (see section 3.4.4 and Figure 25), evidencing that *orf06*'s gene product is responsible for most of the lost activity. The deletion causing the activity loss was at 6 aa position (135 aa in total) of the gene designated *orf06*. No putative conserved domains could be detected when performing Blastx search, but similarities to twelve hypothetical proteins of different *Beta*-, *Gamma*- and *Zetaproteobacteria* were found (see Figure 19), with highest resemblance to the hypothetical protein Tcr0420 of *T. crunogena* XCL-2 (97% aa similarity) (see Table 11). In the metagenomic DNA fragment *orf06* is juxtaposed and upstream of the carbonic anhydrase, *cbbOQ-m* and *cbbM*, which is the same gene structure as described for *T. crunogena* XCL-2 (Scott *et al.* 2006) (see Figure 6). Genome data of other *Thiomicrospira* species e.g.: *Thiomicrospira arctica* DSM 13458 (Knittel *et al.* 2005) and *Thiomicrospira halophila* DSM 15072 (Sorokin *et al.* 2006) (genome papers of both organism are not yet published) has recently become available and show nearly the same gene arrangement with *orf06* as part of a RubisCO gene cluster (see Figure 6), implying that RubisCO gene regulation may be similar in the *Thiomicrospira* genus. However, some minor differences exist for both strains since *T. arctica* have two genes with homologies to *orf06* encoded on its genome, where both genes flank the carbonic anhydrase encoding gene and that gene with minor similarities (26%) is adjacent to *cbbO-m* while *T. halophila* has only one *orf06* homologue which is adjacent to *cbbO-m* without the gene encoding the carbonic anhydrase in between. A metagenome derived sequence of an uncultured SUP05 cluster bacterium belonging to the *Oceanospirillales* (Walsh *et al.* 2009) and the genome data of a symbiont of the *Calyptogena* clam, namely candidatus *Vesicomysocius okutanii* HA (Kuwahara *et al.* 2011) reveal that both organisms encode only the structural gene of the RubisCO form II, but the *orf06*

relative hypothetical protein is again adjacent to *cbbO-m* (see Figure 6), indicating that *orf06* might be associated to the expression of a fully functional form II RubisCO rather than to form I RubisCO expression. Four other genomes from two *Beta*-, one *Gamma*- and one *Zetaproteobacteria*, which also harbor a hypothetical protein with similarities to *orf06* reveal that the corresponding gene is not located near the RubisCO structural genes with ranges of 201 kb in the zetaproteobacterial *Mariprofundus ferroxydans* PV-1 (Emerson and Moyer 2002) to 2,117 kb in the betaproteobacterial *Gallionella capsiferiformans* ES-2 (Emerson *et al.* 2013) (see Figure 6). This suggested that *orf06* not essentially need to be located adjacent to RubisCO structural genes, but it has to be experimentally proven if *orf06* has an impact on RubisCO activity if it is located so far away from the structural gene. However, the closest related protein of known function is the ArsR regulatory protein of *Micromonospora aurantiaca* ATCC 27029 (accession: PRJNA42501) (30% similarity, e-value 0.6), which operates as a metal-sensing transcriptional repressor (Wu and Rosen 1991). Thus one further experiment, pairing the transcription of *cbbL* and *cbbM* with the function of possible RubisCO associated genes, was conducted to elucidate if Orf06 is really involved in the transcription of RubisCO form I and/or form II structural genes. Here the transcription amounts of RubisCO form I (*cbbL*) and II (*cbbM*) of clone 96 (Δ *orf06*) were compared with those of the original fosmid clone 71C2 (pCC1FOS::35.2 kb) where *orf06* is unimpaired. This comparison makes it obvious that the transcription of both structural genes (*cbbL* and *cbbM*) is down-regulated (see Figure 23) and demonstrates that the hypothetical protein Orf06 is directly or indirectly involved in regulating the transcription of *cbbM* and *cbbL*. As above mentioned *orf06* relative genes have been recognized to be also present in organisms harboring only form II RubisCO (see Figure 6). Thus it might be possible that *orf06* regulates the transcription of form II RubisCO. Since the experiments shown that the transcription of both genes is down-regulated one might conclude, that form II RubisCO then again regulates the transcription of form I RubisCO. However, further experiments will be needed to elucidate the role of form II RubisCO for the expression of a fully functional form I RubisCO. In order to determine whether RubisCO associated genes were co-transcribed or not TH-55's RNA was investigated, uncovering that the *orf06* homolog of *T. crunogena* TH-55 was transcribed separately from the Carbonic anhydrase and the YeeE/YedE family protein (see Figure 22). The increased activity of transposon clone 231, with an insertion in the non-coding region between *orf05* and *orf06* (located 166 nucleotides in front of the start codon of *orf06*) may also indirectly result from *orf06* reasoned by an impaired promoter region which inhibits the transcription of *orf06* and lower the RubisCO activity of clone 231 in consequence.

4.2.2.2 Knock-outs of *orf05*, *orf23*, *orf24* and *orf28* led to significantly decreased RubisCO activities

The specific RubisCO activity of clone 58 with an insertion in *orf05* (homolog to Tcr0419, YeeE/YedE family protein) comes with 351 ± 28 nmol 3-PGA*min⁻¹*mg⁻¹ to no more than 73% of the total activity measured for the intact, original metagenome derived clone 71C2 (481 ± 3 nmol 3-PGA*min⁻¹*mg⁻¹). Since *yeeE/yedE* encodes for a hypothetical protein predicted to have a function in the transport of sulphur-containing molecules which directly not seems to be linked to RubisCO functioning, it might be possible that the insertion is not really associated to RubisCO functioning, but also disturbing transcription of the adjacent *orf06*, which is deciding for the decreased RubisCO activity. However, due to the measured significantly lower RubisCO activity, yet one cannot rule out the possibility that the predicted function of the hypothetical protein is wrong and that *orf05* actually is related to RubisCO regulation, maturation or activation. Further experiments are required to clarify *orf05*'s role in expressing a fully functional RubisCO. An insertion upstream of *orf05* and thus possibly in the promoter region (clone 236) might affect the transcription of *orf05* encoding protein (YeeE/YedE) negatively, maybe reasoning the significant loss of RubisCO activity (384 ± 3 nmol 3-PGA*min⁻¹*mg⁻¹ \cong 80% of the total activity measured for the intact clone 71C2). This in turn corroborate the belief of *orf05* being involved in maximizing RubisCO form I and/or form II activities rather than an interruption of *orf06* transcription causing the significant activity loss.

Other transposon clones with insertions outside of the RubisCO gene cluster resulting in a significant loss of RubisCO activity were: clone 89 with an insertion in *orf23* (homolog to Tcr0437, *comFB*, late competence development protein), clone 68 with an insertion in *orf24* (homolog to Tcr0438, *purD*, Phosphoribosylamine-glycine ligase) and clone 91 with an insertion in *orf28* (homolog to Tcr0442, 50S ribosomal proteinL11 Methyltransferase) (see Figure 18). However, none of these three gene products has ever been linked to the expression of a fully functional RubisCO and current annotations not suggest that they encode enzymes with functions associated to RubisCO assembling, regulation or activation mechanism. Nevertheless conducted experiments indicate that in *T. crunogena* RubisCO activity is lowered when one of these three genes were knocked out, suggesting that gene products may interplay with RubisCO associated genes or that they might have an impact on intercellular conditions like e.g.: CO₂ concentrations, pH-values or ATP/ADP ratio, which might be important factors to reach maximal RubisCO activity (Portis 1990, Toyoda *et al.* 2005) applicable in a heterologous host as well, like e.g. in *E. coli*. In order to understand the

full extent of RubisCOs way of functioning it would be necessary to further validate if *orf23*, *orf24* and *orf28* take a role in expressing a fully functional RubisCO.

4.2.3 The consequence of transposon insertions within the RubisCO gene cluster relative to the RubisCO activity

4.2.3.1 Significantly changed RubisCO activities involving *lysR* knock-outs

*Significantly increased RubisCO activities caused by interrupted *lysR2* transcription.*

Transposon clone 149II ($\Delta lysR2$) with an insertion at aa position 271 (of 315 aa) of *lysR2* (*orf11*) displays with 102 ± 19 nmol 3-PGA*min⁻¹*mg⁻¹ a significantly increased RubisCO activity (see section 3.4.1.2, Figure 18C and D), which is roughly twice as much as has been measured for the intact, original subclone 71C2II (55 ± 8 nmol 3-PGA*min⁻¹*mg⁻¹). Since Blastp of the intact metagenome derived *lysR2* (*orf11*) demonstrates that the substrate binding domain is located between 99 aa and 303 aa, the LysR binding capability in the respective transposon clone 149II ($\Delta lysR2$) is considerably impaired, which might be an explanation for the significantly raised RubisCO activity. These results however suggest, that *lysR2* (*orf11*) on the metagenomic fragment encodes for a repressor for *cbbM* and/or *cbbL* expression. Complementation experiments with transposon clone 149II (pCC1FOS:: $\Delta lysR2$) and puc19::*lysR2-1* then again resulted in a decrease of RubisCO activity below to the level of the intact metagenomic fragment (29 ± 2 nmol 3-PGA*min⁻¹*mg⁻¹ \cong 53%) (see section 3.4.4 and Figure 25). This discrepancy between the complemented and the intact version might result from the diverging copy number of pCC1FOS (after autoinduction, 10+ fosmid copies) and pUC19 (without any induction 100+ plasmid copies) and thus for clone 149II in an overcompensation referable to higher copies of the possible repressor *lysR2*, but furthermore support the presumption of *lysR2* acting as a repressor for form I (*cbbL*) and/or form II (*cbbM*) RubisCO. To further investigate whether *lysR2* is involved in transcriptional processes of *cbbL*, *cbbM* or even of both transcription experiments were conducted (see section 3.4.2). Comparisons of the amounts of generated *cbbL* and *cbbM* transcripts of transposon clone 149II with those of the intact clone 71C2II (pCC1FOS::13kb) suggests that the transcription level of *cbbL* is unaffected, but that *cbbM* transcription was up-regulated (see Figure 23). Thus all experiments conducted with transposon clone 149II makes it highly likely that *lysR2* (*orf11*) acts on *cbbM* transcription as a repressor.

Significantly decreased RubisCO activities caused by interrupted lysR1 transcription.

Three insertions in *orf12* (*lysR1*) at the aa positions (i) 220, (ii) 264 and (iii) 285 (of total 308 aa) led to a significant loss of RubisCO activity for clones (i) 169, (ii) 6II and (iii) 161, respectively (see section 3.4.1.2 and Figure 18) equating (i) 90%, (ii) 56% and (iii) 85% of the activity measured for the intact versions. In order to validate this activity change complementation experiments were performed exemplarily for transposon clone 6II (pCC1FOS:: Δ *lysR1*) with pUC19::*lysR2-1*. Since the RubisCO activities of the complemented clone 6II (pCC1FOS::*lysR1* + pUC19::*lysR2-1*) and the intact version 71C2II not differ significantly one can conclude that the measured activities are substantial. In accordance to Blastp search of the intact metagenomic *lysR1* (*orf12*), the LysR substrate binding domain is located between 93 and 298 aa. Thus the insertions in transposon clones (i) 169, (ii) 6II and (iii) 161 affect the last (i) 78 aa, (ii) 34 aa and (iii) 13 aa of the LysR substrate binding domain. Since activity is lost it is likely that *LysR1* functions as an activator. To substantiate this thesis transcription experiments were performed exemplarily for transposon clone 6II (Δ *lysR1*), revealing that *cbbL* transcription level remains stable, but *cbbM* appears to be down-regulated relative to the respective intact metagenomic fragment (see section 3.4.2 and Figure 23), which again led to the conclusion that *lysR1* (*orf12*) probably activates *cbbM* transcription.

The *LysR1* and *LysR2* of the metagenomic fosmid clone 71C2 resembled the transcriptional regulators LysR family homologues *CbbR1* and *CbbRm* from *H. marinus* by 72% and 78% aa identity, respectively (Yoshizawa *et al.* 2004). For *H. marinus* it is suggested that *cbbR1* and *cbbRm* are bi-functional regulators, where *CbbR1* likely activates the expression of *CbbL* but in parallel represses the expression of *CbbM*, while *CbbRm* seems to be the activator of *CbbM* expression and coinstantaneous the repressor of *CbbL* expression (Toyoda *et al.* 2005). On the one hand the conducted experiments with the metagenomic clone 71C2 support the presumptions of Toyoda and colleagues (2005) for form II RubisCO (*CbbM*), because *cbbM* seems to be activated by *lysR1* and repressed by *lysR2*. This might be indicative for an equitable transcriptional regulation of *cbbM* in *H. marinus* and *T. crunogena*. However, under the provided conditions RubisCO form I (*cbbL*) transcription seems to be unconnected with *lysR1* and *lysR2*, speculating that the regulation of *cbbL* transcription differ in both organism or that the *lysR1* and *lysR2* regulation mechanism correlates with the prevailing atmospheric conditions (e.g.: O₂ vs. CO₂ content), where under atmospheric conditions *cbbL* might be transcribed constitutively while *cbbM* transcription is regulated by *lysR1* and *lysR2*, varying when e.g. the oxygen concentration change.

Transcription experiments with TH-55 RNA demonstrated that *lysR2*, *lysR1* and *cbbLS* are co-transcribed (see section 3.4.2, Figure 22 and Figure 23). However, because *lysR2* and *lysR1* pair is juxtaposed to *cbbLS* this transcription is somewhat disconcerting, because either only the *lysR2* and *lysR1* pair or the *cbbLS* gene product would result in a functioning enzyme while the other would result in an unfinished protein. Nevertheless, in case of an overlapping promoter region of the *lysR* promoter and the promoter of the structural gene *cbbL*, it might be possible that the conducted RT-PCR revealed a false positive DNA band and the one acquired, putative transcript (*lysR2R1cbbLS*) might encapsulate two real transcripts (i.e. *lysR2R1* and *cbbLS*).

4.2.3.2 The consequence of impaired *cbbO* and *cbbQ* gene expression

*Interrupting *cbbO-m* expression causes significantly lowered RubisCO activities.* Four transposon clones (17II, 19II, 14II and 21II) with insertions scattered across *orf08* (*cbbO-m*) reach only 58% to 75% of the total activity measured for the original, intact clone 71C2 (see section 3.4.1.2, Figure 18C and 18D). The full RubisCO activity was restored (see 3.4.4 and Figure 25) when one of these transposon clones, namely 17II (pCC1FOS:: Δ *cbbO-m*), was complemented with pUC19::*cbbO-m*, validating that the lowered RubisCO activity was really caused by the lack of CbbO-m. Transcription experiments with transposon clone 17II (Δ *cbbO-m*) illustrate that an impaired *cbbO-m* furthermore has no effect on the transcription amount of *cbbL* and *cbbM* (see 3.4.3 and Figure 23). This leads to the conclusion that CbbO-m is most likely not involved in the regulation of *cbbM* and/or *cbbL* transcription but rather in post-transcriptional processes, which support previously published thesis (Hayashi *et al.* 1997, Hayashi *et al.* 1999, Scott *et al.* 2006). However, further experiments will be needed to understand the actual function of CbbO-m and its exact way of functioning on RubisCO.

*Knock-outs of *cbbO-1* and *cbbQ-m* do not affect RubisCO activity.* Deletions at three different positions of the *orf16* (*cbbO-1*) gene did not result in a significant loss of RubisCO activity (transposon clones 4, 8 and 4II) (see section 3.4.1.2, Figure 18C and 18D) and may indicate that CbbO-m can substitute the role of CbbO-1. This phenomenon also holds true for the *orf09* (*cbbQ-m*), because three insertions scattered across *cbbQ-m* (clones 11II, 8II and 12II) did also not result in a significant RubisCO activity loss, suggesting that CbbQ-m might be substitutable by CbbQ-1. However, it might also very well be possible that CbbO-1 and CbbQ-m in *T. crunogena* relatives are simply not as essential for expressing a fully functional RubisCO. Further experiments are required to unravel the real functioning of *T. crunogena*'s CbbO-1 and CbbQ-m, which might verify the substitutability of both genes but could just as well

confirm that CbbO-1 and CbbQ-m operate in a generalized function unrelated to RubisCO functioning like previously suggested for the *Solemya* symbiont by Schwedock and colleagues (2004).

Insertions in cbbQ-1 offer contrasting results. Curiously, the insertions in *orf15* ($\Delta cbbQ-1$) of (i) clone 23II harboring the RubisCO gene cluster (13 kb) and (ii) clone 3 equipped with the entire metagenomic DNA fragment (35.2 kb) offer contrasting results. Thus clone 23II show significantly increased activities, while the RubisCO activity of transposon clone 3 remained unchanged, relative to the intact metagenomic clones 71C2II (pCC1FOS::13 kb) and 71C2 (pCC1FOS::35.2 kb), respectively (see Figure 18, Table 12 and Table 13). These results may appear conflicting at first appearance but even so they can be explained. Thus CbbQ-1 potentially needs to interact with proteins encoded on the flanking DNA regions. This interplay would not be possible if these regions are not expressed. Thus the lack of flanking genes in transposon clone 23II may be responsible for inhibited protein-protein interactions causing the increased RubisCO activity, but this needs to be proven by further experimental investigations.

4.2.3.3 The consequence of structural gene knock-outs for RubisCO activity

Additionally, three transposon clones with insertions in the RubisCO structural genes *cbbS*, *cbbL* and *cbbM* were investigated. The specific RubisCO activity was significantly lowered when RubisCO form I structural genes, *cbbL* (clone 24II) or *cbbS* (clone 38), were impaired (see Figure 18). By contrast an insertion in the RubisCO form II structural gene *cbbM* (clone 22II) resulted in an increase of the RubisCO activity by a factor of 6.2 relative to the intact version of clone 71C2II (pCC1FOS::13 kb) (see Figure 18, Table 12 and Table 13). This raised activity may indicate that RubisCO form II (*cbbM*) represses RubisCO form I (*cbbL*). Investigations on the received clone 7II support this presumption. Specific RubisCO activities of this clone 7II were also roughly six-fold higher ($301 \pm 21 \text{ nmol 3-PGA} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$) compared to the intact metagenomic clone 71C2II (pCC1FOS::13 kb) (see Figure 20A). Sequencing however revealed, that the kanamycin cassette was inserted in the *lysR2* gene at position 185 aa (of 315 aa in total), but that the adjacent genes *cbbM* and *cbbQ-m* as well as a part of the *cbbO-m* gene accidentally were cut out (see Figure 20B). The measured activity of clone 7II is slightly lowered (89%) compared to the activity measured for $\Delta cbbM$ clone 22II ($338 \pm 7 \text{ nmol 3-PGA} \cdot \text{min}^{-1} \cdot \text{mg}^{-1}$). Thus it is highly likely that transposon clone 7II managed to combine several effects like e.g.: (i) the increased activity caused by the lack of *cbbM* and the impaired *lysR2* but then again (ii) the loss of activity reasoned by the deletion of *cbbO-m*. As expected no transcript for *cbbM* was received for transposon clone 7II

reasoned by the lack of the corresponding gene. However, the transcription level of *cbbL* is highly up-regulated instead (see Figure 24), suggestive for CbbM acting as transcriptional regulator for CbbL. However, further experiments will be needed to verify and understand this unexpected interaction.

5 Conclusion

Within this study it was shown that the envisaged HPLC-based RubisCO screen was successfully established and is qualified to identify RubisCO active fosmid clones from metagenomic libraries based on functionality alone. Furthermore it was demonstrated that detected fosmid clones can be utilized to elucidate the importance of flanking genes and respective enzymes for a fully functional RubisCO. Thus it was shown that the expression of a fully functional *T. crunogena* RubisCO is a complex and highly regulated system, involving much more gene products than structural genes (*cbbLS* and *cbbM*) alone. This may hold true for the RubisCO enzymes of other microorganisms, too. The here established screening procedure enables us to explore RubisCOs and RubisCO associated enzymes of these organisms and even include those of uncultivable ones, which would otherwise remain inaccessible due to the limits of current cultivation techniques. In summary the newly established RubisCO screen circumvents time-consuming cultivation and the inherent bias associated with sequence-dependent methods. This screen makes it possible that the tremendous metagenomic resource of any environment becomes available and hence enables to discover species not previously associated with RubisCO activity. However further experiments will be needed to observe the whole spectra of detectable RubisCOs.

References

- Amann RI, Ludwig W and Schleifer KH** (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.
- Bachmann BJ** (1983). Linkage map of *Escherichia coli* K-12, edition 7. *Microbiol Rev* 47: 180-230.
- Badger MR and Bek EJ** (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J Exp Bot* 59: 1525-1541.
- Bar-Even A, Noor E and Milo R** (2011). A survey of carbon fixation pathways through a quantitative lens. *J Exp Bot* 63: 2325-2342.
- Berg IA, Ramos-Vera WH, Petri A, Huber H and Fuchs G** (2010a). Study of the distribution of autotrophic CO₂ fixation cycles in Crenarchaeota. *Microbiology* 156: 256-269.
- Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M et al** (2010b). Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* 8: 447-460.
- Berg IA** (2011). Ecological aspects of the distribution of different autotrophic CO₂ fixation pathways. *Appl Environ Microbiol* 77: 1925-1936.
- Bertani G** (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62: 293-300.
- Böhnke S and Perner M** (2014). A function-based screen for seeking RubisCO active clones from metagenomes: novel enzymes influencing RubisCO activity. *Isme J*: accepted.
- Bradford MM and Williams WL** (1976). New, rapid, sensitive method for protein determination. *Fed Proc* 35: 274-274.
- Brinkhoff T, Muyzer G, Wirsén CO and Kuever J** (1999). *Thiomicrospira kuenenii* sp. nov. and *Thiomicrospira frisia* sp. nov., two mesophilic obligately chemolithoautotrophic sulfur-oxidizing bacteria isolated from an intertidal mud flat. *Int J Syst Bacteriol* 49 Pt 2: 385-392.
- Campbell BJ and Cary SC** (2004). Abundance of reverse tricarboxylic acid cycle genes in free-living microorganisms at deep-sea hydrothermal vents. *Appl Environ Microbiol* 70: 6282-6289.
- Campbell BJ, Engel AS, Porter ML and Takai K** (2006). The versatile ϵ -proteobacteria: key players in sulphidic habitats. *Nat Rev Microbiol* 4: 458-468.
- Chow J, Kovacic F, Antonia YD, Krauss U, Fersini F, Schmeisser C et al** (2012). The metagenome-derived enzymes LipS and LipT increase the diversity of known lipases. *PLoS one* 7.
- de Boer AP, van der Oost J, Reijnders WN, Westerhoff HV, Stouthamer AH and van Spanning RJ** (1996). Mutational analysis of the *nor* gene cluster which encodes nitric-oxide reductase from *Paracoccus denitrificans*. *Eur J Biochem* 242: 592-600.
- Delwiche CF and Palmer JD** (1996). Rampant horizontal transfer and duplication of RubisCO genes in Eubacteria and plastids. *Mol Biol Evol* 13: 873-882.

- Dick GJ, Anantharaman K, Baker BJ, Li M, Reed DC and Sheik CS** (2013). The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Front Microbiol* 4: 124.
- Dobranski KP, Longo DL and Scott KM** (2005). The carbon concentrating mechanism of the hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol* 187: 5761-5766.
- Dubbs JM, Bird TH, Bauer CE and Tabita FR** (2000). Interaction of CbbR and RegA* transcription regulators with the *Rhodobacter sphaeroides cbbI* Promoter-operator region. *J Biol Chem* 275: 19224-19230.
- Ellis RJ** (1979). Most abundant protein in the world. *Trends Biochem Sci* 4: 241-244.
- Elsaied H and Naganuma T** (2001). Phylogenetic diversity of ribulose-1,5-bisphosphate carboxylase/oxygenase large-subunit genes from deep-sea microorganisms. *Appl Environ Microbiol* 67: 1751-1765.
- Emerson D and Moyer CL** (2002). Neutrophilic Fe-oxidizing bacteria are abundant at the Loihi Seamount hydrothermal vents and play a major role in Fe oxide deposition. *Appl Environ Microbiol* 68: 3085-3093.
- Emerson D, Field EK, Chertkov O, Davenport KW, Goodwin L, Munk C et al** (2013). Comparative genomics of freshwater Fe-oxidizing bacteria: implications for physiology, ecology, and systematics. *Front Microbiol* 4: 254.
- Epstein SS** (2013). The phenomenon of microbial uncultivability. *Current Opinion in Microbiology* 16: 636-642.
- Field CB, Behrenfeld MJ, Randerson JT and Falkowski P** (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281: 237-240.
- Fouquet Y, Cambon P, Etoubleau J, Charlou JL, Ondreas H, Barriga FJAS et al** (2010). Geodiversity of hydrothermal processes along the Mid-Atlantic Ridge and ultramafic-hosted mineralization: A new type of oceanic Cu-Zn-Co-Au volcanogenic massive sulfide deposit. *Geophys Monogr Ser* 188: 321-367.
- Fuchs G** (2011). Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annu Rev Microbiol* 65: 631-658.
- Gabor EM, Alkema WB and Janssen DB** (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* 6: 879-886.
- Guiral M, Prunetti L, Aussignargues C, Ciaccafava A, Infossi P, Ilbert M et al** (2012). The hyperthermophilic bacterium *Aquifex aeolicus*: from respiratory pathways to extremely resistant enzymes and biotechnological applications. *Adv Microb Physiol* 61: 125-194.
- Haase KM, Petersen S, Koschinsky A, Seifert R, Devey CW, Keir R et al** (2007). Young volcanism and related hydrothermal activity at 5°S on the slow-spreading southern Mid-Atlantic Ridge Geochemistry, Geophysics, Geosystems Volume 8, Issue 11. *Geochem Geophys Geosyst* 8: n/a.
- Handelsman J, Rondon MR, Brady SF, Clardy J and Goodman RM** (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5: R245-R249.
- Hayashi NR, Arai H, Kodama T and Igarashi Y** (1997). The novel genes, *cbbQ* and *cbbO*, located downstream from the RubisCO genes of *Pseudomonas*

hydrogenothermophila, affect the conformational states and activity of RubisCO. *Biochem Bioph Res Co* 241: 565-569.

Hayashi NR, Arai H, Kodama T and Igarashi Y (1999). The *cbbQ* genes, located downstream of the form I and form II RubisCO genes, affect the activity of both RubisCOs. *Biochem Bioph Res Co* 265: 177-183.

Heijnen JJ and Vandijken JP (1992). In search of a thermodynamic description of biomass yields for the chemotropic growth of microorganisms. *Biotechnol Bioeng* 39: 833-858.

Hügler M and Sievert SM (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Ann Rev Mar Sci* 3: 261-289.

Jakob R and Saenger W (1985). Reversed phase ion pair chromatographic separation of ribulose1,5-bisphosphate from 3-phosphoglycerate and its application as a new enzyme assay for Rubp carboxylase oxygenase. *FEBS Lett* 183: 111-114.

Jannasch HW, Wirsen CO, Nelson DC and Robertson LA (1985). *Thiomicrospira crunogena* sp. nov. a colorless, sulfur-oxidizing bacterium from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* 35: 422-424.

Joshi GS, Romagnoli S, Verberkmoes NC, Hettich RL, Pelletier D and Tabita FR (2009). Differential accumulation of form I RubisCO in *Rhodospseudomonas palustris* CGA010 under photoheterotrophic growth conditions with reduced carbon sources. *J Bacteriol* 191: 4243-4250.

Kelley DS, Baross JA and Delaney JR (2002). Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annu Rev Earth Pl Sc* 30: 385-491.

Knittel K, Kuever J, Meyerdierks A, Meinke R, Amann R and Brinkhoff T (2005). *Thiomicrospira arctica* sp. nov. and *Thiomicrospira psychrophila* sp. nov., psychrophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacteria isolated from marine Arctic sediments. *Int J Syst Evol Micr* 55: 781-786.

Kusian B and Bowien B (1997). Organization and regulation of *cbb* CO₂ assimilation genes in autotrophic bacteria. *FEMS Microbiol Rev* 21: 135-155.

Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T et al (2011). Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of *Calyptogena* clams. *BMC Evol Biol* 11: 285.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML and Pace NR (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *P Natl Acad Sci USA* 82: 6955-6959.

Lane DJ (1991). 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds). *Nucleic acid techniques in bacterial systematics*. John Wiley and Sons: Chichester, England. pp 115–175.

Li LA, Gibson JL and Tabita FR (1993). The Rubisco activase (*rca*) gene is located downstream from *rbcS* in *Anabaena* sp. strain CA and is detected in other *Anabaena/Nostoc* strains. *Plant Mol Biol* 21: 753-764.

Lorimer GH and Andrews TJ (1973). Plant photorespiration - inevitable consequence of existence of atmospheric oxygen. *Nature* 243: 359-360.

Marcus Y and Gurevitz M (2000). Activation of cyanobacterial RuBP-carboxylase/oxygenase is facilitated by inorganic phosphate via two independent mechanisms. *Eur J Biochem* 267: 5995-6003.

- McCollom TM and Shock EL** (1997). Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. *Geochim Cosmochim Acta* 61: 4375-4391.
- McCollom TM** (2007). Geochemical constraints on sources of metabolic energy for chemolithoautotrophy in ultramafic-hosted deep-sea hydrothermal systems. *Astrobiology* 7: 933-950.
- Melchert B, Devey CW, German CR, Lackschewitz KS, Seifert R, Walter M et al** (2008). First evidence for high-temperature off-axis venting of deep crustal/mantle heat: The Nibelungen hydrothermal field, southern Mid-Atlantic Ridge. *Earth Planet SC Lett* 275: 61-69.
- Minic Z and Thongbam PD** (2011). The biological deep sea hydrothermal vent as a model to study carbon dioxide capturing enzymes. *Mar Drugs* 9: 719-738.
- Mueller-Cajar O, Stotz M, Wendler P, Hartl FU, Bracher A and Hayer-Hartl M** (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* 479: 194-199.
- Nelson DC, Wirsén CO and Jannasch HW** (1989). Characterization of large, autotrophic *Beggiatoa* spp. abundant at hydrothermal vents of the Guaymas Basin. *Appl Environ Microbiol* 55: 2909-2917.
- Nigro LM and King GM** (2007). Disparate distributions of chemolithotrophs containing form IA or IC large subunit genes for ribulose-1,5-bisphosphate carboxylase/oxygenase in intertidal marine and littoral lake sediments. *FEMS Microbiol Ecol* 60: 113-125.
- Pereto JG, Velasco AM, Becerra A and Lazcano A** (1999). Comparative biochemistry of CO₂ fixation and the evolution of autotrophy. *Int Microbiol* 2: 3-10.
- Perner M, Kuever J, Seifert R, Pape T, Koschinsky A, Schmidt K et al** (2007a). The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15 degrees N on the Mid-Atlantic Ridge. *FEMS Microbiol Ecol* 61: 97-109.
- Perner M, Seifert R, Weber S, Koschinsky A, Schmidt K, Strauss H et al** (2007b). Microbial CO₂ fixation and sulfur cycling associated with low-temperature emissions at the Lilliput hydrothermal field, southern Mid-Atlantic Ridge (9.S). *Environ Microbiol* 9: 1186-1201.
- Perner M, Bach W, Hentscher M, Koschinsky A, Garbe-Schonberg D, Streit WR et al** (2009). Short-term microbial and physico-chemical variability in low-temperature hydrothermal fluids near 5°S on the Mid-Atlantic Ridge. *Environ Microbiol* 11: 2526-2541.
- Perner M, Ilmberger N, Köhler HU, Chow J and Streit WR** (2011). Emerging fields in functional metagenomics and its industrial relevance: Overcoming limitations and redirecting the search for novel biocatalysts. In: F.J. de Bruijn (ed). *Handbook of Moleculare Microbial Ecology II*. Wiley-Blackwell: New Jersey. pp 484-485.
- Perner M, Gonnella G, Hourdez S, Böhnke S, Kurtz S and Girguis P** (2013a). In situ chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field. *Environ Microbiol* 15: 1551-1560.
- Perner M, Hansen M, Seifert R, Strauss H, Koschinsky A and Petersen S** (2013b). Linking geology, fluid chemistry, and microbial activity of basalt- and ultramafic-hosted deep-sea hydrothermal vent environments. *Geobiology* 11: 340-355.

- Perner M, Gonnella G, Kurtz S and LaRoche J** (2014). Handling temperature bursts reaching 464 degrees C: different microbial strategies in the Sisters Peak hydrothermal chimney. *Appl Environ Microbiol*.
- Pichard SL, Campbell L and Paul JH** (1997). Diversity of the ribulose biphosphate carboxylase/oxygenase form I gene (*rbcL*) in natural phytoplankton communities. *Appl Environ Microbiol* 63: 3600-3606.
- Pieulle L, Magro V and Hatchikian EC** (1997). Isolation and analysis of the gene encoding the pyruvate-ferredoxin oxidoreductase of *Desulfovibrio africanus*, production of the recombinant enzyme in *Escherichia coli*, and effect of carboxy-terminal deletions on its stability. *J Bacteriol* 179: 5684-5692.
- Portis AR, Jr.** (1990). Rubisco activase. *Biochim Biophys Acta* 1015: 15-28.
- Portis AR, Jr.** (2003). Rubisco activase - Rubisco's catalytic chaperone. *Photosynth Res* 75: 11-27.
- Rabausch U, Jürgensen J, Ilmberger N, Böhnke S, Fischer S, Schubach B et al** (2013). Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Appl Environ Microbiol* 79: 4551-4563.
- Raven JA** (2009). Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat Microb Ecol* 56: 177-192.
- Raven JA** (2013). Rubisco: still the most abundant protein of Earth? *New Phytol* 198: 1-3.
- Rondon MR, Raffel SJ, Goodman RM and Handelsman J** (1999). Toward functional genomics in bacteria: Analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *P Natl Acad Sci USA* 96: 6451-6455.
- Scherf U, Sohling B, Gottschalk G, Linder D and Buckel W** (1994). Succinate-ethanol fermentation in *Clostridium kluyveri*: purification and characterization of 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA delta(3)-delta(2)-isomerase. *Arch Microbiol* 161: 239-245.
- Schwedock J, Harmer TL, Scott KM, Hektor HJ, Seitz AP, Fontana MC et al** (2004). Characterization and expression of genes from the RubisCO gene cluster of the chemoautotrophic symbiont of *Solemya velum*: *cbbLSQO*. *Arch Microbiol* 182: 18-29.
- Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA et al** (2006). The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biol* 4: 2196-2212.
- Sorokin DY, Tourova TP, Kolganova TV, Spiridonova EM, Berg IA and Muyzer G** (2006). *Thiomicrospira halophila* sp. nov., a moderately halophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacterium from hypersaline lakes. *Int J Syst Evol Micr* 56: 2375-2380.
- Staley JT and Konopka A** (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 39: 321-346.
- Streit W, Bjourson AJ, Cooper JE and Werner D** (1993). Application of subtraction hybridization for the development of a *Rhizobium leguminosarum* biovar *phaseoli* and *Rhizobium tropici* group specific DNA-probe. *FEMS Microbiol Ecol* 13: 59-67.

- Streit WR and Schmitz RA** (2004). Metagenomics: the key to the uncultured microbes. *Current opinion in microbiology* 7: 492-498.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D et al** (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333: 1296-1300.
- Tabita FR, Hanson TE, Li H, Satagopan S, Singh J and Chan S** (2007). Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev* 71: 576-599.
- Tabita FR, Satagopan S, Hanson TE, Kreeel NE and Scott SS** (2008). Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *J Exp Bot* 59: 1515-1524.
- Takai K, Campbell BJ, Cary SC, Suzuki M, Oida H, Nunoura T et al** (2005). Enzymatic and genetic characterization of carbon and energy metabolisms by deep-sea hydrothermal chemolithoautotrophic isolates of *Epsilonproteobacteria*. *Appl Environ Microbiol* 71: 7310-7320.
- Toyoda K, Yoshizawa Y, Arai H, Ishii M and Igarashi Y** (2005). The role of two CbbRs in the transcriptional regulation of three ribulose-1,5-bisphosphate carboxylase/oxygenase genes in *Hydrogenovibrio marinus* strain MH-110. *Microbiology* 151: 3615-3625.
- Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG et al** (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326: 578-582.
- Williams S** (2004). Ghost peaks in reversed-phase gradient HPLC: a review and update. *J Chromatogr A* 1052: 1-11.
- Winterberg H** (1898). Zur Methodik der Bakterienzählung. *Zeitschr f Hygiene* 29: 75-93.
- Witte B, John D, Wawrik B, Paul JH, Dayan D and Tabita FR** (2010). Functional prokaryotic RubisCO from an oceanic metagenomic library. *Appl Environ Microbiol* 76: 2997-3003.
- Wu J and Rosen BP** (1991). The ArsR protein is a trans-acting regulatory protein. *Mol Microbiol* 5: 1331-1336.
- Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J et al** (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *Isme J* 5: 414-426.
- Yamamoto M, Arai H, Ishii M and Igarashi Y** (2006). Role of two 2-oxoglutarate:ferredoxin oxidoreductases in *Hydrogenobacter thermophilus* under aerobic and anaerobic conditions. *FEMS Microbiol Lett* 263: 189-193.
- Yoshizawa Y, Toyoda K, Arai H, Ishii M and Igarashi Y** (2004). CO₂-responsive expression and gene organization of three ribulose-1,5-bisphosphate carboxylase/oxygenase enzymes and carboxysomes in *Hydrogenovibrio marinus* strain MH-110. *J Bacteriol* 186: 5685-5691.
- Zhang N, Kallis RP, Ewy RG and Portis AR, Jr.** (2002). Light modulation of Rubisco in *Arabidopsis* requires a capacity for redox regulation of the larger Rubisco activase isoform. *P Natl Acad Sci USA* 99: 3330-3334.

Appendix A: Abbreviations and accession numbers

Appendix Table A 1: Abbreviations and GenBank accession numbers of strains used for Figure 5 and Figure 6.

abbreviation	strain	accession number
metagenomic fragment (this study)	metagenome derived uncultured bacterium	KJ639815
<i>T. crunogena</i>	<i>Thiomicrospira crunogena</i> XCL-2	NC_007520
<i>T. arctica</i>	<i>Thiomicrospira arctica</i> DSM13458	PRJNA200374
<i>T. halophila</i>	<i>Thiomicrospira halophila</i> DSM15072	PRJNA201111
<i>H. marinus</i>	<i>Hydrogenovibrio marinus cbbM</i> gene cluster	AB122071
	<i>Hydrogenovibrio marinus cbbL</i> gene cluster	AB122069
<i>A. ehrlicheii</i>	<i>Alkalilimnicola ehrlicheii</i> MLHE-1	NC_008340
<i>A. ferrooxidans</i>	<i>Acidithiobacillus ferrooxidans</i> ATCC 23270	NC_011761
<i>M. ferrooxydans</i>	<i>Mariprofundus ferrooxydans</i> PV-1	PRJNA13615
<i>T. denitrificans</i>	<i>Thiobacillus denitrificans</i> ATCC25259	NC_007404
<i>N. europaea</i>	<i>Nitrosomonas europaea</i> ATCC 19718	NC_004757
<i>M. capsulatus</i>	<i>Methylococcus capsulatus</i> str. Bath	NC_002977
<i>V. okutanii</i>	Candidatus <i>Vesicomysocius okutanii</i> HA	NC_009465
uncultured SUP05	uncultured SUP05 cluster bacterium	PRJNA34785
Hiromi1	gamma proteobacterium Hiromi1	AP012273
<i>G. capsiferriformans</i>	<i>Gallionella capsiferriformans</i> ES-2	YP003847769
<i>Gallionella</i> sp.	<i>Gallionella</i> sp. SCGC AAA018,	NC_014394
<i>N. halophilus</i>	<i>Nitrosococcus halophilus</i> Nc4	NC_013960
<i>R. gelatinosus</i>	<i>Rubrivivax gelatinosus</i> IL144	NC_017075
<i>B. xenovorans</i>	<i>Burkholderia xenovorans</i> LB400	NC_007952
<i>N. oceani</i>	<i>Nitrosococcus oceani</i> ATCC 19707	NC_007484
<i>N. watsonii</i>	<i>Nitrosococcus watsonii</i> C-113	NC_014315
<i>N. multiformis</i>	<i>Nitrosospira multiformis</i> ATCC 25196	NC_007614
<i>N. winogradskyi</i>	<i>Nitrobacter winogradskyi</i> Nb-255	NC_007406
<i>R. palustris</i>	<i>Rhodopseudomonas palustris</i> BisA53	NC_008435
<i>R. sphaeroides</i>	<i>Rhodobacter sphaeroides</i> ATCC 17029 Chr.1	NC_009049
	<i>Rhodobacter sphaeroides</i> ATCC 17029 Chr.2	NC_009050
<i>R. capsulatus</i>	<i>Rhodobacter capsulatus</i> SB 1003	NC_014034
<i>R. ferrireducens</i>	<i>Rhodoferrax ferrireducens</i> T118	NC_007908
<i>D. aromatic</i>	<i>Dechloromonas aromatica</i> RCB	NC_007298

Appendix Table A 2: Gene abbreviations used in this study (modified from Böhnke and Perner 2014, Supplementary Information).

abbreviation	function of gene product
<i>YeeE/YedE</i>	YeeE/YedE family Protein
<i>CA</i>	Carbonic anhydrase
<i>cbbO-m</i>	von Willebrand factor type A
<i>cbbQ-m</i>	ATPase, AAA-type
<i>cbbM</i>	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, form II
<i>lysR2</i>	transcriptional regulator, LysR family
<i>lysR1</i>	transcriptional regulator, LysR family
<i>cbbRm</i>	transcriptional regulator, LysR family
<i>cbbR1</i>	transcriptional regulator, LysR family
<i>cbbR</i>	transcriptional regulator, LysR family
<i>cbbL</i>	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, form I
<i>cbbS</i>	Ribulose-1,5-bisphosphate carboxylase small subunit
<i>cbbQ-1</i>	ATPase, AAA-type
<i>cbbO-1</i>	von Willebrand factor
<i>cbbA</i>	Fructose-1,6-bisphosphat aldolase
<i>cbbT</i>	Transketolase
<i>cbbF</i>	Fructose-1,6-phosphosphatase
<i>cbbZ</i>	Phospoglycolate phosphatase
<i>cbbG</i>	Glyceraldehyd-3-phosphate dehydrogenase
<i>cbbP</i>	Phosphoribulokinase
<i>cbbX</i>	ATPase, AAA-type, probable RubisCO expression protein
<i>hyp</i>	hypothetical protein
<i>ABC-trans</i>	ABC transporter
<i>purD</i>	Phosphoribosylamine--glycine ligase
<i>purH</i>	Phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase
<i>FIS</i>	Fis family transcriptional regulator
<i>nifR3</i>	Dihydrouridine synthase TIM-barrel protein nifR3
<i>prmA</i>	50S ribosomal protein L11 Methyltransferase

Appendix Table A 3: Abbreviations and GenBank accession numbers of RubisCO encoding genes used for Figure 19 (Böhnke and Perner 2014, Supplementary Information).

abbreviation	strain	accession number
metagenome derived (this study)	uncultured bacterium	KJ639815
<i>M. aurantiaca</i> ,	<i>Micromonospora aurantiaca</i> ATC27029	YP003833177
<i>T. crunogena</i>	<i>Thiomicrospira crunogena</i> XCL-2	YP390690
<i>T. arctica</i> (I),	<i>Thiomicrospira arctica</i> DSM13458	WP019557142
	zeta proteobacterium SCGC AB-137-C09	WP018282693
<i>M. ferrooxydans</i>	<i>Mariiprofundus ferrooxydans</i> PV-1	WP009850905
	zeta proteobacterium SCGC AB-137-G06	WP018288597
	zeta proteobacterium SCGC AB-137-I08	WP018287271
<i>T. nivea</i>	<i>Thiothrix nivea</i> DSM5205	WP002707908
<i>T. denitrificans</i>	<i>Thiobacillus denitrificans</i> ATCC25259	WP019557144
	gamma proteobacterium Hiromi1	BAO45458
	sulfur oxidizing symbiont	WP005958517
<i>G. capsiferriformans</i>	<i>Gallionella capsiferriformans</i> ES-2	YP003847769
<i>Gallionella</i> sp.	<i>Gallionella</i> sp. SCGC AAA018 N21	WP018292968
<i>T. arctica</i> (II)	<i>Thiomicrospira arctica</i> DSM13458	WP019557144
<i>T. halophila</i>	<i>Thiomicrospira halophila</i> DSM15072	WP019895617
	uncultured SUP05 cluster bacterium	ACX30510
Candidatus <i>V. okutanii</i>	Candidatus <i>Vesicomysocius okutanii</i> HA, complete genome from 677709 to 683028	NC009465

Appendix B: Primers used in this study

Appendix Table B 1: Primers used for primer walking (Böhnke and Perner 2014, Supplementary Information).

Primer description	Sequence 5'-3'
<u>sequencing of fragment A</u>	
T7 promoter primer ¹¹	TAATACGACTCACTATAGGG
for_pw2	GCCGAGGCATAAGTGTATTC
for_pw2_inv	CCGCTTAAGAGACGTCATCG
for_pw4	ACGTTGTTGCGTCTTGAGTC
for_pw4_inv	TAAATGCCCCGACTCAAGACG
for_pw5	TGGCGATGTTCAAGTTGTCAC
for_pw5_inv	CCCTCTGACTTGTTCAATGG
for_pw6	AACAAGAAGGGTCCGTCATC
for_pw6_inv	GTCGCACCAATACAGTGAAG
<u>sequencing of fragment B</u>	
tcr_0431_pw1	ACAACAGGAAGGGCCGACAG
CbbO_pw3	AGGAGGCGGCAATCTTGGTG
CbbO_inv_pw5	ACGCTCCGGCTTTCCATTAG
CbbO_pw2	CCGTTTGGTTTGGGCATCCG
CbbO_inv_pw3	TTTCATTCCCGCAAGACGTG
CbbO_pw1	AGCGTAAAGGCTCGTTGTTT
CbbO_inv_pw2	AAAGCTCCGAGCCTAGTACC
CbbO_1-2_LS1	CAGCTCGGAGGCTTCCATAC
CbbQ_pw2	AAGTGTGCCACAACCTGGAAG
CbbO_inv_pw1	AACGAGCCTTTACGCTGTTG
CbbQ_inv_pw2	TCCAGTTGTGGCACACTTTC
CbbQ_pw1	GAAGCCCGTCAAGACACCAC
CbbQ_inv_pw1neu	CGACGGTGATCCGTTAATGG
CbbS_pw1	TCTCGCCCTGGTTTACTTAG
CbbLF_inv_pw2	CGGTGAAACTGACGCTGATG
CbbQ_inv_pw2	TCCAGTTGTGGCACACTTTC
CbbLS-R_LS1	GCGACCTACCAAATCTGAAG
CbbS_inv_pw1	TAAGTAAACCAGGGCGAGAG
CbbLF-R_LS1	CTCAGCCTTACGAGATTGTG
CbbLFn_inv_pw1	TCCACGTGAACGATCTTCAG
CbbLR-revcompl_pw1	GCCTGGTGTATGCCAGTTG
CbbLR_pw1	TGGCATAACACCAGGCATCG
CbbL_Start_inv	CCAGACGCTTAACTTGAACG
CbbL R ¹²	TCGAACCTTGATTTCTTTCCA
CbbL_Start	CATACGCTAGAGGGAAACG
CbbO_inv_pw5	ACGCTCCGGCTTTCCATTAG
1_6098bp_inv	TAACCGCCCAGCAGATAAGG
2_6319bp_inv	TCGGGTATCCGTATTGCAACAG
3_5684bp_inv	GGGCAGCGTCAAATTTACGG
4_5072bp_inv	CAGCGAACAGGCGTTTAACC
5_4556bp_inv	GATCGTAACGGGCTTGACGG
6_3920bp_inv	TACACGGCCTTCGTTCTTGC
7IIneu	ATCGTCACCGCTCATGTTTG
8_2598bp_inv	GGGAAACCGCTCGTAATCTG
9_2060bp_inv	TGGCCGACTACGGCTATAAG

10_1400bp_inv	CCTTTGAAGACTGCCGAGTG
11_865bp_inv	TGAACGCTTACACCCTTCTG
12_217bp_inv	CCATAAAGCGGTCAAGAAC
13_inv_pw1	CTGCAAACGAGCTGGTTAAG
14 inv pw1	CAGCGGTCAAAAACGCATGG

sequencing of fragment C

pCC1FOS_rev ^[1]	CTCGTATGTTGTGTGGAATTGTGAGC
rev_pw2	ATTGTTACCGCGTGCAGGTG
rev_pw2_inv	ACTTATGCCAGCCGGATTTG
rev_pw3_inv	CGCGAGCTGATAACCTATTG
rev_pw4	CACACCAGTGTGCGGATACG
rev_pw4_inv	CGCGAGCTGATAACCTATTG
rev_pw5	TGTAAGCGTGGGTGACCAAG
rev_pw5_inv	TAGGTTAGCATCGGCTGTGG

gap closure between fragment A and B

pw4F	GGGTGATACCGTTCATGTTG
pw10F	GACCACTGCAAACAGATGAG
pw15kb vorne	ATGCATCCTCAACGGCTTTC
pw11F	TGTTGAAGGTGCCTGATGTG

gap closure between fragment B and C

pw15kb_hinten	GTATAACTGGCTTGCCAGAG
LS2_pw1	TACACCGCACGGTCAATTTT
pw3R	GTAAGCCGACTAACCCCTAAG
pw4R	GGTACCTTCTGCCATTAACG
pw5R	TTTCAGTTTACGCGCCATGC
pw6R	AAGACTTGCTCGGGTAACAG
pw7R	GGCCTTAAGATTCCCGTGAC
pw8R	CTGCTTCTGACCCGGTTTCG
pw9R	TGTCATGTTGCGTTGGTAGC
pw10R	GAACGGATTGTAGCTCAGAC
LS3pw1F	TTAGGCGTTCCATCGGATGC
pw12R	GGCCAACAATGGTGAGTTTCG
LS4pw1F	ATGGTCATGCCGTCTTTTAC
pw14R	CCGTCATAGCGTGCTGTTTG
pw15R	GGCGAACAGGTTTCATAAGG
pw16R	GGCTCGACCAATCATAATGC
pw17R	CACCGGATTCTCGACAAATG
pw18R	ATAAACGCGAGCCAAACTGC
pw19R	CGTACGATCTGGGCCATAAC
pw1R	CCCATTAAGCGATGACGAAG

^[1] for details see the manual for the CopyControl™ Fosmid Library Production Kit, epicentre®

^[2] primer was published by Campbell and Cary (2004)

Appendix Table B 2: Primers used for cloning and validation procedures.

primer description	sequence 5'-3'	T_{annaeling}	product length
<u>cloning of <i>cbbLS</i></u>			
CbbL_ncr690_F	ATTCCGGTGGTACTCTTCCC	52°C	2,623 bp
CbbL_ncr139_R	GCCGTTACTTAAGCAGGAGG		
<u>cloning of <i>cbbM</i></u>			
cbbM_ncr332_F	GGTAAGCGAGCGGCATAAGC	49°C	2,269 bp
cbbM_ncr557_R	ATGCGCGAAACGGGTTTCAGG		
<u>sequence based screening of the genomic library</u>			
rbcL+S_for	GCTGGGCGGTTATGTGTAAG	57°C	1,834 bp
rbcL+S_rev	GCTGGGTTCCAACCTCTATC		
<u>proof of successful subcloning</u>			
KO_FOSsite primer	TACCGCACAGATGCGTAAGG	54°C	651 bp
KO_Insertsite primer	TTGCTCCGATAGCCGAGTTC		
<u>verify the successful insertion of the <TET-1> transposon</u>			
KOTet_Kas_F	CTTGTTTCGGCGTGGGTATG	58°C	647 bp
KOTet_Kas_R	CAGCGGTCCAATGATCGAAG		

Appendix Table B 3: Primer pairs used to analyzed (co)-transcription (Böhnke and Perner 2014, Supplementary Information).

primer description	sequence 5'-3'	T _{annealing} [°C]	product length [bp]
4n_CbbOII-QIIF 4n_CbbOII-QIIR	TACGCCCATGCCTATCTTAC TACCGACGATGCCGATATTC	50	329
6n_Rull-LysR2F 6n_Rull-LysR2R	GAAGATTCCGCTGCGATGTG CAGCGAACAGGCGTTTAACC	51	395
7n_CbbL-SF 7n_CbbL-SR	GTACAAGCACGTAACGAAGG ATGTTTCGATTGCAGGGTTCC	48	336
8n_CbbQI-OIF 8n_CbbQI-OIR	TGTATGCGGCACAGCTAATC TCCAGTTGTGGCACACTTTC	49	298
9n_CbbQm-RullF 9n_CbbQm-RullR	CAAAGCGGGATTTACCACAG ATCAGGTGCAGATCCAATCG	51	503
10n_RullF 10n_RullR	TCGATGTGACCGTAAGAACC ACTGCACGTCGTAACATCC	51	383
11n_LysR2-LysR1F 11n_LysR1-LysR2R	ATTGGCGGAATCGTAGAGTG GAGTCGCGATTGGTACATTG	51	373
12n_LysR1-CbbLF 12n_LysR2-CbbLR	ATACGGCAGGTTGAGAAAGG CTGGCATCCAATACGTTTCG	49	391
13n_CbbS-QIF 13n_CbbS-QIR	GTCTCAAGGTGCGAACATGC CCCGATTCAAACAGCTCTAC	50	387
14n_YeeE-tcr0420F 14n_YeeE-tcr0420R	TAGGCTGGCACCCAACAAAG ACTCGGTTCAAAGGCCTGAC	50	404
15n_tcr0420-CAF 15n_tcr0420-CAR	TATTCTTGGCGCGTCTTTGG TTGCTCCGATAGCCGAGTTC	52	599
16n_CA-cbbO-mF 16n_CA-cbbO-mR	TCGGAGCAACCTATCCAGAG GCCGATGAATACGTTGAGAC	50	362

Appendix Table B 4: Primer pairs used to analyze relative transcript abundance of RubisCO form I and II.

primer description	sequence 5'-3'	T _{annaeling} [°C]	product length [bp]
cbbL_810_F cbbL_1115_R	AGGTCTTGCGAACTACTGTC CCAGAA GCAACTGGCATAAC	50	306
cbbM_647_F cbbM_976_R	TCTGCACGGTAGCACATT TC ATTTGACGGTCCTGCTGTTG	52	330

Appendix Table B 5: Primer pairs used for complementation experiments.

primer description	sequence 5'-3'	T _{annaeling} [°C]	product length [bp]
<u>orf06 amplification</u>			
orf06_ncr444_F orf06_ncr161_R	CGCTGCTTAATGCAAGCATC ATGTACCAGCGGTCAGAAAC	50	1,013
<u>lysR2 and lysR1 amplification</u>			
lysR2-1_ncr193_F lysR2-1_ncr383_R	GACAACAGCTCAGGAAAGTG TGAAGATTCAGCCGCAACAG	50	2,457
<u>cbbO-m amplification</u>			
cbbOm_for2 cbbOm_rev2	GGCTATCGGAGCAACCTATC GCACGTTATGGTGCCATCTG	53	3,137

Appendix C: Measured RubisCO activities

Appendix Table C 1: Specific RubisCO activities of TH-55 and recombinant RubisCOs expressed in *E. coli* (Böhnke and Perner 2014, Supplementary Information).

tested organisms	specific RubisCO activity [nmol 3-PGA*min ⁻¹ *mg ⁻¹]
TH-55 (after 15.5 h)	210 ± 12
TH-55 (after 25 h)	252 ± 19
TH-55 (after 47 h)	228 ± 23
pCC1FOS:: <i>cbbLS</i> ^[1]	266 ± 23
pCC1FOS:: <i>cbbM</i> ^[1]	55 ± 9
genomic fosmid clone with RubisCO gene cluster (13 kb) + 25.1 kb flanking DNA ^[1]	
6F3	455 ± 30
metagenomic fosmid clones with RubisCO gene cluster (13 kb) + 22.2 kb flanking DNA	
71C2	481 ± 8
74E1	443 ± 38
74C10	432 ± 25
77H1	448 ± 34
77F4	461 ± 4
77D9	467 ± 37
78G5	483 ± 21
78E10	427 ± 16
80G3	473 ± 6
81G7	484 ± 44
81E1	457 ± 24
84G4	462 ± 39
fosmid subclones only containing the RubisCO gene cluster (13 kb) (<i>cbbOQM_lysR2_lysR1_cbbLSQO</i>)	
71C2II	55 ± 8
84G4II	48 ± 14

^[1] cloned from TH-55

Declaration on oath

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the Acknowledged resources and aids.

Stefanie Böhnke

Hamburg, 05. September 2014

Gedenke der Quelle, wenn du trinkst.

(chinesisches Sprichwort)

In diesem Sinne möchte ich mich zunächst bei Jun. Prof. Mirjam Perner dafür bedanken, dass sie mir dieses überaus interessante Thema überlassen hat und mir das Vertrauen geschenkt hat daran zu arbeiten. Ihre Ideen, die anregenden und überaus hilfreichen Diskussionen, haben wesentlich zum Fortgang dieser Arbeit beigetragen. Auch das andauernde Interesse gegenüber der Thematik und die motivierenden Worte waren ansteckend und haben mich stets angetrieben und nach vorne schauen lassen.

Herrn Prof. Dr. Wolfgang Streit möchte ich besonders dafür danken, dass er mir über den gesamten Zeitraum der Arbeit seine HPLC zur Verfügung gestellt hat, was die in dieser Arbeit durchgeführten Analysen erst ermöglicht hat. Mehr noch möchte ich ihm jedoch dafür danken, dass er mir immer mit einem offenen Ohr beratend zur Seite stand und auch insbesondere bei komplizierten Fragestellungen sein Wissen mit mir geteilt hat.

Danken möchte ich darüber hinaus Dagmar Svensson und Wenke Bahnsen nicht nur für die exzellente technische Zuarbeit zu diesem Projekte sondern auch dafür, dass sie die Arbeitsgruppe mit einer mütterlichen Herzlichkeit erfüllt haben.

Meiner Master Studentin Nicole Adam möchte ich ebenfalls danken, dass sie sich mit einem unbändigen Fleiß und großem thematischem Interesse in diese Arbeit eingebracht hat. Auch meinen Bachelor Studenten und BTA Schülern möchte ich dafür danken, dass sie durch ihr Engagement zu dem Gelingen dieser Arbeit beigetragen haben.

Meinen Laborkollegen Moritz Hansen, Nicolas Rychlik und Yuchen Han möchte ich für die tolle Zeit im Labor danken. Ich habe sehr von eurem Wissen und euren kreativen Ideen profitiert. Einen besonderen Dank möchte ich an dieser Stelle allerdings Moritz zukommen lassen, der sich mit einer bemerkenswerten Ausdauer alle meine fachlichen Probleme angehört hat, sich in die Thematik hinein gedacht hat und letztlich immer mit nützlichen Anregungen geblänzt hat.

Ein herzliches Dankeschön geht zudem an meine "Zellen-Kolleginnen" Dagmar Krysciak, Jennifer Chow, Janine Maimanakos und Claudia Hornung die stets mit einem offenen Ohr und wertvollen Beiträgen für mich und meine Probleme da waren. Ich bin sehr stolz darauf ein Teil des Mädchen-Büros gewesen zu sein. Nele Ilmberger und Angela Jordan ihr komplettiert die Mädels Runde und ich danke euch für jegliche Hilfe

und Unterstützung in meinem Laboralltag, aber vor allem für die Kraft die ihr mir mit eurer positiven Ausstrahlung jeden Morgen entgegen gebracht habt.

Danken möchte ich zudem Ursula Reinitz die mir bei jeder verwaltungstechnischen Frage tatkräftig zur Seite stand und die viele grundlegende Dinge für die gesamte Abteilung organisiert und erledigt.

Ich danke darüber hinaus allen aktuellen und ehemaligen Mitarbeiter der Abteilung Mikrobiologie und Biotechnologie der Universität Hamburg für die hilfreichen Gespräche, den Zusammenhalt und die schönen Momente am Arbeitsplatz und im privaten Bereich. Stellvertretend genannt seien Frederike Haak, Jessica Grote, Christian Utpatel, Ulrich Rabausch und Julia Jürgensen. Es war eine schöne und lehrreiche Zeit!

Von Herzen danke ich Mathias, der mir in den letzten Jahren eine wundervolle Hilfe und beständige Stütze war, indem er immer für mich da war und mich mit seinen liebevollen Worten gestärkt und motiviert hat. Ich danke dir, dass du so bedingungslos hinter mir stehst und mir alle Freiheiten lässt meinem forscherschen Interesse nachzugehen.

Ebenso möchte ich mich bei meinen Eltern Heiko und Sybille, meiner Schwester Kathrin, meinen Nichten Stella und Luna, meinen Großeltern Inge und Klaus sowie bei meinem Lieblingspärchen Steffi und Jens für die Zuversicht, das große Verständnis und die wohltuenden Ablenkungen vom Laboralltag bedanken. Es ist schön Menschen wie euch an meiner Seite zu haben.