Heinrich Heine
Universität
Düsseldorf

Establishment of novel model systems for synthetic microbial communities

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Maurice Mager
aus Fulda

Düsseldorf, Mai 2024

aus dem Institut für Synthetische Mikrobiologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Ilka M. Axmann
Institut für Synthetische Mikrobiologie
Heinrich-Heine-Universität Düsseldorf

2. Prof. Dr. Sebastian Fraune
Institut für Zoologie und Organismische Interaktion
Heinrich-Heine-Universität Düsseldorf

Tag der mündlichen Prüfung:
02.09.2024

Erklärungen und eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass die eingereichte Dissertation selbständig und ohne unzulässige fremde Hilfe unter Beachtung der “Ordnung über die Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf” verfasst worden ist. Weiter versichere ich andere als die in der Dissertation angegebene Literatur nicht benutzt zu haben und dass ich alle ganzen oder annähernd übernommenen Textstellen sowie verwendete Grafiken, Tabellen und Auswertungsprogramme kenntlich gemacht habe. Außerdem versichere ich, dass die vorgelegte elektronische mit der schriftlichen Version der Dissertation übereinstimmt und die Abhandlung in dieser oder ähnlicher Form noch nicht anderweitig als Promotionsleistung vorgelegt und bewertet wurde.

I hereby declare an oath that the submitted dissertation has been written independently and without undue assistance, considering the “Rules and Principles for Safeguarding and Good Scientific Practice at HHU”. Further, I assure that I did not use other references than the ones specified in the dissertation and that I have highlighted complete or partially adopted texts as well as graphics, tables, and analysis programs used in this work. Furthermore, I assure that the submitted electronic version is in accordance with the written version of the dissertation and that the dissertation in this or similar form has not yet been presented and evaluated as a doctoral thesis.

Marburg, Maurice Mager

_____

Erklärung der Betreuer*in über die Dissertation

Die Angaben zum Beitrag des Promovierenden zu den verwendeten Manuskripten sind zum Besten meiner Kenntnis glaubhaft. Dieses Promotionsvorhaben sowie der Vorschlag zur Berichterstattung wurden den hauptamtlichen Professorinnen und Professoren dieses Fachs zur Kenntnis gebracht.

Unterschrift Ilka M. Axmann

_____

Erklärung über die Disputation und den erlangten Grad

Die Disputation wird in englischer Sprache abgelegt. Es wird kein Ausschluss der Öffentlichkeit nach §9(4) der Promotionsordnung beantragt. Es soll der Grad „Doktor der Naturwissenschaften“ (doctor rerum naturalium – Dr. rer. nat.) verliehen werden.

Abstract (German)

Während der Großteil mikrobieller Forschung auf der Untersuchung einzelner, isolierter Spezies beruht, ist die Erforschung von mikrobiellen Konsortien erst seit kurzem im Fokus. Bakterien koexistieren jedoch in ihrer natürlichen Umgebung hauptsächlich als Teil von solchen Konsortien, die von der Tiefsee bis zu heißen Quellen, von der Haut bis zum menschlichen Darm überall gefunden werden können. Diese Gemeinschaften sind an essentiellen geo- und biochemischen Transformationsprozessen beteiligt. Änderungen in ihren Zusammensetzungen konnten bei Menschen schon mit Krankheiten wie Alzheimer in Verbindung gebracht werden. Jedoch sind die fundamentalen Prinzipien und Dynamiken solcher mikrobiellen Konsortien kaum verstanden. Um eine einzelne Spezies tiefgehend zu verstehen, muss diese im Kontext ihrer Interaktionen mit ihrer Umgebung inklusive anderer Bakterien untersucht werden. Ziel dieser Arbeit ist die Weiterentwicklung bestehender und die Entwicklung neuer Modellsysteme zur Erforschung mikrobieller Konsortien. Im ersten Projekt wurde die Domestizierung von *Curvibacter* AEP1-3, einem natürlichen Bewohner von *Hydra vulgaris*, durch die Entwicklung neuer Genexpressionsysteme vorangetrieben. Mehrere neue Systeme wurden entdeckt und deren Genaktivität während verschiedener Wachstumsphasen gemessen. In einem weiteren Projekt wurde die Reproduzierbarkeit von Forschung mit Cyanobakterien anhand einer Genexpressionsstudie ermittelt, die in mehreren teilnehmenden Laboren in Europa durchgeführt wurde. Ziel war es, auf die mangelnde Reproduzierbarkeit von Ergebnissen in diesem Bereich aufmerksam zu machen, Gründe dafür offenzulegen und das Arbeiten mit Cyanobakterien zugänglicher für Forschende außerhalb des eigenen Felds zu machen, um zur Entwicklung neuer cyanobakterieller Symbiosen mit neuen Spezies anzuregen. In einem weiteren Projekt wurden synthetische Konsortien zwischen *Escherichia coli* und *Vibrio natriegens* basierend auf Dependenz durch Aminosäure-Auxotrophien entwickelt. Es wurde eine starke Regulation der relativen Zelldichte beider Partner beobachtet, die resistent gegen externe und interne Störungen ist.

Abstract (English)

While most of microbiological research focuses on the study of individual species, research on microbial consortia is still in its infancy. However, microbes in nature exist almost exclusively as part of such communities which can be found anywhere from the deep sea to hot springs, from the skin to the gut of humans. They are responsible for vital geo- and biochemical transformations and have been linked to Alzheimer's and other diseases. Yet, very little is known about the fundamental principles and dynamics of a microbial consortium. In order to truly understand individual species and their ecosystem it is essential to study them in the context of their interactions with their partners. This thesis aims to advance the field of microbial community research by advancing existing and creating novel model systems to investigate. In the first project we continued domestication of the natural colonizer of *Hydra vulgaris*, the bacterium *Curvibacter sp.* AEP1-3. *Curvibacter* is in an intimate cross talk with its host but also other members of the *Hydra vulgaris* microbiome. We established several new *in trans* expression systems and assessed their activity levels across different growth phases. In the second project we assessed the reproducibility of cyanobacterial experiments in an interlaboratory study across Europe. The aim was to raise awareness for reproducibility issues in cyanobacterial research, to make cyanobacteria more reproducible and to allow researchers outside of the phototroph field better access to cyanobacteria as potential partners for symbiosis. In the third project we created synthetic inter- and intraspecies mutualistic relationships based on amino acid auxotrophy dependencies between *Escherichia coli* and *Vibrio natriegens*. We observe a strict maintenance of a cell-to-cell abundance ratio in these communities that is resistant to external and internal perturbations.

Danksagung

I would like to thank my supervisor Ilka for hosting me in her lab, guiding me through three years of PhD, helping me see the value of my work when I could not and gently pushing me along on my journey. I also thank my secondary supervisor Sebastian for great conversations and support around my *Curvibacter* project. Thanks also go out to the SynMibi team for providing the infrastructure on which I based my work, for fruitful input on technical questions and their companionship. A big thanks is also owed to the students I supervised over the years. Nina, Dave and Tijn, I could not have picked better students to aid me in my journey and I am grateful that I was also able to tread along your individual paths for a while as well. I wish all of you the best of success and hope that you keep your time with me in good memory.

Big thanks also go out to Mo. Your positivity and pragmatical view have made dealing with bureaucratic madness slightly more manageable. I would have certainly missed out on a lot of joy and tea sessions if it wasn't for Hettie. I am grateful for your friendship as well as all our honest and open conversations about anything and everything that had me eventually regain the ability to comprehend (your) British dialect (to an extent).

It is no exaggeration to say that without Lukas I may never have come this far in my PhD. Thanks to your support and deep friendship, in bad and in good times. Thanks for your genuine kindness and for helping me see the world from a better perspective. I have absolute confidence that you will do just fine in your journey.

Thanks also go out to the friends that accompanied me on the many stages of this PhD rollercoaster. Thank you to Yannik, Rada, Aleks and Josh for taking my mind off work until they needed to endure another rant about my project.

Finally, I would like to thank Alyssa for the mental support throughout difficult times, for being the "good-angel-voice" on my shoulder when I am hot headed and for accompanying me on this journey.

Contents

List of abbreviations

BG11: Blue Green Media

CRISPR: Clustered Interspaced Palindromic Repeats

DART: DNA-editing All-in-one RNA-guided CRISPR-Cas Transposase

DMAPP: Dimethylallyl di-phosphate

DNA: Desoxyribonucleic Acid

GAP: Glyceraldehyde 3-phosphate

GFP: Green Fluorescent Protein

HEPES: 2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethanol sulfonic acid

HSL: Homoserine lactone

IPP: Isopentyl di-phosphate

LB: Lysogeny broth (media)

NADP: Nicotinamide adenine dinucleotide phosphate

PCR: Polymerase Chain Reaction

pH: *potentia hydrogenii*

QS: Quorum sensing

Introduction

Since the beginning of microbiology, the dogma persists that in order to study a microbe it needs to be cultivated under axenic conditions. This rationale was adopted from experimental traditions of physics and chemistry, fields where scientists aimed to minimalize experimental setups to rule out conflicting explanations of their results. The isolation of individual microbes like *E.coli* by Theodor Escherich in 1886 opened up a way to study individual species and to understand basic principles of life. With the help of *E.coli*, the regulatory mechanisms of the lac-operon as one of the first genetic circuits were elucidated. *E.coli* also aided in understanding DNA replication, bacterial conjugation, phage infection and several other basic concepts of microbiology. This highlights the importance of well understood, and isolated model organisms in the research of fundamental concepts in biology. Using highly tractable and domesticated model systems is not only an accelerator of research, it also works as an enabler for a variety of scientific endeavors that rely on the minimalization and simplification of observed systems, such as quantitative measurement methods and bioinformatic modeling of living systems. The use of individual species has brought the field of microbiology to the point where it is today and will remain irreplaceable for answering many open questions in the field.

However, microbes in nature exist almost exclusively as communities. In fact, these communities are highly abundant and have conquered most habitable surfaces, from the human skin and gut to the deep sea. Their interactions are diverse, ranging from positive interactions such as nutrient exchange to competitive interactions such as antibiotic warfare and competing over central nutrients. As a community, they play important roles in critical ecological processes such as nutrient cycling and decomposition. Yet, the understanding of microbial consortia is still in its infancy. Simple concepts like labor division are well established: a process in which one type of microbe supplies one type of service and in return profits from services provided by other members of the community. The advantage of such a labor division is an

increased efficiency of individual service due to specialization and genomic reduction due to redundancy (Louca *et al.* 2018). Also, microbial consortia exhibit a much higher functional diversity compared to the contained individual species, allowing them to perform complex biochemical transformations that individuals would not be able to carry out.

Microbial communities additionally display emergent properties which often cannot be predicted from the characteristics of individual species (Osburn *et al.* 2021). Such emergent properties include quorum sensing, spatial organization, and labor division. On top of this, interactions of species in consortia also subjects these species to evolutionary drives which can only be explained in the context of their interactions, adding another layer of complexity to the emergent dynamics. But, due to the sheer complexity and diversity of microbial communities, elucidation of their interactions and organization proves a difficult task. The lack of technologies to monitor the sum of exchanges between microbes limit the output of experimental work and the resulting lack of data makes any model of microbial communities incomplete. In summary, there is a need for simple, tractable microbial community model systems as well as associated methodology. Only under these conditions, full scale models of complex interactomes can be informed.

Studies of microbial consortia are also a promising field for applied microbiology: the utilization of labor division may enhance the productivity of biotechnological production processes. Additionally, it may open ways to produce chemicals that were previously considered infeasible for biotechnological production due to the complexity or toxicity of the compounds or intermediates. In recent years, Alzheimer's disease has been linked to compositions of the human gut microbiome (Jiang *et al.* 2017). Yet, these observations currently only give phenomenological evidence to correlations and we are merely scratching the surface of the underlying mechanics.

To truly understand individual species, it is therefore important to also study them in their natural context of their respective microbiota. Vice versa, studying a species solely as an isolate may never reveal "blind spots" in our understanding. Inference of

information about their natural life cycle can only occur via the abstraction of behavior observed in isolation, additionally there is no information about whether certain properties of a species are simply artifacts derived from the isolated cultivation.

Hydra vulgaris AEP1-3 as a model system for inter kingdom symbiosis

Hydra vulgaris is a well-established and highly interesting model organism. It belongs to the Basal Metazoan, a diverse group that is unique for showing a point symmetric- instead of the bilateral symmetry found in almost all bilaterian animals. A secondary unique feature is that while most bilaterian animals possess three germ layers (endo-, ecto- and mesoderm) the basal metazoan are all diploblastic, meaning that during the development of the blastula only two germ layers are developed: an outer ectodermal layer and an inner endodermal layer. *Hydra vulgaris* therefore belongs to an ancient clade of animals; with the exceptions of Porifera and Ctenophora they represent the animal ancestors most distantly related to humans. This makes *Hydra vulgaris* of special interest to fundamental animalia research as any conserved function between mammalians and *Hydra*s is 1) likely to be conserved between most other Bilateria as well and 2) represents an ancient function of animal life. An example for such an ancient gene class is the presence (and similar function) of Toll-like receptors (which will be mentioned in closer detail further below), that has partially conserved its function in Mammalia and Basal Metazoans alike, despite the long parallel evolution. Such well conserved functions give insight into the evolutionary development of such a function. Another highly conserved feature of all animalia is the outer extracellular layer of lipids, proteins and carbohydrates called glycocalyx: the thick glycocalyx layer around the epithelia of *Hydra vulgaris* serves as a protective layer but also as a habitat for its microbiome. The viscose layer serves as a sticky habitat for bacteria as well as a food source.

A microbiome comparison of the glycocalyx of free living and laboratory *Hydra vulgaris* and *H. oligactis* individuals revealed, that the microbial composition pattern of free living and laboratory individuals of the same species were more similar compared to individuals of both species that have been cultivated in the laboratory for >30 years under the same cultivation conditions (Bosch *et al.* 2009). This finding is interesting for multiple reasons: first, it shows a remarkable stability of microbial community compositions on the epithelia of *Hydra vulgaris* even when the animals are grown under laboratory conditions, making them optimal candidate systems for the study of natural host-microbiome interactions under experimental conditions. Second, it shows that the composition of the microbiome is strongly dependent on the host and less on the environmental conditions, assuming a strong evolutionary link between *Hydra vulgaris* and their associated commensal bacteria. This in turn indicates that the bacteria increase the fitness of their host in a symbiotic relationship.

The absence of neurons in *H. magnipapillata* leads to a drastic change in microbial composition (Fraune *et al.* 2009), indicating that epithelial homeostasis somehow influences the development of the microbiome. The absence of other cell lines such as nematocytes or interstitial cells had no impact on the microbial composition.

For *Hydra vulgaris*, the microbiome is known to play a vital role in the protection against fusarium sp., a fungal pathogen (Fraune *et al.* 2015). While individual species of the microbiome are capable of conferring fungal protection to the host, the combination of multiple species leads to synergistic or additive effects and in some cases even antagonistic effects. This shows that the microbiome is not just a commensal community that happens to continuously survive on the *Hydra* glycocalyx but is indeed in a mutually beneficial symbiotic relationship with the host. Thus, the evolutionary advantages for a host in actively managing and controlling a beneficial microbiome become evident. The Toll-like receptor MyD88 has been identified in *Hydra vulgaris* as an ancient form of bacterial signaling (Franzenburg *et al.* 2012) and is assumed to be involved in bacterial homeostasis on the epithelial surface.

Observing the microbial composition of *Hydra vulgaris* during development elucidates colonization dynamics that follow a pattern from early juvenile to mature stage (Franzenburg *et al.* 2013). Principal component analysis of these bacterial compositions shows that individual juvenile samples taken from *Hydra vulgaris* hatchlings display very similar bacterial communities. Further, their composition shifts for all samples in a similar manner, indicating a directed pattern that follows the development of the *Hydra vulgaris* hatchlings. Interestingly, the adult community composition shows high similarity to a transient state of the community found on 3 weeks old *Hydra vulgaris* individuals, assuming that the adult composition is already achieved early in development but only in a transient form. This further highlights the interplay of the developmental stage of the host and its effect on the microbiome composition. Notably dominant colonization of a particular species seems to coincide with this adult composition: a *Curvibacter* species dubbed *Curvibacter sp.* AEP1-3 after its associated host *Hydra vulgaris* AEP1-3.

Curvibacter sp. AEP1-3 is a key member of the *Hydra vulgaris* microbiome

Curvibacter sp. AEP1-3 (further just *Curvibacter*) is a gram negative betaproteobacteria. It is heterotrophic, rod shaped and possesses a flagellum. Many *Curvibacter* species are found as commensal bacteria, but free-living species exist as well. As a closely associated symbiont of the freshwater polyp *Hydra vulgaris* AEP1-3, *Curvibacter* is a strictly salt sensitive freshwater bacteria.

When associated with its host, *Curvibacter* is in constant chemical communication with *Hydra vulgaris* via quorum sensing (QS) molecules (Pietschke *et al.* 2017). *Hydra vulgaris* specifically modifies bacterial 3-oxo-homoserine lactones (HSL), a class of QS molecules produced by multiple *Hydra vulgaris* colonizers, including *Curvibacter*. The sensing of the modified HSL is known to lead to transcriptional changes in *Curvibacter*, closing the circle on this chemical communication loop. Transcriptional changes include the inhibition of flagellar synthesis. This is assumed to aid the host associated symbiotic

lifestyle, as the lack of motility might lead to better glycocalyx attachment and the loss of the flagellum allows the bacterial partners to avoid *Hydra*s immune responses targeting the flagellin protein (Bosch *et al.* 2009). Hence, the interspecies communication seems to directly change the lifestyle of *Curvibacter* from a motile, free-living bacteria to a sessile symbiont.

A *Curvibacter* associated phage TJ1 further modifies the microbiome of its host (Lange *et al.* 2019). While the host colonization of *Curvibacter* and *Acidovorax sp.* is less affected by the presence of the TJ1 phage, other natural *Hydra vulgaris* colonizers like *Duganella sp*., *Undibacterium sp.* and *Pelomonas* AEP1-3 are significantly inhibited in their growth on *Hydra vulgaris*. This leads to a competitive advantage of *Curvibacter* and *Acidovorax sp.* over other colonizers in the presence of the TJ1 phage. While the complete context of such phages for microbial community composition and the effect on the host are yet to be elucidated, it is evident that this phage seems to play a vital role in the homeostasis of the microbial community on *Hydra vulgaris*.

Advantages and challenges of using *Curvibacter sp.* AEP1-3 and *Hydra vulgaris* as a model system for symbiotic interactions

Curvibacter and its natural host *Hydra vulgaris* make an interesting model system in the context of their inter-kingdom symbiosis for several reasons. First, the basic body plan and simple cultivation of *Hydra vulgaris* makes it an easy host for biological studies compared to animals with much higher degrees of complexity, such as fruit flies and mice. Second, compared to higher animals a lot of advanced methods are already established for *Hydra vulgaris* such as germline editing via CRISPR Cas9 based knockouts and siRNA knockdowns (Lommel *et al.* 2017) as well as single cell transcriptomics (Siebert *et al.* 2019). A high level of amenability to manipulation and good tools for the observation of both symbiotic partners are very helpful to elucidate their symbiotic interaction. Third, the model system has a good expandability, as more natural symbionts of *Hydra vulgaris* have already been utilized to create more complex

symbiotic networks. *Curvibacter* stands out as a key player in the natural community and is so far the only genetically amenable candidate to my knowledge.

The *Curvibacter* genome is sequenced and this data is deployed in the NCBI RefSeq database as assembly GCF_002163715.1. This is a crucial step as without a genome neither comparative bioinformatic approaches can be used to infer information about the species nor genome editing is accessible. A transcriptome of free living *Curvibacter* is provided with the manuscript 1. To the best of my knowledge, no proteome data has been published so far.

Cultivation of *Curvibacter* is currently standardized with R2A media, a complex LB-like media with low salinity that is commonly used to isolate freshwater bacteria. An M9 based minimal media is currently under development (unpublished results). This is a crucial step for the chemical isolation of any compound of interest from *Curvibacter* as complex media like R2A are full of trace elements and compounds of unknown amounts, making analytical methods very difficult.

Curvibacter can also be used to colonize germ-free *Hydra vulgaris* polyps (Wein *et al.* 2018; Hemmrich *et al.* 2007). Here, *Curvibacter* can either be used as a sole colonizer or in competition with another species or as a mix of a conventionalized microbiome (Deines *et al.* 2020). The relative abundance of the species in such a microbiota is currently determined by outgrowth of the isolated microbiome on solid media and counting colony forming units or more sophisticated methods such as PCR. Microscopy imaging of *Curvibacter* on *Hydra vulgaris* epithelia is also viable, as well as fluorescence imaging using GFP tagged *Curvibacter* (Wein *et al.* 2018).

However, compared to other more established counterparts like *Escherichia coli* and *Saccharomyces cerevisiae* it lacks the proper tractability and availability of tools. Genomic modifications are possible but very cumbersome. Simple and scarless integration or knockout of genes via CRISPR-Cas systems (Jinek *et al.* 2012) have so far not been established, the same accounts for knockdown strategies of essential genes via depletion assays or CRISPRi (Larson *et al.* 2013).

Significant challenges of *Curvibacter* include its low tolerance for salinity. This makes manipulation of growth media fairly difficult as the addition of any ionic compound in higher molarities will lead to an increase in osmolarity and may therefore interfere with *Curvibacter*s growth. R2A as well as our provisional M9 derivative (unpublished results) are both weakly buffered media for this reason. This is an issue in prolonged cultivation, as *Curvibacter,* like many heterotroph species, tends to acidify the medium during growth. Without any buffering system to counteract this, the pH decreases during cultivation until the growth of *Curvibacter* is inhibited. The same accounts for the availability of carbon and nitrogen sources: as chemicals like glucose and ammonium chloride also possess osmotic activity, the *Curvibacter* media needs to strike a fine balance between nutrient availability and osmolarity.

The low salt tolerance may also be the reason that so far, no protocol for chemical competence has not been established. As chemical competence for eg. *E.coli* relies on the destabilization of membranes with strong ions to facilitate DNA uptake, high levels of strong ions like Calcium and chloride are required to make cells chemically competent. Such levels of salinity may be already way too high to yield viable, competent *Curvibacter* cells.

Another issue is the high toxicity of several chemicals commonly used to induce gene expression (Bachelor Thesis Nina Schulten, HHU). This limits the development of induction systems based on commonly used systems like the *ptet* system.

In manuscript 1, our goal was to find suitable expression systems for *Curvibacter sp.* AEP1-3. To work around the issue that classical inducible systems are incompatible due to inducer toxicity we screened native expression systems using a curated oligonucleotide library of native promoter sites.

Cyanobacteria – phototrophic organisms with remarkable properties

Cyanobacteria are phototrophic bacteria that represent the only prokaryotic clade that performs oxygenic photosynthesis. Due to their large variety in morphology and metabolic capabilities they have conquered diverse biomes and are present in liminal as well as marine biomes.

A stunning feature of cyanobacteria is that many species are both oxygenic as well as diazotroph. This is remarkable as nitrogenases, an enzyme class involved in the reduction of N2 to NH3, are extremely oxygen sensitive, making oxygen exclusion an absolutely essential prerequisite for successful nitrogen fixation. All diazotroph organisms employ some form of strategy to avoid oxygen availability for nitrogenases but cyanobacteria have developed unique ways to overcome this challenge.

In higher plants, nitrogen fixation is often outsourced to symbiotic partners such as mycorrhizal fungi or in some cases root nodule bacteria. Nitrogen and phosphate compounds are exchanged for carbohydrates between higher plants and their symbiotic partners. Cyanobacteria on the other hand use a unique strategy of either temporal or spatial separation of nitrogen fixation and oxygenic photosynthesis to be able to perform both processes. A good example of spatial separation is the well-studied heterocyst formation in *Nostoc Anabaena* PCC 7120 which consists of filaments of attached cells. Under nitrogen-limited conditions, these filaments can undergo a differential development pattern in which individual cells are turned into so-called heterocyst (Golden und Yoon 2003). In these cells, the photosynthesis apparatus is degraded and oxygenic photosynthesis is completely inhibited, which is visible by their brownish color instead of the otherwise green color of Anabaena filaments. Instead, these cells start expressing nitrogenases to perform nitrogen fixation. Fixed nitrogen compounds are then exchanged with carbohydrates among neighboring photosynthetic cells, making heterocyst formation a strategy similar to higher plants' labor division, with the notable difference that no symbiotic partner species is required.

A second strategy is the temporal separation of oxygenic photosynthesis and nitrogen fixation over the day and night cycle. While at day time the sunlight is used to perform synthesis, at night photosynthesis is inhibited and nitrogen fixation is performed instead. This separation is regulated by the circadian clock of cyanobacteria, a molecular machine that acts as a global key regulator for many cellular processes (Köbler *et al.* 2021). The minimal molecular machine consists of only three protein components, KaiABC, and functions as a timekeeper, essentially dictating the time of day on a cellular level. One cycle of this clock lasts roughly 24 hours and is indirectly entrailed by the presence of light by sensing the global redox pool.

Cyanobacteria as perfect fits for phototroph – heterotroph communities

While cyanobacteria are generally prototroph and capable of growing without the need for symbiotic partners, many marine cyanobacteria can be found as partners of a highly symbiotic structure: microbial mats in the ocean are sheaths of bacteria embedded in a network of extracellular polysaccharides. These sheaths can measure from a few micrometers to multiple centimeters in depth (Ward *et al.* 1998) and offer a variety of microenvironments ranging in oxygen, nitrogen, and sulfide levels to suit a highly diverse variety of different bacteria.

The top level is often dominated by cyanobacteria, using photosynthesis to fix carbon dioxide and generate carbohydrates for the rest of the mat. Beneath an intermediate layer of heterotrophs, anoxic conditions can be found in which fermenting and nitrogen fixing bacteria as well as sulfate reducing bacteria dwell. Cyanobacteria are therefore a naturally good fit for phototrophic heterotrophic symbiotic communities. Through evolutionary adaptation in their symbiotic relations, cyanobacteria show a multitude of properties that make them very suitable for (engineered) symbiotic interactions, a few of which are highlighted below.

First, the obvious advantage is that carbohydrates generated from cyanobacterial photosynthesis can be directly fed into the growth of heterotrophic partners. The synthesis of carbohydrates from photosynthesis makes heterotroph partners independent from the presence of carbon sources in the environments and in return they can provide high complexity compounds and/or fix nitrogen.

One reoccurring issue of cultivating purely heterotrophic bacteria is media acidification, especially during fermentation under oxygen depleted conditions, when carbohydrates instead of oxygen are used as electron acceptors and carbonic acids like lactate and acetate are created in the process. But even under oxygen rich conditions, slight increases in pH need to be countered with adequate buffering. Here Cyanobacteria are a natural fit as they tend to alkalize the media during growth. This in turn limits the need for buffering agents in the media which could further reduce the running cost of large-scale cultivation processes.

Some cyanobacteria are known to secret lipid vesicles from their membrane (Lima *et al.* 2020). This could be a valuable avenue for a form of direct product exchange between symbiotic partners that does not rely on excretion and uptake and is therefore not only universal (in terms of the exchanged good) but potentially also circumvents rate limiting steps like export/import rates for compounds.

Their complex secondary metabolism extends the spectrum of engineerable substances dramatically. For example, a structurally and functionally diverse class of metabolites called terpenoids which use a 5-carbon isoprene body as a basic building block are already produced by many cyanobacterial species. The direct precursor of terpenoid biosynthesis in cyanobacteria, isopentyl di-phosphate (IDP) and dimethylallyl di-phosphate (DMADP) are produced over the MEP-pathway from pyruvate and glyceraldehyde 3-phosphate (GAP) and further “linked” to chains of different lengths. The MEP pathway is the more common pathway in eubacteria and not in eukaryotes but is also present in plant plastids, suggesting that MEP pathway genes in plastids were transferred from bacteria to plants during the degradation of ancestral cyanobacteria to plastid organelles. An advantage of using cyanobacteria to produce terpenoids is that

many terpenoid enzymes require reduction equivalents like NADPH which are abundant in these photosynthetically active bacteria. Additionally, many plant plastid originating enzymes seem to work with high activity in cyanobacteria, in some cases even better than in their native environment, for example the expression of CP450 enzymes of Sorghum bicolor thylakoid membranes in the *Synechococcus* PCC 7002 (Lassen *et al.* 2014).

The isoprene building block itself is a platform compound for chemical synthesis of many polymers such as rubber. Monoterpenes (C10) include many plant-based fragrant molecules such as limonene from citrus fruits and linalool from several flowering plants. For larger triterpenoid (C30) molecules, squalene acts as a precursor. Squalene itself is also a relevant compound used in pharmaceutical and engineering applications. Recently, squalene production was maximized in *Synechocystis* PCC 6803 by knockdown of the squalene-hopene cyclase gene *shc* and overexpression of the native squalene synthase *sqs* (Germann *et al.* 2023).

A challenge for a cyanobacteria based Green Biotechnology and for cyanobacterial research is the current reproducibility crisis. A rise in reports (M. Baker 2016) about reproducibility issues within psychological research swiftly drew attention to that matter across various other research domains, including biology. Among other factors, improper documentation of equipment and protocols are cited as the main culprits of this issue and are therefore at the center of innovations in publishing culture. These problems are especially dominant in cyanobacterial research due to the complexity of the cultivation equipment (including light quality and quantity, carbon dioxide availability and shaking speed) and the lack of media standardization. Our manuscript aims to highlight that these issues are already dominant in a community of mainly experienced investigators solely focusing on cyanobacteria. If in the future cyanobacteria are increasingly used in microbial community composition, the lack of reproducibility is likely to become a more complicated challenge when investigators without a cyanobacterial background enter the research field. Hence, this poses an

unnecessary barrier, prohibiting a new target audience in the scientific community from bringing a fresh new perspective into the field.

In the interlaboratory study of manuscript 2 we quantified the reproducibility of a simple cyanobacterial experiment. Further we discussed methodical error sources to carry this knowledge not only to researchers in the field but also to make cyanobacteria more accessible to investigators outside of the field.

Vibrio natriegens – a rapidly growing novel model organism

Vibrio natriegens is a halophile proteobacteria isolated from salt marshes. It is well known for having the fastest doubling time of any non-parasitic microbial species. In its natural habitat, it is assumed to follow a "feast and famine" strategy which includes rapid outgrowth and competition for limited carbon sources (Hoff *et al.* 2020). When rainfall in salt marshes leads to a habitable salinity, *V.natriegens* quickly outgrows competing species and captures the limited pool of available nutrients. During dry periods, when the salinity rises to unbearable conditions, cells remain dormant until the next rain period. A dormant state similar to spores for gram positive bacteria has not been discovered for *Vibrio natriegens* so far and little is known how *V.natriegens* has adapted to survival under these conditions.

Even during rain periods, the high salinity of salt marshes would result in an osmotic pressure that would drive water out of cells, leading to a loss of cell turgor. Since inner pressure is a driving factor of cell division and therefore absolutely essential, cells need to adapt to high salinity. Adaptations to osmotic stress in bacteria is facilitated by two major strategies, the "salt in" and the "salt out" strategy (Bremer und Krämer 2019). The salt in strategy involves the import of ions like sodium and chloride to rebalance the inner and outer osmotic pressure (Ma *et al.* 2010). High concentrations of charged ions however lead to cellular damage such as protein misfolding. Bacteria using the salt in strategy have therefore adapted their protein to utilize more structurally robust but metabolically costly amino acids like aspartate, threonine, and glutamine (Paul *et al.*

2008). This higher cost for protein production leads to a reduced fitness and therefore this salt in strategy is limited to habitats with a permanently high salinity and low frequency of rain falls. A much more common strategy is the salt out strategy, which is also used by *V.natriegens*. In this strategy, ions are only temporarily imported as a stringent reaction to increasing salinity. In a second step, ions are exported again and replaced with so called compatible solutes; biomolecules that include a net charge and therefore contribute to overall osmotic pressure but additionally contain a large net neutral group that limit their interaction potential with proteins, DNA and RNA to inhibit cytotoxic damage in a phenomenon described as “preferential exclusion” (Bolen und Baskakov 2001). In fact, many compatible solutes have stabilizing properties and are also used by thermophilic bacteria to prevent heat degradation. Examples of compatible solutes include molecules with multiple cellular functions such as glutamate (Dinnbier *et al.* 1988) but also molecules with mainly osmoprotective functions such glycin-betaine and ectoine (Grammann *et al.* 2002). The protective property of ectoine has been noted by the pharmaceutical industry and is used in many skin care products. As part of the salt in strategy, the production and import of compatible solutes is increased as a secondary response to high osmotic pressure. If the osmotic pressure happens to fall again, water starts entering the cell over water channels and the inner cell pressure rises. To avoid a rupture of the membrane, mechanical channels open and release water and cytosolic components, including osmoprotectants, into the environment. This salt out strategy is more common than the salt in strategy and can also be used in habitats that are subject to varying salinities such as salt marshes, brackish water, and any ground with frequent rain periods. *V.natriegens* produces ectoine among other osmoprotectants to protect against high osmotic pressure (Gregory und Boyd 2021). As a follower of the salt out strategy, *V.natriegens* is technically capable of surviving in low salinity environments but the presence of a genomic prophage leads to cell death under low salt conditions: GFP tagging of viral coating proteins showed that this phage enters the lytic phase under low salt conditions, leading to the cell death of *V.natriegens* (Pfeifer *et al.* 2019). When the prophage is removed from *V.natriegens* genome, it can grow in media with

significantly lower salinities. Compatible solutes hold interesting implications for interspecies interactions given their frequent excretion and uptake: they may serve as feed stock for potential partners and their release can be triggered by external stimuli such as decreases in osmolarity. Due to its fast growth and excellent protein expression capabilities, researchers aim to establish *Vibrio natriegens* as a platform organism that could dethrone *Escherichia coli* as the central workhorse of microbiological laboratory work (González et al. 2021).

Synthetic Inter- and intra- species communities of *Vibrio natriegens* and *Escherichia coli* shed light on fundamental aspects of labor division

In nature, microbes can often be found as part of microbial communities in which they coexist through a variety of interactions. These interactions may be beneficial such as symbiotic relationships over nutrient exchange but can also be competitive by scavenging similar nutrients or even antagonistic e.g. through antibiotic warfare. Generalizing positive interactions in microbial communities, functional "guilds" are proposed as building blocks of functional microbial communities: Individual members of a community are part of a functional guild if they perform a characteristic "service" of this guild, such as denitrification (Burke *et al.* 2011). It has been discovered that on a species level the composition of natural microbial communities undergoes constant changes, while they maintain a higher stability on guild level. Communities seem to be regulated rather by the maintenance of functionalities/services rather than the maintenance of individual species.

Such a reduction of complexity of microbial communities from hundreds of individual species to a few guilds may simplify modeling in the future and offer a new way of designing synthetic consortia (Mee et al. 2014; Wintermute und Silver 2010). Yet, little is known about how the interactions shape the composition and dynamic of the community: how does the type of offered service and the capacity to which this service

can be offered by individual guild members play into the community dynamics? How relevant are individual growth rates?

In our manuscript 3 we create synthetic interactions between *E.coli* and *V.natriegens* mutants based on mutually exclusive amino-acid based auxotrophies. With various perturbations, we aim to probe the composition stability of these communities. The goal is to contribute to the understanding of the underlying mechanisms and balancing factors that make a microbial community stable with respect to its composition.

Aim of this thesis

The study of individual species in microbiology is far advanced regarding a selection of model organisms, yet recently microbial consortia are moving into the focus of study. The advantages of labor division make them a potential platform for biotechnological processes and their impact on health and environment calls for a thorough understanding of their inner workings. Yet, the fundamental principles governing their interactions and the shape of their compositions are barely understood. Aim of this thesis was to aid in the paradigm shift from classical microbiology focused on individual species towards complex microbial consortia as the object of interest. The following aspects were touched as part of this thesis:

Further domestication of an inter-kingdom symbiotic system involving *Hydra vulgaris* AEP1-3 and its native bacterial colonizers via the development of expressions tools for *Curvibacter sp.* AEP1-3 – **Manuscript 1**

Standardization of cyanobacterial cultivation and assessment of their reproducibility to improve the accessibility of cyanobacteria for scientists outside of this respective field – **Manuscript 2**

Bottom-up development of a minimal and easy to use synthetic microbial consortia from *Escherichia coli* and *Vibrio natriegens* to study how these synthetic communities determine and regulate cell to cell abundance ratio – **Manuscript 3**

Manuscripts

The following manuscripts have been attached as part of this cumulative thesis:

1) "**Oligonucleotide Library Assisted Sequence Mining Reveals Promoter Sequences With Distinct Temporal Expression Dynamics For Applications In *Curvibacter***" by Mager and Becker *et al.* has been published on biorxiv at the time of submission of this thesis. DOI: https://doi.org/10.1101/2024.03.24.586450

Contributions: Conceptualization, Methodology, Investigation, Validation, Data Curation, Writing – Initial Draft, and Writing – Revising and Editing

2) "**Interlaboratory Reproducibility in Growth and Reporter Expression in the Cyanobacterium *Synechocystis* sp. PCC 6803**" by Mager and Hernandez *et al.* has been published at ACS Synthetic Biology. DOI: 10.1021/acssynbio.3c00150

Contributions: Conceptualization, Methodology, Resources, Investigation, Validation, Data Curation, Supervision, Writing – Initial Draft, and Writing – Revising and Editing

3) "**Synthetic inter- and intraspecies symbiosis between *E.coli* and *V.natriegens* shows robust cell-to-cell abundance ratio**" by Mager and Wilken *et al.* is being published as part of this thesis.

Contributions: Conceptualization, Methodology, Investigation, Validation, Writing - Initial Draft, and Writing – Revising and Editing

OLIGONUCLEOTIDE LIBRARY ASSISTED SEQUENCE MINING REVEALS PROMOTER SEQUENCES WITH DISTINCT TEMPORAL EXPRESSION DYNAMICS FOR APPLICATIONS IN *Curvibacter* SP. AEP1-3

Maurice Mager[1 *], Lukas Becker[1 2 *], Nina Schulten[1], Sebastian Fraune[2], and Ilka M. Axmann[1 +]

[1]Institute for Synthetic Microbiology, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[2]Institute for Zoology and Organismic Interactions, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[+]corresponding author
[*]contributed equally

ABSTRACT

The *ß-proteobacterial* species *Curvibacter* sp. AEP1-3 is a model organism for the study of symbiotic interactions as it is the most abundant bacterial colonizer of the basal metazoan *Hydra vulgaris*. Yet, genetic tools for *Curvibacter* are still in an infancy: few promoters have been characterized for *Curvibacter*. Here we employ an oligonucleotide based strategy to find potential expression systems derived from the genome of *Curvibacter*. Potential promoters were systematically mined from the genome in silico. The sequences were cloned as a mixed library into a mCherry reporter gene expression vector and single positive candidates were selected through Flow Cytometry based sorting to be further analyzed through bulk measurements. From 500 candidate sequences, 25 were identified as active promoters of varying expression strength levels. Bulk measurements revealed unique activity profiles for these sequences across growth phases. The expression levels of these promoters ranged over two orders of magnitudes and showed distinct temporal expression dynamics over the growth phases: while 3 sequences showed higher expression levels in the exponential phase than in the stationary phase, we found 12 sequences saturating expression during stationary phase and 10 that showed little discrimination between growth phases. From our library, promoters the genes encoding for DnaK, RpsL and an AHL synthase stood out as the most interesting candidates as their expression profiles fit a variety of applications. Examining the expression levels of successful candidates in relation to RNAseq read counts revealed only weak correlation between the two datasets. This underscores the importance of employing comprehensive high-throughput strategies when establishing expression systems for newly introduced model organisms.

Keywords *Curvibacter* sp. AEP1-3 · *Hydra vulgaris* AEP · Promoter · FACS · Genome Mining

1 Introduction

Curvibacter sp. AEP1-3 (hereafter *Curvibacter*) is a rod-shaped *ß-proteobacterial* species best known for its symbiotic interaction with *Hydra vulgaris* [1], a freshwater polyp of the basal metazoan family *Cnidaria*, a sister group to the *Bilateria*. Together with other members of *Hydras* microbiota, they form a complex system of bacteria-bacteria as well as bacteria-host interactions [2, 3, 4]. While the host provides an ecological niche to its colonizers, the microbiota affects mobility, asexual reproduction as well as protection against fungi [1]. This symbiotic relationship provides an invaluable avenue for the study of inter kingdom interactions and allows for the exploration of general principles of symbiosis in the natural world such as the remarkable host-microbe communication between *Curvibacter* and *Hydra vulgaris* established through the exchange of N-acetyl homoserine lactone [2].

The limited genetic accessibility of *Curvibacter* restricts advancements in genetically manipulating its cells, thereby impeding progress in the field of interkingdom symbiosis. Genomic modifications over homologous recombination [5] are cumbersome [6] and have a low success rate. *Curvibacter* cells are amenable to transformation using RSF1010 vector [7] constructs through conjugation with *E. coli* donor cells but only a limited number of promoters are accessible for use and none of them have been characterized to date. In this study, we set out to develop a strategy to create tool kit of viable promoters for *Curvibacter* to promote its use as a model organism.

The Anderson Collection of Synthetic Promoters is a good reference for the development of orthogonal constitutive expression systems (`http://parts.igem.org/Promoters/Catalog/Anderson`). The collection provides a range of expression levels that have been well characterized in many species such as *E.coli*, *V. natriegens* and some *Cyanobacteria* [8, 9] but as orthogonal promoters they usually show the same temporal expression dynamics in the form of a stable, constant activity over growth conditions, providing the same transcript level over the entire time of cultivation. However, when expressed from a plasmid, the total expression level of most promoters often increases significantly in the stationary phase due to changes in the copy number of most plasmids: the copy number generally increases with slower growth during stationary phase, leading to higher transcript expression and increased protein levels [10, 11, 12]. While this may be either desirable for some applications or irrelevant in experimental setups where cells are only observed during logarithmic growth, such accumulation may prove detrimental, for example, in long-term experiments. Therefore, for the development of plasmid-based expression systems with stable expression in different growth phases, orthogonal promoters that cannot absorb this burst of expression may not be the most suitable.

While hand picking or designing individual sequences with a predicted expression strength is a valid strategy to develop expression systems, the use of entire sequence libraries provides a promising alternative due to the high throughput of tested sequences [13]. Such libraries are generated by synthesizing oligonucleotide sequences on highly sophisticated commercial DNA synthesis platforms that allow the simultaneous generation of many sequences at the same time. These libraries are collected and purified in a single sample and can be used for cloning applications to create a library of plasmids each containing a different synthesized sequence, as well as, in our case, a reporter sequence such as the Green Fluorescence Protein (GFP) or mCherry. Downstream, the use of flow cytometry and cell sorting can aid in picking positive candidates from such libraries to avoid the extensive effort of manually picking and analyzing individual colonies. The aforementioned libraries can consist of sequences with varying degrees of randomization, generated by using mixed nucleotides during synthesis which allows for the incorporation of any base by chance [14]. This strategy is necessary for projects in which the investigators aim to obtain the best suited sequences from a bias free sequence space or if there is no information available that could reduce the degree of freedom. The GeneEE library of Lale *et al.* [15] follows exactly this approach by using long stretches of randomized nucleotides to find novel promoter sequences de novo.

An alternative we employ here is the generation of highly curated libraries. Limiting the pool only to sequences with a high probability of success simplifies downstream processes and can yield many more positive candidates in significantly smaller libraries, making it an easier and more cost efficient method. The generation of such libraries can be facilitated by neural networks, trained on existing promoter sequences, extracted from curated libraries such as the Prokaryotic Promoter Database (PDD) [16] or as in our case simply by using existing sequences harvested directly from the target species genome [17]. The latter approach results in finding promoter sequences that won't be orthogonal, but it is a valid approach to also find expression systems which are either inducible or show a desired temporal expression dynamic in certain growth phases. Moreover, these sequences already inherit the genomic context for specific regulation, to a degree. By extracting those sequences, it allows the identification of different regulatory elements depending on the extracted length and culture conditions.

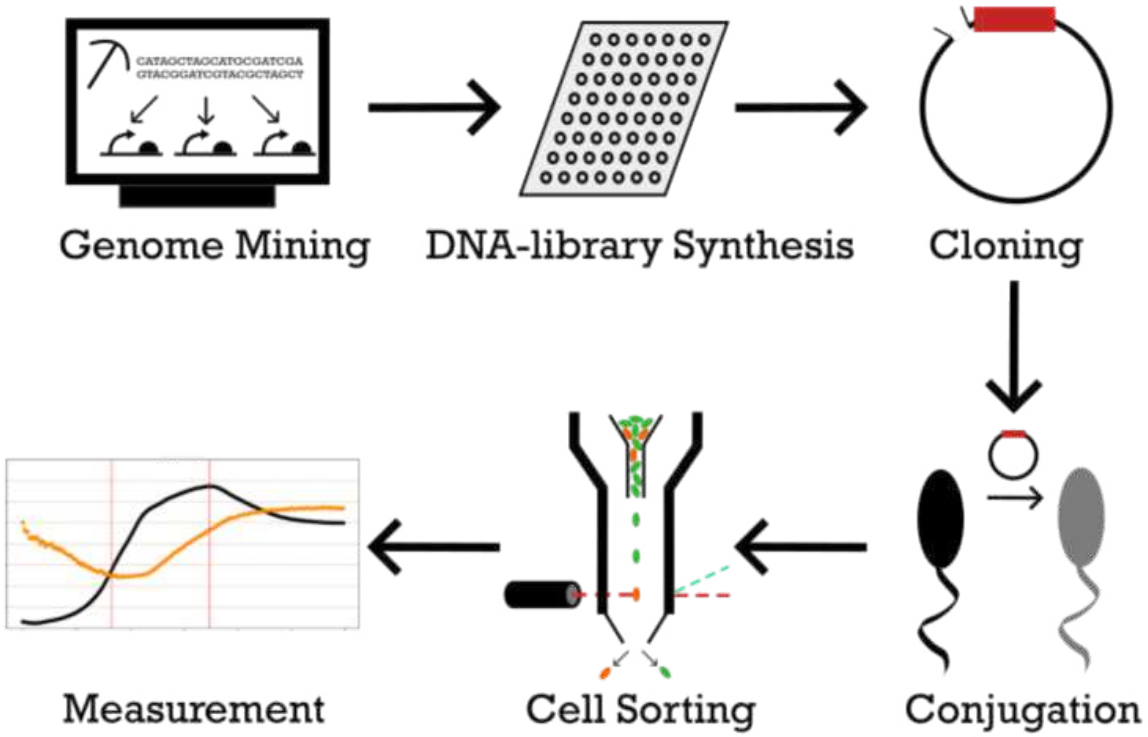


Figure 1: **Schematic overview of the applied workflow.** Design of the oligonucleotide library starts by mining the host genome for suitable promoter sequences. Sequences are synthesized, cloned and imported into the host species and sorted for activity by Flow Cytometry Cell Sorting. In depth assays of individual sequences characterize each candidate promoter in detail.

Aim of this study is the discovery and characterization of novel expression systems for the use in expression vectors for *Curvibacter*, with a special focus on promoters that provide stable expression independently on growth phases in liquid media. To increase the odds of individual promoter sequences, promoter and ribosomal sequences were not further discriminated, but entire $5'$ upstream non-coding regions were used instead. Candidate sequences for this study were directly harvested (Figure 1) from the *Curvibacter* genome sequence (GCF_002163715.1). Sequences were ordered as single stranded oligonucleotides and cloned into an expression vector with mCherry as reporter gene. mCherry expression serves here as quantifiable proxy for expression strength of a given candidate sequence. Positive candidates were selected with the aid of flow cytometry and cell sorting and were subsequently analyzed via bulk fluorescence measurement. The extracted $5'$ upstream non-coding regions displayed a wide range of activity levels that can be used for different applications. We discovered different behavior of the selected candidates in terms of their activity over growth phases: while most candidate promoters showed typical increased activity in late exponential- to stationary phase, we also found some promoters to be slightly more active in exponential than in stationary phase. Several sequences were found to show similar expression levels during exponential- as well as stationary phase. We mapped the strength of our characterized expression systems against normalized RNAseq read counts and found weak correlation between the two data sets, showing that RNAseq does not generally serve as a reliable tool to predict promoter strength, at least in an *in trans* context, showing how useful oligonucleotide library based approaches are to find promoter sequences with desired features.

2 Material & Methods

2.1 Extraction of Promoter Sequences in the Genome of *Curvibacter* sp. AEP1-3

To extract candidate promoter sequences suitable for our experiments, first we retrieved the GenBank file for the *Curvibacter* genome from the GenBank FTP site, serving as our primary source of genomic data. All subsequent data processing and analysis was performed in the Python 3 programming environment within the Jupyter Notebook development platform. To extract and manage genomic data, we utilized the BioPython library for parsing the GenBank file. The scripts and all relevant data files can be found on our github repository (`https://github.com/Kanomble/curvibacter_promotor_studies`).

Local start and stop positions as well as the genomic orientation of all sequence elements labeled as "genes" within the *Curvibacter* genome were structured into a dictionary. Gene sequence identifiers were employed as keys but simultaneously also stored in a list object, preserving their exact positions within the genome. This list was instrumental for tracking sequence location and order. To provide context for the gene sequences, we implemented a parsing process to retrieve information about the current gene sequence, the previous gene sequence, and the next gene sequence based on the current identifier. To refine our dataset, we filtered out sequences that didn't meet length criteria, specifically sequences with more than 170 base pairs and less than 60 base pairs, ensuring that only appropriately sized $5'$ upstream non-coding regions were included in the analysis. Further filtering steps involve removing sequences with opposing orientations and overlapping segments. To remove potential tRNA and rRNA promoter regions or promoter regions without any activity the resulting data frame was merged with a data frame of a previously conducted transcriptome analysis of an RNAseq experiment with *Curvibacter* wildtype cells (Further details s. Pietschke *et al.* 2017 [2]). The merging step eliminated all potential promoter sequences from genes that are not part of the transcriptome analysis, and as a result, these genes do not exhibit any expression levels in the standard R2A growth media of *Curvibacter*. In addition, it excluded all genes labeled as tRNA or rRNA since they were not present in the transcriptome data frame. This step was vital in eliminating conflicting or redundant information within the analyzed gene sequences. In a next step five sequences have been filtered out for containing restriction binding sites for BsaI, BsmBI and BbsI. Sequences containing these binding sites were excluded. After this 722 candidates remained in the data frame. As the sequence synthesis order was capped to 500 sequences this selection was cut further: the 350 smallest sequences included all sequences from 60 to 98 bp were included first, as the sequence synthesis order was capped to a maximum length of 150 bp including added restriction sites on each site. The remaining sequences were sorted by their read counts derived from the RNAseq experiment mentioned above. As the read counts encompassed three orders of magnitudes of read count levels, the 50 $5'$ upstream regions from genes with reads of each order of magnitude were picked to cover a wide range of potential expression levels. All sequences were subsequently trimmed from the $5'$ end to reach a length of 98 base pairs. Restriction sites for golden gate cloning were then added, followed by the insertion of random bases behind the restriction site to meet the synthesis specifications, ensuring that all sequences reached a uniform length of exactly 150 base pairs.

2.2 Transcriptome Read Mapping

Raw RNAseq reads were obtained as described in Pietschke *et al.* 2017 [2]. Before mapping the obtained sequences against the reference genome of *Curvibacter*, the sequences were subjected to quality control and preprocessing. Sequences were trimmed using trimmomatic (Version 0.39, [18]). Trimmed FASTQ files were analyzed for quality using FASTQC (Version 0.11.9, [19]). Subsequently, the preprocessed sequences were mapped against the reference genome of *Curvibacter* utilizing the kallisto mapper (Version 0.50.0, [20]). RAW sequences are uploaded to NCBI as BioProject PRJNA1082616. They will be made public after full publication and are available as of now upon request. The relevant code can be found in this GitHub repository: (`https://github.com/Kanomble/curvibacter_transcriptomics`).

2.3 Library Golden Gate Cloning

40 ng/kb of DNA library (but a minimum of 5 ng per reaction) were added to 20ng/kb of entry vector to maximize yield of successful integration without compromising efficiency. Golden Gate Reaction was performed according to standard protocol, but the final digest duration was increased to 1 h to reduce entry vector religation. 50 µl of highly competent Dh5 Alpha cells were transformed with 10 µl of Golden Gate reaction and plated on 4 selection plates to reduce colony crowding. After 24 h of growth, a minimum of 5.000 (for a library with 500 sequences) colonies of equal size were obtained and scraped off the plates. Plasmid DNA was extracted from the cell mixture using a standard MiniPrep kit from Macherey Nagel. The *E. coli* donor strain for conjugation into *Curvibacter* was retransformed with the plasmid library to yield a minimum of 5.000 colonies.

2.4 Conjugation of Library Vectors into *Curvibacter* sp. AEP1-3 glmS::GFP

Curvibacter with a GFP insertion in the glmS locus was inoculated from a fresh plate into R2A+ media and grown for 36h to stationary phase prior to conjugation. DAP auxotroph *E. coli* donor cells were directly scraped off transformation plates, washed in LB media and used for conjugation.

3 ml of *E. coli* donor cells at OD1 and 5 ml of *Curvibacter* stationary phase at OD2 were mixed and the conjugation mix was centrifuged at 5000 rpm for 5 min. Cells were washed in 1 ml of R2A+ and centrifuged as before. Cells were resuspended in 100 µl of R2A+ and spotted on 4 plates of R2A media without the addition of DAP or antibiotics.
Plates were incubated overnight at 30°C and cell spots were scraped off and washed in 1 ml of R2A+. 50 µl of the conjugation mixture was separately plated on a R2A plate containing the respective antibiotic for quality control. The rest of the mixture was centrifuged, resuspended in 200 µl of the remaining media and spread equally over 8 R2A plates containing the respective antibiotic. If the colony count on the quality control plate was higher than 250, the total conjugation yielded over 5000 conjugation events and the plates could be used further. Conjugation plates were scraped and *Curvibacter* cells were washed in 1 ml of R2A+. Cells were diluted to OD 0.02 in R2A and sorted as described below.

2.5 Flow cytometry and Cell sorting

Curvibacter cells were sorted using the CytoFlex SRT Benchtop Cell Sorter. Forward scatter (FSC) and side scatter (SSC) was measured using a 488 nm laser and a 488/8 nm Bandpass filter. Violet side scatter (VSSC) was measured using a 405 nm laser and a 405/5 nm Bandpass filter. GFP fluorescence was measured using a 488 nm laser and a 525/40 nm Bandpass filter. mCherry fluorescence was measured using a 561 nm laser and a 610/20 nm Bandpass filter. Following gain settings were used:

Table 1: **Gain settings for flow cytometry**

Filter	**Gain setting (X/3000)**
FSC	76
SSC	299
VSSC	106
GFP	196
mCherry	1216

The cell population of interest was sorted based on a mCherry fluorescence higher than the background signal. To determine the gate for background fluorescence, the background strain *Curvibacter* AEP1-3 glmS::GFP was used and the gate was set to exclude 99% of this population. The subpopulation was split into four quadrants based on their GFP and mCherry signal: green fluorescence from the genomic GFP over 100.000 AU was considered "high green" (HG), below was considered "low green" (LG) fluorescence. Equally, red fluorescence above 7.000 AU was considered "high red" (HR), below was considered "low red" fluorescence. The quadrant of each sorted cell is stated in the inventory list (s. Supplementary Table 2). 2000 Cells were sorted into 4 tubes containing 100 µl of R2A based on their combined red and green fluorescence signal.
This mixture was plated on R2A plates and incubated at 30°C for 48 h. Colonies of surviving cells were tested for successful promoter integration in the expression vector using cPCR and Sanger Sequencing. Colonies were inoculated in 1 ml of R2A+ containing the respective antibiotic and grown for 48 h. 1 ml of 50% (v/v) glycerol in distilled water was added and cultures were frozen at -80°C for further use.

2.6 Bulk Fluorescence Intensity measurements

Curvibacter cells containing one of the selected $5'$ upstream non-coding regions were inoculated from a fresh R2A+ plate into R2A+ media and grown for 36 h at 30°C in 24 well plates in a BMG labtech Clariostar plate reader until stationary phase. From these pre-cultures, main cultures were inoculated to an OD of 0.05. Growth and fluorescence was monitored over 36 h. mCherry fluorescence from the reporter constructed was monitored at 570/15 nm bandwidth excitation and 620/20 nm bandwidth emission and GFP fluorescence from the genomically integrated GFP was monitored at 470/15 nm bandwidth excitation and 515/20 nm bandwidth emission.

2.7 *Curvibacter* Growth Media

R2A+ media was prepared by adding additional nutrients to premixed R2A from Carl Roth.

Table 2: **R2A+ media composition.**

Ingredient	Amount
R2A (premixed)	3 g
Peptone	4 g
Glucose	2.5 g
Yeast extract	1 g
distilled water	up to 1 L

2.8 Mathematical Operations for RFU Assessment

Fluorescence intensity measurements were adjusted to eliminate background signals by employing a media-only control well. The corrected fluorescence intensity for a specific fluorophore was obtained by subtracting the signal in the presence of media control from the signal of that fluorophore alone.

$$FI(fluorophore) = signal(fluorophore) - signal(media\ control) \tag{1}$$

The fluorescence intensity FI(fluorophore) is the value for the Relative Fluorescence Unit (RFU) of the measured fluorophore (mCherry or GFP, signal(fluorophore)) substracted by the RFU of the media control (R2A+, signal(media control).

The datasets obtained from the plate reader (BMG labtech clariostar), which included Biomass and FI data points for mCherry (RFP) and GFP, underwent a filtering process to remove outliers and reduce noise. This smoothing step utilized the savgol_filter function from the Python scipy package, implementing the Savitzky-Golay smoothing technique [21]. The Savitzky-Golay smoothing filter is a data processing method commonly employed in signal processing and data analysis. Its purpose is to smooth noisy data while preserving essential signal features. This step was taken to enhance the accuracy of determinations regarding stationary and exponential growth phases.

The time points for defining the exponential and stationary phases were determined by applying the Savitzky-Golay smoothing technique to the raw OD600 values and calculating the maximum slope for the exponential phase, as well as the maximum OD600 value for the stationary phase. To account for variations in biomass, relative fluorescence units (RFU) were further normalized using the GFP intensity as a reliable proxy for biomass. This normalization was carried out as follows:

$$Normalized\ FI\ values = FI(RFP)/FI(GFP) \tag{2}$$

In order to assess changes in promoter activity across different growth phases, FI values were compared between the exponential and stationary phases. This calculation was performed using the formula:

$$Differential\ activity = \frac{Normalized\ FI\ values(exponential\ growth\ phase)}{Normalized\ FI\ values(stationary\ growth\ phase)} \tag{3}$$

Values below 0.7 indicated promoters more active during the stationary phase, values above 1.3 suggested greater activity in the exponential phase, and values falling in between were indicative of promoters that did not show a significant preference for either growth phase. This method allowed for a comprehensive assessment of promoter behavior in relation to different growth phases.

2.9 Inference of Transcription Factor Binding Motifs

To identify potential transcription factor binding sites and motifs (TF-Motifs), the nucleotide sequences of the 33 $5'$ upstream non-coding regions identified via Flow Cell Cytometry were used as input sequences for the XSTREME algorithm ([22]). The XSTREME algorithm conducts a comprehensive motif analysis. It was configured with default settings to search for binding motifs within the prokaryotic CollectTF database (`http://www.collectf.org/browse/home/`), which contains known bacterial transcription factor binding sites.

3 Results

3.1 Development of a streamlined workflow for promoter mining

We designed a synthetic biology approach for promoter mining in sequenced bacterial species. The details of the workflow to filter potential promoter sites are described in the first methods section in detail. Briefly, candidate promoter sequences for this study were directly harvested from the *Curvibacter* genome sequence. As the *Curvibacter* genome contains 4096 predicted genes, nearly the same amount of intergenic regions (as some genes overlap) exist as potential promoter sites. Filtering steps removed intergenic regions of divergent genes; with a length <60 and >170bp; promoter sites of tRNA and rRNA genes as well as those intergenic regions containing restriction sites of BsaI, BsmBI and BbsI. A table of all ordered sequences is available in Supplementary Table 1. The complete workflow is available on Github (`https://github.com/Kanomble/curvibacter_promotor_studies`) and can be adapted to any bacterial genomic sequence. In the following we show how representative our selection is for the genome of *Curvibacter* sp. AEP1-3.

3.2 Representativeness of extracted 5′ upstream non-coding regions in *Curvibacter* sp. AEP1-3

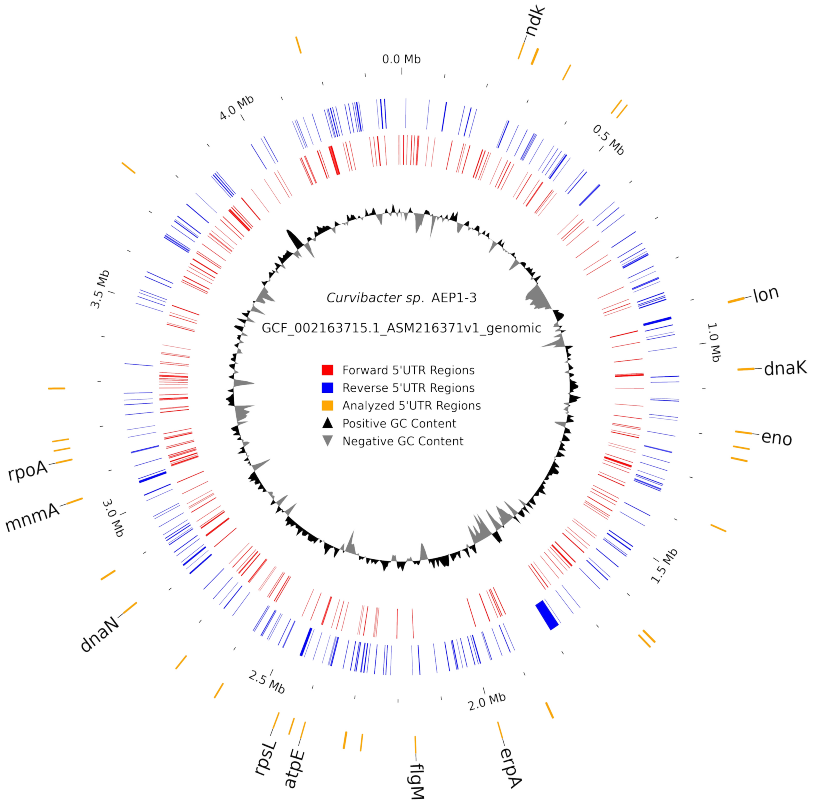


Figure 2: **Distribution of extracted 5′ upstream non-coding regions within the genome of *Curvibacter*.** The innermost circle is composed of a density plot that showcases the GC content of the respective genome regions. The two following red and blue circles highlight the initially extracted 500 5′ upstream non-coding regions. Blue lines correspond to genes with a forward orientation (clockwise), red lines vice versa. The outer circle represents the 33 via Flow Cytometry sorted 5′ upstream non-coding regions. 5′ upstream non-coding regions of CDS regions labeled as "hypothetical protein" or with protein names that are too long are not labeled.

The 33 5′ upstream non-coding regions that were confirmed by Flow Cytometry (see next section) are evenly distributed throughout the entire *Curvibacter* genome (s. Figure 2). The length of the confirmed 5′ upstream non-coding regions ranges from 60 bp to 146 bp. Further, sequences of these confirmed candidate sequences were analyzed for the occurence of common motifs (s. Figure 3).

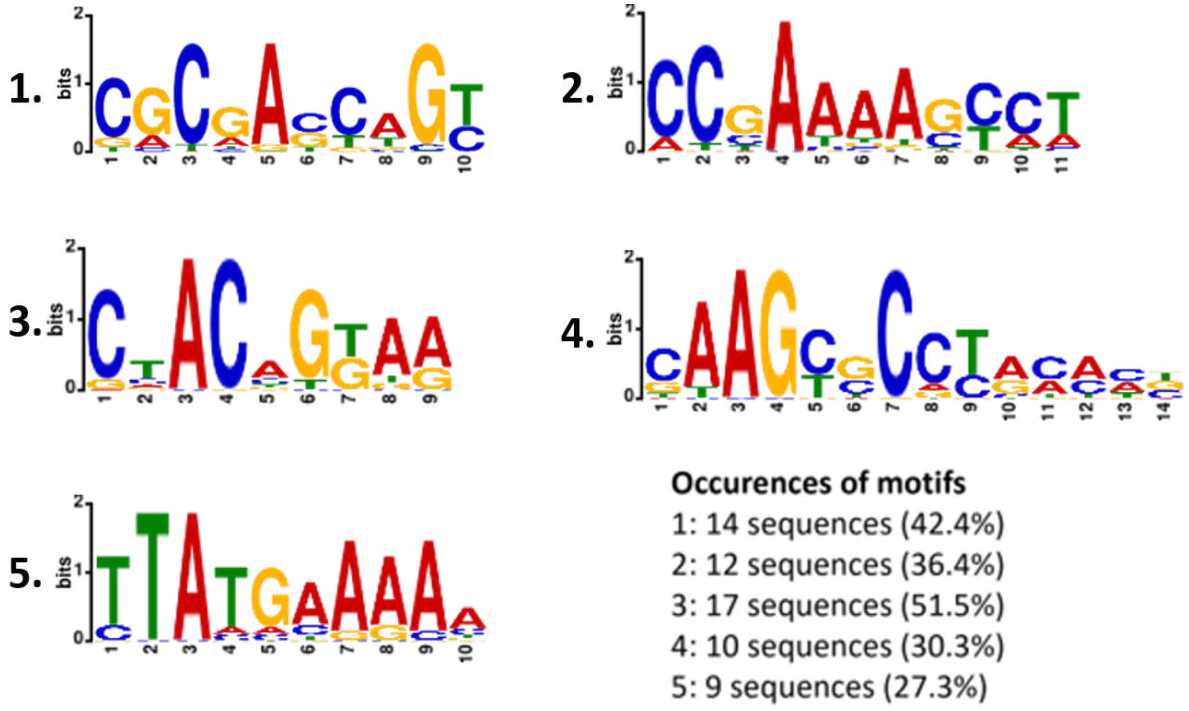


Figure 3: **Conserved sequence motifs within the analyzed 33 5′ upstream non-coding regions of *Curvibacter*.** Sequence motifs are part of the CollectTF database. Figure highlights the 5 most prevalent motifs, additional motifs can be found in Supplementary Table 3.

Among the 33 5′ upstream non-coding regions, five distinct transcription factor binding site motifs (TF-Motifs) have been identified using the XSTREME algorithm from the MEME-suite portal, with a setup that enables searching for TF-Motifs within the CollectTF database for bacterial transcription factor binding sites [22]. The identified motifs are found in several transcriptional regulators. For instance, the first motif (1) exhibits similarities to TF-Motifs of the AmrZ and LasR genes of *Pseudomonas aeruginosa* [23, 24]. AmrZ serves as a transcriptional activator and/or repressor of virulence factors, as well as genes involved in environmental adaptation. LasR, on the other hand, serves as a transcriptional activator of the elastase structural gene LasB [25] and it is considered as a transcriptional activator for virulence genes in *Pseudomonas aeruginosa* [26]. In addition, LasR is a homolog to the CurR1 and CurR2 genes in *Curvibacter*, which are described in Pietschke *et al.* [2]. In 23 of 25 5′ upstream non-coding region strains that were further analyzed by bulk fluorescence measurement (further referred to as candidate promoters) (s. Figure 4a) at least one of the inferred TF-Motifs has been identified (s. Supplementary Table 3). A detailed description of all identified motifs can be found in Supplementary Table 3.

3.3 *Curvibacter* strains carrying functional reporter constructs show a range of expression levels.

33 unique candidate promoters sorted by flow cytometry were further analyzed for their expression level throughout different growth phases (s. Figure 4a) using bulk fluorescence measurement in a plate reader (full list of candidates s. Supplementary Table 1). The candidates vary in length and GC content. For instance, CPL0025 has a length of 76 bp with a GC content of 40%. In contrast, CPL0095 is 113 bp long with a GC content of 42%, and CPL0022 spans 123 bp with a GC content of 46%. From 33 total sorted candidates, 25 showed detectable expression levels and were therefore included in the following analysis (further referred to as candidate promoters).

All strains are based on the same *Curvibacter* background strain containing a genomic GFP integration in the *glmS* locus with a constant expression level relative to biomass until early stationary phase (s. Supplementary Figure 2). As this GFP signal was less noisy compared to OD measurements at optical densities near OD 0.1 and as the fluorophore

accumulation in the late stationary phase due to protein aggregation was nearly identical for both the reporter mCherry construct and the genomic GFP integration, the GFP signal was further used as a normalization factor for the RFP signal (s. equation three in Mathematical Operations for RFU Assessment). This reduces noise in low optical density cultures and leads to a stable signal in the later stationary phase, making it easier to determine reporter activity during exponential- and stationary phase. Relative fluorescence units are therefore given as the fraction of RFP/GFP intensity. Supplementary Figure 2 shows the correlation between GFP and biomass of the background strain, showing that GFP expression remains constant relative to biomass until stationary phase is reached and linearly increases after reaching stationary phase in the same way that it does in in trans expression systems, effectively negating this drift. Figure 5 a-d shows a linear relation for RFP/GFP during the stationary phase as a result of this.

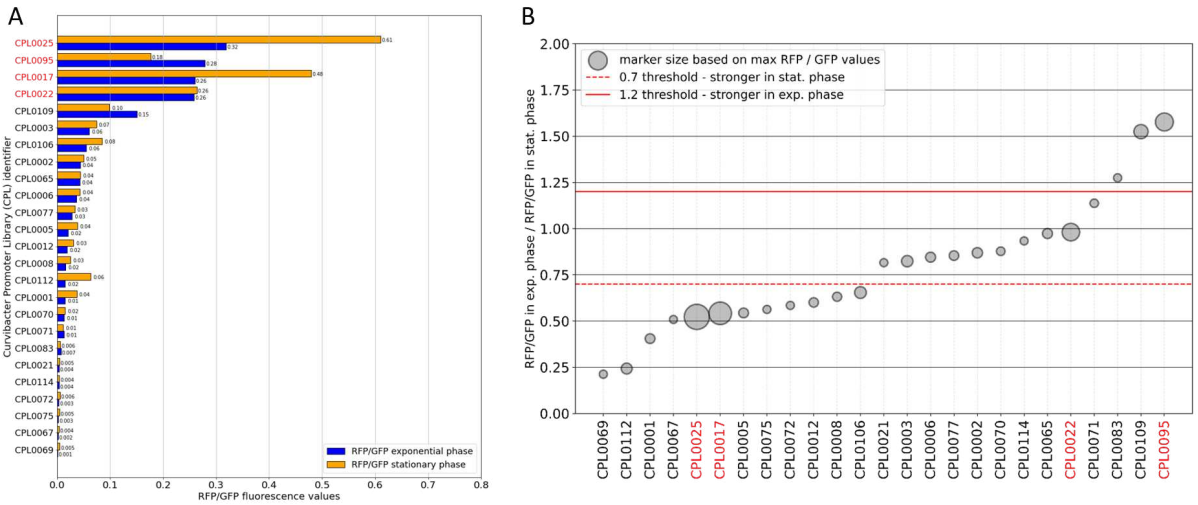


Figure 4: **Left: Promoter expression levels as measured by RFP/GFP levels within the exponential growth phase and stationary growth phase.** Promoter sequences are sorted in decreasing order based on the RFP/GFP value during the exponential growth phase. The red marked CPL identifiers represent the highlighted sequences in the section below. **Right: Exponential phase and stationary phase proportion of promoter expression strengths.** Promoters with values under 0.7 are more active during the stationary growth phase, while promoter sequences with values above 1.2 are considered to be more active during the exponential growth phase. Dot size correlates with overall expression level. The red marked CPL identifiers represent the highlighted sequences in the section below.

Figure 4a shows the relative fluorescence units (RFUs) of mCherry normalized to GFP for all candidates, which showed expression levels above the background noise. The above values serve as a proxy for the relative expression levels of their corresponding promoter sequences during the exponential phase and the stationary phase and should guide investigators in picking expression systems for their specific use case. Expression levels range from 0.61 to 0.005 relative to GFP in the stationary phase, encompassing two orders of magnitude in terms of expression strength. Among the 25 analyzed candidate sequences, 12 show less than 75% of activity during the exponential phase compared to the stationary phase, while 10 display relatively consistent expression strength regardless of the current growth phase (Figure 4b). Additionally, three candidate sequences demonstrate at least 1.25-fold higher activity during the exponential growth phase than in the stationary phase.

3.4 Activity level of candidate promoters shows distinct temporal expression dynamics over growth phases

In this section we provide detailed information of the measured fluorescence activity over time of three (plus CPL0017 as control) selected candidate promoters of *Curvibacter*. We recommend these promoters for further experimental use as they cover a range of different temporal expression patterns and strengths.

The CPL0025 (Figure 5B) sequence is the promoter of the gene AEP_RS11205, annotated as the acyl-homoserine-lactone (AHL) synthase (RefSeq protein identifier: WP_232459811) and described as Curl2 in Pietschke *et al.* [2]. The full promoter region (519 bp) of the AHL synthase Curl2 is activated by homoserine lactones, a bacterial quorum sensing molecules which plays a crucial role in regulating gene expression in response to population density [27]. Here we show that even the smaller promoter region of 76 bp could drive the expression of our reporter construct. The 5'UTR shows lower expression levels during exponential growth and elevated expression levels after entering stationary phase. CPL0025 was the strongest candidate among all tested promoters, surpassing even the expression level of the

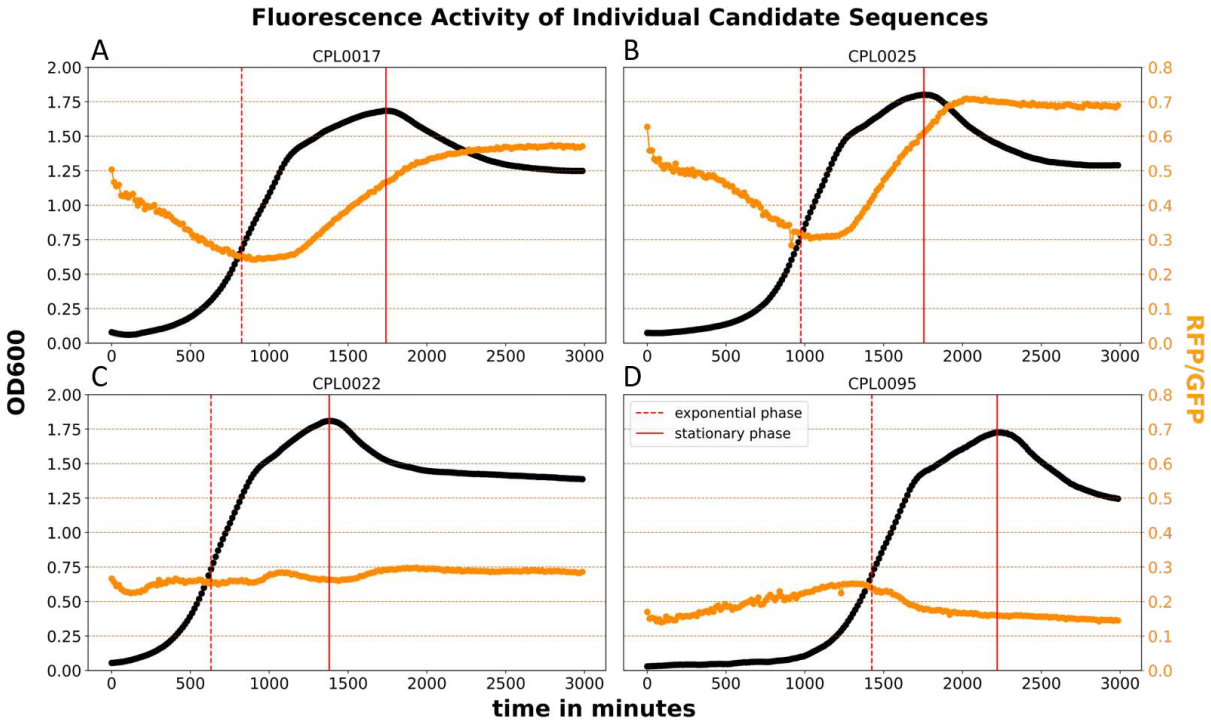


Figure 5: **Promoter expression levels for the highlighted sequences CPL0017 (control), CPL0025, CPL0022 and CPL0095**.Full and dotted lines represent time point of exponential and stationary phase, respectively, which was used for Figure 5B. RFU (orange line) equals RFP intensity (candidate promoter) normalized to GFP intensity (which is located on the genome).

highly active J23100-RBS* promoter (CPL0017), which was used as reference (Figure 5a). A TF-Motif similar to the AHL activated transcriptional regulator LasR (s. Supplementary Table 3) was not detected within the 76 bp long 5'UTR of CPL0025. The only motif found in this 5'UTR is similar to TF-Motifs found in Gram-positive bacteria, the motif is located within the positions 20 - 29 (TTACAAGAAA) of the 5'UTR. Specifically, similar to motifs of the global transcriptional regulator CodY from *Lactococcus lactis* and *Streptococcus pyogenes* as well as for CcpA from *Streptococcus pneumoniae* (s. Figure 3 5. and Supplementary Table 3) [28, 29, 30, 31].

Sequence CPL0022 (Figure 5c) is the promoter of the gene AEP_RS05045 expressing DnaK (RefSeq protein identifier: WP_087494375), a molecular chaperone protein of the (Heat-shock-protein 70) Hsp70 family. The *dnaK* candidate promoter shows a very constant expression level throughout all growth phases in *Curvibacter* compared to all other tested promoters, with minor bursts of transcriptional activity during late exponential and early stationary growth phases. In comparison with other sequences in this study, the *dnaK* shows a very constant expression level throughout all growth phases shows a relatively high expression level and very little bias towards growth phases. CPL0022 contains 9 bp long TF-Motif for LasR binding within the positions 12 - 21 (CACAACCAGC) of the 5'UTR sequence. Additionally, CPL0022 contains a CodY motif from *Bacillus anthracis* and a ExpR motif from *S. melliloti* (s. Supplementary Table 3).

CPL0095 (Figure 5d) is the promoter of the gene AEP_RS11420 (RefSeq protein identifier: WP_011466063) expressing RpsL, a 12S protein component of the 30S ribosomal subunit. This candidate promoter displays high activity during the exponential phase, with a steady increase in activity until the mid-exponential phase. The activity then decreases to approximately half of its maximum during the stationary phase. CPL0095 contains a range of sequence motifs, such as a LasR motif from *P. aeruginosa*, CodY from *B. anthracis* as well as *S. pyogenes*, and a LexA motif from *V. parahaemolyticus* (s. Supplementary Table 3).

3.5 Comparison of RNAseq- and RFU- assessed promoter strength only shows a weak correlation

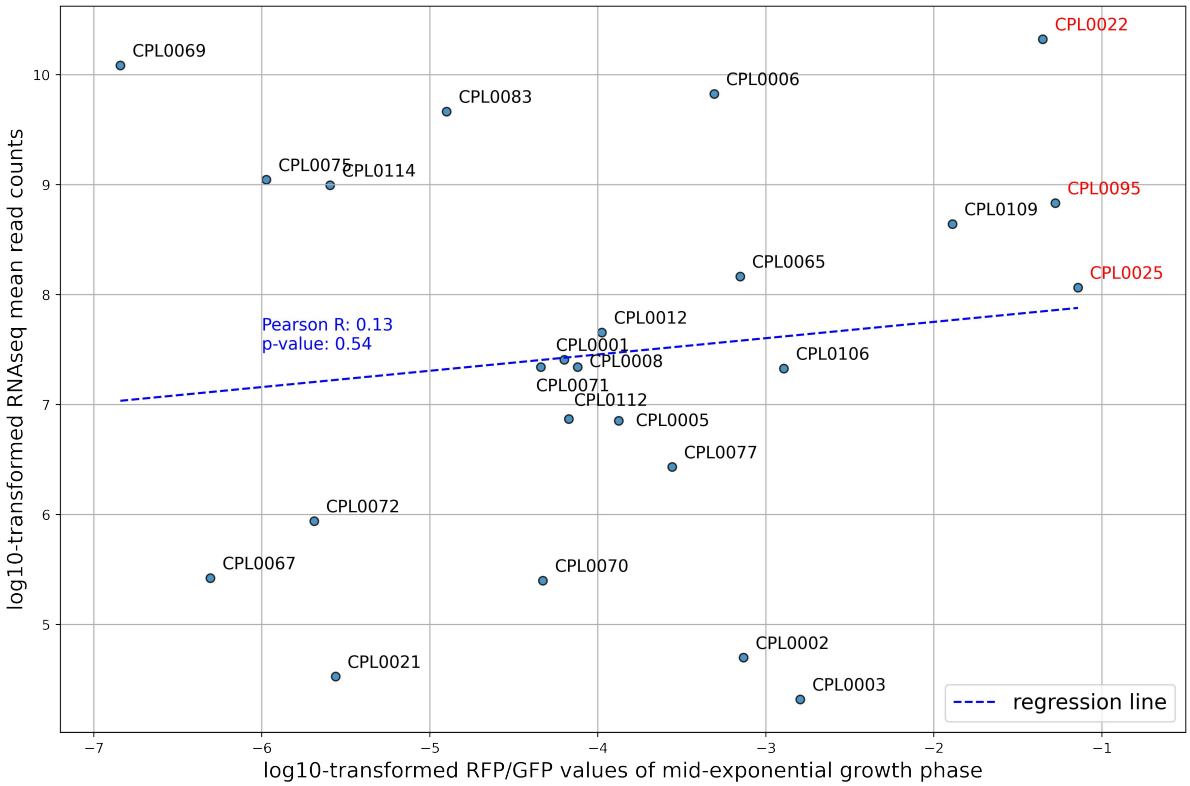


Figure 6: **Comparison of promoter strengths in the exponential phase with their corresponding mean RNAseq read counts for candidate promoters.** The dotted blue line represents linear regression across all data points. Red marked CPL identifiers represent sequences highlighted in Figure 5.

Figure 6 presents a comparison between the measured and normalized RFU values of the promoter sequences and their corresponding mean read counts across *Curvibacter* RNA-seq samples. When comparing our data on the promoter strength of the candidate sequences to the RNA-seq read counts of the same genes, significant disparities in relative strength within these datasets become evident. The Pearson correlation analysis resulted in an R-value of 0.13, indicating only a weak positive correlation between both datasets. Data points close to the regression line represent sequences that show expression levels in our assays matching closely to the read counts gathered from the RNA-seq analysis, relatively to all data points. For data points falling below the separation line, promoters were stronger in the experimental assessment with respect to their mean read counts in the RNA-seq, and vice versa for data points above the regression line.

4 Discussion

4.1 25 novel promoters for the use in *Curvibacter* show distinct temporal expression dynamics

As *Curvibacter* is a promising model organism we set out in project this project to extract novel expression systems for this species from a self designed oligonucleotide library. This library was generated by mining the *Curvibacter* genome for potential promoter sites in an automated fashion. Positive candidates were first picked via Flow Cytometry and subsequently individual sequences were analyzed by bulk fluorescence measurement. From our 500 initial candidate sequences we found 25 positive candidates that showed expression based on our reporter plasmid. Among these, we could find expression levels over two orders of magnitude and a variety of different temporal expression dynamics over growth phases (s. Figure 4a). We found 12 candidate promoters which show a higher expression level in the stationary phase compared to the exponential phase. 10 candidate promoters showed very little discrimination between growth phases, maintaining a stable level of expression throughout the observed duration and three candidate promoters showed a higher expression level in the exponential phase compared to the stationary phase (s. Figure 4b).

Not only can these new expression platforms be used as tools for expression during different growth phases in liquid medium, but the expression strength assay may also indicate the temporal expression dynamics of these genes in their native genomic context: As expected, many of the candidate promoters with higher activity levels in the exponential phase belong to genes expressing proteins involved in central metabolism and proliferation (50S ribosomal protein L25/general stress protein Ctc, RpsL, Ndk (s. Figure 4b)).

This is in accordance with previous findings which show that bacterial cells are able to recall distinct global expression patterns based on their stage of growth by the spatio-temporal regulation of chromosomal macrodomains [32]. While replication induced transient changes in actual copy numbers are a factor directing genomic transcription biases along the oriC/ter axis [33], the regulation of macrodomains occurs for functionally similar genes through direct DNA topology and transcriptional control. While *in trans* expression systems are per definition not affected by positional effects of the promoter of interest (as they are taken out of their natural, genomic context), they are partially affected by DNA topology [34] (e.g. plasmid supercoiling) and fully affected by transcriptional modulation, under the condition that the entire sequence relevant for regulation is included in the expression system. On the other hand, we see a variety of (often hypothetical proteins) gene functions associated with the candidate promoters where the expression levels are higher in the stationary compared to the exponential phase. This is a result of the general expression bias in plasmid-based expression systems, which tend to exhibit higher expression levels in the stationary phase. Consequently, a bias towards stationary phase expression can be observed, complicating the interpretation of the native context of these genes and their temporal expression dynamics. This effect is primarily attributed to the enrichment of plasmid copy numbers in the stationary phase relative to the number of cells [10, 11, 12]. While saturation of protein density was normalized in our assay by utilizing GFP FI values as a normalization factor for mCherry FI values, a bias introduced due to plasmid copy number enrichment is not. We showed that by harvesting $5'$ upstream non-coding regions from the target species genome we were able to create expression systems that behave differently from most synthetic, orthogonal *in trans* expression systems. These expression systems can now be used to further study *Curvibacter* sp. AEP1-3.

The intial library encompassed 500 $5'$UTR sequences from the *Curvibacter* genome. As 25 of these showed detectable expression levels, the discovery rate is therefore at a minimum of five percent (5%). Many potential promoters may not be active under the artificial laboratory environment and hence show little activity, especially considering that R2A is a complex media that already serves a lot of metabolites and thus requires less *de novo* synthesis of many compounds. To eventually raise the success rate of promoter prediction before manually curating the oligonucleotide library, stretches of sequences around the extracted loci could be used as input sequences for a neural network trained by known promoter sequences such as sequences from the PDD [16]. A similar approach was recently conducted by Seo *et al.* for the cyanobacterial species *Synechocystis* sp. PCC 6803 [35]. The AI generated prediction could further be used to extract and construct more efficient oligonucleotide sequences. These sequences can be based not only on a continuous DNA-sequence between gene regions but also on specific k-mers of $5'$UTR sites. Thus motifs responsible for RNA-polymerase recruitment can be located upstream of the sequences ranging into the next gene sequence, which our approach currently does not cover.

4.2 Temporal expression dynamics of highlighted promoters may correspond with their biological functions

For applications where a stable expression level is essential or accumulation of protein aggregates is a known issue, we recommend the use of the CPL0022 promoter, which drives expression of the *dnaK* gene in *Curvibacter* [36]. DnaK in *E. coli* is constitutively expressed throughout all of its life cycle and the same seems to account for the DnaK equivalent in *Curvibacter* (s. Figure 5c). The DnaK protein is a molecular chaperon, a class of enzyme involved in guiding correct folding after translation as well as for already matured proteins. While this maintenance is required constantly, it is generally upregulated when bacteria face external stresses that lead to rapid protein degradation such as

heat shocks. Thus, DnaK in *E. coli* is part of the Hsp70 protein group. In *Curvibacter*, the CPL0022 promoter also showed a relatively stable expression level throughout all growth phases (s. Figure 5c). It would be interesting to see whether this promoter could be utilized as an inducible expression system by applying heat shocks to the cells as a stimulus, effectively acting as an inducible promoter. As *Curvibacter* is studied due to its symbiotic partnership with its host *Hydra vulgaris*, it would be interesting to see whether this promoter also maintains stable activity when growing on the glycocalyx of *Hydra*.

Alternatively, protein aggregation can also be prevented by using a promoter with lower expression levels during the stationary phase such as the CPL0095 promoter. In its native context, this CPL0095 expresses RpsL, a 12S ribosomal protein of the 30S subunit. This 12S subunit is added late in the biogenesis of the 30S subunit and is essential [37]. Due to the high demand for protein expression during the exponential phase, genes involved in protein translation are upregulated during that phase, explaining the unusual temporal expression dynamic of this candidate promoter (s. Figure 5d). While the expression level of CPL0095 during the exponential growth phase is equal to the strongest sequence CPL0025, it is almost five fold weaker during the stationary phase in comparison (compare to Figure 5b). Further modification to reduce or enhance the general expression level of this promoter could fine tune its function for application in continuous cultivation systems such as chemostats. It is likely that other promoters driving gene expression of proteins involved in the assembly of ribosomal subunits or translation are upregulated in a similar fashion and could be a first avenue to find more promoters that behave similarly to CPL0095.

For high levels of expression we recommend the use of CPL0025 (s. Figure 5b) or CPL0017 (s. Figure 5a) which has been used in this study as a reference sequence. CPL0025 had the highest level of expression among all tested sequences. Both promoters are well suited for the expression of proteins with very little burden on the host cell metabolism, such as fluorophore proteins for imaging. The gene expressed from CPL0025 functions as an AHL synthase (Curl2), as described by Pietschke et al. [2]. They have shown that the full promoter region (519 bp) of CPL0025 is activated by AHLs produced by *Curvibacter*, as well as by AHLs modified by *Hydra vulgaris*. Here we show, that the shortened promoter region of 76 bp is able to drive a strong expression of our reporter (s. Figure 5b). Within the identified transcription factor binding motifs, a TF-Motif similar to a motif discovered for the transcriptional regulator LasR of *Pseudomonas aeruginosa* has been identified [23]. LasR is a LuxR-type regulatory protein and a key component in the quorum sensing system of *Pseudomonas aeruginosa*. LasR binds to AHLs activating the expression of genes involved in various virulence factors and genes important for the adaptation to the environment [38]. However, no LasR TF-Motif can be found within the CPL0025 5'UTR but a binding motif for the transcriptional regulator CodY from *Lactococcus lactis* and *Streptococcus pyogenes* as well as for CcpA from *Streptococcus pneumoniae*. Those transcriptional regulators are known to regulate the expression of a wide range of genes, e.g. genes responsible for carbohydrate and (p)ppGpp metabolism or virulence factors [29, 30, 31]. Regarding the expression of the reporter driven by CPL0025 and the previous finding that the promoter region of Curl2 is activated by AHLs, it is possible that the motif responsible for AHL induction is located within the 76 bp long 5'UTR of CPL0025. It is unclear whether the high expression level of CPL0025 in the stationary phase are result of transcriptional changes directly related to the growth phase or a result of increasing levels of AHLs in the media or both.

4.3 RNAseq data alone is insufficient to predict the usability of promoters for expression platforms but can aid in selecting best candidates

For the 25 promoters that showed notable activity in our assay (s. Figure 4a), we compared the expression level to the RNAseq mean read counts (s. Figure 6). While some promoters, especially around the midrange of transcriptional strength in the RNAseq data set, align very well with the experimentally verified expression strength, in many cases both data sets do not seem to match very well, indicating that the RNAseq results could only partially predict the expression levels in our assay. While for some sequences the experimentally assessed expression level was higher than expected given the low read counts (eg. sequences CPL0003, CPL0002, CPL0025 in Figure 6) for other sequences the opposite was the case (eg. sequences CPL0069, CPL0075, CPL0114, CPL0083, CPL0006 Figure 6).

As the growth conditions between the RNAseq experiment and the reporter activity assay deviated (18°C growth in R2A for the RNAseq data set and 30°C in R2a+ media), we can assume that in some cases temperature sensitivity of gene activity may play a part in the overall noise between data sets. However, the activity of the *dnaK* promoter is comparably strong in the RNAseq data set (s. Figure 6) which was taken at 18°C, indicating that heat stress induced signals may not be a factor affecting expression levels too much. We can't exclude transcriptional changes due to heat stress responses or elevated growth rates as a reason for the deviations observed between RNAseq data and our bulk fluorescence measurements. However, other factors may also contribute to the low comparability of the two data sets. It is important to mention that while RNAseq data reports on the relative abundance of cellular mRNA level, it acts as a proxy for transcriptional strength levels of genes. Reporter gene activity however is measured as the amount of active reporter protein; mCherry in our case; and acts therefore as a coupled proxy for transcriptional as well as translational

strength levels. Comparison of such data always has to consider the fact that large discrepancies in translational strength levels of different mRNAs due to different ribosomal binding site affinities may have a huge impact on the abundance of gene activity on mRNA and protein level. Lale *et al.* also reported discrepancies between measured mRNA levels and protein levels of promoters from their GeneEE library [15].

We could find some sequences to have almost no reporter activity despite high read counts in our RNAseq data. While we cant exclude truncation of sequence elements vital for transcription in some cases, a weak translation level may be another reason for the low reporter activity. On the other hand, CPL0025 for instance shows a surprisingly high reporter activity despite moderate read counts in the RNAseq data. Even though unlikely, it is entirely possible that a regulatory site may actually be located upstream of the intergenic region, either as part of the proceeding gene or even further, repressing this promoter in its native context, which has not been integrated into the reporter expression system.

Concluding, we can therefore not claim a strong correlation between RNAseq data and performance of the reporter plasmid. To reinform our pipeline based on these results, we can conclude that while picking candidates based on high RNAseq read counts may improve the chance to obtain reporter constructs of high activity, it is definitely not a guarantee for success. It also shows that simply picking a strong promoter from the genome and trying to utilize it in an *in trans* context is not guaranteed to yield a strong expression system for a new model organism, again highlighting how essential broader high throughput strategies such as the one employed in our case are for tool development for lesser researched species.

5 Outlook

We created a scalable pipeline for the semi-automated discovery of expression systems that can be applied to any bacterial species of interest with an available genome sequence and expression vector. Our workflow was able to find novel expression systems for *Curvibacter* (s. Figure 5) which can now be utilized in a variety of applications. We also showed, that RNAseq results were insufficient to predict a promoter candidates performance in an *in trans* context. It can be extended to specifically search for inducible promoters, another very relevant manipulation tool for novel model species. First, the entire library could be sorted once under "induced" and once under default conditions and sorted in bulk. After sequencing all promoter regions in the expression vector of both populations, the promoters that appear only in the induced population serve as a list of potential candidates for inducible expression systems. In the same way, this method can be used to search for promoters that are active under any condition of interest e.g., for *Curvibacter* in the presence of the host species *Hydra vulgaris*.

In our curated oligonucleotide library approach, we approximate the rate of RNA polymerase activity by assessing the fluorescence of a reporter protein. In contrast, traditional RNA-seq transcriptome analysis involves counting and comparing mapped read abundances of expressed genes. Nevertheless, both experimental designs share the common goal of detecting the rate of transcriptional activity under specific environmental conditions. Comparing these strategies, we argue that both methods show different scalability towards distinct scenarios: while for some applications, it is desired to find transcriptional changes of a few genes under a multitude of different circumstances, in other cases the focus may be on global changes in transcription levels for only a handful of environmental circumstances. Traditional RNAseq workflows allow for a global analysis of a transcriptome but the amount of samples increases linearly with the amount of observed conditions. To maintain an adequate sequencing depth of each individual sample, this leads to scaling sequencing costs based on the amount of samples. Vice versa, our oligonucleotide based screening workflow can be very well adapted to screen many conditions in a single workflow, as cultivation capacities and the scalability of flow cytometry based sorting are the only rate limiting factors. Another advantage is that once positive candidates are sorted and sequenced, the vectors for further studies are already available.

6 Contributions

6.1 Author Contributions

Maurice Mager: Conceptualization, Methodology, Investigation, Writing - Original Draft, Writing - Review and Editing, Visualization

Lukas Becker: Conceptualization, Software, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review and Editing, Visualization

Nina Schulten: Methodology, Resources

Sebastian Fraune: Writing - Review and Editing, Funding Acquisition

Ilka Axmann: Supervision, Writing - Review and Editing, Funding Acquisition

6.2 Acknowledgements

We thank Timo Minten, Petra Kolkhof and Jay Bathia for their support in this project. The implementation of the cell sorter (CytoFLEX SRT) was kindly supported by Dennis Hasenklever.

6.3 Funding

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB1535 - Project ID 458090666, and Major Research Instrumentation INST 208/808-1.

6.4 Conflict of interest statement

The authors declare no conflict of interest.

References

[1] Sebastian Fraune, Friederike Anton-Erxleben, René Augustin, Sören Franzenburg, Mirjam Knop, Katja Schröder, Doris Willoweit-Ohl, and Thomas C G Bosch. Bacteria–bacteria interactions within the microbiota of the ancestral metazoan hydra contribute to fungal resistance. *The ISME Journal*, 9:1543–1556, 7 2015.

[2] Cleo Pietschke, Christian Treitz, Sylvain Forêt, Annika Schultze, Sven Künzel, Andreas Tholey, Thomas C. G. Bosch, and Sebastian Fraune. Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts. *Proceedings of the National Academy of Sciences*, 114(40):E8488–E8497, 2017.

[3] Timo Minten-Lange and Sebastian Fraune. *Hydra and Curvibacter: An intimate crosstalk at the epithelial interface*, pages 79–90. CRC Press, 1 2020.

[4] Jan Taubenheim, Doris Willoweit-Ohl, Mirjam Knop, Sören Franzenburg, Jinru He, Thomas C G Bosch, and Sebastian Fraune. Bacteria-and temperature-regulated peptides modulate β-catenin signaling in hydra.

[5] Emma R. Holden, Gregory J. Wickham, Mark A. Webber, Nicholas M. Thomson, and Eleftheria Trampari. Donor plasmids for phenotypically neutral chromosomal gene insertions in enterobacteriaceae. *Microbiology*, 166(12):1115–1120, 2020.

[6] Tanita Wein, Tal Dagan, Sebastian Fraune, Thomas C. G. Bosch, Thorsten B. H. Reusch, and Nils F. Hülter. Carrying capacity and colonization dynamics of curvibacter in the hydra host habitat. *Frontiers in Microbiology*, 9, 2018.

[7] E Scherzinger, M M Bagdasarian, P Scholz, R Lurz, B Rückert, and M Bagdasarian. Replication of the broad host range plasmid rsf1010: requirement for three plasmid-encoded proteins. *Proceedings of the National Academy of Sciences*, 81(3):654–658, 1984.

[8] Lisa Tietze, Antonia Mangold, Maria W. Hoff, and Rahmi Lale. Identification and cross-characterisation of artificial promoters and 5′ untranslated regions in vibrio natriegens. *Frontiers in Bioengineering and Biotechnology*, 10, 2022.

[9] Ravendran Vasudevan, Grant A.R. Gale, Alejandra A. Schiavon, Anton Puzorjov, John Malin, Michael D. Gillespie, Konstantinos Vavitsas, Valentin Zulkower, Baojun Wang, Christopher J. Howe, David J. Lea-Smith, and Alistair J. McCormick. CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax. *Plant Physiology*, 180(1):39–55, 02 2019.

[10] Nathalie Turgeon, Christian Laflamme, Jim Ho, and Caroline Duchaine. Evaluation of the plasmid copy number in b. cereus spores, during germination, bacterial growth and sporulation using real-time pcr. *Plasmid*, 60(2):118–124, 2008.

[11] Naoki Akasaka, Wiwik Astuti, Yuri Ishii, Ryota Hidese, Hisao Sakoda, and Shinsuke Fujiwara. Change in the plasmid copy number in acetic acid bacteria in response to growth phase and acetic acid concentration. *Journal of Bioscience and Bioengineering*, 119(6):661–668, 2015.

[12] Bertram M. Berla and Himadri B. Pakrasi. Upregulation of plasmid genes during stationary phase in synechocystis sp. strain pcc 6803, a cyanobacterium. *Applied and Environmental Microbiology*, 78(15):5448–5451, 2012.

[13] Michele A Cleary, Kristopher Kilian, Yanqun Wang, Jeff Bradshaw, Guy Cavet, Wei Ge, Amit Kulkarni, Patrick J Paddison, Kenneth Chang, Nihar Sheth, Eric Leproust, Ernest M Coffey, Julja Burchard, W Richard McCombie, Peter Linsley, and Gregory J Hannon. Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis. *Nature Methods*, 1:241–248, 12 2004.

[14] Arnold R Oliphant, Alexander L. Nussbaum, and Kevin Struhl. Cloning of random-sequence oligodeoxynucleotides. *Gene*, 44(2):177–183, 1986.

[15] Rahmi Lale, Lisa Tietze, Maxime Fages-Lartaud, Jenny Nesje, Ingerid Onsager, Kerstin Engelhardt, Che Fai Alex Wong, Madina Akan, Niklas Hummel, Jörn Kalinowski, Christian Rückert, and Martin Frank Hohmann-Marriott. A universal approach to gene expression engineering. *Synthetic Biology*, 7(1):ysac017, 08 2022.

[16] Wei Su, Meng-Lu Liu, Yu-He Yang, Jia-Shu Wang, Shi-Hao Li, Hao Lv, Fu-Ying Dao, Hui Yang, and Hao Lin. Ppd: A manually curated database for experimentally verified prokaryotic promoters. *Journal of Molecular Biology*, 433(11):166860, 2021. Computation Resources for Molecular Biology.

[17] Ye Wang, Haochen Wang, Lei Wei, Shuailin Li, Liyang Liu, and Xiaowo Wang. Synthetic promoter design in Escherichia coli based on a deep generative network. *Nucleic Acids Research*, 48(12):6403–6412, 05 2020.

[18] Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics*, 30:2114–2120, 2014.

[19] Simon Andrews. Fastqc: a quality control tool for high throughput sequence data. available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. 2010.

[20] Nicolas L Bray, Harold Pimentel, Páll Melsted, and Lior Pachter. Near-optimal probabilistic rna-seq quantification. *Nature Biotechnology*, 34:525–527, 5 2016.

[21] Abraham. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

[22] Timothy L Bailey. STREME: accurate and versatile sequence motif discovery. *Bioinformatics*, 37(18):2834–2840, 03 2021.

[23] Gerardo Croda-García, Victoria Grosso-Becerra, Abigail Gonzalez-Valdez, Luis Servín-González, and Gloria Soberón-Chávez. Transcriptional regulation of pseudomonas aeruginosa rhlr: role of the crp orthologue vfr (virulence factor regulator) and quorum-sensing regulators lasr and rhlr. *Microbiology*, 157(9):2545–2555, 2011.

[24] Christopher J. Jones, Cynthia R. Ryder, Ethan E. Mann, and Daniel J. Wozniak. Amrz modulates pseudomonas aeruginosa biofilm architecture by directly repressing transcription of the psl operon. *Journal of Bacteriology*, 195(8):1637–1644, 2013.

[25] M J Gambello and B H Iglewski. Cloning and characterization of the pseudomonas aeruginosa lasr gene, a transcriptional activator of elastase expression. *Journal of Bacteriology*, 173:3000–3009, 5 1991.

[26] Pattarachai Kiratisin, Kenneth D. Tucker, and Luciano Passador. Lasr, a transcriptional activator of pseudomonas aeruginosa virulence genes, functions as a multimer. *Journal of Bacteriology*, 184(17):4912–4919, 2002.

[27] Michael G. Surette, Melissa B. Miller, and Bonnie L. Bassler. Quorum sensing in escherichia coli, salmonella typhimurium, and vibrio harveyi: A new family of genes responsible for autoinducer production. *Proceedings of the National Academy of Sciences*, 96(4):1639–1644, 1999.

[28] Balasubramanian Ganesan and Bart C. Weimer. Chapter 19 - amino acid catabolism and its relationship to cheese flavor outcomes. In Paul L.H. McSweeney, Patrick F. Fox, Paul D. Cotter, and David W. Everett, editors, *Cheese (Fourth Edition)*, pages 483–516. Academic Press, San Diego, fourth edition edition, 2017.

[29] Angelica Pellegrini, Germana Lentini, Agata Famà, Andrea Bonacorsi, Viola Camilla Scoffone, Silvia Buroni, Gabriele Trespidi, Umberto Postiglione, Davide Sassera, Federico Manai, Giampiero Pietrocola, Arnaud Firon, Carmelo Biondo, Giuseppe Teti, Concetta Beninati, and Giulia Barbieri. Cody is a global transcriptional regulator required for virulence in group b streptococcus. *Frontiers in Microbiology*, 13, 2022.

[30] Jacqueline Abranches, Marcelle M. Nascimento, Lin Zeng, Christopher M. Browngardt, Zezhang T. Wen, Mercedes F. Rivera, and Robert A. Burne. Ccpa regulates central metabolism and virulence gene expression in streptococcus mutans. *Journal of Bacteriology*, 190(7):2340–2349, 2008.

[31] José A. Lemos, Marcelle M. Nascimento, Vanessa K. Lin, Jacqueline Abranches, and Robert A. Burne. Global regulation by (p)ppgpp and cody in streptococcus mutans. *Journal of Bacteriology*, 190(15):5291–5299, 2008.

[32] Patrick Sobetzko, Andrew Travers, and Georgi Muskhelishvili. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proceedings of the National Academy of Sciences*, 109(2):E42–E50, 2012.

[33] Marc Teufel, Werner Henkel, and Patrick Sobetzko. The role of replication-induced chromosomal copy numbers in spatio-temporal gene regulation and evolutionary chromosome plasticity. *Frontiers in Microbiology*, 14, 2023.

[34] Carlo A. Klein, Marc Teufel, Carl J. Weile, and Patrick Sobetzko. The bacterial promoter spacer modulates promoter strength and timing by length, tg-motifs and dna supercoiling sensitivity. *Scientific Reports*, 11:24399, 12 2021.

[35] Euijin Seo, Yun-Nam Choi, Ye Rim Shin, Donghyuk Kim, and Jeong Wook Lee. Design of synthetic promoters for cyanobacteria with generative deep-learning model. *Nucleic Acids Research*, 51(13):7071–7082, 05 2023.

[36] Florence Arsène, Toshifumi Tomoyasu, and Bernd Bukau. The heat shock response of escherichia coli. *International Journal of Food Microbiology*, 55(1):3–9, 2000.

[37] Anke M. Mulder, Craig Yoshioka, Andrea H. Beck, Anne E. Bunner, Ronald A. Milligan, Clinton S. Potter, Bridget Carragher, and James R. Williamson. Visualizing ribosome biogenesis: Parallel assembly pathways for the 30s subunit. *Science*, 330(6004):673–677, 2010.

[38] Kathirvel Brindhadevi, Felix LewisOscar, Eleftherios Mylonakis, Sabarathinam Shanmugam, Tikendra Nath Verma, and Arivalagan Pugazhendhi. Biofilm and quorum sensing mediated pathogenicity in pseudomonas aeruginosa. *Process Biochemistry*, 96:49–57, 2020.

ACS SyntheticBiology

pubs.acs.org/synthbio

Research Article

Interlaboratory Reproducibility in Growth and Reporter Expression in the Cyanobacterium *Synechocystis* sp. PCC 6803

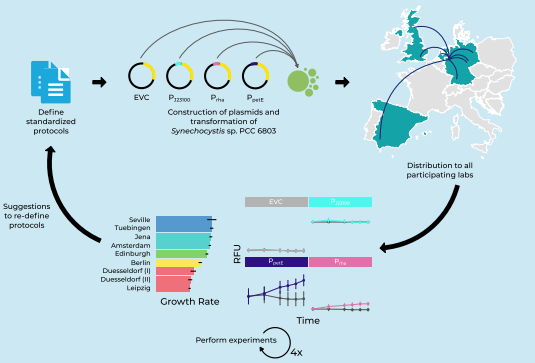
Maurice Mager,[¶] Hugo Pineda Hernandez,[¶] Fabian Brandenburg, Luis López-Maury, Alistair J. McCormick, Dennis J. Nürnberg, Tim Orthwein, David A. Russo, Angelo Joshua Victoria, Xiaoran Wang, Julie A. Z. Zedler, Filipe Branco dos Santos,* and Nicolas M. Schmelling*

Cite This: *ACS Synth. Biol.* 2023, 12, 1823–1835

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In recent years, a plethora of new synthetic biology tools for use in cyanobacteria have been published; however, their reported characterizations often cannot be reproduced, greatly limiting the comparability of results and hindering their applicability. In this interlaboratory study, the reproducibility of a standard microbiological experiment for the cyanobacterial model organism *Synechocystis* sp. PCC 6803 was assessed. Participants from eight different laboratories quantified the fluorescence intensity of mVENUS as a proxy for the transcription activity of the three promoters P_{J23100}, P_{rhaBAD}, and P_{petE} over time. In addition, growth rates were measured to compare growth conditions between laboratories. By establishing strict and standardized laboratory protocols, reflecting frequently reported methods, we aimed to identify issues with state-of-the-art procedures and assess their effect on reproducibility. Significant differences in spectrophotometer measurements across laboratories from identical samples were found, suggesting that commonly used reporting practices of optical density values need to be supplemented by cell count or biomass measurements. Further, despite standardized light intensity in the incubators, significantly different growth rates between incubators used in this study were observed, highlighting the need for additional reporting requirements of growth conditions for phototrophic organisms beyond the light intensity and CO_2 supply. Despite the use of a regulatory system orthogonal to *Synechocystis* sp. PCC 6803, P_{rhaBAD}, and a high level of protocol standardization, ∼32% variation in promoter activity under induced conditions was found across laboratories, suggesting that the reproducibility of other data in the field of cyanobacteria might be affected similarly.



KEYWORDS: *promoter, cyanobacteria, reproducibility, interlab*

INTRODUCTION

As oxygenic phototrophs, cyanobacteria can potentially mitigate climate change when utilized by a carbon-neutral biotechnological industry. However, robust molecular biology tools are needed to study their physiology, enable their manipulation, and tap into their potential as green biotechnology platforms. Over the past decades, a plethora of new molecular tools like molecular manipulation tools and CRISPR/Cas systems have been developed for various model species.[1−3] Those tools have been adapted and further extended for cyanobacteria at an unprecedented rate in recent years. RSF1010-based replicative plasmid for various cyanobacterial species[4] and chromosomal integration for introducing transgenes have been used for decades to introduce heterologous genes into cyanobacteria. Many constitutive and inducible promoters as well as novel terminators have been characterized for transgenic expression in cyanobacteria.[5−7] Inducible promoters have been successfully implemented in more sophisticated applications such as CRISPR and CRISPRi systems as well as for metabolic engineering.[3,8] However, to ensure a seamless transfer of inducible promoter tools to other investigators, their performance needs to be optimally characterized under different experimental conditions to give users a sense of their robustness and limitations.

Usually, those promoters are described and tested in cyanobacterial model strains like *Synechocystis* sp. PCC 6803 (hereafter PCC 6803) or *Synechococcus elongatus* PCC 7942. However, Cyanobacteria are a diverse phylum, represented by multiple model strains, which show differences in their optimal cultivation conditions, their ability for genetic manipulation, and the behavior of genetic parts. While *lacI*-based inducible expression systems have been commonly used in *Synechococ-*

Received: March 11, 2023
Published: May 29, 2023

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

cus[9] and *Anabaena*,[10] attempts to design similar systems for PCC 6803 showed very little tunability and high leaky expression.[11] Further, different investigators have reported extreme variability in promoter activity among PCC 6803 substrains. Reported fold changes for those inducible promoters range from 3- to 32-fold for the P_{petE} promoter and 30- to 55-fold for variations on the P_{rhaBAD} promoter in PCC 6803[3,6,12] (see Table S1). For the P_{petE} promoter, varying basal activity levels have been reported[13−15] (see Table S2). Due to the lack of a standardized measurement unit for promoter activity, comparisons among different reports in the literature are nearly impossible. Thus, despite the efforts to create robust genetic tools for cyanobacteria, the implementation of inducible promoters often requires readaptation of the promoter sequence and sometimes also the protocols, even when used in the same cyanobacterial model strain under seemingly identical cultivation conditions.

Reproducibility of research results has been noted as a problem in various fields and dubbed as the Reproducibility Crisis, including improper documentation of equipment and detailed methods.[16] This problem is much more extensive for cyanobacterial research as culture conditions for phototrophs are very complex. Variations in light intensity and quality, as well as CO_2 availability in different incubators, are additional factors that influence growth rates and cellular metabolism. Furthermore, it has been shown that optical density (OD), a standard reporting unit for the growth of bacteria, is not comparable across devices, strains, and, thus, laboratories.[17] This fact is especially an issue when using dose-responsive inducible promoters, as the ratio of inducer molecules to the number of cells will determine the actual induction rather than a defined inducer concentration.[6] However, OD is still the most dominant form used to report bacterial growth and define the starting point for experiments.

Further, the respective mRNA sequence can influence the observed promoter activity,[18] as well as the type of assay used to measure promoter activity, such as fluorescence or luminescence-based reporters, protein quantification, or Northern blot. All these factors lead to different reports on promoter strengths and fold changes of inducible promoters.

Interlaboratory comparisons (ILCs) are, according to the European Commission Science Hub, organized to either "check the ability of laboratories to deliver accurate testing results" or to "find out whether a certain analytical method performs well and is fit for its intended purposes".[19] Within an ILC, the same analytical method is performed by multiple laboratories on the same samples. Subsequently, the results of each independent laboratory are compared in terms of conformity and deviation. As such, ILCs are used to assess each laboratory's accuracy and the method's accuracy in general. The term "interlab study" was recently popularized within the synthetic biology community by the international Genetically Engineered Machines (iGEM) competition as an ILC study to assess the reproducibility of properties from single genetic elements inside a defined context.[20−23] Such studies are a valuable tool to investigate the robustness of methods and expression platforms and identify possible sources of variation. However, those studies are seldom reported and, to our knowledge, have not been performed with cyanobacteria to test the reproducibility of widely used techniques such as promoter activity or growth.

We, therefore, set up an interlab study in eight different laboratories to investigate the reproducibility of a relatively simple growth and promoter activity quantification experiment in the cyanobacterial model strain PCC 6803. We designed a protocol reflecting state-of-the-art methods reported in the literature and performed the experiment four times in each laboratory. Three commonly used promoters for heterologous expression in PCC 6803, P_{J23100}, P_{petE}, and P_{rhaBAD} were chosen, and their transcriptional activity was quantified using the fluorescent reporter protein mVENUS as a proxy. While actual transcript levels and to a lesser extent protein levels may fluctuate in response to minute changes in environmental conditions, we chose mVENUS expression as a simplified form to approximate transcript and resulting protein levels, which we further refer to as "mVENUS expression". Experimental conditions were standardized, leaving equipment and investigators as the primary sources of variability. OD measurements were highly reproducible within replications in a single laboratoy. However, we noticed that dilutions of the initially identical cultures to the starting OD resulted in vastly different cell biomass concentrations across laboratories. This effect results from taking OD values as a proxy for cell concentration instead of cell count. In the following, we observed significant differences in growth rate across laboratories that could not be found to correlate with the initial difference in cell biomass concentration. Even when expression of the promoters of interest was induced with concentrations above saturation level, high variability of promoter strengths was observed across different laboratories. With this information, we aim to formulate best practices for reporting these parameters to ensure better reproducibility and robustness of research results in the future.

RESULTS AND DISCUSSION

The aim of this interlab study was to assess the reproducibility of routinely performed microbiological experiments for the cyanobacterial model organism PCC 6803. The experiment of choice was a time series of the transcription strength of three promoters in PCC 6803 using the fluorescence reporter mVENUS, representing common procedures used frequently in molecular biology laboratories. Those experiments include the growth of PCC 6803 under "standard" cultivation conditions and fluorescence measurements in a plate reader and should not require overly sophisticated equipment. Furthermore, the flask type as well as the flask cap, light intensity, and shaking speed were defined to reduce the number of confounding factors. Thus, we performed the same predefined experiments (see Methods) under conditions as identical as possible across all eight independent laboratories routinely working with cyanobacteria. This study is hence aimed to give a broad, unbiased picture of the current state of reproducibility of some of the published methods in cyanobacterial research.

Selection and Design of Genetic Parts and Constructs. Two widely used inducible promoters have been chosen as representative candidates. The rhamnose-dependent *rhaBAD* system is based on the *E. coli* native *rhaBAD* operon regulated by the AraC-like positive transcription regulator RhaS. Upon addition of rhamnose, RhaS dimers bind to the RhaS regulon and recruit RNA polymerases by interaction with *E. coli* sigma 70 factor RpoD.[24] Based on sequence similarity, it has been hypothesized that in PCC 6803, RhaS recruits the main sigma factor SigA instead. This is further supported by conserved sites between RpoD and SigA, which were previously described as relevant for RhaS-RpoD interaction.[13]

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

RhaS-based regulation of the *rhaBAD* promoter was utilized in PCC 6803 to reach a 55× induction when RhaS was expressed from a strong constitutive promoter J23119.[6]

The copper-dependent *petE* system is based on the PCC 6803 native promoter of the *petE* gene, which is responsible for plastocyanin expression. The *petE* promoter is regulated by an interplay of PetR and PetP: PetR represses transcription of *petE* by binding to the *petE* promoter, while PetP is a protease that degrades PetR in the presence of copper.[14] Englund et al. observed a 5× induction of the *petE* promoter in PCC 6803 in the presence of copper[15] using the same experimental approach as the one used in this manuscript.

The J23100 promoter is a constitutive promoter from the Anderson promoter collection.[25] This collection is a small combinatorial library of J23119 derivatives and covers multiple orders of magnitude in transcription strength in *E. coli*. The entire collection has been characterized in PCC 6803 by Vasudevan et al.[5]

Each individual promoter was cloned upstream of the bicistronic design (BCD2) ribosomal binding site (RBS).[25] This module consists of a ribosomal binding site followed by a short reading frame, a stop codon, and a secondary ribosomal binding site. Through translational coupling, this sequence has been reported to minimize coding sequence-based bias to translation activity. Additionally, the insulation effect of the BCD2 ribosomal binding site reduces background activity from upstream genes, improving the fold changes of inducible promoter systems. We chose this RBS to reduce deviation through genetic design further. In addition, each individual promoter was cloned upstream of the *mVENUS* coding sequence on the RSF1010 origin of replication, a commonly used shuttle vector for PCC 6803.

As an empty vector control (hereafter: EVC), the RSF1010 background only harboring the chloramphenicol resistance gene was used. A single designated laboratory was chosen to clone the desired constructs. Further, a defined PCC 6803 background strain was selected to mitigate any effects that different PCC 6803 background strains might have on the results. The designated laboratory transformed the PCC 6803 background strain and later shipped cryopreserved stock cultures on dry ice to each participating laboratory to ensure that each laboratory had genetically identical strains for the experiments.

Experimental Setup of the Interlab Study. We created a set of detailed protocols to standardize the experimental conditions and data collection across all participants. The protocols contained the required information to handle the (frozen) stocks of the strains, prepare the growth medium, set up the incubator conditions, handle the cultures during the experiment, and perform the measurements.

In short, glycerol stocks of the four strains, EVC, P_{J23100}, P_{petE}, and P_{rhaBAD} (Table 1), were inoculated in liquid culture and grown for 36 to 48 h. The day before the assay, each preculture was diluted to OD_{730} of 0.3 to ensure cells grew exponentially at the onset of the assay. On the next day, each culture was divided into two flasks (induced and uninduced), the OD_{730} was adjusted to 0.5 when necessary, and inducers were added accordingly. From this moment, samples were taken during the first seven hours and after 24 h (see Figure S1 for a graphical representation of the protocol). In these samples, growth was recorded by measuring OD_{730} in a benchtop spectrophotometer and promoter activity by measuring fluorescence and OD_{730} in a plate reader. Additionally, full absorption spectra were measured as a control for the vitality of the cultures and as an internal quality control. However, these data were not included in this analysis because no notable changes in absorption spectra were observed. The raw data can be found at 10.6084/m9.figshare.21525747.v5.

Table 1. Strains Used in This Study

strain name	fluorescent reporter	reporter promoter	inducer
EVC	–	–	–
P_{J23100}	mVENUS	P_{J23100}	constitutive
P_{petE}	mVENUS	P_{petE}	$CuSO_4$
P_{rhaBAD}	mVENUS	P_{rhaBAD}	rhamnose

OD_{730} Measurements Are Highly Reproducible. The measurements performed by all participating laboratories, spectrophotometer OD_{730} and the normalized relative fluorescence units (nRFU) were used as proxies for biomass concentration and mVENUS expression, respectively (see the section Fluorescence Analysis and Normalization in Methods for a detailed explanation of our normalization strategy). We estimated the coefficient of variation (CV) to assess the reproducibility of our protocols by calculating the ratio of the standard deviation to the mean for each time point, strain, and induction regime in the first seven hours of the assay, either for each participant (intralab) or over all locations (interlab) (Figure 1). The CV, inversely proportional to the precision of replicate measurements, was lower at both the intra- and interlab level for the spectrophotometer OD_{730}. Measurements from this data set presented a median CV of 6.1% and 11.5% (Figure 1), with 95.4% and 100% of measurements with a CV lower than 20% for intra- and interlab, respectively (Figure S2). On the other hand, the nRFU showed a median CV of 23.9% and 60.6% at the intra- and interlab levels, respectively (Figure 1). In this data set, 43% of intralab replicates had a CV lower than 20%, but at the interlab level, all replicates showed a CV higher than 20% (Figure S2). Even though the nRFU values were less reproducible than the spectrophotometer measurements, this normalization strategy clearly improved the comparability of results across laboratories. This can be seen from the much higher interlab CV values calculated from not normalized values as the background-corrected fluorescence units (FU_{bc}) or the relative fluorescence units (RFU) (Figure S3). Interestingly, including the OD_{730} in the normalization procedure did not lead to lower CV at the interlab level (Figure S3). This analysis shows that the measurements performed in the spectrophotometer were the most reproducible in our protocol, which can be partially explained by the fact that the starting conditions of the assay were based on measurements from these devices.

Discrepancies in Growth Rates Are Not Explained by Different Initial Biomass Concentrations. Even though the spectrophotometer results showed the highest reproducibility, we further evaluated if this resulted from highly reproducible culture conditions. To assess this, we used the spectrophotometer OD_{730} to estimate growth rates for each biological replicate over 24 h (Figure 2). This metric is not dependent on the actual OD_{730} values per se but rather on their relationship. The growth rate is known to be influenced by environmental conditions. We performed ANOVA to test if significant differences were found at the strain, induction regime, or laboratory level. The latter was the only factor

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

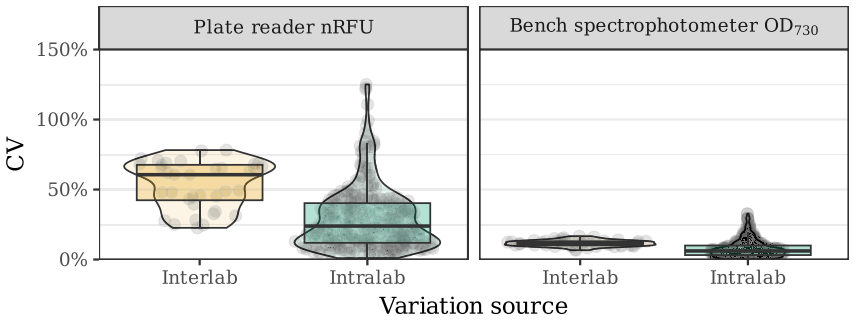


Figure 1. Coefficient of variation (%) of nRFU in EVC, P_{rhaBAD}, and P_{petE} (left) and spectrophotometer OD_{730} (right) data sets. The coefficient of variation was calculated for either all the replicates within a laboratory (intralab, in green) or for all the replicates across all laboratories (interlab, orange). In each panel, a boxplot summarizes all data points by showing the median as a horizontal line, the 25th and 75th percentiles as the bottom and top of the box, respectively, and whiskers extending 1.5 × IQR from the box margins. In addition, all data points are shown and distributed over the area of a violin plot.

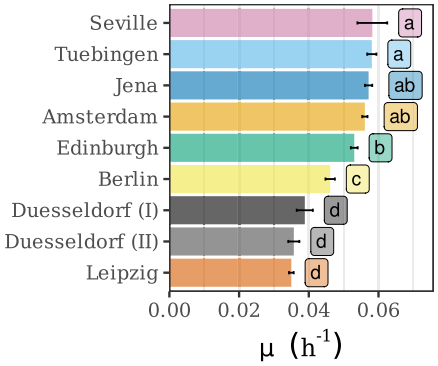


Figure 2. Measured growth rates in each laboratory. Bars represent the mean growth rate and error bars 95% CI (n = 32 or 24). Text boxes on the right side of the bars show the results from Tukey's test. Laboratories with significantly different growth rates (P-value < 0.05) are labeled with different letters.

showing a significant influence (P-value < 0.0001) on the measured growth rates.

Therefore, we compared the average growth rate between laboratories using a post hoc Tukey's test. The results of this analysis showed significant differences (P-value < 0.05) between most of the participants. However, the most striking and biologically relevant differences were found between the laboratories, with the highest (Seville, Tuebingen, Jena, and Amsterdam) and lowest growth rates (Duesseldorf and Leipzig). In the first group, growth rates varied between 0.056 and 0.058 h^{-1}, while in the second, it ranged between 0.035 and 0.039 h^{-1}. This translates into a 36% reduction in the average growth rate.

Our protocols precisely defined the growth conditions (temperature, shaking speed, light intensity) and media composition. However, since we used the spectrophotometer OD_{730} as a reference value to set the initial biomass concentration in the assay, the starting amount of photons per cell could have differed among laboratories. For example, this could have been the case if spectrophotometers used across laboratories had different relationships between OD_{730} and the number of cells.

To test this hypothesis, we performed additional measurements in each spectrophotometer by preparing a dilution of the four glycerol stocks and directly measuring it without allowing cells to grow. The results (Figure S4A) showed that, indeed, there were significant differences (ANOVA, P-value < 0.05) across spectrophotometers when measuring the same amount of cells. With this information, we could determine if the differences in growth rates we observed during the assays resulted from the initial biomass concentration. Therefore, we normalized the initial OD_{730} values of the assays by the measurements obtained from the glycerol stocks (Figure S4B). Next, we tested whether growth rates correlated with this relative OD_{730} but found no significant correlation (Pearson's coefficient: 0.089, P-value: 0.834, Figure S4C). Thus, we can conclude that even if the starting amount of photons per cell was not the same across laboratories, this could not explain the observed differences in growth rates.

P_{petE} Expression Across Laboratories Is Less Reproducible than P_{rhaBAD} When No Inducer Is Added. The uninduced P_{J23100} ($P_{J23100-}$, see the Promoter Quantification Assay section) RFU was used as an internal standard to compare the expression of mVENUS across laboratories (Figure 3). As expected, we observed an average increase in fluorescence during the first seven hours after induction in P_{petE} and P_{rhaBAD} strains. However, these promoters' expression patterns differed in reproducibility across laboratories, leakiness, or magnitude of induction.

The reproducibility of the nRFU within each laboratory was comparable for both strains and induction regimes (Figure 4). The median CV of the cultures without induction was 19.3% and 29.3% for P_{rhaBAD} and P_{petE}, respectively. Interestingly, the reproducibility across laboratories was higher when an inducer was added (median CV of 16.4% and 17.1% for P_{rhaBAD} and P_{petE}, respectively).

Since we used the $P_{J23100-}$ RFU to normalize the RFU values of the other strains, it was impossible to use the nRFU CV to compare the reproducibility of fluorescence measurements across all strains. Thus, we additionally calculated the CV of the not normalized RFU at the intralab level (Figure S5). P_{J23100} measurements were the most reproducible among all strains and induction regimes. When no inducer was added, EVC, P_{petE}, and P_{rhaBAD} showed similar reproducibility. However, when the inducer was added, the P_{petE} and P_{rhaBAD} measurements were more reproducible than the EVC RFU.

As expected, the observed variation of nRFU at the interlab level was higher than within each laboratory (Figure 4). When cultures were induced, increased reproducibility was observed compared to uninduced. A pattern that was already observed on the intralab level comparison. However, while CV values were comparable between both induced strains (median CV of 32% and 31.2% for P_{rhaBAD} and P_{petE}, respectively), the

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

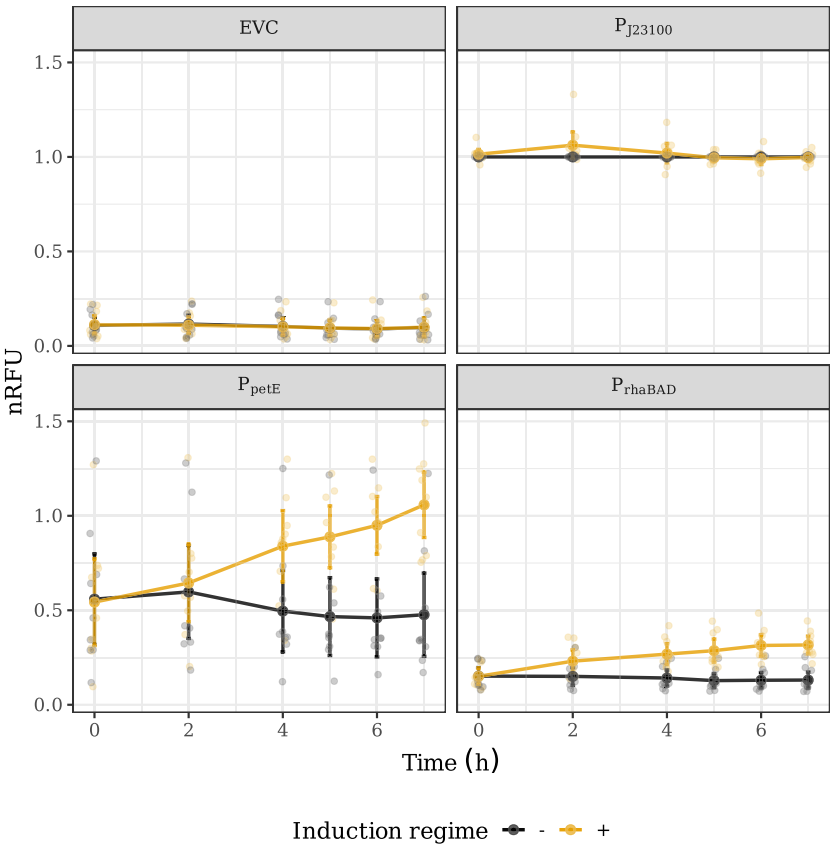


Figure 3. Time series of promoter assay. The nRFU over time is shown for the four strains, each depicted in an individual panel. Cultures, where an inducer was added, are shown in yellow, and those without are shown in black. Smaller data points represent the average values for a single laboratory (n = 4 or 3). Larger data points show the overall average and error bars of the overall 95% CI (n = 7). The induced and uninduced conditions for EVC and P_{J23100} refer to specific aspects of the assay preparation. See the Promoter Quantification Assay section in Methods for further details.

difference between induced and uninduced was much more pronounced. Further, the interlab variability observed in the uninduced P_{petE} cultures was much higher than for the uninduced P_{rhaBAD} cultures. We observed nonoverlapping distribution of CV values between the two strains, where P_{petE}'s median CV value was 67% compared to the 46.8% P_{rhaBAD}'s median CV.

Inducible promoters can be characterized by measuring a change in expression strength between induced "ON" and uninduced "OFF" states. We, therefore, investigated how these parameters varied between the P_{rhaBAD} and P_{petE} strains and how they varied between laboratories. We calculated the changes in transcription activity upon induction by dividing the nRFU at seven hours by the value at the beginning of the assay for both induced and uninduced cultures (Figure 5A). When the inducer was absent, the transcriptional activity did not change over the experiment as expected. Changes for all strains in the absence of an inducer were close to 1, with a mean change of 0.86, 95% CI [0.77, 0.95] for P_{rhaBAD} and 0.93, 95% CI [0.73, 1.13] for P_{petE}. When an inducer was added, P_{petE} cultures showed the largest average change (2.87, 95% CI [1.46, 4.28]) while also showing high variability between laboratories (Figure 5A). On the other hand, P_{rhaBAD} cultures showed an average change of 2.35, 95% CI [1.84, 3.76] with less variability across laboratories.

Regarding expression strength, we observed that after seven hours of induction and with the addition of 1 μM of $CuSO_4$, the P_{petE} promoter led to similar mVENUS expression levels as the P_{J23100} promoter (average 106% of P_{J23100}, 95% CI [89%, 123%]) (Figure 5B). Under the same conditions and with 10 mM rhamnose, the average expression of the P_{rhaBAD} was 32% of P_{J23100} (95% CI [27%, 37%]).

The leakiness of an inducible promoter refers to its basal expression level in the absence of the inducer. To assess the P_{petE} and P_{rhaBAD} promoters' leakiness, we calculated the ratio of RFU OD_{730}^{-1} at seven hours in the uninduced cultures to the EVC (Figure 5C). This ratio is equal to 1 if no signal from the fluorescent reporter is measured in the uninduced condition and will increase proportionally to the leakiness of the promoter. The uninduced P_{rhaBAD} cultures showed an average ratio of 1.6, 95% CI [1.3, 1.9]. Conversely, the mean ratio in the P_{petE} cultures was 8.4, 95% CI [0.9, 15.9] when no inducer was added, suggesting that either the promoter was very leaky or the environment contained too many residual copper ions (Figure 5C). The high variability observed across the different laboratories (Figure 5D), with an average ratio within laboratories ranging from 1.2 to 38.5 and a median value of 4.8, suggests that residual copper present in the medium of some of the participating laboratories is responsible.

Across eight different laboratories, the growth of PCC 6803 was quantified via OD_{730} measurements, and the transcriptional activity of three promoters (P_{J23100}, P_{rhaBAD}, and P_{petE}) was quantified using fluorescence intensity from mVENUS as a

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

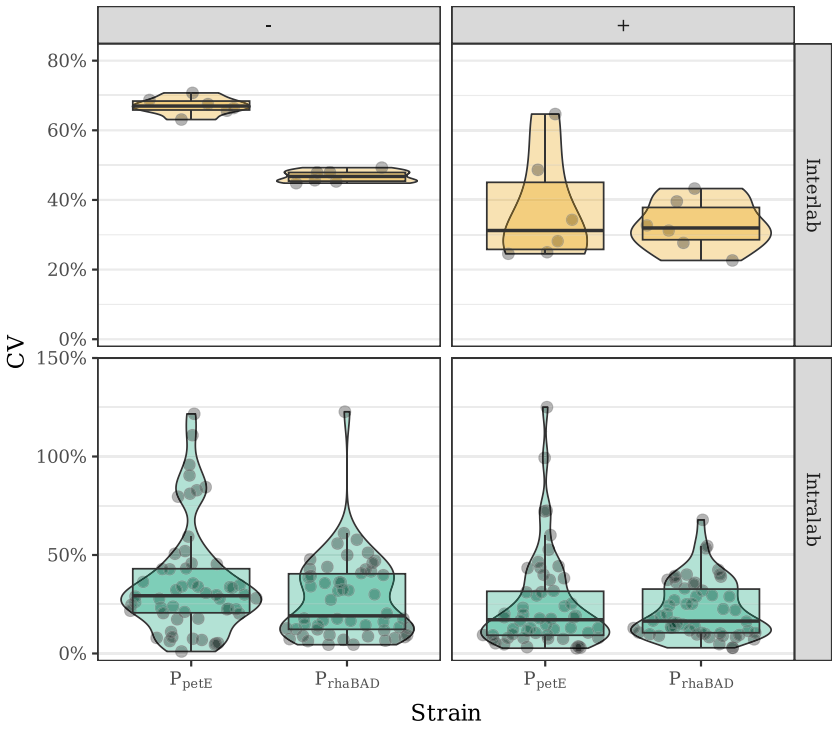


Figure 4. Coefficient of variation (%) of the nRFU measured for all time points in P_{petE} and P_{rhaBAD} cultures without induction (−, left column) and with induction (+, right column) at the interlab (top row) and intralab level (bottom row). The coefficient of variation was calculated for either all the replicates within a laboratory (intralab, in green) or for all the replicates across all laboratories (interlab, orange). In each panel, a boxplot summarizes all data points by showing the median as a horizontal line, the 25th and 75th percentiles as the bottom and top of the box, respectively, and whiskers extending 1.5 × IQR from the box margins. In addition, all data points are shown and distributed over the area of a violin plot. Notice that the *Y*-axis range is different in the top and bottom rows.

proxy. Reproducibility was evaluated by using inter- and intralaboratory coefficients of variation. We discovered that OD_{730} values showed high reproducibility, albeit the absolute OD_{730} values were not normalized between different laboratories. The fluorescence intensity values of reporters were less reproducible than OD_{730}. Generally, fluorescence intensity values of induced promoters were more reproducible than uninduced promoters.

Reproducibility Is Not Homogeneous Across the Different Measurement Methods. The spectrophotometer measurements in the growth assay showed the highest reproducibility (Figure 1). However, this was expected since the initial OD_{730} of the assay was determined from this device. Nevertheless, the growth rates estimated from these measurements showed significant differences across the participating laboratories (Figure 2). It is well-known that OD measurements are only comparable across devices with additional calibrations.[17] However, as this study is intended to investigate experimental variation according to commonly used experimental practices, we were interested in the effect of this variation on the experimental results. Thus, we chose not to calibrate the starting OD_{730} by cell counts across laboratories. Instead, calibration measurements were initially performed to later infer the influence of differences in OD_{730} measurements on the overall reproducibility of the study. In these calibration measurements, the reported OD_{730} varied more than 5-fold between participants, confirming the high variability of bulk OD_{730} measurements between spectrophotometer devices. Although this shows that internally, OD_{730} measurements were very reproducible (Figure 1), they were practically incomparable across laboratories without normalization. Therefore, the starting OD_{730} target at 0 h does not represent the same biomass concentration across different laboratories.

Initially, we expected to find a correlation between the growth rate and the normalized initial biomass concentration. However, our data did not reveal such a correlation (Figure S4C), and it remains a matter of speculation which factor caused the differences in the growth rates. It is possible that these differences arose from variations in the light spectra since these are known to affect this physiological parameter.[26,27] In our protocol, the light color was specified as "white", which is how it is commonly defined in scientific literature. White light, however, is a combination of different wavelengths, and light spectra across laboratories could have been different. Thus, we compared the light spectra across laboratories and incubators. However, most of the laboratories did not have the instruments to measure the light spectra accurately; thus, we relied, in most cases, on manufacturer specifications. Unfortunately, those specifications were, at best, if at all available, a picture of the light spectrum. After aligning all available spectra and comparing the peaks, we did not find any conclusive trend that explains the observed differences in growth rate.

The plate-reader measurements resulted in less reproducible data at the intra- and interlab level than the spectrophotometer measurements (Figure 1). This observation can be partially explained by the fact that the initial conditions of the assay were set based on the latter device. In addition, several other factors could have influenced the reproducibility of the estimated RFU. First, two different measurements were performed to obtain these values and, thus, two different sensors (OD and fluorescence), increasing the chances of

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

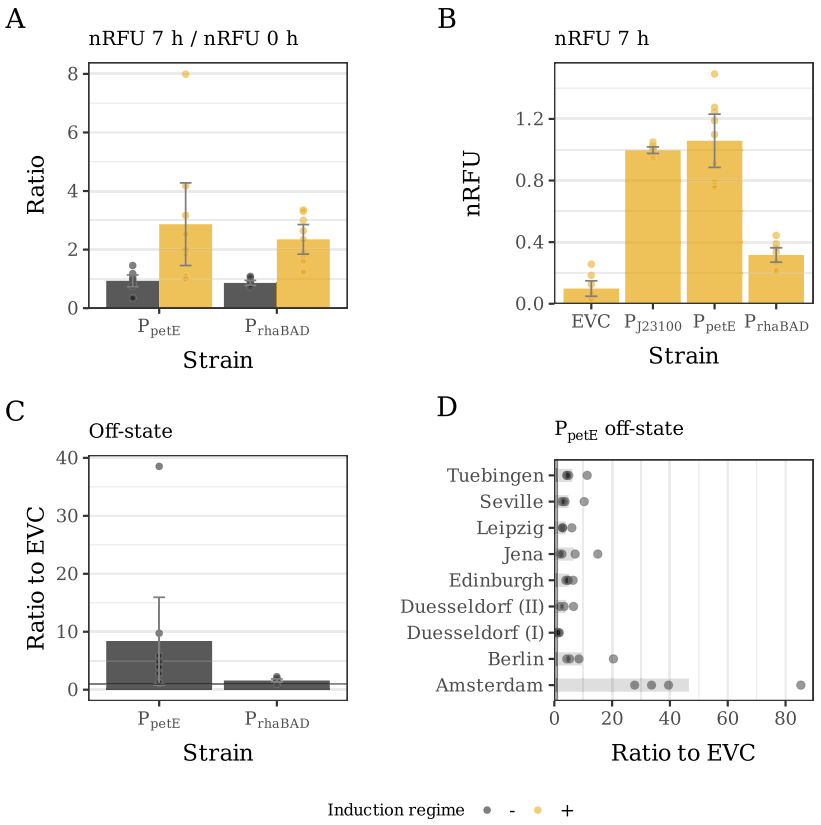


Figure 5. Characterization of P_{petE} and P_{rhaBAD} promoters. (A) nRFU ratio of time 7 h over time 0 h. (B) nRFU values at time 7 h of induced cultures. (C) nRFU ratio of uninduced PpetE and PrhaBAD to EVC culture at 7 h. (D) nRFU ratio of uninduced PpetE to EVC culture at 7 h in each laboratory. From A to C, points represent the mean value per laboratory (n = 3 or 4), bars show the average across all laboratories, and error bars show 95% CI (n = 7). In D, points show the mean value of technical replicates from a single experiment (n = 3), and bars show the average from each laboratory (n = 3 or 4). The color of bars and points indicate the induction regime, black for uninduced and yellow for induced cultures. In C and D, the horizontal and vertical black lines, respectively, represent a ratio of 1.

measurement errors. Next, it has been shown that fluorescence measurements of weak promoters (such as P_{rhaBAD}) are less reproducible due to the poor sensitivity of plate readers at low signals.[21,22] Lastly, the inclusion of two different induction regimes, which certainly affect the fluorescent output but not the growth behavior, might have contributed to the lower reproducibility of the estimated RFU × OD_{730}^{-1} compared to the spectrophotometer measurements.

Uninduced Reporter Systems Show a Higher Level of Variation Compared to Systems Induced above Saturation Levels. Interestingly, the promoter activity of inducible promoters was more reproducible in the presence than in the absence of the respective inducer. Introducing an additional experimental step (addition of the inducer) did not lead to a higher level of variation. Further, the observed significant differences between measured growth rates did not seem to translate into high variation in promoter activity, highlighting the robustness of those promoters across different physiological stages of cells. Lastly, the differences in initial cell biomass, inferred by differences in OD_{730} values of the identical stock between laboratories, resulted in differences in the inducer/cell count ratio. However, as induction was performed above saturation concentration for the inducer based on previous experiments,[6] the investigated promoters both seem to perform relatively robustly toward variations in inducer/cell count ratio.

It is important to note that the expression level of the P_{rhaBAD} promoter in the OFF state is extremely low and barely above background signals. Thus, much of the intralab CV of the P_{rhaBAD} can be attributed to the background noise and not variations in expression level (compare Supplemental Figure S5). The P_{petE} system shows higher location-dependent variation. Whereas residual copper could explain the higher variance in the P_{petE} promoter system, this can be excluded for P_{rhaBAD} as rhamnose is unlikely to be naturally present in trace concentrations. Varying copper residuals on glassware or in filtered water are expected to be the main culprit of this issue. Copper is a ubiquitously present metal ion that is difficult to remove altogether. Acid-washing glassware prior to experiments[28] could help here. Additionally, the use of chelators such as bathocuproinedisulfonic acid disodium salt[14] in the medium is a way to reduce copper availability efficiently during the experiment. For future studies employing the P_{petE} promoter in real applications, we recommend supporting the data with a fluorescence activity-based assay, as done in this study. This additional experiment should be done in the same strain used in the application and under similar experimental conditions. This process control not only helps to approximate the activity of the P_{petE} promoter in the actual experiment but also serves as a process control that can be used by investigators trying to replicate the results. As our results not only show high activity and variance in the OFF state of the P_{petE} promoter, we recommend carefully observing experimental results of a strain

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

with a seemingly uninduced P_{petE} promoter. Furthermore, these native problems of the P_{petE} promoter might be mitigated in the future by further engineering the promoter system: heterologous coexpression of the petR/petP regulators may positively impact the sensitivity to residual copper.

Strategies to Improve Comparability and Reproducibility. With the increasing throughput of experiments, improving reporting and reproducibility standards becomes increasingly important. None of the following recommendations are definite answers to this question. Instead, they are intended to provoke discussions and reflections about the reproducibility of cyanobacterial research.

A good strain-handling policy is important for keeping research results reproducible over time. Especially for bacterial strains, this is essential since genetic changes occur rapidly. Those changes resulting from improper strain maintenance lead to genotypic diversity and consequently to conflicts in research findings. The effect of the genetic diversity of PCC 6803 and its impact on phenotype has been analyzed in the past.[29,30] Cryopreservation techniques have been established for long-term microbial evolution experiments to reduce genotypic variability between experiments with the same strain.[31] To our knowledge, such techniques are not standard practice in the cyanobacterial field. In this study, we aimed to minimize strain variability not only between participants but also between experimental runs within a single location. Therefore, we adapted a cryopreservation and inoculation protocol based on glycerol conservation from Price et al.[31] to the PCC 6803 strains used in this interlab study. With this protocol, we hope to contribute to standardizing strain handling for cyanobacteria, especially for PCC 6803.

Next to varying practices in strain-keeping, the preparation of BG11 medium is a source for variations. During the establishment of the experimental procedures, we encountered that the preparation and final mass concentration of almost all solutes in BG11 media differed across participating laboratories. In addition, we obtained similar outcomes when comparing BG11 media protocols published in the scientific literature. Regarding standardization, we agreed on the formula developed by van Alphen et al.,[32] except for removing copper from the media for our particular purposes. Additionally, as outlined in the methods, a weak HEPES buffer was added to standardize starting conditions after inoculation. As cyanobacterial cultures rapidly increase the pH of their environment during cultivation, this buffer is assumed to be nonsignificant in prolonged cultivation. However, we assumed it would be better to buffer initial pH disruptions during inoculation. It remains open if those changes in BG11 media would substantially affect the experiments. However, we would encourage everyone also to standardize the preparation of BG11 as most differences in the preparation have no fundamental reason other than long-lasting traditions in respective laboratories.

In addition, when working with phototrophs, reporting the exact specification of light spectra used would be an important addition to reporting growth conditions. However, measuring light spectra requires a light meter, which is not standard laboratory equipment. Furthermore, manufacturers of light bulbs do not disclose those specifications in most cases or in an inappropriate format. Thus, we encourage manufacturers of light bulbs to report the light spectra as a CSV file with raw values. Further, we would encourage scientists to ask for the light spectra when buying a new incubator or contact the manufacturer to request the light spectra of your current incubators.

The iGEM interlab studies have gathered a considerable amount of data and protocols that support the use of external calibrants for both optical density and fluorescence for promoter characterization.[20−23] There are indeed clear benefits in defining such protocols. By calibrating plate readers (or flow cytometers) with external standards for fluorescence and cell concentration, promoter activity can be reported in absolute units, allowing direct comparisons across laboratories and facilitating the detection of biological and technical errors.[20−23] However, these protocols are not universally applicable. Regarding fluorescence reporters, GFP is not the only fluorescent protein used in cyanobacterial research, and YFP derivatives such as mVENUS have become popular in recent years.[33] Therefore, this requires that new calibrants are found for each new fluorescent protein, matching its excitation and emission wavelengths.

It has been known that OD measurements differ between instruments and, thus, laboratories.[17,34] Hence, only reporting OD values as a measure for cell biomass is problematic but still common practice in microbiology. The Stevenson et al. and later the iGEM interlab studies proposed using materials that match the refractive index of *E. coli* to calibrate OD.[20−23,34] In addition to being species-specific, this approach presents the disadvantage of assuming that cells of a given species are always the same size and shape. An option to overcome this problem would be to report cell counts and size with the starting OD values so that individual researchers can calibrate their OD measurements to the respective cell count.[35] We are aware that cell counters are not always available in all laboratories, as was the case among the participants of this study. However, we encourage reporting cell counts in combination with OD for better reproducibility in the future.

Our alternative to overcome these issues was using an internal biological standard. In our experimental design, we employed P_{J23100} as the normalizing promoter within each location, therefore accounting for differences between instruments. In addition, this approach also excludes factors such as the intracellular abundance of RNA polymerases and sigma factors, ribosomes, and other components of protein production impacting the mVENUS expression and maturation. Increased mVENUS expression from one of the inducible promoters related to any of these factors was expected to affect the mVENUS expression in the J23100 strain equally. As all our RFU values are reported as $RFU_{strain}/RFU_{J23100}$, these factors are presumably either buffered or completely negated. Considering that no significant differences in growth rates between the strains were found within each laboratory and identical absorption spectra across all the used strains in all laboratories, highly similar growth conditions can be assumed for all data sets. By correcting these factors, this normalization approach enables enhanced comparability across laboratories. This is demonstrated by the reduced interlab CV in nRFU, compared to the non-normalized metrics (Figure S3). It is worth noting that we observed a slightly lower coefficient of variation (CV) between laboratories when employing FU_{bc} instead of RFU for our normalization method (Figure S3). Essentially, including OD_{730} in the procedure led to an increased variation due to the intrinsic error of these measurements propagating into nRFU. Although this small increase in variation is present, we consider it justified as it addresses potential variation in fluorescence due to cultures

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

with different optical densities. Therefore, we recommend such a normalization strategy to report more reproducible data.

Lastly, our experimental setup used the native copper repressible promoter P_{petE} and the heterologous P_{rhaBAD} system as a somewhat orthologous regulation system. Even though the BG11 medium used in this study was prepared without adding copper salts, copper is hard to remove from glassware, and even a slight amount of residual copper could have caused the differences in basal activity observed in our experiments. Additionally, as a native promoter, P_{petE} may be subjected to more levels of regulation than so far known, which could lead to greater variability across different laboratories. Thus, we encourage the use of orthogonal systems like the P_{rhaBAD} promoter, which overall performed better in this study in terms of reproducibility.

Considering the orthogonal nature of the P_{rhaBAD} promoter in PCC 6803 and assuming no residual rhamnose in uninduced media, we hypothesize that the interlaboratory reproducibility of this experimental procedure and the multitude of measures taken to ensure maximum comparability of results across this study, we propose that the coefficients of variation of promoter activity for this system could indicate where the baseline for maximum reproducibility of quantitative data in PCC 6803 lies. Essentially, the coefficient of variation for the P_{rhaBAD} promoter serves as a call for caution when it comes to the overall reproducibility of fluorescence activity—and potentially other—data in the field of cyanobacterial research and a call for action to aid in improving comparability by following the discussed strategies: standardized and rigid strain keeping policies, standardized media composition, thorough reporting of cultivation conditions and the correct use of normalization agents.

CONCLUSION

In conclusion, the reproducibility of promoter activity in cyanobacteria within this study across multiple laboratories was better than expected. However, we also identified some causes for errors that could be improved in future studies. The remaining problem when using cyanobacteria is that most options to correct for those errors are understudied or completely missing, and those tools would first need further development to increase reproducibility.

METHODS

The strains for this study were generated in a single laboratory (Duesseldorf) and distributed to all participants. The interlab experiment was performed following a detailed set of protocols, which can be found at 10.17504/protocols.io.3byl4j69rlo5/v1. See Figure S1 for a visual description of the protocol.

Plasmid and Strain Construction. The plasmids were constructed via Golden Gate Cloning.[1] Complete and annotated sequences of all plasmids are available in supplementary data on Figshare (10.6084/m9.figshare.21525747.v5). According to standard procedures, *Escherichia coli* DH5α cells were transformed with Golden Gate mixes according to the MoCloFlex protocol.[36] Sequences were verified via Sanger sequencing. PCC 6803 was conjugated via triparental mating,[37] and conjugants were confirmed via colony PCR.

The PCC 6803 strain (glucose tolerant, nonmotile) was obtained from D. Bhaya (Carnegie Institution for Science, Stanford, USA).

Media. All participants prepared a BG11 medium without $CuSO_4$ supplemented with 10 mM $NaHCO_3$ and 5 mM HEPES-NaOH (pH 8). The exact composition of BG11 and the detailed recipe for the stock's preparation can be found in the supplements (Table S3 and S4) and was extracted from the Supporting Data S3 document of van Alphen et al.[32] It is hereafter referred to as BG11. For each repetition of the assay, 1 L of BG11 was prepared fresh for each round of experiments: First, approximately 500 mL of ultrapure H_2O were autoclaved in a 1 L volumetric bottle. Then stock solutions were acclimated to room temperature and added in the following order: 5 mL of 1 M HEPES-NaOH, 2.5 mL BG11 S1, 2.5 mL BG11 S2 (without $CuSO_4$), 2.5 mL BG11 S3, and 10.5 mL 0.95 M $NaHCO_3$. After adding the solutions, sterile ultrapure H_2O was added up to the 1 L mark of the bottle.

Cultivation for Strain Conservation. For the preparation of cryoconserved cultures, PCC 6803 was grown under the same conditions outlined in Culture Conditions in Interlab Experiment, except for increasing the light intensity to 80 μ mol photons $\times$ m^{-2} $\times$ s^{-1}. When the OD measured at 730 nm wavelength (hereafter OD_{730}) reached a value of 3, cells were centrifuged at 12,000*g* (fixed angle rotor) for 15 min, washed with fresh BG11 medium, centrifuged once again, and resuspended in 10% of the initial volume with BG11 and 15% (v/v) final concentration glycerol. Cells were stored at −80 °C and shipped to participating laboratories on dry ice. The incubator and photometer listed in Table S5 under "Duesseldorf I" were used to prepare cryoconserved cultures.

Participants. Eight different laboratories participated in the study and are described throughout the study by the geographical location of the research facility. In alphabetical order, the participants were from Amsterdam (University of Amsterdam), Berlin (Freie Universität Berlin), Duesseldorf (Heinrich Heine University), Edinburgh (University of Edinburgh), Jena (Friedrich Schiller University), Leipzig (Helmholtz Centre for Environmental Research), Seville (University of Seville), and Tuebingen (University of Tübingen), as included in the author list. In Duesseldorf, the complete experiment was independently performed by two researchers, indicated in the text as Duesseldorf (I) and Duesseldorf (II). Strains were prepared and grown in Duesseldorf and shipped to participants on dry ice.

Culture Conditions in Interlab Experiment. Cultures were grown in 100 mL Erlenmeyer nonbaffled glass flasks with cotton plugs in shaking incubators set at 100 rpm (see Table S5) under 50 μmol photons $\times$ m^{-2} $\times$ s^{-1} constant white light illumination, ambient CO_2, and 30 °C (except for Tuebingen where the temperature was 28 °C). Cultures were grown in BG11 medium, as outlined above. Cultures were always grown with 10 μg mL^{-1} chloramphenicol, which was added individually to each flask before inoculation.

Promoter Quantification Assay. PCC 6803 strains harboring plasmids containing the investigated promoter-reporter cassette were inoculated from cryoconserved cultures by adding 330 μL of inoculum to 10 mL of copper-free BG11 medium supplied with chloramphenicol and cultivated for 48 h. Cultures were then diluted to an OD_{730} of 0.3 in 35 mL and grown overnight to OD_{730} of 0.5−0.6. The following day, the complete volume of each preculture was transferred to a sterile tube and diluted to OD_{730} 0.5 if necessary. The flasks where

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

the precultures had been grown were washed twice with copper-free BG11. Subsequently, cultures were separated into two times 20 mL of OD_{730} 0.5 in two separate sets of flasks, which were labeled as induced and uninduced. The washed flasks were used for the uninduced cultures and were supplied with 200 μL of ultrapure H_2O. In the induced set, 200 μL of 200 μM $CuSO_4$ or 1 M rhamnose, were supplied to the P_{petE} and P_{rhaBAD} flasks, respectively. In the EVC and P_{J23100} cultures, 200 μL of ultrapure H_2O was supplied. Throughout the text, we refer to induced and uninduced cultures of EVC and P_{J23100}, although they do not carry an inducible promoter, and ultrapure H_2O was supplied in all cases. Therefore for these strains, this nomenclature should be interpreted as cultures that were grown in the washed flasks (uninduced) and those that grew in the new flasks (induced). Reusing the preculture flask was an attempt to remove residual copper from the flask where the uninduced P_{petE} culture would be grown. The rationale behind it was that by growing a culture in a free-copper medium, cells could potentially take up residual copper from the glassware. To maintain a homogeneous protocol for all strains, this step was also performed on the other three strains.

Sampling and Measurements. In total, seven samples were taken from each flask. Samples were taken at 0, 2, 4, 5, 6, 7, and 24 h after adding the inducers in the promoter quantification assay. In most cases, 1 mL of culture was taken at each sampling point except for those laboratories that performed chlorophyll extraction at times 0, 7, and 24 h, in which case the sample volume was 2 mL.

As mentioned above, OD_{730} was measured in a spectrophotometer at a 730 nm wavelength (see Table S5 for models at each laboratory). For spectrophotometer measurements, samples were diluted 1:2 (500 μL:500 μL) for the same-day measurements and 1:5 (200 μL:800 μL) for the 24-h measurements. For plate reader measurements, 100 μL of liquid cultures were used without further dilution for same-day measurements and 100 μL of 1:5 (20 μL:80 μL) dilution for 24 h measurements.

The mVENUS fluorescence intensity was measured in a plate reader (see Table S5 for models at each laboratory) using an excitation window of 506−518 nm, and emission was detected in a 542−562 nm window.

The 24 h time point was excluded from the data analysis for technical reasons. To ensure the comparability of measurements within each laboratory, we used the same plate reader settings for all measurements. While this worked well for all time points, including the 7 h measurement, culture cell densities at the 24 h time point were beyond the linear sensitivity range. Simultaneously the dilution of the culture at 24 h caused some measurements to be at the lower plate reader sensitivity level for the low-density cultures. In both cases, measurements produced unreliable results that made data analysis impossible. We encountered this issue within the running study and did not have enough identical starting cultures in each laboratory to revise the study's design. This issue highlights the problems of reproducing experiments in other laboratories with different equipment if no further precautions and tests are implemented before doing the actual study. Thus, reproducing other results requires substantial modifications to the setup of instruments and workflow across laboratories in some circumstances. However, for the purpose of the study, the included data points are sufficient to analyze the reproducibility of the promoters used in this study.

Chlorophyll Extraction. One milliliter of cyanobacterial culture was centrifuged at 10,000*g* for 5 min. 900 μL of supernatant was discarded, and the pellet was resuspended in the remaining 100 μL. Next, 900 μL of 100% methanol was added and mixed thoroughly by vortexing. Samples were incubated in the dark at 4 °C for 5 min and centrifuged again at 10,000*g* for 5 min. The supernatant was transferred to a cuvette, and extinction was measured at 665 nm using a spectrophotometer. 900 μL methanol mixed with 100 μL BG11 was used as the reference solution. To estimate chlorophyll concentration from the absorbance at 665 nm, eq 1 was used. The extraction protocol and eq 1 were adapted from Ritchie.[38]

$$\mathrm{Chl}\left(\frac{\mu g}{mL}\right) = A_{665nm} \times 12.9447 \frac{\mu g}{mL} \times \text{Dilution factor} \quad (1)$$

Data Collection. The assay was performed four times independently in each laboratory. The four strains used in each experimental run were inoculated from an individual glycerol stock and, therefore, considered a biological replicate from each strain, induction regime, and experimental run. Each biological replicate included seven measurement points at 0, 2, 4, 5, 6, 7, and 24 h after inoculation. In the case of the plate reader, each biological replicate was measured in three independent wells at each time point.

Data Analysis. Data from all participants were submitted in a standardized spreadsheet file containing all measurements from a single experimental run. All subsequent analyses to process the data and create the figures in this manuscript were carried out in R.[39] The following R packages were used in the analysis: agricolae,[40] broom,[41] ggforce,[42] ggthemes,[43] janitor,[44] patchwork,[45] readxl,[46] and tidyverse.[47] The data, as well as the code of the complete analysis and figures, are available at https://github.com/hugo-pH/cyano_interlab. The raw data are also available at 10.6084/m9.figshare.21525747.v5.

Growth Rates. Growth rates were estimated from the spectrophotometer OD_{730} data by linear regression, using the log-transformed equation of exponential growth ($\ln OD_{730}\ t = \mu \times t + \ln OD_{730}\ t_0$), where μ denotes the growth rate. All time points were used in this analysis.

Fluorescence Analysis and Normalization. Both raw OD_{730} and fluorescence units (FU) from the plate reader data set were first background-corrected by subtracting the average value of the blank wells at each time point (eqs 2 and 3). In addition, the smallest FU value measured in each experimental run and location was added to all data points to avoid negative FU values. Next, relative fluorescence units (RFU) were calculated by dividing the background-corrected FU by the background-corrected OD_{730} for each technical replicate (eq 4), followed by averaging all technical replicates. Since the scale of OD_{730} and FU recorded by plate readers greatly varies between devices, a normalization method was implemented to compare results across different devices. Instead of using an external calibrant, the RFU of P_{J23100} cultures without an inducer was used as an internal biological standard for plate reader measurements. This strain was chosen because it expresses mVENUS constitutively. This approach can correct not only the different measuring ranges of plate reader devices but also possible differences in the physiological state due to discrepancies in growth conditions across the experimental runs (and across laboratories). Finally, the RFU of each biological replicate and time point was divided by the

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

corresponding P_{J23100} RFU value (eq 5) to apply this method, obtaining the normalized RFU (nRFU).

$$OD_{730bc} = OD_{730raw} - \overline{OD_{730blank}} \quad (2)$$

$$FU_{bc} = FU_{raw} - \overline{FU_{blank}} \quad (3)$$

$$RFU = \frac{FU_{bc}}{OD_{730bc}} \quad (4)$$

$$nRFU = \frac{RFU}{RFU_{J23100-}} \quad (5)$$

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.3c00150.

Additional analysis of variations of plate reader and spectrophotometer measurements, an equipment list for each participating laboratory, and a detailed description of the medium preparation (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Filipe Branco dos Santos − *Molecular Microbial Physiology Group, Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, Amsterdam 1098 XH, The Netherlands*; orcid.org/0000-0002-4268-8080; Email: F.BrancodosSantos@uva.nl

Nicolas M. Schmelling − *Institute for Synthetic Microbiology, Heinrich Heine University Duesseldorf, 40225 Duesseldorf, Germany*; Email: Nicolas.Schmelling@hhu.de

Authors

Maurice Mager − *Institute for Synthetic Microbiology, Heinrich Heine University Duesseldorf, 40225 Duesseldorf, Germany*

Hugo Pineda Hernandez − *Molecular Microbial Physiology Group, Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, Amsterdam 1098 XH, The Netherlands*

Fabian Brandenburg − *Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany*

Luis López-Maury − *Instituto de Bioquímica Vegetal y Fotosíntesis, University of Seville − CSIC, 41092 Sevilla, Spain; Departamento de Bioquímica Vegetal y Biología Molecular, Facultad de Biología, University of Seville, 41012 Sevilla, Spain*

Alistair J. McCormick − *Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, EH9 3BF Edinburgh, U.K.*; orcid.org/0000-0002-7255-872X

Dennis J. Nürnberg − *Department of Physics, Experimental Biophysics, Freie University Berlin, 14195 Berlin, Germany; Dahlem Centre of Plant Sciences, Freie Universität Berlin, 14195 Berlin, Germany*

Tim Orthwein − *Interfaculty Institute of Microbiology and Infection Medicine, University of Tuebingen, 72076 Tübingen, Germany*

David A. Russo − *Institute for Inorganic and Analytical Chemistry, Bioorganic Analytics, Friedrich Schiller University Jena, 07743 Jena, Germany*

Angelo Joshua Victoria − *Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, EH9 3BF Edinburgh, U.K.*

Xiaoran Wang − *Department of Physics, Experimental Biophysics, Freie University Berlin, 14195 Berlin, Germany*

Julie A. Z. Zedler − *Matthias Schleiden Institute for Genetics, Bioinformatics and Molecular Botany, Synthetic Biology of Photosynthetic Organisms, Friedrich Schiller University Jena, 07743 Jena, Germany*; orcid.org/0000-0002-0462-8810

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.3c00150

Author Contributions

[¶]MM and HPH contributed equally.

Author Contributions

NMS developed the initial concept of the study. All authors together revised the initial study design and determined the final study design. MM prepared strains and performed initial experiments prior to the study. MM, HPH, FB, LLM, DJN, XW, TO, DAR, AV, JAZZ, and NMS performed the experiments. HPH performed the data analysis. MM, HPH, and NMS wrote the manuscript. All authors read, corrected, and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dennis Dienst for valuable input on the plasmid design and the supply of BCD2 as well as the P_{petE} promoter and Anna Behle for the P_{rhaBAD} promoter. Further, we would like to thank Alberto Scarampi and Moritz Koch for their comments and discussion in the early phase of the project. Furthermore, we thank Ilka Axmann for providing lab space and support to the authors in Düsseldorf and for discussions and valuable comments on the manuscript. In addition, we thank Anna Frederike Schulte-Huermann for her help in the lab in Düsseldorf. Finally, the authors thank the "Gesellschaft von Freunden und Förderern der HHU" for the funding required to design and ship the strains and constructs to all participating laboratories. FB was cofinanced from funds of the European Regional Development Fund (EFRE) and by means of taxation based on the budget adopted by the representatives of the Landtag of Saxony (Nr. 100361842). Further, we acknowledge the use of the facilities of the Centre for Biocatalysis (MiKat) and the Helmholtz Centre for Environmental Research (UFZ), which is supported by the Helmholtz Association. DAR was supported by the Alexander von Humboldt Foundation. JAZZ was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy − EXC 2051 − Project-ID 390713860 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) − CRC 1127/2−239748522. HPH and FBS received funding from the European Union's Horizon 2020 Research, an innovation program under Grant Agreement No. 760994 (ENGICOIN project). LLM was supported by Grant PID2020-112645GB-I00 funded by MCIN/AEI/10.13039/501100011033. DJN was funded through the Deutsche Forschungsgemeinschaft − Emmy Noether Project Award No. NU 421/1. AJV was funded by a postgraduate research scholarship from the Darwin Trust of Edinburgh. AJM acknowledges funding

from the UK Biotechnology and Biological Sciences Research Council (BBSRC) grants [BB/S020128/1].

■ ABBREVIATIONS

nRFU, normalized relative fluorescence units; BCD2, bicistronic design.

■ REFERENCES

(1) Weber, E.; Engler, C.; Gruetzner, R.; Werner, S.; Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS One* **2011**, *6*, 1−11.
(2) Ungerer, J.; Pakrasi, H. B. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Sci. Rep.* **2016**, DOI: 10.1038/srep39681.
(3) Liu, D.; Johnson, V. M.; Pakrasi, H. B. A Reversibly Induced CRISPRi System Targeting Photosystem II in the Cyanobacterium *Synechocystis* sp. PCC 6803. *ACS Synth. Biol.* **2020**, *9*, 1441−1449.
(4) Sode, K.; Tatara, M.; Takeyama, H.; Burgess, J. G.; Matsunaga, T. Conjugative gene transfer in marine cyanobacteria: *Synechococcus* sp., *Synechocystis* sp. and *Pseudanabaena* sp. *Appl. Microbiol. Biotechnol.* **1992**, *37*, 369.
(5) Vasudevan, R.; Gale, G. A. R.; Schiavon, A. A.; Puzorjov, A.; Malin, J.; Gillespie, M. D.; Vavitsas, K.; Zulkower, V.; Wang, B.; Howe, C. J.; Lea-Smith, D. J.; McCormick, A. J. CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax. *Plant Physiology* **2019**, *180*, 39−55.
(6) Behle, A.; Saake, P.; Germann, A. T.; Dienst, D.; Axmann, I. M. Comparative Dose−Response Analysis of Inducible Promoters in Cyanobacteria. *ACS Synth. Biol.* **2020**, *9*, 843−855.
(7) Gale, G. A. R.; Wang, B.; McCormick, A. J. Evaluation and Comparison of the Efficiency of Transcription Terminators in Different Cyanobacterial Species. *Front. Microbiol.* **2021**, *11*, 11.
(8) Englund, E.; Pattanaik, B.; Ubhayasekera, S. J. K.; Stensjö, K.; Bergquist, J.; Lindberg, P. Production of Squalene in *Synechocystis* sp. PCC 6803. *PLoS One* **2014**, *9*, 1−8.
(9) Geerts, D.; Bovy, A.; de Vrieze, G.; Borrias, M.; Weisbeek, P. Inducible expression of heterologous genes targeted to a chromosomal platform in the cyanobacterium *Synechococcus* sp. PCC 7942. *Microbiology* **1995**, *141*, 831−841.
(10) El-Shehawy, R. M.; Kleiner, D. Effect of controlled expression of the hetR gene on heterocyst formation in the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Physiologia Plantarum* **2003**, *119*, 44−48.
(11) Ferreira, E. A.; Pacheco, C. C.; Pinto, F.; Pereira, J.; Lamosa, P.; Oliveira, P.; Kirov, B.; Jaramillo, A.; Tamagnini, P. Expanding the toolbox for *Synechocystis* sp. PCC 6803: validation of replicative vectors and characterization of a novel set of promoters. *Synth. Biol.* **2018**, DOI: 10.1093/synbio/ysy014.
(12) Giner-Lamia, J.; López-Maury, L.; Florencio, F. J. Global Transcriptional Profiles of the Copper Responses in the Cyanobacterium *Synechocystis* sp. PCC 6803. *PLoS One* **2014**, *9*, 1−16.
(13) Kelly, C. L.; Taylor, G. M.; Hitchcock, A.; Torres-Méndez, A.; Heap, J. T. A Rhamnose-Inducible System for Precise and Temporal Control of Gene Expression in Cyanobacteria. *ACS Synth. Biol.* **2018**, *7*, 1056−1066.
(14) García-Cañas, R.; Giner-Lamia, J.; Florencio, F. J.; López-Maury, L. A protease-mediated mechanism regulates the cytochrome c_6/plastocyanin switch in *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2017898118.
(15) Englund, E.; Liang, F.; Lindberg, P. Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* **2016**, DOI: 10.1038/srep36640.
(16) Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **2016**, *533*, 452.
(17) Myers, J. A.; Curtis, B. S.; Curtis, W. R. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys.* **2013**, DOI: 10.1186/2046-1682-6-4.
(18) Tietze, L.; Mangold, A.; Hoff, M. W.; Lale, R. Identification and Cross-Characterisation of Artificial Promoters and 5 Untranslated Regions in *Vibrio natriegens*. *Front. Bioeng. Biotechnol.* **2022**, DOI: 10.3389/fbioe.2022.826142.
(19) Interlaboratory Comparisons. *EU Science Hub*. European Commission, 2022. https://joint-research-centre.ec.europa.eu/reference-measurement/interlaboratory-comparisons_en (Accessed April 19, 2022).
(20) Beal, J.; Haddock-Angelli, T.; Gershater, M.; de Mora, K.; Lizarazo, M.; Hollenhorst, J.; Rettberg, R. Reproducibility of Fluorescent Expression from Engineered Biological Constructs in *E. coli*. *PLoS One* **2016**, *11*, 1−22.
(21) Beal, J.; Haddock-Angelli, T.; Baldwin, G.; Gershater, M.; Dwijayanti, A.; Storch, M.; de Mora, K.; Lizarazo, M.; Rettberg, R. Quantification of bacterial fluorescence using independent calibrants. *PLoS One* **2018**, *13*, 1−15.
(22) Beal, J.; Farny, N. G.; Haddock-Angelli, T.; Selvarajah, V.; Baldwin, G. S.; Buckley-Taylor, R.; Gershater, M.; Kiga, D.; Marken, J.; Sanchania, V.; Sison, A.; Workman, C. T. Robust estimation of bacterial cell count from optical density. *Commun. Biol.* **2020**, DOI: 10.1038/s42003-020-01127-5.
(23) Beal, J.; et al. Comparative analysis of three studies measuring fluorescence from engineered bacterial genetic constructs. *PLoS One* **2021**, *16*, 1−15.
(24) Egan, S. M.; Schleif, R. F. A Regulatory Cascade in the Induction of rhaBAD. *J. Mol. Biol.* **1993**, *234*, 87−98.
(25) Mutalik, V. K.; Guimaraes, J. C.; Cambray, G.; Lam, C.; Christoffersen, M. J.; Mai, Q.-A.; Tran, A. B.; Paull, M.; Keasling, J. D.; Arkin, A. P.; Endy, D. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **2013**, *10*, 354−360.
(26) Bland, E.; Angenent, L. T. Pigment-targeted light wavelength and intensity promotes efficient photoautotrophic growth of Cyanobacteria. *Bioresour. Technol.* **2016**, *216*, 579−586.
(27) Luimstra, V. M.; Schuurmans, J. M.; Verschoor, A. M.; Hellingwerf, K. J.; Huisman, J.; Matthijs, H. C. P. Blue light reduces photosynthetic efficiency of cyanobacteria through an imbalance between photosystems I and II. *Photosynthesis Research* **2018**, *138*, 177−189.
(28) Elomaa, H.; Seisko, S.; Junnila, T.; Sirviö, T.; Wilson, B. P.; Aromaa, J.; Lundström, M. The Effect of the Redox Potential of Aqua Regia and Temperature on the Au, Cu, and Fe Dissolution from WPCBs. *Recycling* **2017**, *2*, 14.
(29) Trautmann, D.; Voß, B.; Wilde, A.; Al-Babili, S.; Hess, W. R. Microevolution in Cyanobacteria: Re-sequencing a Motile Substrain of *Synechocystis* sp. PCC 6803. *DNA Research* **2012**, *19*, 435−448.
(30) Zavřel, T.; Očenášová, P.; Červený, J. Phenotypic characterization of *Synechocystis* sp. PCC 6803 substrains reveals differences in sensitivity to abiotic stress. *PLoS One* **2017**, *12*, 1−21.
(31) Price, C. E. Adaption to glucose limitation is modulated by the pleotropic regulator CcpA, independent of selection pressure strength. *BMC Evol. Biol.* **2019**, DOI: 10.1186/s12862-018-1331-x.
(32) van Alphen, P.; Abedini Najafabadi, H.; Branco dos Santos, F.; Hellingwerf, K. J. Increasing the Photoautotrophic Growth Rate of *Synechocystis* sp. PCC 6803 by Identifying the Limitations of Its Cultivation. *Biotechnol. J.* **2018**, *13*, 1700764.
(33) Yokoo, R.; Hood, R. D.; Savage, D. F. Live-cell imaging of cyanobacteria. *Photosynthesis Research* **2015**, *126*, 33−46.
(34) Stevenson, K.; McVey, A. F.; Clark, I. B. N.; Swain, P. S.; Pilizota, T. General calibration of microbial growth in microplate readers. *Sci. Rep.* **2016**, *6*, 38828.
(35) Hays, S. G.; Yan, L. L. W.; Silver, P. A.; Ducat, D. C. Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. *J. Biol. Eng.* **2017**, DOI: 10.1186/s13036-017-0048-5.
(36) Klein, C. A.; Emde, L.; Kuijpers, A.; Sobetzko, P. MoCloFlex: A Modular Yet Flexible Cloning System. *Front. Bioeng. Biotechnol.* **2019**, DOI: 10.3389/fbioe.2019.00271.

(37) Behle, A. Triparental mating of Synechocystis. *protocols.io*. December 20, 2016. DOI: 10.17504/protocols.io.ftpbnmn.
(38) Ritchie, R. J. Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and total chlorophylls in natural assemblages of photosynthetic organisms using acetone, methanol, or ethanol solvents. *Photosynthetica* **2008**, *46*, 115−126.
(39) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
(40) de Mendiburu, F. *agricolae: Statistical Procedures for Agricultural Research*; R Package version 1.3.1; 2019.
(41) Robinson, D.; Hayes, A.; Couch, S. *broom: Convert Statistical Objects into Tidy Tibbles*; R package version 0.7.10; 2021.
(42) Pedersen, T. L. *ggforce: Accelerating ggplot2*; R package version 0.4.1; 2022.
(43) Arnold, J. B. *ggthemes: Extra Themes, Scales and Geoms for ggplot2*; R package version 4.2.4; 2021.
(44) Firke, S. *janitor: Simple Tools for Examining and Cleaning Dirty Data*; R package version 2.1.0; 2021.
(45) Pedersen, T. L. *patchwork: The Composer of Plots*; R package version 1.1.2; 2022.
(46) Wickham, H.; Bryan, J. *readxl: Read Excel Files*; R package version 1.3.1; 2019.
(47) Wickham, H.; et al. Welcome to the tidyverse. *Journal of Open Source Software* **2019**, *4*, 1686.

https://doi.org/10.1021/acssynbio.3c00150
ACS Synth. Biol. 2023, 12, 1823−1835

Manuscript 3

Synthetic inter- and intraspecies symbiosis between *E.coli* and *V.natriegens* shows robust cell-to-cell abundance ratio

Maurice Mager*[1], St. Elmo Wilkens*[2], Oliver Ebenhöh[2], Ilka M. Axmann[1]

*Contributed equally

1: Synmibi Institute of Synthetic Microbiology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

2: Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

Abstract

Most microbes coexist as part of communities comprised of many different species. Such microbial consortia have conquered a variety of habitats and are part of essential geo- and biochemical transformation processes. Yet, research on microbial communities is still in its infancy. The recent emergence of studies on hypercomplex consortia such as the human gut microbiome, spurred by increasing availability of omic technologies, has revealed interesting findings such as the connection of Alzheimer's disease and the gut microbiome composition. Yet, most of these works operate on a phenomenological level and few dwell into the fundamental mechanics behind microbial consortia. Hence, due to the sheer complexity of natural consortia, there is a need for simple, amenable consortia for the study of the dynamics that govern communication, nutrient exchange, and composition of microbial networks. In this work, we created artificial inter- and intraspecies symbiotic communities based on mutual dependency using *E.coli* and *V.natriegens*. Mutants of each species were generated that show amino acid auxotroph and are hence incapable of growing in minimal media individually, yet can grow when mixed in one culture. We discover that these communities maintain a stable relative cell density that is resistant to a variety of internal and external manipulation attempts. We also observe a dynamic establishment of this stable relative cell density that is independent from initial conditions.

Introduction

Most microbes in nature are organized as part of consortia. These microbial networks have conquered nearly every habitat, from hot springs to the gut of humans. They coexist in an intricate web of interactions, in relationships that have been shaped by evolution over millions of years and play a crucial role in biogeochemical transformation cycles, environmental processes, and human health. Yet, most microbiological research has so far focused on the study of individual, isolated species and the study of microbial networks is still in its infancy despite its potential for bioproduction, bioremediation and healthcare.

With the rise of omics technologies in the last decades, large scale analysis of highly complex microbial networks such as the human gut microbiome has become highly popular in recent years, linking changes in the microbiome to Alzheimer disease on a phenomenological level (Pistollato et al. 2016). However, the elucidation of fundamental dynamics and mechanism that describe the relationships, interactions and changes in microbial networks is difficult in a system of such complexity.

Microbial communities from the same soil isolate assemble in minimal media to different stable communities in each replicate, however these communities showed a higher degree of similarity on a family taxa level (Goldford et al. 2018). Additionally, the composition of the taxa could be changed by using a different carbon source. Louca *et al.* interprets the observation that in most microbial community at any given time multiple species can perform the same function by suggesting that functions seem to be more conserved than species in microbial networks (Louca et al. 2018). Species regularly extinct or immigrate in consortia, functionalities, like denitrification or methanogenesis, remain constant as an effect of the functional redundancy. Observations on the microbiome of the algal surface of *U. australis* confirm this theory: microbial samples taken from these algae overlap in only 15% on a phylogenetic level, however a comparison of the present functionalities shows a 70% overlap with some

functionalities being present on each sample (Burke et al. 2011). As this functional theory seems to contradict the ecological dogma that niches cannot be occupied by a multitude of species, Louca *et al.* suggest that functional redundancy is favored by the multifactorial nature of the niches: the multitude of metabolic services is vast and other functions unrelated to metabolism such as predation, growth kinetics and antibiotic warfare further split niches and allow for functional redundancy in most metabolic services. It becomes apparent, that the theoretical background of community stability and the underlying governing factors is still discussed and more experimental work is required to answer the remaining questions.

Hence, the need for simple microbial networks for the study of their fundamental laws becomes apparent. In recent years, synthetic biology has emerged as a powerful tool for engineering microbial systems with tailored functionalities (Schwille 2011). Inspired by natural mutualistic relationships, researchers have begun to explore and create design principle to form synthetic symbiotic networks. Further, microbial consortia capable of performing complex tasks beyond the capabilities of individual partners have been developed (Fang et al. 2018; Zhang et al. 2015) but to our knowledge little effort has been made to engineer the stability and predictability of such networks. A popular approach involves the creation of dependency-based mutualisms, where two or more microbial species rely on each other's metabolic outputs for growth and survival (Hosoda et al. 2011). This strategy not only enhances the stability of the synthetic community but also allows for the implementation of division of labor and metabolic specialization among the co-cultured organisms. One target for dependencies is the generation of amino acid auxotrophies, as amino acids are essential and the auxotrophies easy to achieve, often requiring a single mutation. Mee *et al.* and Silver *et al.* probed pairing for such amino acid dependencies in *E.coli* and assessed their ability to form mutualistic relationships in combinatorial assays (Mee et al. 2014; Wintermute und Silver 2010). As *E.coli* is a well understood host, this poses a great platform to create minimal and controllable microbial communities.

Aim of this project was to generate inter- and intraspecies symbiotic networks of *E.coli* and *V.natriegens* via amino acid based mutualistic dependencies. Further, the stability of cell-to-cell abundance ratios in these networks was assessed by tracking relative cell densities over time. We discover a remarkable robustness of relative ratios and a change in the ratio corresponding to the type of amino acid dependency. Further, we aimed to manipulate the ratio of individual networks via internal and external manipulation and discovered a remarkable resilience of these communities.

Materials and Methods

Generation of KO mutants

Knockout mutants of *E.coli* and *V.natriegens* were generated using the CRISPR SWAPnDROP system according to standard protocol (Teufel et al. 2022). Following strains were used in this project:

Table 1: Strains used for this manuscript

Strain name	Annotation	Origin
V.natriegens	Vibrio natriegens ATCC 14048 Δvnp12 Δdns	Obtained from Patrick Sobetzko
E.coli	Escherichia coli MG1655	Strain collection
V.natriegens ΔilvA	Vibrio natriegens ATCC 14048 Δvnp12 Δdns ilvA::GFP + pCR3	This work
V.natriegens ΔmetA	Vibrio natriegens ATCC 14048 Δvnp12 Δdns metA::mCherry + pCR3	This work
E.coli ΔilvA	Escherichia coli MG1655 ilvA::GFP + pCR2	This work
E.coli ΔmetA	Escherichia coli MG1655 metA::mCherry +pCR2	This work

Media

E.coli was grown in LB media or M9 media, unless specified otherwise.

V.natriegens was grown in LBv2 media or VN minimal media (Hoffart et al. 2017), unless specified otherwise.

Inoculation of microbial communities

For intraspecies communities of E.coli, overnight cultures of each strain were prepared in 5ml of LB containing appropriate antibiotics. Next day, 2ml of each culture were washed twice in M9 media and adjusted to a final OD of 1. Communities were inoculated in M9 media from washed cells to a total starting OD of 0.02 at a ratio of 1:1 from each strain, unless specified in the experiment. Antibiotics and other additives were supplied as specified in each experiment.

For intraspecies communities of V.natriegens, 5ml of LBv2 cultures of each strain were prepared either over night or in the morning and appropriate antibiotics were added. 2 ml of each culture were washed twice in VN minimal media (Hoffart et al. 2017) and adjusted to a final OD of 1. Communities were inoculated in VN minimal media from washed cells to a total starting OD of 0.02 at a ratio of 1:1 from each strain, unless specified in the experiments. Antibiotics and other additives were supplied as specified in each experiment.

For interspecies communities of *V.natriegens* and E.coli, precultures were prepared and washed as outlined above. Communities were inoculated in a mix of 70% VN minimal media and 30% M9 media as outlined for intraspecies communities above.

Measurement of relative cell densities

For the measurement of relative cell densities, a 1 µl sample was taken from a microbial community and diluted in 200µl of the same media used for cultivation. Cells were counted using a Cytoflex Flow cytometer. Populations were distinguished due to their green (ilvA::GFP) or red (metA::RFP) fluorescence and gated based on cell counts from each individual strain grown in complex media as parallel positive controls. Relative cell densities were calculated as the fraction of green or red cells from the sum of all green or red cells.

Results

Intraspecies mutualistic communities converge towards a specific cell-to-cell abundance ratio

E.coli mutants lacking ilvA or metA were mixed to a community and grown for 7 days. Four communities were started, at initial relative ratios for Δilva:ΔmetA of 1:9, 1:2, 2:1, and 9:1, respectively. Communities were diluted every 24h to OD 0.1 and before dilution a sample was taken to measure the relative cell densities.

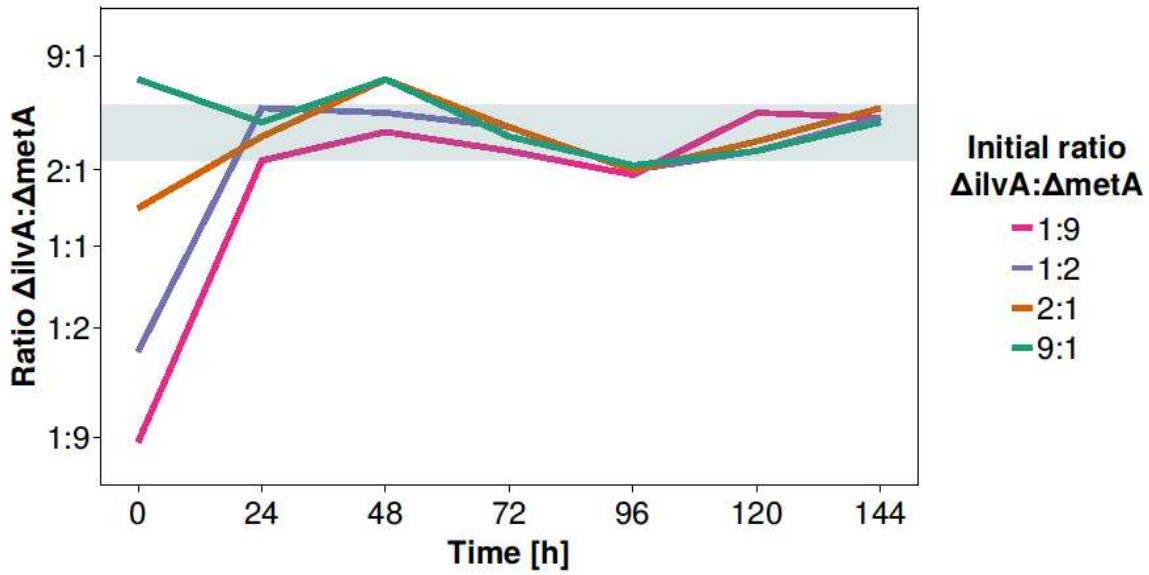


Figure 1: Relative cell density of *E.coli* Δilva and ΔmetA cells in a mixed community. Each line shows technical triplicates of communities that was inoculated at a specific initial ratio. Cultures were diluted 1:40 every 24h, right after measuring relative cell density. Gray band shows final relative density of 73% +/-5%.

We observed that, no matter the initially established ratio, all communities converged towards the same relative cell density over night (24h) and were able to maintain this ratio for the time of observation (144h) (Figure 1). This occured independent of which strain was underrepresented (compare 2:1 and 9:1 ratio). The final relative density was at about 73% of ilvA cells with a standard deviation of 5% (average of all data points from 24h to 144h, highlighted as grey bar).

Dynamic establishment of preferred ratio occurs during growth

In the experiment shown in Figure 1 we observed a considerably long lag phase of ~16 hours (compared to ~4 hours when wild type *E.coli* strains are inoculated in M9 media, data not shown). Hence, we assumed some form of adaptation to the new media in conjunction with the symbiotic interaction. We noticed that the lag phase was longer, the more the initial relative cell density deviated from the final ratio found in the stable communities. This led to the question whether the partners had to first establish the preferred relative cell density before notable growth could occur. Therefore, we monitored a freshly mixed community in their first 24 hours in higher resolution with respect to the relative cell density (Figure 2). An *E.coli* community with an initial ratio for Δilva:ΔmetA of 1:9 was inoculated to a starting OD of 0.1 and relative cell densities were measured every 4h for 36h until stationary phase was reached.

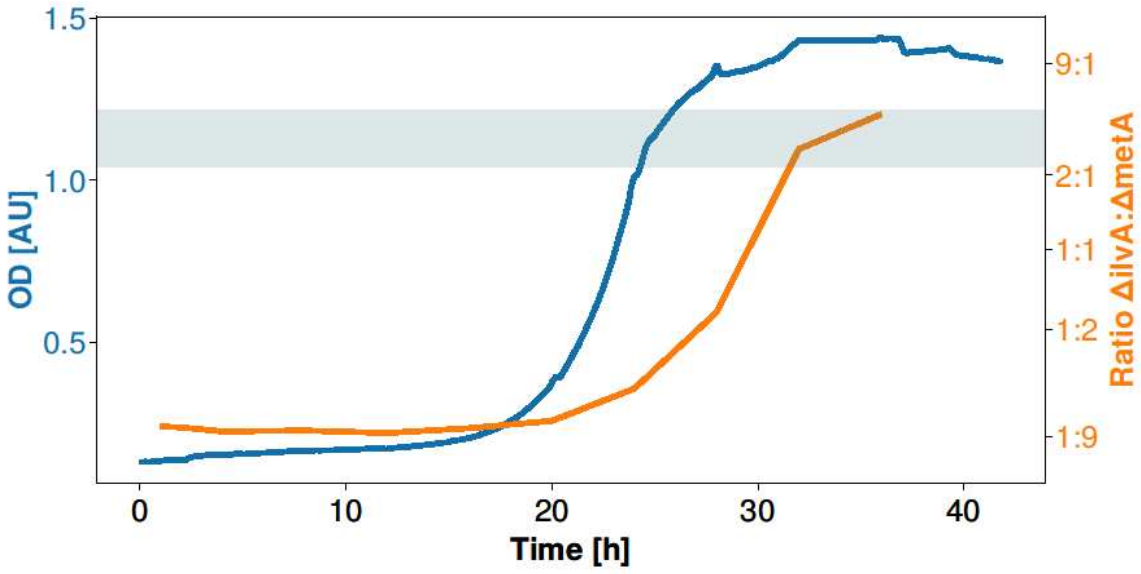


Figure 2: Relative cell density of *E.coli Δilva* and *ΔmetA* cells in a mixed community in first batch culture. Blue line indicates biomass growth measured by optical density and orange line describes relative cells density. Gray band shows final relative density of 73% +/-5% from the experiment in figure 1.

We observed that establishment of the preferred ratio happens neither before nor after growth phase, but in fact occurs together with growth.

Preferred ratio is resilient towards perturbation by genetic manipulation

Next, we asked the question whether it was possible to manipulate the preferred ratio with a series of genetic manipulation in individual strains (Figure 3). First, we attempted to impose a metabolic burden on cells by high expression levels of *lacZ in trans*. Next, we expressed *metA* in individual strains as earlier observations by the Kost lab suggested the establishment of nanotubes among amino acid auxotroph *E.coli* strains to exchange cytosolic goods directly (Pande et al. 2015) . We assumed that expression of *metA* in the Δ*ilvA* strain would lead to a reduced dependency of the Δ*metA* strain and thus change the final observed relative cell density. Finally, we expressed a chloramphenicol resistance in individual strains and grew the community in presence of chloramphenicol. Communities were grown for two consecutive batch cultures for 36h (48h for communities containing chloramphenicol) each time and the relative cell densities were measured.

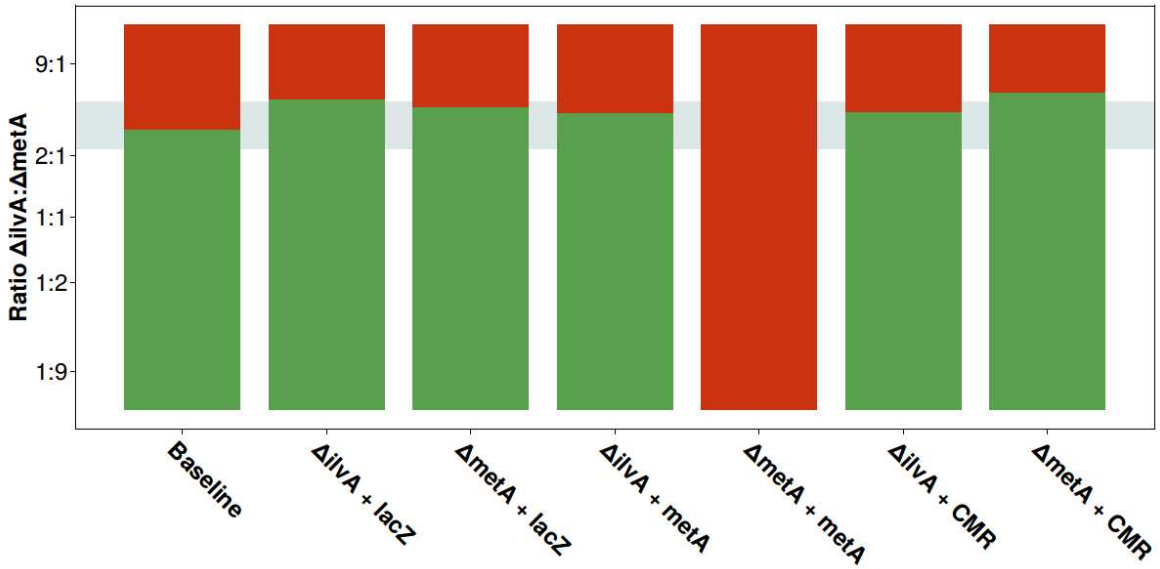


Figure 3: Relative cell density of *E.coli Δilva* and *ΔmetA* cells subjected to genetic manipulations. Red section describes relative abundance of the Δ*metA* strain and green section describes relative abundance of the Δ*ilvA* strain. Gray band shows final relative density of 73% +/-5% from the experiment in figure 1.

The observed preferred ratios did not differ from the communities without further genetic manipulations. Only expression of *metA* in the Δ*metA* strain leads to an out competition of the Δ*ilvA* strain, as the Δ*metA* strain is not dependent on its partner anymore. To our surprise, communities were able to grow in the presence of chloramphenicol despite that only one of the community partners is chloramphenicol resistant.

Preferred ratio is also established in interspecies communities between *E.coli* and V.natriegens

Our next aim was to create stable interspecies communities of *E.coli* and *V.natriegens* mutants. For this, we first generated auxotrophic mutants of *V.natriegens* with equivalent auxotrophies to their *E.coli* counterparts (Δ*ilvA* and Δ*metA*). When mixing these mutants in VN minimal media (Hoffart et al. 2017), they were able to grow, showing that the ability to form symbiotic relationships is not just an oddity among *E.coli* strain but in fact extends to other species such as *V.natriegens* as well (Figure 4). To find a suitable cocultivation media, we inoculated *V.natriegens* and *E.coli* wildtype strains individually into media compositions that were comprised of a mixture of M9- and VN minimal media in varying ratios. The final optical densities of these cultures were measured after 36h to estimate an optimal mixture for cocultivation (Table 2).

Table 2: Final Optical densities of individual wildtype strains grown in cocultivation media mixtures.

VN:M9 media	0:100	10:90	20:80	30:70	40:60	50:50	60:40	70:30	80:20	90:10	100:0
E.coli OD	2.37	2.85	3.20	3.40	3.60	3.80	3.70	3.60	3.30	3.00	3.00
V.natriegens OD	0.14	0.16	0.33	0.56	0.59	2.87	2.68	2.84	2.10	2.10	2.00

We observed that both strains were able to grow in a mixture of 70:30 of VN:M9 media and hence continued to use this formulation for our cocultivation experiments.

Next, we mixed the *E.coli* Δ*ilvA* and a *V.natriegens* Δ*metA* to create a *V.natriegens* intraspecies- as well as an interspecies community, respectively (Figure 4). As a cocultivation media the VN:M9 mixture established in the previous experiment (Table 2) was used. Preferred ratios were measured after 36h and continuously monitored to confirm their stability (time course data not shown).

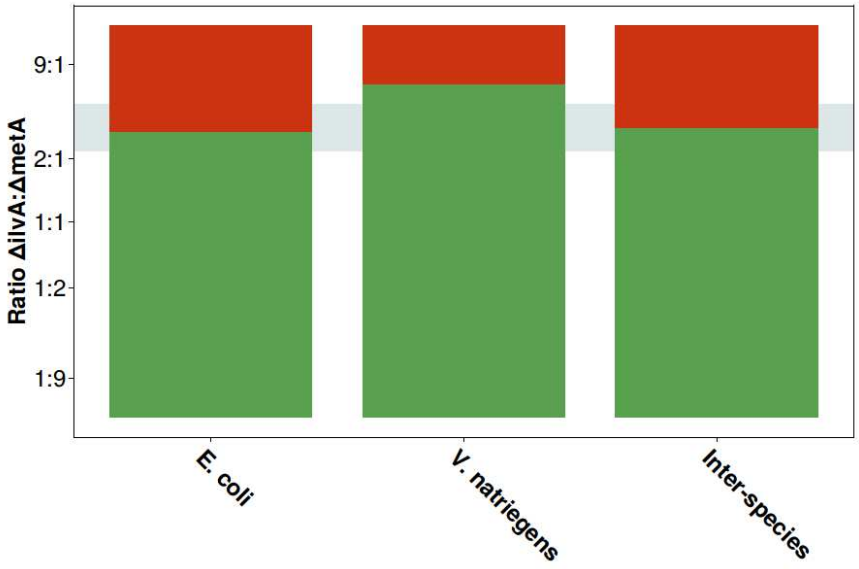


Figure 4: Relative cell density of *E.coli* and *V.natriegens Δilva* and *ΔmetA* intra- and interspecies communities. Red section describes relative abundance of the Δ*metA* strain and green section describes relative abundance of the Δ*ilvA* strain. Gray band shows final relative density of 73% +/- 5% from the experiment in figure 1. First bar equals to average relative density of 73% from figure 1.

We observed that both new communities also showed a stable preferred ratio. Moreover, these ratios were very close to the original *E.coli* community with respect to the *metA* or *ilvA* mutation.

Discussion

In this work, a synthetic symbiotic relationship has been established between *E.coli* and *V.natriegens* amino acid auxotroph mutants based on mutualistic dependency. The communities show a remarkable long-term stability, as they were able to maintain their relative cell density for a week of continuous batch cultivation. This stability is dynamically established during the growth of initiated communities. Raw flow cytometry data (not shown) suggest, that not a reduction of cell density (eg. cell death) but a constant cell density (stalling of growth) for the "overrepresented" partner and increasing cell density (growth) are contributing to the regulation of the relative density. Hence, modulation of growth rate, most likely tuned by the availability of the essential amino acid that each partner is lacking, seems to be the driving factor. The relatively long lag phase of ~16h (Figure 2) after inoculation suggests that potentially an extracellular pool of amino acids needs to be built up before exchange (more about the mode of exchange between partners will be discussed below).

The preferred ratio for the Δ*ilvA* and Δ*metA* combination of *E.coli* mutants was about 73% of Δ*ilvA* cells with a standard deviation of 5% (Figure 1). This ratio was independent of the initial ratio when communities were mixed and was very resilient towards perturbation by genetic manipulation. Interestingly, the overexpression of *lacZ* lead to no notable change in relative density (Figure 3). We expected that, given the burden on the metabolism, sink of biomass and energy, and resulting change in growth speed would decrease the relative cell density of the burdened partner. Apparently, the regulation that enforces the balance outweigh the impact this burden. The overexpression of *metA* in the Δ*ilvA* strain led to no notable change in relative cell density as well. *metA* expression alone is unlikely to impact cellular methionine expression (Gruzdev et al. 2023), as many cofactors are involved in methionine synthesis, but the previous discovery of intercellular nanotubes between *E.coli* auxotroph mutants suggests that *E.coli* auxotroph mutants may be able to transfer cytosolic goods directly (Pande et al. 2015). Hence, we assumed that MetA proteins could directly be transferred this way. Yet, we see no influence of *metA* overexpression

in this case. Additionally, microscopy imaging of the community members so far showed no structures similar to those reported by Pande *et al.* (Supplementary Figure 1A).

However, in our flow cytometry data we were able to observe minor cell populations which were “double labeled” with red and green color at fluorescence levels comparable to single red or single green cells (Supplementary Figure 1B). Potentially, these cells have exchanged cytosolic goods by nanotubes and exchanged GFP and mCherry proteins. Additionally, the “transfer” of the antibiotic resistance favors the theory of nanotubes: as chloramphenicol transferases act intracellularly, it is difficult to explain how a chloramphenicol resistance can be transferred without transfer of the chloramphenicol acetyl transferases as a protein (Shaw und Leslie 1991). The only potential explanation is that the chloramphenicol pool in the media is continuously acetylated during the lag phase, allowing both partners to grow without the need for a direct chloramphenicol resistance. The fact that these communities had an extended lag phase and required ~48h to grow (data not shown) favors this theory.

Given that *E.coli*, a member of the human gut microbial community , and *V.natriegens*, a halophile dwelling in salt marshes (Eagon 1962; Payne et al. 1961), are from two entirely different and separated biomes, they have unlikely undergone previous symbiotic interactions in their recent evolutionary past. Yet, both species managed to establish a symbiotic relationship with very little incentive given aside from the lack of a single amino acid. It is unlikely, that the lack of the given amino acid triggered a response that would favor this specific interaction, yet possibly there is a global trigger that favors exchange with the environment when growth is stalled due to missing metabolites. So far, to our knowledge no such global trigger has been discovered. Either way, it is remarkable how two species that can be considered “strangers” enter a symbiosis that easily. If there is such a thing as a general openness to metabolite exchange, the question remains how such random interactions deal with the emergence of “cheaters”. Perhaps, this form of exchange only applies to metabolites that are excreted as part of overflow metabolism and do not really impose a

disadvantage on their producer and this form of symbiosis just emerges by chance from two species who happen to mutually supply their needs. Many of the services supplied by guilds are not inherently altruistic; in fact, processes like denitrification can generate energy (Koike und Hattori 1975). Potentially, the symbiosis emerges from inherently selfish functions that just happen to be beneficial to the community, thus benefiting its growth, thus further benefiting the growth of the individual species. Seen from a game theory perspective in a system divided between a species serving such a function and the rest of the community merely being part of its environment, the species already gains a net benefit even if their services are not rewarded by the community/environment. Hence, this could explain how such functions are exerted by individual species without controlling for a direct compensation from community members. In conclusion, too little is known about the establishment and stability of such microbial consortia and more work on model systems like the one established in this manuscript is required to answer these questions.

Outlook

We have created stable inter- and intraspecies communities of *E.coli* and *V.natriegens* have shown that they are capable of balancing their relative cell density. Additionally, these communities have shown to be remarkably resistant towards perturbations.

To confirm the results of the *metA* overexpression experiments, it would be necessary to repeat it with a strain producing and expressing larger quantities of methionine: if this has an impact on the preferred ratio, this would confirm the hypothesis that the ratio correlates with the buildup of amino acid pools. This would also speak against the exchange of proteins over the nanotubes, as this should have had the same effect in this case (Pande et al. 2015).

More combinations of knockouts are required to understand the impact of a specific amino acid dependency on the preferred ratio. Mee *et al.* and Silver *et al.* reported successful combinations of mutualistic dependent *E.coli* strains but only published "fold changes of growth" based on qPCR results rather than actual relative cell densities (Wintermute und Silver 2010; Mee et al. 2014). The data sets of these authors are useful in guiding towards more feasible combinations of amino acid auxotrophies.

Mee *et al.* also constructed communities of more than two partners by mixing several auxotroph strains and cultivating this mix until not further partner drops out of this consortium. While this is a valid strategy and a great guide on potential combinations of amino acid auxotrophies for polycultures, the partners in this community are technically not "mutualistic dependent" as any partner more than two is not necessarily required to satisfy the amino acid requirements of each other. Hence, for the creation of a polyculture with n partners, n - 1 mutations are required per strain to create true mutualistic dependency. However, with a scaling number of members this rule probably does not suffice, as such extreme levels of specialization are also not observed in natural, complex communities (Goldford et al. 2018). The impact of an increased complexity of the community on the stability of the preferred ratio may hold interesting implications for the fundamental mechanisms behind it.

Author Contributions

Maurice Mager: Conceptualization, Methodology, Investigation, Writing

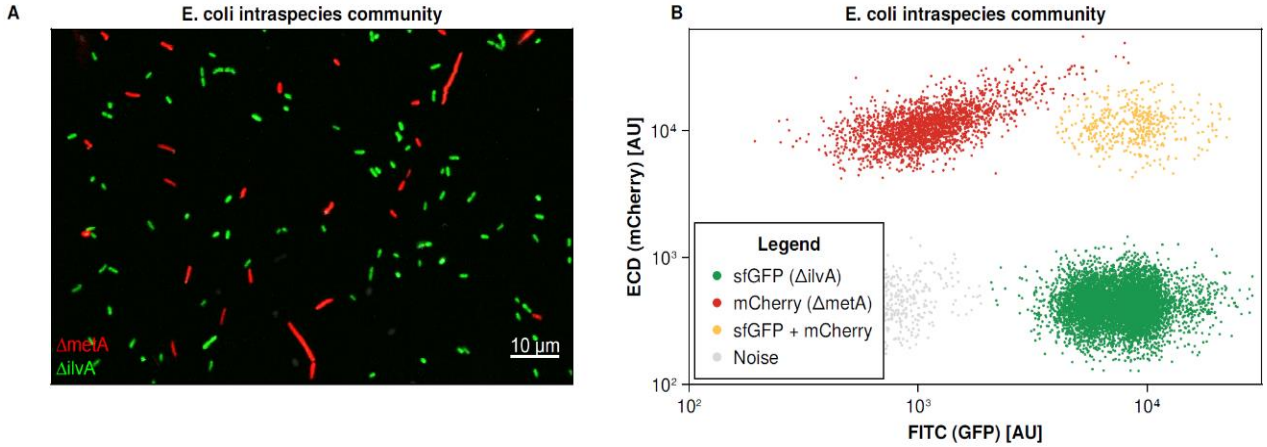
St. Elmo Wilkens: Conceptualization, Investigation, Data Analysis, Visualization, Project administration

Oliver Ebenhöh: Supervision, Funding Acquisition

Ilka Axmann: Supervision, Funding Acquisition

We thank Nina Schulten and Alexa Grebel for their support in this project.

Supplementary Material



Supplementary Figure 1: Fluorescence microscopy image (A) and Flow cytometry dot plot (B) of *E.coli* interspecies communities.

References

Burke, Catherine; Steinberg, Peter; Rusch, Doug; Kjelleberg, Staffan; Thomas, Torsten (2011): Bacterial community assembly based on functional genes rather than species. In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (34), S. 14288–14293. DOI: 10.1073/pnas.1101591108.

Eagon, R. G. (1962): Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. In: *Journal of bacteriology* 83 (4), S. 736–737. DOI: 10.1128/jb.83.4.736-737.1962.

Fang, Zhen; Jones, John A.; Zhou, Jingwen; Koffas, Mattheos A. G. (2018): Engineering Escherichia coli Co-Cultures for Production of Curcuminoids From Glucose. In: *Biotechnology journal* 13 (5), e1700576. DOI: 10.1002/biot.201700576.

Goldford, Joshua E.; Lu, Nanxi; Bajić, Djordje; Estrela, Sylvie; Tikhonov, Mikhail; Sanchez-Gorostiaga, Alicia et al. (2018): Emergent simplicity in microbial community assembly. In: *Science (New York, N.Y.)* 361 (6401), S. 469–474. DOI: 10.1126/science.aat1168.

Gruzdev, Nadya; Hacham, Yael; Haviv, Hadar; Stern, Inbar; Gabay, Matan; Bloch, Itai et al. (2023): Conversion of methionine biosynthesis in Escherichia coli from trans- to direct-sulfurylation enhances extracellular methionine levels. In: *Microbial cell factories* 22 (1), S. 151. DOI: 10.1186/s12934-023-02150-x.

Hoffart, Eugenia; Grenz, Sebastian; Lange, Julian; Nitschel, Robert; Müller, Felix; Schwentner, Andreas et al. (2017): High Substrate Uptake Rates Empower Vibrio natriegens as Production Host for Industrial Biotechnology. In: *Applied and environmental microbiology* 83 (22). DOI: 10.1128/AEM.01614-17.

Hosoda, Kazufumi; Suzuki, Shingo; Yamauchi, Yoshinori; Shiroguchi, Yasunori; Kashiwagi, Akiko; Ono, Naoaki et al. (2011): Cooperative adaptation to establishment of a synthetic bacterial mutualism. In: *PloS one* 6 (2), e17105. DOI: 10.1371/journal.pone.0017105.

Koike, I.; Hattori, A. (1975): Energy yield of denitrification: an estimate from growth yield in continuous cultures of Pseudomonas denitrificans under nitrate-, nitrite- and oxide-limited conditions. In: *Journal of general microbiology* 88 (1), S. 11–19. DOI: 10.1099/00221287-88-1-11.

Louca, Stilianos; Polz, Martin F.; Mazel, Florent; Albright, Michaeline B. N.; Huber, Julie A.; O'Connor, Mary I. et al. (2018): Function and functional redundancy in microbial systems. In: *Nat Ecol Evol* 2 (6), S. 936–943. DOI: 10.1038/s41559-018-0519-1.

Mee, Michael T.; Collins, James J.; Church, George M.; Wang, Harris H. (2014): Syntrophic exchange in synthetic microbial communities. In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (20), E2149-56. DOI: 10.1073/pnas.1405641111.

Pande, Samay; Shitut, Shraddha; Freund, Lisa; Westermann, Martin; Bertels, Felix; Colesie, Claudia et al. (2015): Metabolic cross-feeding via intercellular nanotubes among bacteria. In: *Nature communications* 6, S. 6238. DOI: 10.1038/ncomms7238.

Payne, W. J.; Eagon, R. G.; Williams, A. K. (1961): Some observations on the physiology of Pseudomonas natriegens nov. spec. In: *Antonie van Leeuwenhoek* 27 (1), S. 121–128. DOI: 10.1007/BF02538432.

Pistollato, Francesca; Sumalla Cano, Sandra; Elio, Iñaki; Masias Vergara, Manuel; Giampieri, Francesca; Battino, Maurizio (2016): Role of gut microbiota and nutrients in amyloid formation and pathogenesis of Alzheimer disease. In: *Nutr Rev* 74 (10), S. 624–634. DOI: 10.1093/nutrit/nuw023.

Schwille, Petra (2011): Bottom-up synthetic biology: engineering in a tinkerer's world. In: *Science (New York, N.Y.)* 333 (6047), S. 1252–1254. DOI: 10.1126/science.1211701.

Shaw, W. V.; Leslie, A. G. (1991): Chloramphenicol acetyltransferase. In: *Annual review of biophysics and biophysical chemistry* 20, S. 363–386. DOI: 10.1146/annurev.bb.20.060191.002051.

Teufel, Marc; Klein, Carlo A.; Mager, Maurice; Sobetzko, Patrick (2022): A multifunctional system for genome editing and large-scale interspecies gene transfer. In: *Nature communications* 13 (1), S. 3430. DOI: 10.1038/s41467-022-30843-1.

Wintermute, Edwin H.; Silver, Pamela A. (2010): Emergent cooperation in microbial metabolism. In: *Molecular systems biology* 6, S. 407. DOI: 10.1038/msb.2010.66.

Zhang, Haoran; Pereira, Brian; Li, Zhengjun; Stephanopoulos, Gregory (2015): Engineering Escherichia coli coculture systems for the production of biochemical products. In: *Proceedings of the National Academy of Sciences of the United States of America* 112 (27), S. 8266–8271. DOI: 10.1073/pnas.1506781112.

Discussion

Aim of this thesis was to support the advent of microbiology towards the analysis of complex microbial networks. The projects focused on the establishment of existing microbial model systems (such as the native colonizers of *Hydra vulgaris*), improving the handling of promising candidates for synthetic consortia (by establishing stricter and transparent handling and investigation routines for cyanobacteria) and ultimately by creating a novel synthetic microbial symbiosis between *Escherichia coli* and *Vibrio natriegens* from scratch. By providing tractable microbial systems for other investigators, this thesis aims to pave the way for further research on complex microbial networks. The following discussion lays out how this aim was pursued by the works summarized in this thesis.

We successfully established constitutive promoter expression systems for *Curvibacter sp.* AEP1-3

In the first manuscript, an oligonucleotide library of potential promoter systems was used to find suitable expression systems for *Curvibacter sp.* AEP1-3. We discovered 25 new promoters that show notable expression levels. From these, three promoters were highlighted due to the variety of use cases they cover. The promoter dubbed CPL0025 regulates the gene AEP_RS11205, annotated as an AHL synthase. CPL0025 was the strongest discovered promoters, surpassing the previous strongest expression system using the J23100 promoter from the Anderson Collection and the RBS* synthetic ribosomal binding site (used as reference in the manuscript). CPL0025 shows a classic activity dynamic for an *in trans* expression system, with a maximum activity during the extended stationary growth phase. CPL0025 may be used when a high activity of a gene of interest is required such as the overexpression of a protein of interest.

For a more growth phase independent expression system we recommend the use of CPL0022, the upstream region of the *dnaK* gene in *Curvibacter*. As DnaK acts as a molecular chaperone that is constantly active and upregulated under stress conditions, this promoter sequence shows a steady activity level over the course of all growth

phases. CPL0022 may be used in cases where excessive expression of a gene of interest during the stationary phase may be toxic to *Curvibacter*.

These promoters add to the repertoire of expression systems for *Curvibacter*. Previously, only the tac promoter has been used (Wein et al. 2018) as well as a J23100-RBS* combination that has been developed as part of this thesis. Additionally, this is the first time that expression systems have been comparatively characterized in this species. Such foundational work is essential for the well guided design of more complicated genetic circuits. Similar works have been performed for *Escherichia coli, Vibrio natriegens* (Stukenberg *et al.* 2021) and *Synechocystis* PCC 6803 (Vasudevan *et al.* 2019) and have been well received by the scientific communities as anchor points for genetic designs.

Native systems are a pragmatic first step to establish expression systems. They are likely to be functional as they already have in their native context, however the decontextualization (that is in this case: using the expression system *in trans* instead of in their native genomic location) may alter their performance. In our manuscript we showed that from 500 native candidates, 25 candidates were discovered that showed activity out of their native context. While the high level of duplicate discoveries in the high activity area of promoters suggests statistical saturation of positive hits, the high diversity of positive hits in the low activity section of promoters suggests that with more extensive cell sorting more promoter sequences are likely to appear (raw cell sorting and sequencing data, unpublished, full list of discovered promoters can be found in the Supplementary Files of Manuscript 1). Hence the real number of active promoters may be well above 50 but further sorting effort would be required to find these. However, even when considering 50 potentially active promoter sequences, this would still mean that around 90% of the promoter sequences were unable to maintain their activity after decontextualization. 10% positive hits can still be considered a significant amount when compared to completely randomized libraries, as the GeneEE library for example yields around 3-4% active sequences (Lale *et al.* 2022). However, it remains a surprise that so many native sequences lose their activity.

Sources of errors for the reproducibility of cyanobacterial experiments remain elusive but progress towards standardization has been made

In this second manuscript, the reproducibility of a simple promoter strength assay with *Synechocystis* PCC 6803 across several laboratories in Europe was assessed. Idea and form of the Cyano Interlab study were inspired by the ongoing iGEM Interlab study. In this study hosted by the iGEM competition, members supplied data of a simple fluorescence measurement experiment to be compared for the statistical deviation between their results. The study has been hosted annually for a few years now (Beal *et al.* 2021) and is continuously being refined alongside with the insights gained (Beal *et al.* 2020; Beal *et al.* 2022). While the iGEM Interlab study focuses on experiments around *E.coli*, we wanted to have a closer look at the reproducibility of experiments on cyanobacteria and decided to host a similar study around the cyanobacterium *Synechocystis* PCC 6803.

Prior to launching the Interlab study, an extensive study design phase was launched to eliminate as many protocols related error sources of our experimental procedure. Publication of these protocols were proclaimed by us as “recommended standards” for working with cyanobacteria and we hope that the scientific community incorporates these standards to improve the comparability across the field.

One notable standard was the establishment of a standardized cryo-conservation protocol that allows for proper strain-keeping, a standard that is widespread for most other model organisms in microbiology. Without a good long-term strain keeping strategy, laboratory evolution is bound to alter laboratory strains. This has already been reported for *Synechocystis* PCC 6803, where the Hess lab retraced the origins of multiple “lineages” of this widely used laboratory strain in an extensive effort involving literature study and resequencing of strains (Trautmann *et al.* 2012). A second relevant standard was the proposition of a standard BG11. Being used as the most common minimal media for several cyanobacterial species, it was previously commonly believed that the BG11 formulation was identical globally. However, when participants of the

study compared their own BG11 recipes, differences in preparation and final concentrations as well as addition or omission of individual reagents became evident. Hence, a standardized BG11 recipe was proposed to further improve comparability between results of different laboratories. For this recipe, a mild HEPES buffer was introduced to establish consistent initial conditions. Given the propensity of cyanobacterial cultures to elevate pH levels during growth, this buffer was considered inconsequential for extended cultivation periods.

Despite these precautions, among several others, the standard deviation of results among participants of the Interlab study was alarmingly high. Additionally, due to organizational complications we were unable to assess the effect our precautions had on the standard deviation. It is nonetheless evident, that additional error sources have not been covered that lead to the lack of reproducibility in this study. This study is a stern reminder that an acceptable level of reproducibility in the field of cyanobacterial research has not been reached and that for the sake of its credibility and accessibility to investigators outside of the cyanobacteria field more effort on this issue is crucial

The underlying mechanics of consortia homeostasis remain unknown but simple dependencies are sufficient to create very robust cell to cell abundance ratios in synthetic consortia

In this third manuscript, stable inter- and intra-species communities of *Escherichia coli* and *Vibrio natriegens* mutants were created. Species were made codependent by knocking out genes involved in the synthesis of essential amino acids. Individually, these mutant strains were unable to grow in minimal media but continued to grow when a mutant lacking a different amino acid synthesis gene was added to the same culture. Growth as well as relative cell to cell abundance ratio was measured over extended growth periods to observe homeostasis dynamics in this community. The synthetic consortia were able to restore strain specific cell to cell abundance ratios independent of initial ratios and maintain this ratio over the course of an entire week.

The specific knockouts were chosen based on the works of two groups, Silver *et al.* and Wang *et al.* (Mee *et al.* 2014; Wintermute und Silver 2010), who independently created similar synthetic interactions for *E.coli* that were also based on auxotrophy dependencies. Wang *et al.* created a total of 14 different amino acid auxotroph strain and tested each possible combination for potential symbiotic interaction. They discovered that only a minority of combinations was able to exchange metabolites and grow and the knockout combinations tested in our manuscript agree with the data from Wang *et al.*. Both groups reported successful growth as "change in growth rate" determined by qPCR, in our manuscript we were able to report actual cell to cell abundance ratios with the use of flow cytometry. This revealed a remarkable stability of the composition in these consortia that has not been reported anywhere prior. Our observations of the growing communities and their cell-to-cell abundance ratios over time revealed that the rebalancing of the ratio happens simultaneous with growth. Additionally, flow cytometry raw data (unpublished) indicates that the absolute cell density of the culture does not decline for overrepresented partners but that instead faster growth of the underrepresented partner leads to restoration of the preferred ratio. Therefore, we assume modulation of growth rates as the main dynamic of rebalancing the ratio, but a normalization of absolute cell densities for flow cytometry needs to be applied before this hypothesis can be confirmed with full certainty.

The preferred ratio of a given synthetic community changes with the type of amino acid auxotrophy of each partner. The ratios of the combinations we tested for interspecies *E.coli* combinations match closely to the growth rate fold changes reported by Wang *et al.*, as far as they can be compared. However, the "rule set" behind these changes in the ratio remains elusive. A few potential factors for rules have been proposed by us:

1. The synthesis cost of a given amino acid. If the energy cost to synthesize an amino acid is high, this may impact the growth rates of both partners and hence the final abundance ratio. The lack of production of this given amino acid may also act as a benefit from an energy cost perspective, as shutting down the flux over a pathway means that energy is conserved.

2. The cellular production- and excretion rate of a given amino acid. When viewed from a game theory point of view, both mutants in our consortia may be considered as partners in a “altruism” dilemma: while they need their partner to survive on a global level, each individual cell can gain a growth advantage by favoring faster growth over the symbiotic interaction with little expected penalty. As in our system both partners interact only for a couple of generations, evolutionary adaptations can be neglected for this theory. Instead, cellular “over”-production- and excretion rates are likely to be an artifact of internal regulation rather than an active form of gambling inside this highly artificial symbiosis. Hence, we assume that certain amino acids are excreted at constant rates, dictating how much the growth of their dependent partner is favored and vice versa.
3. Cellular requirements of a given amino acid. Adding to point 2, not only the given excretion, but also the cellular requirements for a given amino acid may play into determining the final ratio for the same reasons as outlined above.

While the synthesis cost (calculated as the amount of required phosphate bounds per synthesized molecules in a metabolic pathway) and the amino acid requirements (approximated by total amino acid content in the *E.coli* proteome) are publicly available, very little is known about excretion rates of amino acids and these values would need to be verified with data from our consortia. Unfortunately, we were not able to correlate neither synthesis cost nor amino acid requirements to our preferred ratios. In a second series of experiments, we attempted to manipulate preferred ratios by manipulation of the consortia on a genetic level. Neither additional metabolic burdens (*lacZ* overexpression in individual symbiotic partners) nor changes in amino acid synthesis (overexpression of the gene lacking in their symbiotic counterpart) had any impact on the preferred ratio whatsoever. Hence, to date we do not know of any of the rules that determine these ratios, but we do know that the strong forces driving these ratios are very resilient towards external (manipulation of initial cell to cell ratio) and internal (genetic manipulation) perturbation and should pose an interesting field for future research.

Closing thoughts

Aim of this thesis was to support the advent of microbiology towards the analysis of complex microbial networks. For this we developed novel expression systems for *Curvibacter sp.* AEP1-3, a native colonizer of *Hydra vulgaris*. We employed a large-scale bioinformatics guided approach to test 500 candidate sequences for their ability to effectively drive gene expression in this species and found 25 positive sequences. The methodology is easily adaptable towards other species as well, thus serving as a blue print to discover novel expression systems in other less established model organisms as well. In this project we learned that most promoter sequences on the hosts genome were not utilizable as expression systems and that their behavior in an *in trans* context differed from their genomic behavior. This work aids in better understanding the intimate cross talk between *Curvibacter* and its host by providing new tools to genetically manipulate this species.

Additionally, we helped in standardizing the cultivation- and experimental conditions when working with *Synechocystis* PCC 6803 and inferred the expected reproducibility of studies when using this organism in an interlaboratory study. We found severe reproducibility issues and discussed solutions to mitigate these challenges going further. This should aid future researchers outside of the field working with this organism and thus allow for new collaborations leading to novel synthetic communities involving photosynthetic organisms.

Last, we created a synthetic microbial consortium that is easy to replicate, monitor and engineer. The codependent mutualistic relationship between two auxotroph mutants of *Escherichia coli* and *Vibrio natriegens* shows remarkable stability in terms of the relative cell density of both species. On this excellent platform we have shown for the first time, that relative cell densities in synthetic communities can be highly robust towards perturbations. We learned that, despite our extensive effort to manipulate these relative cell densities, these communities remained unchanged, showing that our understanding of their balancing mechanisms remains lacking.

Outlook

Towards a fully controllable high complexity model system of *Hydra vulgaris* and its natural colonizers

A next step in this project is the search for inducible expression systems. For this, the experimental work needs to be slightly modified: starting from a library of *Curvibacter* cells carrying a variety of promoter sequence reporter plasmids, instead of sorting the cells directly for active promoters (which would be constitutive expression systems) cells are instead sorted for low to undetectable promoter activity. This filters the library for cells containing promoter sequences which are inactive under "default" growth conditions. For this, it is best to pick a liquid culture in the early stationary phase to reduce the amount of false positive candidates, given the results from our time course assay. Next, the sorted cells are pooled and recovered. In a final step, this cell culture can be inoculated in media containing a "inducing agent" of choice. These are chemicals that are predicted to likely influence the expression level on any of the remaining promoter sequences. The choice of chemicals can be informed by the annotation of the genes where promoter sequences are taken from (eg. the promoter of a gene involved in rhamnose metabolism may respond positively to the presence of rhamnose) but also from the lifestyle of the species (eg. for *Curvibacter* the presence of *Hydra* associated molecules may influence the expression level of individual genes). After treatment with a candidate inducing agent, cells are sorted again, but this time for cells with high promoter sequence activity. Individual cells can now be analyzed similarly to our manuscript and tested for their response to their respective inducing agent.

A lot of recent research efforts in the field of microbial communities is directed towards hyper complex systems such as the human gut microbiome. Encompassing many thousand species, such a multidimensional system is incomprehensible without the aid of omic technologies and big data analysis. While technological advances like artificial

intelligence may at some point allow to translate hyper complex omic data into concepts which are graspable for humans, this may either take a while and/or will require a better understanding of fundamental mechanisms to be fully decipherable. Hence, the need for simple, easy to understand microbial community systems becomes evident.

Hydra vulgaris and its associated microbiome has the potential to become such a system: It is based on a simple eukaryotic host (a simple basal metazoan) and can host a "conventionalized" (minimalized) microbiome (Fraune *et al.* 2015). Communication between the host and the microbiome as well as inside the microbiome has already been confirmed. A fully modifiable *Curvibacter sp.* AEP1-3 strain as one of the main colonizers of *Hydra vulgaris* would be a big first step towards a controllable microbiome. Other species living on *Hydra vulgaris* such as *Undibacterium sp., Duganella sp.* and *Acidovorax sp.* also need to be made accessible to allow for extensive manipulation of the network. Fluorophore tracking of multiple community members simultaneously would open new ways to observe recolonization dynamics in hourly resolution. The manipulation of all partners allows investigator to find all nodes in the complex communication network between individual members of the community as well as the host. Manipulation of communities in the future via DART may be approaches that become suitable once the model system has reached a critical stage of complexity (Ha und Devkota 2022).

Towards more reproducibility in cyanobacterial research

While the Cyano Interlab Study has contributed towards proposing more standardized protocols for cyanobacterial research and raised attention for a lack of reproducibility, it could not pin down the practical causes of this issue. A repetition of the Interlab Study may have to focus on additional factors. As the current instance of the study focused on experimental procedures and the cyanobacterial cells, in the future the focus could be directed towards the environment: for example, standardization of light quality was

proposed in the study but not implemented. Preparation of the media was standardized but composition of ingredients may vary when purchased from different suppliers. Water quality differs between locations and this may impact cells even after filtration. Residuals on glassware have been neglected even though glassware is regularly subject to cleaning chemicals and unfiltered water containing trace metals that bind to glass.

Once sufficient reproducibility is established, individual factors can be changed to observe their effect. If the study is for example once performed with standardized chemicals and once with local equivalent, the impact of that can be seen in the change of reproducibility. In this way, insights can be generated that lead to direct advice for better laboratory practices.

Towards a better understanding of the fundamental mechanics and dynamics of microbial communities

More advancements in the experimental as well as the modeling of microbial communities are required to advance our understanding of the fundamental principles behind microbial communities. Our manuscript has shown that the implementation of novel techniques such as flow cytometry can give new insights into existing experimental setups: without cell counting, the remarkable stability of auxotrophy based symbiotic interaction would not have been as apparent.

Kost *et al.* proposed the establishment of intercellular nanotubes across partners as a means to exchange cytosolic goods directly in a symbiotic relationship of *E.coli* mutants very similar to our setup (Pande *et al.* 2015). So far, we were neither able to observe similar structures in our setup nor to prove their existence experimentally, but further research needs to be made.

Given the strong resilience of communities to metabolic burden and perturbation, they can be leveraged to produce complex chemical compounds using division of labor (Roell *et al.* 2019). Production can be spread across partners by dividing genes for metabolic pathways and exchanging metabolites. Cocultures of *E.coli* mutants were

engineered to produce 4-hydroxybenzoic acid, bisdemethoxycurcumin and anthocyanin (Fang *et al.* 2018; Jones *et al.* 2017; Zhang *et al.* 2015). Nutritional syntrophy (one species feeding on the byproduct of another) is employed in a range of these works to avoid substrate competition. Roell *et al.* name a few key steps towards improving production of compounds via consortia (Roell *et al.* 2019) . Identification of optimal inoculation ratios and ideal production times are crucial to reach high production rates. With a better understanding of homeostasis processes, eventually these steps may be circumvented. Adequate export- and import strategies need to be developed that do not pose an additional burden on their host. Potentially, nanotubes as shown by Kost *et al.* may be a solution for interspecies consortia (Pande *et al.* 2015). Similar to monoculture processes, “cheaters” are problematic for production in consortia as well. Roell *et al.* propose the use of biosensors that actively favor compound production may aid in population control, as for example naringenin-responsive biosensors have been developed for *E.coli* cocultures (Xiu *et al.* 2017).

References

Literaturverzeichnis

Beal, Jacob; Baldwin, Geoff S.; Farny, Natalie G.; Gershater, Markus; Haddock-Angelli, Traci; Buckley-Taylor, Russell et al. (2021): Comparative analysis of three studies measuring fluorescence from engineered bacterial genetic constructs. In: *PloS one* 16 (6), e0252263. DOI: 10.1371/journal.pone.0252263.

Beal, Jacob; Farny, Natalie G.; Haddock-Angelli, Traci; Selvarajah, Vinoo; Baldwin, Geoff S.; Buckley-Taylor, Russell et al. (2020): Robust estimation of bacterial cell count from optical density. In: *Communications biology* 3 (1), S. 512. DOI: 10.1038/s42003-020-01127-5.

Beal, Jacob; Telmer, Cheryl A.; Vignoni, Alejandro; Boada, Yadira; Baldwin, Geoff S.; Hallett, Liam et al. (2022): Multicolor plate reader fluorescence calibration. In: *Synthetic biology (Oxford, England)* 7 (1), ysac010. DOI: 10.1093/synbio/ysac010.

Bolen, D. W.; Baskakov, I. V. (2001): The osmophobic effect: natural selection of a thermodynamic force in protein folding. In: *Journal of molecular biology* 310 (5), S. 955–963. DOI: 10.1006/jmbi.2001.4819.

Bosch, Thomas C. G.; Augustin, René; Anton-Erxleben, Friederike; Fraune, Sebastian; Hemmrich, Georg; Zill, Holger et al. (2009): Uncovering the evolutionary history of innate immunity: the simple metazoan Hydra uses epithelial cells for host defence. In: *Developmental and comparative immunology* 33 (4), S. 559–569. DOI: 10.1016/j.dci.2008.10.004.

Bremer, Erhard; Krämer, Reinhard (2019): Responses of Microorganisms to Osmotic Stress. In: *Annual review of microbiology* 73, S. 313–334. DOI: 10.1146/annurev-micro-020518-115504.

Burke, Catherine; Steinberg, Peter; Rusch, Doug; Kjelleberg, Staffan; Thomas, Torsten (2011): Bacterial community assembly based on functional genes rather than species. In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (34), S. 14288–14293. DOI: 10.1073/pnas.1101591108.

Deines, Peter; Hammerschmidt, Katrin; Bosch, Thomas C. G. (2020): Microbial Species Coexistence Depends on the Host Environment. In: *mBio* 11 (4). DOI: 10.1128/mBio.00807-20.

Dinnbier, U.; Limpinsel, E.; Schmid, R.; Bakker, E. P. (1988): Transient accumulation of potassium glutamate and its replacement by trehalose during adaptation of growing cells of Escherichia coli K-12 to elevated sodium chloride concentrations. In: *Archives of microbiology* 150 (4), S. 348–357. DOI: 10.1007/BF00408306.

Fang, Zhen; Jones, John A.; Zhou, Jingwen; Koffas, Mattheos A. G. (2018): Engineering Escherichia coli Co-Cultures for Production of Curcuminoids From Glucose. In: *Biotechnology journal* 13 (5), e1700576. DOI: 10.1002/biot.201700576.

Franzenburg, Sören; Fraune, Sebastian; Altrock, Philipp M.; Künzel, Sven; Baines, John F.; Traulsen, Arne; Bosch, Thomas C. G. (2013): Bacterial colonization of Hydra hatchlings follows a robust temporal pattern. In: *The ISME journal* 7 (4), S. 781–790. DOI: 10.1038/ismej.2012.156.

Franzenburg, Sören; Fraune, Sebastian; Künzel, Sven; Baines, John F.; Domazet-Loso, Tomislav; Bosch, Thomas C. G. (2012): MyD88-deficient Hydra reveal an ancient function of TLR signaling in sensing bacterial colonizers. In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (47), S. 19374–19379. DOI: 10.1073/pnas.1213110109.

Fraune, Sebastian; Abe, Yuichi; Bosch, Thomas C. G. (2009): Disturbing epithelial homeostasis in the metazoan Hydra leads to drastic changes in associated microbiota. In: *Environmental microbiology* 11 (9), S. 2361–2369. DOI: 10.1111/j.1462-2920.2009.01963.x.

Fraune, Sebastian; Anton-Erxleben, Friederike; Augustin, René; Franzenburg, Sören; Knop, Mirjam; Schröder, Katja et al. (2015): Bacteria-bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance. In: *The ISME journal* 9 (7), S. 1543–1556. DOI: 10.1038/ismej.2014.239.

Germann, Anna T.; Nakielski, Andreas; Dietsch, Maximilian; Petzel, Tim; Moser, Daniel; Triesch, Sebastian et al. (2023): A systematic overexpression approach reveals native targets to increase squalene production in Synechocystis sp. PCC 6803. In: *Frontiers in plant science* 14, S. 1024981. DOI: 10.3389/fpls.2023.1024981.

Golden, James W.; Yoon, Ho-Sung (2003): Heterocyst development in Anabaena. In: *Current Opinion in Microbiology* 6 (6), S. 557–563. DOI: 10.1016/j.mib.2003.10.004.

González, Sebasthian Santiago; Ad, Omer; Shah, Bhavana; Zhang, Zhongqi; Zhang, Xizi; Chatterjee, Abhishek; Schepartz, Alanna (2021): Genetic Code Expansion in the Engineered Organism Vmax X2: High Yield and Exceptional Fidelity. In: *ACS central science* 7 (9), S. 1500–1507. DOI: 10.1021/acscentsci.1c00499.

Grammann, Katrin; Volke, Angela; Kunte, Hans Jörg (2002): New type of osmoregulated solute transporter identified in halophilic members of the bacteria domain: TRAP transporter TeaABC mediates uptake of ectoine and hydroxyectoine in Halomonas elongata DSM 2581(T). In: *Journal of bacteriology* 184 (11), S. 3078–3085. DOI: 10.1128/JB.184.11.3078-3085.2002.

Gregory, Gwendolyn J.; Boyd, E. Fidelma (2021): Stressed out: Bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from Vibrionaceae. In: *Computational and structural biotechnology journal* 19, S. 1014–1027. DOI: 10.1016/j.csbj.2021.01.030.

Ha, Connie W. Y.; Devkota, Suzanne (2022): DART takes aim at community editing. In: *Nature microbiology* 7 (1), S. 8–9. DOI: 10.1038/s41564-021-01017-4.

Hemmrich, Georg; Anokhin, Boris; Zacharias, Helmut; Bosch, Thomas C. G. (2007): Molecular phylogenetics in Hydra, a classical model in evolutionary developmental biology. In: *Molecular Phylogenetics and Evolution* 44 (1), S. 281–290. DOI: 10.1016/j.ympev.2006.10.031.

Hoff, Josef; Daniel, Benjamin; Stukenberg, Daniel; Thuronyi, B. W.; Waldminghaus, Torsten; Fritz, Georg (2020): Vibrio natriegens: an ultrafast-growing marine bacterium as emerging synthetic biology chassis. In: *Environmental microbiology* 22 (10), S. 4394–4408. DOI: 10.1111/1462-2920.15128.

Jiang, Chunmei; Li, Guangning; Huang, Pengru; Liu, Zhou; Zhao, Bin (2017): The Gut Microbiota and Alzheimer's Disease. In: *Journal of Alzheimer's disease : JAD* 58 (1), S. 1–15. DOI: 10.3233/JAD-161141.

Jinek, Martin; Chylinski, Krzysztof; Fonfara, Ines; Hauer, Michael; Doudna, Jennifer A.; Charpentier, Emmanuelle (2012): A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. In: *Science (New York, N.Y.)* 337 (6096), S. 816–821. DOI: 10.1126/science.1225829.

Jones, J. Andrew; Vernacchio, Victoria R.; Collins, Shannon M.; Shirke, Abhijit N.; Xiu, Yu; Englaender, Jacob A. et al. (2017): Complete Biosynthesis of Anthocyanins Using E. coli Polycultures. In: *mBio* 8 (3). DOI: 10.1128/mBio.00621-17.

Köbler, Christin; Schmelling, Nicolas M.; Pawlowski, Alice; Spät, Philipp; Scheurer, Nina M.; Sebastian, Kim et al. (2021): Two circadian oscillators in one cyanobacterium.

Lale, Rahmi; Tietze, Lisa; Fages-Lartaud, Maxime; Nesje, Jenny; Onsager, Ingerid; Engelhardt, Kerstin et al. (2022): A universal approach to gene expression engineering. In: *Synthetic biology (Oxford, England)* 7 (1), ysac017. DOI: 10.1093/synbio/ysac017.

Lange, Janina; Fraune, Sebastian; Bosch, Thomas C.G.; Lachnit, Tim (2019): The neglected part of the microbiome: Prophage TJ1 regulates the bacterial community of the metaorganism Hydra. In: *bioRxiv*, S. 607325. DOI: 10.1101/607325.

Larson, Matthew H.; Gilbert, Luke A.; Wang, Xiaowo; Lim, Wendell A.; Weissman, Jonathan S.; Qi, Lei S. (2013): CRISPR interference (CRISPRi) for sequence-specific control of gene expression. In: *Nat Protoc* 8 (11), S. 2180–2196. DOI: 10.1038/nprot.2013.132.

Lassen, Lærke Münter; Nielsen, Agnieszka Zygadlo; Olsen, Carl Erik; Bialek, Wojciech; Jensen, Kenneth; Møller, Birger Lindberg; Jensen, Poul Erik (2014): Anchoring a plant cytochrome P450 via PsaM to the thylakoids in Synechococcus sp. PCC 7002: evidence for light-driven biosynthesis. In: *PLOS ONE* 9 (7), e102184. DOI: 10.1371/journal.pone.0102184.

Lima, Steeve; Matinha-Cardoso, Jorge; Tamagnini, Paula; Oliveira, Paulo (2020): Extracellular Vesicles: An Overlooked Secretion System in Cyanobacteria. In: *Life* 10 (8), S. 129. DOI: 10.3390/life10080129.

Lommel, Mark; Tursch, Anja; Rustarazo-Calvo, Laura; Trageser, Benjamin; Holstein, Thomas W. (2017): Genetic knockdown and knockout approaches in Hydra. In: *bioRxiv*, S. 230300. DOI: 10.1101/230300.

Louca, Stilianos; Polz, Martin F.; Mazel, Florent; Albright, Michaeline B. N.; Huber, Julie A.; O'Connor, Mary I. et al. (2018): Function and functional redundancy in microbial systems. In: *Nature ecology & evolution* 2 (6), S. 936–943. DOI: 10.1038/s41559-018-0519-1.

M. Baker (2016): 1,500 scientists lift the lid on reproducibility. Online verfügbar unter https://philpapers.org/rec/BAKSL-2.

Ma, Yanhe; Galinski, Erwin A.; Grant, William D.; Oren, Aharon; Ventosa, Antonio (2010): Halophiles 2010: life in saline environments. In: *Applied and environmental microbiology* 76 (21), S. 6971–6981. DOI: 10.1128/AEM.01868-10.

Mee, Michael T.; Collins, James J.; Church, George M.; Wang, Harris H. (2014): Syntrophic exchange in synthetic microbial communities. In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (20), E2149-56. DOI: 10.1073/pnas.1405641111.

Osburn, Ernest D.; Badgley, Brian D.; Strahm, Brian D.; Aylward, Frank O.; Barrett, J. E. (2021): Emergent properties of microbial communities drive accelerated biogeochemical cycling in disturbed temperate forests. In: *Ecology* 102 (12), e03553. DOI: 10.1002/ecy.3553.

Pande, Samay; Shitut, Shraddha; Freund, Lisa; Westermann, Martin; Bertels, Felix; Colesie, Claudia et al. (2015): Metabolic cross-feeding via intercellular nanotubes among bacteria. In: *Nature communications* 6 (1), S. 6238. DOI: 10.1038/ncomms7238.

Paul, Sandip; Bag, Sumit K.; Das, Sabyasachi; Harvill, Eric T.; Dutta, Chitra (2008): Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. In: *Genome Biol* 9 (4), R70. DOI: 10.1186/gb-2008-9-4-r70.

Pfeifer, Eugen; Michniewski, Slawomir; Gätgens, Cornelia; Münch, Eugenia; Müller, Felix; Polen, Tino et al. (2019): Generation of a Prophage-Free Variant of the Fast-Growing Bacterium Vibrio natriegens. In: *Applied and environmental microbiology* 85 (17). DOI: 10.1128/AEM.00853-19.

Pietschke, Cleo; Treitz, Christian; Forêt, Sylvain; Schultze, Annika; Künzel, Sven; Tholey, Andreas et al. (2017): Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in

bacterial symbionts. In: *Proceedings of the National Academy of Sciences of the United States of America* 114 (40), E8488-E8497. DOI: 10.1073/pnas.1706879114.

Roell, Garrett W.; Zha, Jian; Carr, Rhiannon R.; Koffas, Mattheos A.; Fong, Stephen S.; Tang, Yinjie J. (2019): Engineering microbial consortia by division of labor. In: *Microbial cell factories* 18 (1), S. 35. DOI: 10.1186/s12934-019-1083-3.

Siebert, Stefan; Farrell, Jeffrey A.; Cazet, Jack F.; Abeykoon, Yashodara; Primack, Abby S.; Schnitzler, Christine E.; Juliano, Celina E. (2019): Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. In: *Science (New York, N.Y.)* 365 (6451). DOI: 10.1126/science.aav9314.

Stukenberg, Daniel; Hensel, Tobias; Hoff, Josef; Daniel, Benjamin; Inckemann, René; Tedeschi, Jamie N. et al. (2021): The Marburg Collection: A Golden Gate DNA Assembly Framework for Synthetic Biology Applications in Vibrio natriegens. In: *ACS synthetic biology* 10 (8), S. 1904–1919. DOI: 10.1021/acssynbio.1c00126.

Trautmann, Danika; Voss, Björn; Wilde, Annegret; Al-Babili, Salim; Hess, Wolfgang R. (2012): Microevolution in cyanobacteria: re-sequencing a motile substrain of Synechocystis sp. PCC 6803. In: *DNA research : an international journal for rapid publication of reports on genes and genomes* 19 (6), S. 435–448. DOI: 10.1093/dnares/dss024.

Vasudevan, Ravendran; Gale, Grant A. R.; Schiavon, Alejandra A.; Puzorjov, Anton; Malin, John; Gillespie, Michael D. et al. (2019): CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax. In: *Plant physiology* 180 (1), S. 39–55. DOI: 10.1104/pp.18.01401.

Ward, D. M.; Ferris, M. J.; Nold, S. C.; Bateson, M. M. (1998): A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. In: *Microbiology and molecular biology reviews : MMBR* 62 (4), S. 1353–1370. DOI: 10.1128/MMBR.62.4.1353-1370.1998.

Wein, Tanita; Dagan, Tal; Fraune, Sebastian; Bosch, Thomas C. G.; Reusch, Thorsten B. H.; Hülter, Nils F. (2018): Carrying Capacity and Colonization Dynamics of Curvibacter in the Hydra Host Habitat. In: *Frontiers in microbiology* 9, S. 443. DOI: 10.3389/fmicb.2018.00443.

Wintermute, Edwin H.; Silver, Pamela A. (2010): Emergent cooperation in microbial metabolism. In: *Molecular systems biology* 6, S. 407. DOI: 10.1038/msb.2010.66.

Xiu, Yu; Jang, Sungho; Jones, J. Andrew; Zill, Nicholas A.; Linhardt, Robert J.; Yuan, Qipeng et al. (2017): Naringenin-responsive riboswitch-based fluorescent biosensor module for Escherichia coli co-cultures. In: *Biotechnology and bioengineering* 114 (10), S. 2235–2244. DOI: 10.1002/bit.26340.

Zhang, Haoran; Pereira, Brian; Li, Zhengjun; Stephanopoulos, Gregory (2015): Engineering Escherichia coli coculture systems for the production of biochemical products. In: *Proceedings of the National Academy of Sciences of the United States of America* 112 (27), S. 8266–8271. DOI: 10.1073/pnas.1506781112.