Decoding symbiosis: Adaptive traits of *Curvibacter* within the *Hydra* holobiont

Inaugural-Dissertation

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

vorgelegt von

Lukas Becker
aus Duisburg

Düsseldorf, Mai 2025

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

2. Prof. Dr. Sebastian Fraune

Tag der mündlichen Prüfung: 05.11.2025

Abstract

Symbiotic associations are widespread across all domains of life. These associations exert essential effects on ecology, evolution and physiology of organisms. They impact the natural development of ecosystems and have a decisive influence on their ecological stability. To gain a better understanding of the driving forces of symbiotic associations, I studied the *Hydra* holobiont model system, with a particular focus on the bacterial symbiont *Curvibacter*. To investigate symbiosis-related traits in *Curvibacter*, I compared host-associated genomes with those of free-living species. Among the symbiont-specific genes, I identified a gene cluster linked to extracellular polymeric substance production. Bioanalytical and microbiological analyses confirmed that *Curvibacter* secretes sugar-containing polymers. Knockouts of two members of this gene cluster further revealed that these genes modulate exopolymer composition, bacterial growth, and recolonization efficiency in *Hydra*. To further elucidate the symbiont-associated characteristics and their underlying genetic traits, I analyzed publicly available *Curvibacter* transcriptomes. Integrated with the symbiont-specific gene set, this analysis revealed that transporter proteins are conserved in symbiotic strains and differentially regulated during host colonization. Furthermore, growth experiments in defined media revealed a vitamin B12-dependent methionine auxotrophy in *Curvibacter*. Notably, the bacterium proliferates in association with *Hydra* without external methionine supplementation, implicating the host as the primary source of methionine provision. Together, these findings highlight key genetic and metabolic adaptations that enable *Curvibacter* to thrive in symbiosis, particularly through polymer secretion, nutrient dependency, and host-responsive gene regulation. To support both computational and experimental investigations, I developed a bioinformatics tool to streamline ortholog identification within custom databases, featuring a user-friendly interface and dynamic post-processing for in-depth sequence analysis. Complementing this, we established a high-throughput workflow to functionally characterize endogenous *Curvibacter* promoters using cell sorting and growth assays. Our pipeline facilitates the identification of novel expression systems and enables transcriptome-independent gene activity profiling. Together, these tools enhance genetic analysis by streamlining ortholog comparison and enabling efficient functional characterization of endogenous promoter sequences.

Zusammenfassung

Symbiotische Assoziationen sind in allen Domänen des Lebens weit verbreitet. Diese Assoziationen haben wesentliche Auswirkungen auf die Ökologie, Evolution und Physiologie von Organismen. Sie beeinflussen die natürliche Entwicklung von Ökosystemen und haben einen entscheidenden Einfluss auf deren Stabilität. Um ein besseres Verständnis der treibenden Kräfte symbiotischer Assoziationen zu erlangen, habe ich das Holobionten-Modellsystem *Hydra* untersucht, mit Fokus auf den bakteriellen Symbionten *Curvibacter*. Um Symbiose bezogene Merkmale in *Curvibacter* zu untersuchen, habe ich Wirt-assoziierte Genome mit denen freilebender Arten verglichen. Unter den Wirt-assoziierten Genen identifizierte ich solche, die mit der Produktion von extrazellulären Polymer-Substanzen in Verbindung stehen. Bioanalytische und mikrobiologische Analysen bestätigten, dass *Curvibacter* zuckerhaltige Polymere sekretiert. Gen-Deletionen zeigten weiter, dass diese Gene die Polymerzusammensetzung, das Wachstum der Bakterien und die Kolonisationseffizienz in *Hydra* beeinflussen. Um weitere Symbiose bezogene Merkmale zu erforschen, analysierte ich öffentlich verfügbare *Curvibacter*-Transkriptome. In Kombination mit dem Wirt-assoziierten Gen-Set ergab diese Analyse, dass Transporterproteine in symbiotischen Stämmen konserviert sind und während der Wirt-Kolonisierung unterschiedlich reguliert werden. Wachstumsexperimente in definierten Medien ergaben eine Vitamin B12-abhängige Methionin-Auxotrophie in *Curvibacter*. Bemerkenswerterweise kann *Curvibacter* in Assoziation mit *Hydra* auch ohne externen Methionin-Zusatz wachsen. Diese Ergebnisse heben genetische und metabolische Anpassungen hervor, die es *Curvibacter* ermöglichen in Symbiose zu prosperieren, insbesondere durch die Sekretion von Polymeren, Nährstoffabhängigkeit und Wirt-reaktive Genregulation. Für die Analyse von Protein Sequenzen entwickelte ich ein Bioinformatik-Tool, welches die Identifikation von Orthologen innerhalb benutzerdefinierter Datenbanken vereinfacht. Das Tool bietet eine benutzerfreundliche Oberfläche und dynamische Nachbearbeitungsfunktionen für eine detaillierte Sequenzanalyse. Ergänzend dazu haben wir ein Hochdurchsatz-Verfahren etabliert, um endogene *Curvibacter*-Promotoren funktionell zu charakterisieren, wobei Zellsortierung und Wachstumsexperimente eingesetzt wurden. Unser Ansatz ermöglicht die Identifizierung neuer Expressionssysteme und die Analyse der Genaktivität in *Curvibacter*. Diese Werkzeuge optimieren Gen Analysen, indem sie die Identifizierung von Orthologen erleichtern und eine präzise sowie effiziente Charakterisierung von Promotorsequenzen ermöglichen.

Contents

1 Introduction

1.1 General Introduction

The evolution of species has led to the development of a vast diversity of life forms, each shaped by complex and dynamic interactions with their surrounding environment and organisms. Driven by ecological and evolutionary processes, these interactions have led to the development of mutual dependencies among community members. Ultimately, this has resulted in the formation of symbiotic associations, which can still be observed in the present day. To comprehensively describe such communities of interacting organisms, the term "symbiosis" was introduced by Anton de Bary in the late nineteenth century, which he defined as a "cohabitation of unlike organisms" (Bary 1879). It is a fundamental and integral part of nature and many, if not all, species are involved in organismic interactions. Symbiotic interactions can manifest between macroscopic and microscopic organisms or between both. In this dissertation, I will focus mainly on interactions between animal hosts and bacterial symbionts. To evaluate the impact of symbiosis on the fitness of interacting organisms, scientists introduced three primary categories of symbiosis: parasitism, mutualism, and commensalism (Egerton 2015). These classifications offer a more structured framework for systematically analyzing organismic interactions and their ecological consequences. Parasitism describes a relationship where one species exploits the other to their own benefit, at the others organisms' expense (Roberts et al. 2009). In contrast other species rely on collaborative interactions such as commensalism, where one species benefits without significantly affecting the other, or mutualism, where both species derive mutual advantages (Bronstein 1994; Douglas 2010; Wiesmann et al. 2023). Symbiotic associations are characterized by an ongoing development of specialized traits, driven by selective adaptations within community members, enabling the participating organisms to effectively fulfill their roles within the relationship (Groussin et al. 2020; Wade 2007). For example, pathogenic bacteria possess specific traits that allow them to evade the host's innate immune response (Finlay and McFadden 2006). Conversely, hosts need to counteract these strategies and develop their own defense mechanisms while simultaneously maintaining a suitable habitat for beneficial microbes (Aleru and Barber 2020; Sackton et al. 2007; Zamioudis and Pieterse 2012). While the terms of parasitism, commensalism, and mutualism can effectively describe the current mode of symbiosis for the organisms

under consideration, the boundaries between these categories are not always as rigid as these definitions might suggest (Dale and Moran 2006).

1.2 The holobiont theory of symbiosis

To accurately describe the selective changes within symbiotic associations and interactions, a more precise terminology was introduced. The holobiont theory seeks to explain species evolution as a holistic, multi-level selection process, one that includes all participating organisms, not just isolated individuals (Rosenberg et al. 2007; Zilber-Rosenberg and Rosenberg 2008). The holobiont theory emerged as a generalization of the "coral probiotic hypothesis", which proposes that the rapid adaptation of corals is shaped by the dynamic interplay between the environment and their associated microorganisms, ultimately leading to the evolutionary selection of the most advantageous coral holobionts (Reshef et al. 2006).

The term holobiont is derived from the Greek word *holos*, meaning "whole" or "entire", and *biont*, which refers to a distinct biological entity or lifeform that can be clearly differentiated from its surrounding environment. The holobiont theory integrates the host and all associated symbiotic partners (s. Figure 1.1) into a unified biological entity capable of transmitting evolutionary information in the form of a hologenome (s. Figure 1.2) from one generation to the next (Rosenberg and Zilber-Rosenberg 2018; Zilber-Rosenberg and Rosenberg 2008). The hologenome refers to the collective genetic composition of the holobiont at a given point in time, acknowledging that some associated organisms can dynamically associate with or dissociate from the host (Theis et al. 2016). This association and dissociation is described as facultative (non-host-restricted) symbiosis, where the facultative symbiont can transition between a free-living and a host-associated existence. In contrast, obligate (host-restricted) symbiosis is characterized by a mutual dependence, where both the host and the symbiont are typically unable to survive independently (Kucuk 2020). This selective association may result from factors such as prolonged co-evolution between specific microbes and hosts, or priority effects, where early-arriving microbes, especially those transmitted directly from parent to offspring, gain a competitive advantage in colonization (Chen et al. 2024; Fisher et al. 2017). This mode of transmission is referred to as vertical transmission, and it typically occurs during reproduction. In contrast, horizontal transmission involves the acquisition of symbionts from the surrounding environment (Bright and Bulgheresi 2010). Thus, microbial symbionts and their contributions to the holobiont and its

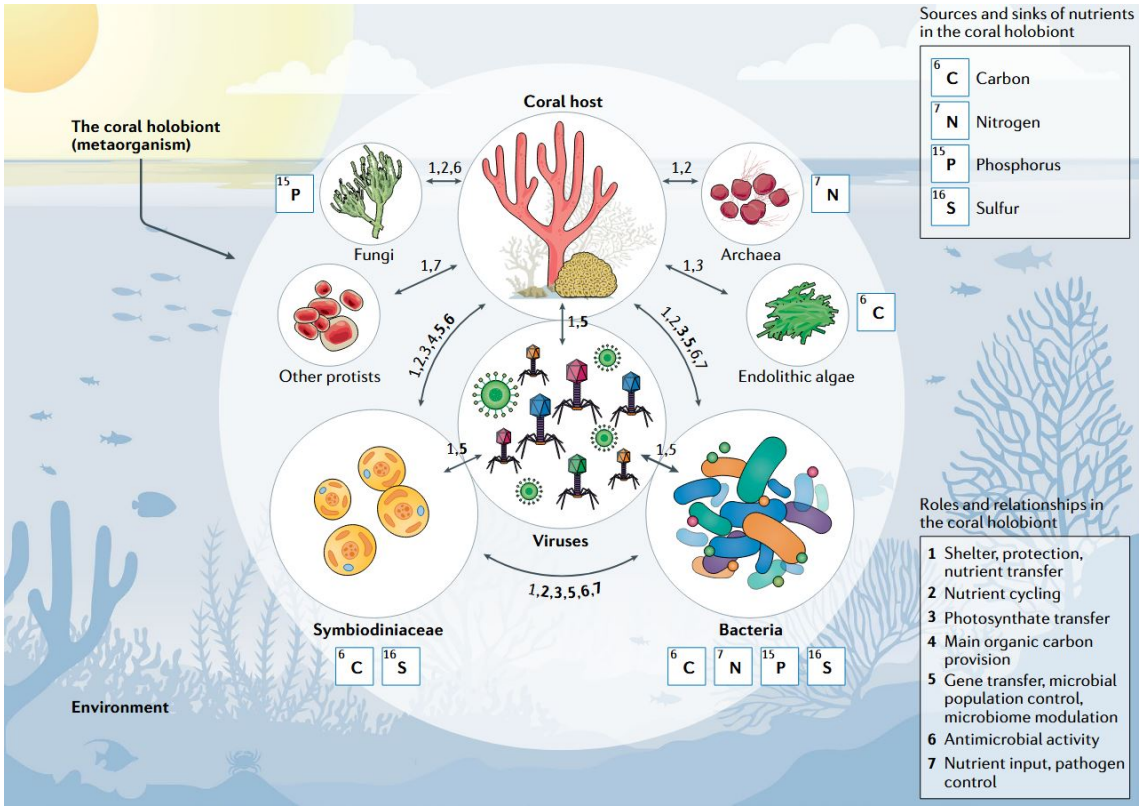

Figure 1.1: **Schematic representation of the coral holobiont as a model for the holobiont theory (Voolstra et al. 2021).** Corals can be considered as "founder species" in the context of reef ecosystems and are of immense ecological importance. The coral host exists within a dynamic environment composed of diverse microorganisms, including fungi, archaea, algae, protists, viruses, and bacteria. These microbial partners interact both with the coral host and with each other through various biological processes, denoted by numbered labels (1-7), explained in the bottom right of the figure. Bold numbers represent interactions with experimentally confirmed functional roles. Letters (C,N,P and S) indicate the contributions of specific organisms to key metabolic cycles, with definitions provided in the upper right corner. The figure was taken from Voolstra et al. (2021).

hologenome can be either stable or transient (Casadevall and Pirofski 2015). The holobiont concept integrates these diverse and dynamic interactions within symbiosis, thereby enabling a more precise description of their various modes of action.

Interacting organisms in the holobiont share a common environment and often exhibit a certain degree of functional integration of their metabolic capacities, thereby influencing the biological processes of both the host and other associated microorganisms. This process tends to create an environment that favors dynamic cospeciation, in which the microbiome reflects the phylogenetic relationships of the host organisms. This phenomenon, known as phylosymbiosis , suggests that microbial communities associated with the host have evolved in parallel with their host species, resulting in a microbial composition that mirrors the host's evolutionary

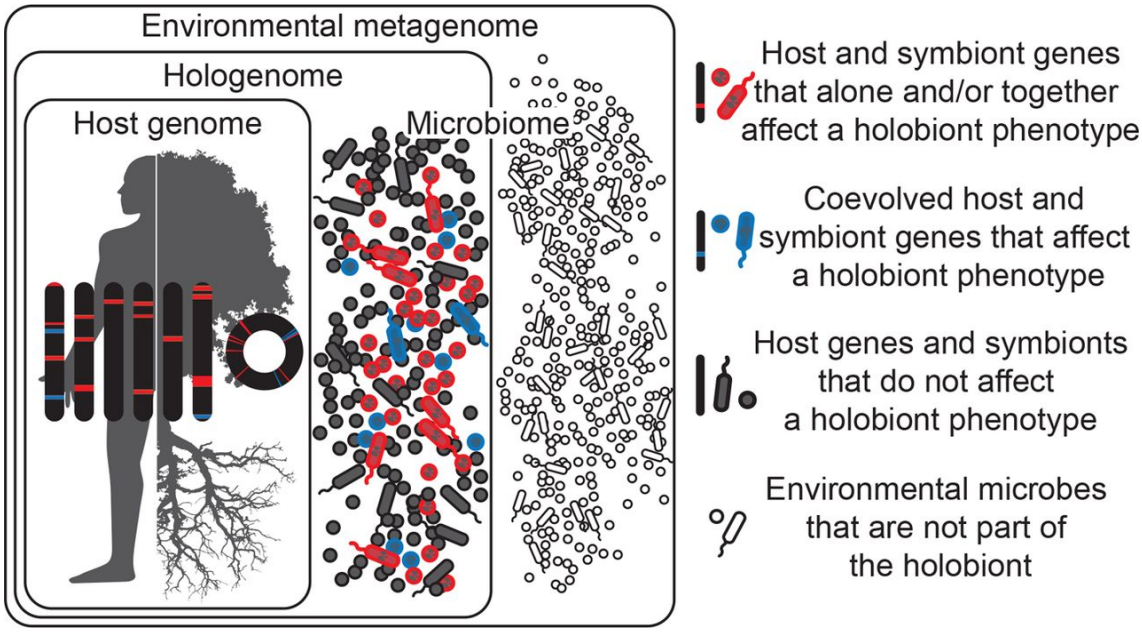

Figure 1.2: **Schematic overview of the hologenome concept (Theis et al. 2016).** The hologenome encompasses the combined genomes of the host and its associated symbiotic microbiome. This collective genomic content has different evolutionary origins, which are represented by the colors red, blue and grey in the figure. Some host and symbiont genes, depicted here as red alleles and microbes, have not co-evolved within the holobiont but can nonetheless influence its phenotype, either independently or through their interaction. Blue alleles and microorganisms represent genes that have co-evolved within the holobiont and directly shape its phenotype. Additionally, certain host genes and microbial symbionts, shown in grey, do not appear to impact the holobiont phenotype. Finally, environmental microorganisms that do not colonize the holobiont have no direct effect on the hologenome. Nevertheless, facultative microbes have the capacity to transition from a free-living environmental state to a host-associated state. At this point, they may become part of the holobiont and contribute to the hologenome. The figure was taken from Theis et al. (2016).

history (Brooks et al. 2016; Liang et al. 2024). To summarize, the concept of the holobiont and its hologenome escapes the boundaries of traditional definitions of symbiosis by incorporating all organisms at a given time that shape the fitness and phenotype of a life form composed of a diverse community.

1.3 The dynamic nature of symbiotic interactions

Holobionts are biological entities, and the interactions among their components represent dynamic processes. The relationships observed today may have emerged from a dynamic and stochastic series of events, involving all participating organisms, driving the evolution of the holobiont (Hussa and Goodrich-Blair 2013; Lood et al. 2015).

Symbiotic interactions can evolve in a way where interaction partners transition from one type of symbiosis to another. For example, under certain conditions, or-

ganisms that were previously mutualists or commensals can become pathogenic (Dey 2024). A prominent example is the opportunistic transmission of *Escherichia coli*, a typically harmless resident of the human gastrointestinal tract, which can cause infections when it colonizes other body sites, such as the urinary tract (Alteri and Mobley 2015). Furthermore, alterations within the microenvironment (i.e. the ecological niche in which host-associated microorganisms reside) can mediate profound effects on the symbiotic status of microorganisms (Kopac and Klassen 2016; Lood et al. 2015). The human gut serves as an ecological niche for a diverse community of microbes (Cresci and Bawden 2015). Inflammatory responses triggered by various factors, including diet, stress, medications, and immune system dysregulation, can alter the chemical properties of this microenvironment. These alterations are associated with an increased proliferation of previously mutualistic or commensal *Enterobacteriaceae*, accompanied by a simultaneous decline in other beneficial microbial species (Zeng et al. 2017). This imbalance in the gut microbiota can in turn have a variety of effects on human health and gut homeostasis (Grundmann 2020; Zhao et al. 2023). Thus, the abundance and population density of species within symbiotic systems can be a critical factor, influencing the balance of fitness costs and benefits across the entire holobiont (Parker 2021).

While harmless or even beneficial bacteria can transition into pathogenic forms, the reverse can also occur. A notable example was described by Jansen et al. (2015) who performed selective evolution experiments with the Gram-negative bacterium *Pseudomonas aeruginosa* (*P. aeruginosa*) and its host nematode *Caenorhabditis elegans (C. elegans)*. Originally pathogenic to *C. elegans*, the *P. aeruginosa* strains evolved into commensal organisms due to mutations in the genes of the global transcriptional regulator *lasR* and the polymerase subunit *rpoB* (Jansen et al. 2015). The observed mutations had pleiotropic effects due to the impact on global transcriptional regulation, which led to an attenuation of virulence and lower mortality of the host *C. elegans*. The evolved *P. aeruginosa* populations exhibited enhanced fitness within *C. elegans*, which in turn produced more offspring compared to colonization with the pathogenic ancestors of *P. aeruginosa* (Jansen et al. 2015).

Interchangeable occurrences of events that support either a shift to parasitism or mutualism have the capacity to alter the symbiotic status of the associated organisms in a more dynamic manner. *Wolbachiae*, a Gram-negative, obligate intracellular symbiont, is associated with over half of the world's described insect species, as well as some nematodes (Hilgenboecker et al. 2008). Different populations of *Wolbachiae*

can confer mutualistic benefits to the host (Hosokawa et al. 2010; Zug and Hammerstein 2015), with some strains also exhibiting varying degrees of reproductive parasitism (Hurst and Frost 2015). It is hypothesized that these association-dependent symbiotic interactions are mediated by the events of lateral gene transfer, through which *Wolbachia* can acquire genetic elements that drive evolutionary shifts toward either parasitism or mutualism (Gerth and Bleidorn 2016; Nikoh et al. 2014; Zug and Hammerstein 2015).

To study the different modes of symbiotic interactions, it is essential to simplify these highly complex holobionts into subunits that allow us to explain symbiotic interactions at a smaller scale. Although the holobiont resembles a community of organisms that collectively shape symbiotic interactions, it is crucial to disentangle these relationships to describe them effectively. This descriptive process requires adopting a holistic perspective, one that considers interactions not only between the host and individual microbes, or among microbes themselves, but with the holobiont as an integrated system.

1.4 The *Hydra* holobiont as a model system for the investigation of symbiotic interactions

For this reason, model systems have been employed to address some of the most intriguing questions in microbiome and organismic interaction research (Douglas 2019). One such model organism is *Hydra*, a freshwater diploblast belonging to the Hydrozoa class within the Cnidarian phylum (Kovačević et al. 2024; McFall-Ngai and Bosch 2021). This section will provide a comprehensive overview of the lifestyle and physiology of the Cnidarian *Hydra* polyp. Following this overview, the next section will introduce *Hydra* as a model system for organismic interactions.

The Cnidarian phylum is phylogenetically positioned as the sister group to the Bilateria (Collins 2002; Collins et al. 2006). Cnidarians, such as jellyfish, corals and sea anemones, share common features. Members of this phylum possess a diploblastic body organization, characterized by the presence of two primary germ layers, the ectoderm and endoderm (Chipman 2024; Nielsen et al. 1996). The third germ layer, the mesoderm, is believed to have first evolved after the divergence of Cnidarians. It first appears within the lineage that gave rise to the triploblastic Bilaterians. Cnidarians are the first animal phylum to develop true epithelia with tight junctions and a simple nervous system. A specialized and complex cell type, the cnidocyte

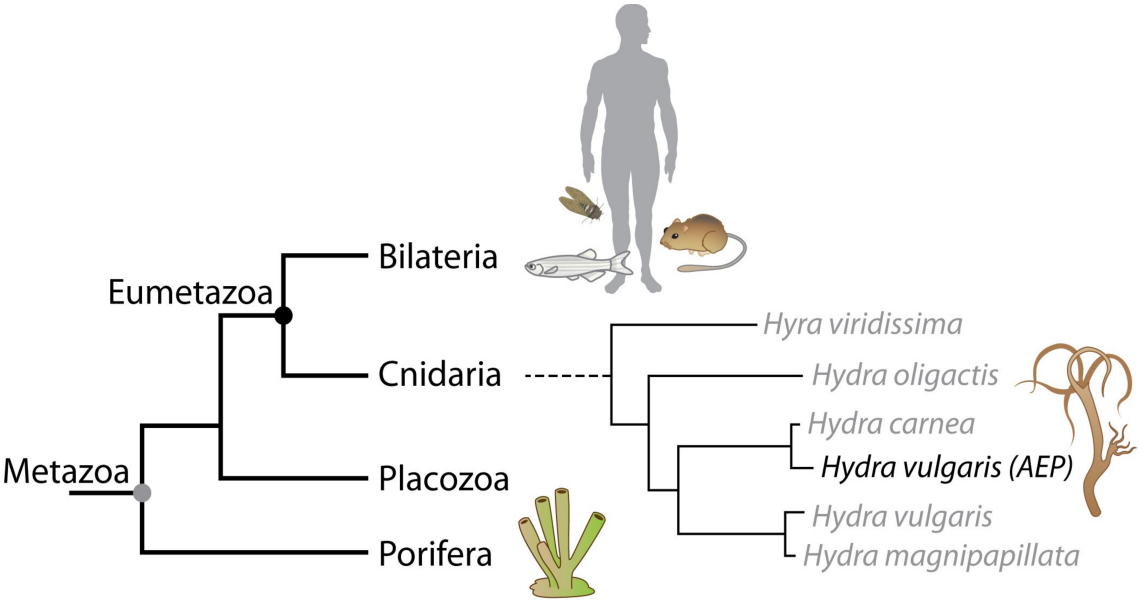

Figure 1.3: **Overview of the phylogenetic position of the phylum Cnidaria, with a specific focus on the genus *Hydra*.** Taken and modified from Bosch (2013).

(or nematocyte), is embedded within the epithelial cell layer (s. Figure 1.4) and is used for capturing prey (Steele 2002). The study of Cnidarians is essential for understanding the complex processes that have shaped the evolution and development of metazoans (Holstein 2022; Lousada et al. 2022).

Most members of this phylum are primarily marine organisms, while some, such as species from the *Hydra* genus, are found in freshwater ecosystems (Bathia et al. 2024; Taubenheim et al. 2022). *Hydra* species live as solitary polyps with no medusa stages or gonophores, which are common in other hydrozoans. Their diet consists of small aquatic animals, which they capture using their tentacles. If one of these small animals touches a tentacle, it gets stuck and is paralysed or killed by the poison of the cnidocyte (Klug et al. 1989). As a result, the tentacles bend and transport the prey towards *Hydra's* hypostome. From there it reaches the gastrovascular tract of the polyp, where it is digested by secretional activity of specialized gland cells and absorbed by nutrient muscle cells (endocytosis). *Hydra* species possess a substantial regenerative ability. The ectodermal and endodermal epithelial cells, along with the interstitial stem cells (s. Figure 1.4), are continuously involved in the maintenance and renewal of the *Hydra* body (Bosch et al. 2010). Although the ectodermal and endodermal epithelial cells are fully differentiated and function as muscle cells, they retain the ability to self-renew within the body column (Buzgariu et al. 2015). Interstitial stem cells are pluripotent and give rise to a variety of specialized cell

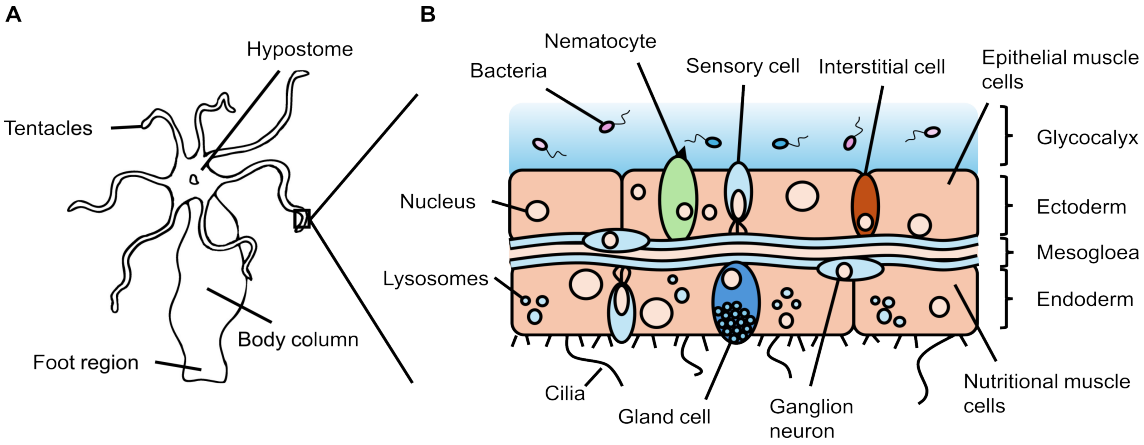

Figure 1.4: **Schematic overview of the *Hydra* holobiont with its body organization and epithelial cell layer.** **(A)** The *Hydra* polyp is anatomically composed of tentacles, a hypostome, which facilitates the transfer of captured prey into the gastrovascular cavity, a body column, and a foot region (peduncle) for substrate attachment. Asexual reproduction occurs through budding, a process that takes place within the body column and gives rise to genetically identical offspring. **(B)** *Hydra* exhibits a diploblastic organization, consisting of an outer ectoderm and an inner endoderm, which are separated by an acellular, gelatinous layer known as the mesoglea. Specialized secretory cells and neurons are embedded within the endoderm, contributing to the production and release of digestive enzymes into the gastrovascular cavity. Ganglion (motor) neurons, which control coordinated body movements, are localized at the basal surface of the epithelial muscle cells. Sensory neurons project outward from this neural network and are responsible for detecting and responding to environmental stimuli. Interstitial stem cells and nematocytes reside within the epithelial muscle layer and contribute to tissue dynamics and defense, respectively. The ectodermal epithelial cells secrete glycoproteins and glycosaminoglycans that form a mucous-like layer termed the glycocalyx. This layer functions not only as a protective barrier but also as a habitat for microorganisms.

types, including neurons, gland cells, and germ cells (Bosch et al. 2010). Initially, *Hydra's* remarkable regenerative ability drew significant research interest, and today, it serves as a model organism for various applications, including developmental biology, animal physiology, environmental toxicology, innate immune system and holobiont research, owing to its association with a diverse microbiome (Kovačević et al. 2024).

1.5 *Hydra* maintains a stable microbiome that influences host functions

Hydra's microbiome plays a crucial role in mediating pathogen resistance and modulating various host processes (Minten-Lange and Fraune 2020). The epithelial cell layer of *Hydra* secretes a glycoprotein and glycosaminoglycan-rich meshwork (s. Figure 1.5C) known as the glycocalyx (Böttger et al. 2012). This glycocalyx functions as an environmental niche for a stable community of species-specific bacterial microbes

(Franzenburg et al. 2013; Fraune and Bosch 2007). It provides a dynamic interface for host-microbe and microbe-microbe interactions (Deines et al. 2017). The depletion of epithelial gland and neuronal cells disrupts the epithelial homeostasis, leading to significant shifts in *Hydra's* microbial community structure (Fraune et al. 2009). *Hydra* polyps can actively shape and maintain their resident microbial species through the spatially controlled secretion of various species-specific antimicrobial peptides (AMPs) (Augustin et al. 2010, 2009; Bosch et al. 2009; Franzenburg et al. 2013; Fraune et al. 2010; Jung et al. 2009). AMPs are short molecules that have the capacity to act against a broad spectrum of microorganisms. This capacity is typically realised through membrane disruption or the penetration and impairment of intracellular functions (Savitskaya et al. 2023). The secretion of these peptides is controlled by a conserved Toll-like receptor (TLR)-mediated immune pathway, which detects bacterial molecules such as flagellin and activates the conserved NF-κB transcriptional pathway (Bosch et al. 2009; Franzenburg et al. 2012; Hemmrich et al. 2007). The repertoire of AMPs is diverse, with each group exhibiting activity against distinct microbial compositions (Klimovich and Bosch 2024).

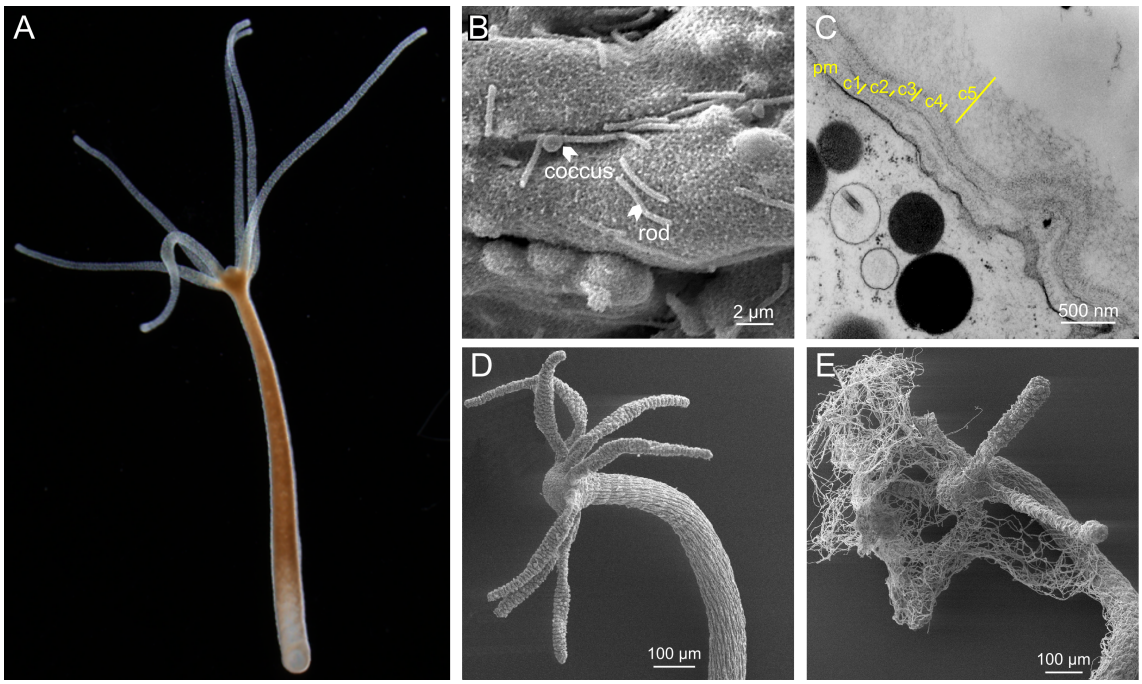

Figure 1.5: **The *Hydra* holobiont. (A)** Phenotypic appearance of the freshwater polyp *Hydra vulgaris* AEP. The figure was taken from Fraune and Bosch (2007). **(B)** Raster electron micrograph of rod and coccus shaped bacterial cells on the ectodermal epithel of *Hydra*. **(C)** The five distinct cell layers of *Hydra's* glycocalyx (c1-c5) and the plasma membrane (pm) of epithelial muscle cells. **(D)** *Hydra* polyps with an intact microbiome exhibit resistance to fungal infection, **(E)** whereas microbiome-depleted *Hydra* show significantly increased susceptibility to infection by the environmental fungus Fusarium. Figures B-E were taken from Fraune et al. (2015).

The natural habitat of *Hydra* polyps are aquatic freshwater systems, where diverse environmental microorganisms such as bacteria and viruses influence and shape the resident *Hydra* microbiome (Fraune and Bosch 2007). In experiments involving *Hydra oligactis* populations collected from various Hungarian lakes, shifts in bacterial beta-diversity were observed when treated with non-native or native lake water samples. Over a period of four weeks, these polyps gained and lost certain bacterial taxa, whereby the changes were influenced by the water source (Bathia et al. 2024). This shift in associated microbes can have profound effects on the host organism and its microbial community. It has been shown that aposymbiotic *Hydra viridissima* polyps, in which endosymbiotic algae of the *Chlorella* genus have been depleted, are more susceptible to colonization by invasive species such as *Legionella* (Bathia et al. 2022). In addition to the biotic factors that influence microbiome dynamics, seasonal variations and environmental perturbations have the capacity to alter abiotic conditions, such as nutrient availability or temperature. These factors can profoundly affect the microbiome of the polyps (Bathia et al. 2024; Bosch et al. 2015; Lachnit et al. 2025; Taubenheim et al. 2022). Consequently, host identity and biogeography play an essential role in shaping the composition of *Hydra's* microbial community.

In Greek mythology, the Lernaean Hydra was a serpentine water monster with multiple heads, renowned for its extraordinary regenerative ability, whenever one head was severed, two would grow in its place (Encyclopaedia Britannica 2024). Interestingly, the modern *Hydra* species exhibits remarkable similarities in terms of regeneration and longevity. *Hydra's* regenerative processes are driven by a population of stem cells that activate proliferation following injury. One of the key regulators of stem cell function is the FoxO transcription factor, which not only controls stem cell proliferation but also modulates the expression of AMPs (Boehm et al. 2012; Bosch 2019). A deficiency in FoxO signaling disrupts the selective maintenance of native, beneficial microbial partners, leading to increased colonization by foreign bacteria (Mortzfeld et al. 2018). This intricate connection between microbial composition and stem cell proliferation underscores the pivotal role of the microbiome in *Hydra's* regenerative processes. Interestingly, not only do regenerative processes influence microbiome composition, but tumorigenesis in *Hydra* is also associated with alterations in the microbiome. Tumor-bearing *Hydra oligactis* individuals were found to be associated with intracellular bacteria from the *Chlamydiales* genus, which formed large vacuoles within the host cells (Boutry et al. 2023). Rathje et al. (2020) identified an environmental *Spirochete* (*Turneriella* sp.) linked to tumor formation in

Hydra, with its virulence being modulated by the presence of a beneficial microbe from the *Pseudomonas* genus. Recent findings demonstrated that supplementing *Hydra's* medium with L-arginine induces a shift in the *Pseudomonas* strain toward a pathogenic phenotype, resulting in severe physiological disruption and host mortality (Lachnit et al. 2025). These interactions have the capacity to influence the structure of the microbial community, mediate developmental processes, impact disease outcomes, and consequently affect the fitness of the *Hydra* host.

With its simple body plan and well-defined cell populations, *Hydra* also serves as a valuable model in neuroscience research. Since AMP production is tightly regulated by neurons, these cells play a crucial role in shaping the microbial composition of *Hydra* polyps (Fraune et al. 2009). *Hydra's* neuronal system is controlled by pacemaker neurons, which generate rhythmic bursts of action potentials that drive neuronal activity (Klimovich et al. 2020). These rhythmic bursts trigger coordinated body contractions and are essential for regulating the polyp's feeding behavior (Giez et al. 2023; Murillo-Rincon et al. 2017; Nawroth et al. 2023). This neuronal activity is also influenced by *Hydra's* microbiome. Germ-free polyps exhibit lower spontaneous contraction frequencies (Murillo-Rincon et al. 2017; Nawroth et al. 2023), shorter feeding response durations, reduced mouth-opening behavior (Giez et al. 2023), and decreased expression of pacemaker-specific genes (Klimovich et al. 2020). These results highlight the crucial interplay between microbial signals and neurophysiological functions in *Hydra*.

1.6 *Curvibacter* AEP is the main colonizer of *Hydra vulgaris* AEP

To unravel the influence of microbial species on the *Hydra* holobiont, both a holistic perspective on microbiome interactions and a focused examination of individual microbial partners are essential. To this end, specific colonizers known for their intimate interactions with the host, such as the green algal symbiont *Chlorella* and the Gram-negative, rod-shaped bacterium *Curvibacter*, were selected for closer investigation. To analyze genetic traits and adaptive processes, I focused on the most abundant bacterial colonizer of *Hydra vulgaris* AEP (hereafter *Hydra* AEP), the bacterial symbiont *Curvibacter* sp. AEP1.3 (hereafter *Curvibacter* AEP) (Fraune et al. 2015). *Curvibacter* AEP belongs to the phylum of β-Proteobacteria from the *Comamonadaceae* family (Fraune et al. 2015; Minten-Lange and Fraune 2020). This species thrives within the ecological niche provided by *Hydra*, with its abun-

dance being regulated by epithelial homeostasis (Fraune et al. 2009). *Curvibacter* is vertically transmitted to the next generation during reproduction (Fraune et al. 2010). The *Hydra-Curvibacter* interaction is bidirectional, with *Hydra* modulating *Curvibacter's* quorum-sensing signals (Pietschke et al. 2017), while *Curvibacter* is inducing the expression of host peptides that function as antagonists of the Wnt signaling pathway (Taubenheim et al. 2020). This evolutionary conserved signaling pathway is involved in a variety of developmental processes, stem cell regulation and tissue homeostasis (Chen et al. 2008).

The *Hydra* microbiota plays a crucial role in protecting the polyp from infections caused by a filamentous fungus of the *Fusarium* genus (Fraune et al. 2015). Interestingly, this antifungal defense is only effective when *Hydra* is co-associated with *Curvibacter* AEP and *Duganella* sp. C1.2 (hereafter *Duganella*). Neither of these symbionts alone is sufficient to confer protection against fungal infection (Fraune et al. 2015). In *in-vitro* co-cultivation experiments, *Duganella* outcompetes *Curvibacter* AEP in terms of abundance (Deines et al. 2020), while both species exhibit frequency-dependent, non-linear growth dynamics (Li et al. 2015). Interestingly, *Duganella* and *Curvibacter* AEP abundances on the polyp were higher in wild-type *Hydra* compared to mono- or di-association experiments, with *Curvibacter* AEP emerging as the dominant colonizer on *Hydra* AEP (Deines et al. 2020). This variation in the bacterial carrying capacity of *Hydra* polyps may be influenced by the presence of a bacteriophage isolated from *Curvibacter* AEP, which can infect and lyse *Duganella* (Li et al. 2017), as well as by interactions between those microbes and other rare members of the native *Hydra* microbiome (Deines et al. 2020, 2017). Not only *Duganella* but also *Curvibacter* AEP may serve as a target for bacteriophages, highlighting the complex interplay between microbial competition and viral predation within the *Hydra* microbiome (Ulrich et al. 2022). Furthermore, nutrient exchange and availability within the holobiont has an effect on the composition of the microbial community (Lachnit et al. 2025). In experiments involving external nutrient supplementation, the relative abundance of *Curvibacter* decreased significantly - from approximately 60% to just 7% - compared to its levels in *Hydra's* standard, nutrient-poor culture medium (Lachnit et al. 2025). In addition, Giez et al. (2023) demonstrated that *Curvibacter* AEP secretes the amino acid glutamate, which is subsequently taken up by other microbial members, such as *Duganella* and *Undibacterium*, highlighting metabolic interactions within *Hydra's* microbiome.

Hydra and *Curvibacter* can be easily maintained in the laboratory, and a variety of methods and techniques are available for the genetic manipulation of both *Hydra* and *Curvibacter*. Genome sequences are available for *Curvibacter* AEP and other members of *Hydra's* microbiome, enabling targeted analysis and manipulation of genetic sequences, such as gene knockouts or the introduction of plasmid systems carrying fluorescent proteins (Wein et al. 2018). Additionally, transcriptional analysis of *Curvibacter* AEP in its different lifestyles has been conducted, allowing the examination of its transcriptional activity both on its *Hydra* host and in free-living states (Giez et al. 2023; Ulrich et al. 2022). In *Hydra*, not only has the genome been fully sequenced, but a single-cell expression atlas is available, allowing transcript abundance analysis at the resolution of individual cells (Cazet et al. 2023; Siebert et al. 2019). Transgenic *Hydra* can be generated through microinjection techniques in the embryo, facilitating functional gene analysis (Juliano et al. 2014). Fluorescent reporter lines can be established with labeled proteins, enabling the tracking of specific cell types (Wittlieb et al. 2006). In situ hybridization experiments allow for the spatial localization of specific RNA transcripts within *Hydra* tissues (Bode et al. 2008; Hansen et al. 2000). Stem cell culture and transplantation experiments provide a platform for studying interstitial stem cell function and enable lineage tracing to investigate stem cell fate and differentiation (Bosch et al. 2010). Furthermore, *Hydra's* neuromuscular activity and signal transduction in nerve cells can be examined using electrophysiological techniques (Ji and Flavell 2017; Szymanski and Yuste 2019).

This fundamental knowledge about *Hydra* and *Curvibacter*, combined with the availability of advanced methods and techniques for studying both, highlights the *Hydra-Curvibacter* relationship as an emerging model system for symbiosis and its broader impact across various research fields. However, *Curvibacter* AEP is a facultative symbiont that can survive independently of its host. Notably, several other species within the *Curvibacter* genus exist as free-living organisms or associated with bacterial communities and other hosts (Ding and Yokota 2004; Ma et al. 2016; McKenzie et al. 2012). This raises the question of the genetic adaptations that drive *Curvibacter* AEP to form a symbiotic relationship with *Hydra* AEP.

1.7 Drivers that influence the transition from free-living to host-associated microbes

In general, the symbiotic status of bacterial genera is influenced by a variety of factors. Within a specific bacterial genus, not all species are associated with animal or plant hosts (Hammer et al. 2019; Itoh et al. 2019). In fact, most known bacterial species exist in free-living states, often interacting with other microbial species to varying degrees, ranging from loose associations to close partnerships and stable communities. The classic distinction between free-living, pathogenic, commensal, or mutualistic bacterial species primarily focuses on comparisons of their genetic adaptations. Despite their diverse interactions in the holobiont, all microbes share a fundamental trait: they adapt to the unique environment shaped by the host's niche (Steinert et al. 2000). Research on the evolutionary mechanisms underlying host-microbe interactions, particularly through comparative genomics, provides key insights into the genetic adaptations that drive these associations.

Genetic traits that drive successful symbiotic associations are those that allow microbes to colonize and persist within the environmental niche provided by the host. These colonization strategies include the use of quorum-sensing signals, two-component regulatory systems (Bélanger et al. 2009; Visick and Skoufos 2001; Williams 2007), and mechanisms that help the microbes suppress host defenses (Miwa and Okazaki 2017), particularly those mediated by innate immune responses (Wiesmann et al. 2023). Following initial colonization, microbes can contribute to an environment that supports the growth and reproduction of their population. The rate of proliferation must be tightly regulated to maintain a balanced, homeostatic relationship with other members of the holobiont. Traits that enable such interactions include pathways for the production of essential nutrients and vitamins (Tarracchini et al. 2024), the secretion and sensing of specialized molecules (Dale and Moran 2006; Mandel et al. 2012; Ortíz-Castro et al. 2009), and the regulation of homeostasis within the micro-environment, such as pH and oxygen availability (Becker et al. 2004; Nourabadi and Nishiguchi 2021). A well-known example, which can be found in various symbiotic model systems, is the production of microbial extracellular polysaccharides (EPS), which are known to interact with the host environment or directly with host cells (Acosta-jurado et al. 2021; Shibata et al. 2012; Yip et al. 2005). EPS contribute to biofilm formation, which facilitates immune evasion through direct interactions with immune cells or by sequestering harmful substances such as AMPs, thereby promoting bacterial persistence within

the host (Alhede et al. 2014; Otto 2006). Moreover, studies have demonstrated that EPS-mediated biofilms can provide a protective niche that serves as a substrate for commensal bacteria, potentially benefiting the host by supporting a balanced microbiota and contributing to overall homeostasis (Acosta-jurado et al. 2021; Ma et al. 2020).

While the host-provided environmental niche is generally stable, microbial species must adapt to rapidly changing conditions at the microscale (Cao and Goodrich-Blair 2017). This requires the ability to sense and communicate with the environment, making genetic pathways involved in such signaling processes key favorable traits for successful symbiotic interactions. Advances in genome comparisons have revealed the presence of genome reductions in mutualistic bacterial genera (Nikoh et al. 2011). These species lose genes unnecessary for independent survival while retaining those essential for interactions with their host organism (Nikoh et al. 2011). In contrast, pathogenic species are known to frequently acquire and exchange virulence factors through horizontal gene transfer (Gyles and Boerlin 2014). By integrating comparative genomics with functional analyses in the laboratory, we can dissect the genetic basis of symbiotic interactions at both the genetic and evolutionary levels, as well as at the functional level by describing and annotating gene functions.

1.8 Evolutionary and comparative genomics

Investigating the various modes of symbiotic interactions often requires the study of species-specific genetic traits, such as AMPs in the *Hydra* genus (Klimovich and Bosch 2024) or nod factors in the legume-*Rhizobia* interaction (Ghantasala and Choudhury 2022). These analyses rely on comparing the identified genes and their products either among each other or against biological sequence databases such as RefSeq, SwissProt, or GenBank. The results of such sequence comparisons can subsequently be used for functional and evolutionary studies, facilitating precise classification and categorization. Furthermore, the bioinformatic analysis of genes and genomes serves as the foundation for further biological research, enabling the formulation of hypotheses and theories that can be experimentally tested in the laboratory. Consequently, comparative and evolutionary genomics play a fundamental role in the study of symbiotic systems.

1.9 A short history of comparative genomics

Deoxyribonucleic acid (DNA) is the primary carrier of genetic information, organized into sequences that form the genome, the combined genetic information of a species. This genome encodes the hereditary blueprint of an organism. DNA is transcribed into messenger ribonucleic acid (mRNA) and then translated into amino acid sequences. These sequences fold into functional polypeptides, undergoing post-translational modifications such as phosphorylation and glycosylation to acquire their three-dimensional structures. This process, from DNA to mRNA to protein is known as the Central Dogma of Molecular Biology (Crick 1958, 1970). Proteins, particularly enzymes, drive essential cellular processes, including metabolism and gene regulation, forming the molecular foundation of life. Variations in amino acid sequences across species provide valuable insights into evolutionary processes. All living organisms share a common ancestor, from which the amino acid sequences observed today have evolved. Through comparative analysis of these sequences, we can infer evolutionary relationships, thereby shedding light on genetic divergence and the ancestral lineage of species (Koonin 2005).

A comparative analysis of protein sequences is defined as a direct comparison of amino acid sequences of interest (Pearson 2013). Various strategies exist for sequence comparison, most of them are using sequence alignment as key method. In general, sequences are aligned using either local (Smith and Waterman 1981) or global (Needleman and Wunsch 1970) alignment algorithms. Global alignments measure the overall similarity across entire sequences, while local alignments focus on identifying regions of local similarity to determine alignment scores. The substitution matrices employed in this process serve as the foundation for deriving alignment scores, as they provide a quantitative representation of the probability of one amino acid mutating into another (Henikoff and Henikoff 1992; Thorne 2000). While aligning two sequences can be challenging, especially within huge sequence databases, the complexity increases significantly when comparing multiple sequences. Multiple sequence alignment (MSA) algorithms address a Non-deterministic Polynomial-time hard (NP-hard) problem (Cai et al. 2000; Feng and Doolittle 1987; Wang and Jiang 1994). To solve such problems, the optimal solution must be found in a huge solution space. For MSA, this means that the sequences must be arranged to maximize similarity based on an evaluation function that takes insertions, deletions and mutations into account and thus reflects their evolutionary relationships. However, advanced algorithms have been developed to handle this complexity efficiently and to yield

consistent and robust alignment results even for large datasets (Altschul et al. 1990, 1997; Buchfink et al. 2015; Kent 2002; Pervez et al. 2014; Reddy and Fields 2024; Steinegger and Söding 2017).

Coupled with advancements in sequencing technologies (Goodwin et al. 2016), it has been possible since the release of the first commercial next-generation sequencing machine in 2005 (Margulies et al. 2005), to compare entire genomes across diverse organisms (Gresham et al. 2008; Saada et al. 2024). Insights derived from these comparisons led to the exploration of (a) evolutionary histories of species, including holobionts, (b) functional similarities among protein sequences, (c) transcription factor motifs and regulatory regions across species and (d) clade specific genetic traits (Fagorzi et al. 2020; Nobrega and Pennacchio 2004; Sivashankari and Shanmughavel 2007).

1.10 Evolutionary reconstructions and functional annotations through homologous sequence categorization

Protein sequences used for such comparisons can be systematically categorized, enabling a more robust and sophisticated reconstruction of evolutionary relationships and functional roles. The foundation of this classification is the concept of homology, which refers to sequences that share a common evolutionary origin (Fitch 1970, 2000). Homologous sequences can be classified into paralogous and orthologous sequences. Paralogous sequences arise from gene duplication events within the same genome, whereas orthologous sequences originate through speciation events from a common ancestral gene in the compared species (Koonin 2005). Sets of orthologous proteins characteristically exhibit analogous functionality. Consequently, a fundamental objective of comparative genomics is the identification of groups of orthologous sequences.

These sequences serve as the foundation for most genetic and evolutionary studies, where functional conservation and phylogenetic accuracy are crucial. Moreover, sets of orthologous sequences can be annotated based on their known or predicted functions, using frameworks such as Gene Ontology (GO) terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) categories, which systematically describe protein function through standardized terminology (Lin et al. 2024). This annotation facil-

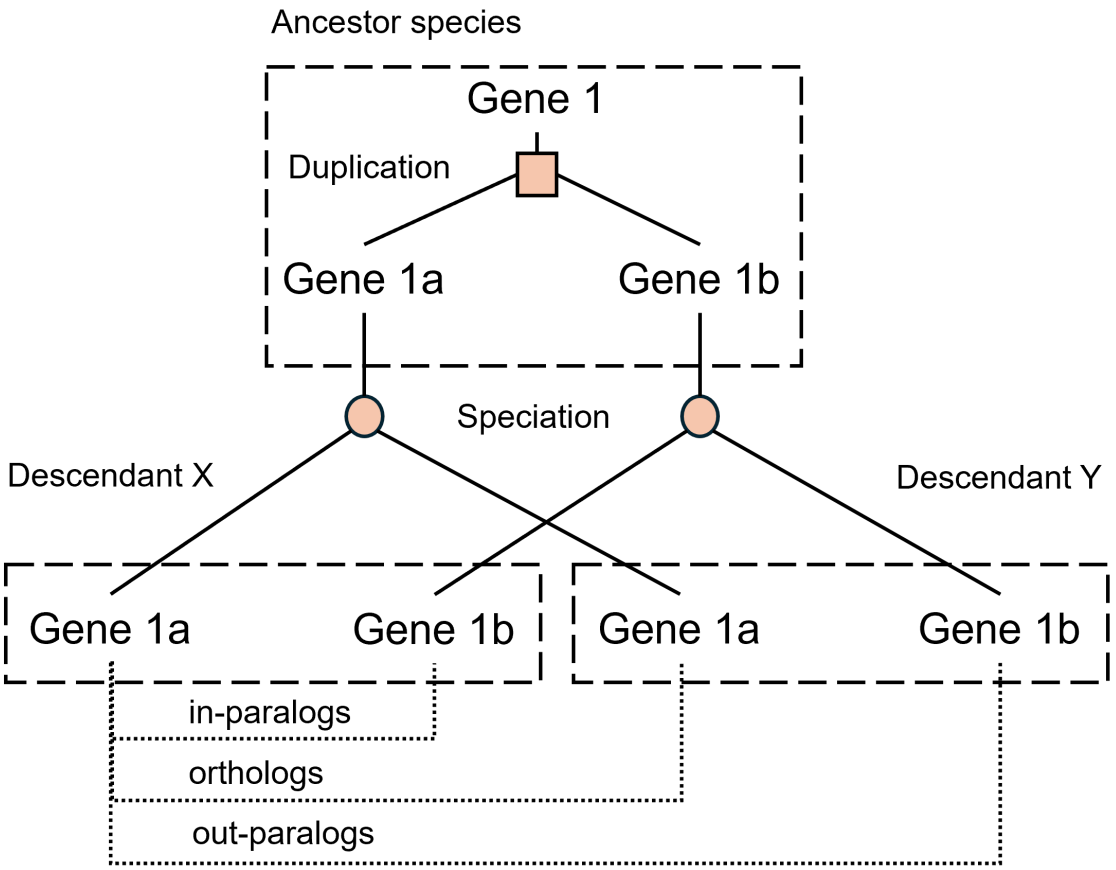

Figure 1.6: **Schematic phylogeny of homologous gene sequences.** The classification of homologous sequences into paralogous and orthologous groups is fundamental to understanding gene evolution. A duplication event of Gene 1 in the common ancestor of species X and Y gave rise to the paralogous genes Gene 1a and Gene 1b. While these paralogs may retain similar functions, they typically undergo functional divergence over time. Following speciation, these duplicated genes are inherited by the descendant lineages, leading to the formation of orthologous genes. For instance, Gene 1a in the descendant X is orthologous to Gene 1a in the descendant Y. Orthologous genes generally maintain their ancestral function, often even across distantly related clades. The illustration is adapted according to Elena Aledo (2025).

itates gene set enrichment analyses (Klopfenstein et al. 2018; Yu et al. 2012), which can be employed to investigate the effects of specific treatments on transcriptional regulation or to assess the impact of evolutionary histories that may lead to the retention or loss of particular genetic traits (Black et al. 2012; Fagorzi et al. 2020).

1.11 The inference of orthologous proteins

Orthologous sequences can be identified using various computer-aided tools for the comparison of entire proteomes of organisms or by searching in special sequence databases (Kristensen et al. 2011). Performing local alignments and using the top hits as putative orthologous sequences is often an insufficient solution,

a more refined approach is required to distinguish true orthologs from paralogous sequences (Hernández-Salmerón and Moreno-Hagelsieb 2020; Tekaia 2016). One reliable method is the bidirectional BLAST technique, where identified homologous sequences are compared against the complete set of sequences from the target organism (Moreno-Hagelsieb and Latimer 2008). A sequence is considered a putative ortholog if it produces the best reciprocal alignment with the original query sequence. Another option is the search within curated biological sequence databases such as the KEGG or Cluster of Orthologous Genes (COG) databases (Kanehisa et al. 2016; Tatusov et al. 2000). These databases contain functional characterized sequences that enable reliable annotations for homologous sequences. The available tools for the classification of sequences are mostly command line based or require special installation routines. Thus, the identification of orthologs is hindered by the complexity of the task. In addition, downstream analysis and post-processing of the results obtained can pose a challenge.

The tasks required to perform homology searches and downstream analyses are ideal for streamlining, as the initial sequence comparisons and post-processing techniques such as assigning conserved domains or inferring phylogenies are in most cases repeatable and similar across different experimental setups. This is ideal for the use of workflow management systems. These systems are used for the sequential execution of tasks, enabling a reproducible, fast and efficient task execution (Köster and Rahmann 2012; Tommaso et al. 2017). With the ever-growing biological databases and the constant updating of available biological sequences, it is becoming increasingly important to develop standardized workflows that enable a differentiated analysis of the genetic information of species. In combination with simple installation routines and graphical user interfaces, bioinformatic workflows can aid in various analyses (Chen et al. 2020; Koh et al. 2023). In the context of this thesis I developed a bioinformatics software tool, for the easy-to-use search of homologous and putative orthologous protein sequences in user-defined databases with the integration of various post-processing steps (Becker et al. 2023). Our Comparative Analysis Tool for Homolog Identification (CATHI) is based on an intuitive web interface that allows users to run bioinformatic pipelines to identify and process homologous protein sequences. In conjunction with other bioinformatic tools, this approach facilitated a detailed analysis of some genetic characteristics in *Curvibacter* AEP, which play a role in the association with its host organism, the freshwater polyp *Hydra* AEP.

2 Aims of this study

Studies of the *Hydra-Curvibacter* symbiotic interaction have already identified several traits that are involved in this interaction, including *Curvibacter's* ability to confer antifungal resistance (Fraune et al. 2015), its role in inter-kingdom communication - where *Hydra* interacts with *Curvibacter* through its quorum sensing system (Pietschke et al. 2017) - and its capacity to induce the expression of host peptides (Taubenheim et al. 2020). However, these factors alone do not fully explain why *Curvibacter* AEP has developed such a close association with *Hydra* AEP, nor do they fundamentally clarify the evolutionary drivers behind the emergence of this symbiotic relationship.

As mentioned earlier, several host-associated and environmental *Curvibacter* species have been identified, and genomic sequences are available for some of them. These genomic resources can be leveraged to investigate genetic adaptations that enable these bacteria to thrive in their specific environmental niches, such as the glycocalyx of *Hydra*. To gain insight into symbiont-associated traits and the genetic adaptations that drive facultative free-living species like *Curvibacter* AEP to establish close interactions with host organisms, I addressed the following research questions:

1. What are genetic traits that enable the transition of *Curvibacter* AEP from a free-living state to host-association?
    - What functional role do these genetic traits possess within the *Hydra* holobiont?
    - Does *Curvibacter* AEP have nutrient dependencies that influence its dynamic interactions within the *Hydra* host environment?
2. How can we bioinformatically improve the investigation of genetic traits and gene products involved in adaptation processes?

    This also raises the question of how these traits are regulated, leading to the emergence of the following sub-question:

    - What are the regulatory mechanisms underlying these genetic traits, and how is their expression controlled?

3 List of Manuscripts and Contributions

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

- **Manuscript 1**: "Extracellular Polymeric Substances Drive *Curvibacter* Host Adaptation and Co-Speciation with *Hydra*". By Becker et al., is being published as part of this thesis.
  Contributions: Sebastian Fraune (SF) made contributions to the supervision, conceptualization, initial writing, reviewing, and editing of this mansucript, in addition to providing resources. IA contributed to the project's conceptualization, the review process, supervision, and the provision of resources. MP contributed to the review process, editing and provided resources. Sylvain Foret was responsible for the sequencing and post-processing of the C.Hmag1.1 and C.Hvul1 genomes. TM conducted the phylogenetic analysis and monocolonization of *Hydra* and *Curvibacter*, as illustrated in Figure 1. JB, MK and NW (formerly NS) contributed to the reviewing process of the manuscript and provided domain-specific insights. LF made contributions to the methodology and reviewing of the manuscript. LB contributed to the conceptualization, methodology, investigation, validation, data curation, initial drafting, and the review and editing of the manuscript.

- **Manuscript 2**: "Host Colonization Triggers Amino Acid Transporter Upregulation in the Cobalamin-Dependent Methionine Auxotroph *Curvibacter*". By Becker et al., is being published as part of this thesis.
  Contributions: SF made contributions to the supervision, conceptualization, initial writing, reviewing, and editing of this mansucript, in addition to providing resources. IA contributed to the conceptualization of the project, the review process, and the provision of resources. SL and LF made contributions to the methodology of this manuscript. JB provided domain-specific insights. LB contributed to the conceptualization, methodology, investigation, validation, data curation, initial drafting, and the review and editing of the manuscript.

- **Manuscript 3**: "CATHI: An interactive platform for comparative genomics and homolog identification". By Becker et al., has been published on bioRxiv at the time of submission of this thesis. DOI: `https://doi.org/10.1101/2023.09.04.556229`

  Contributions: NS made contributions to the supervision, conceptualization, initial writing, reviewing, and editing of this mansucript, as well as the design of Figure 2. IA and SF contributed to the conceptualization of the project, the review process, and the provision of resources. GK and PS contributed to the review process of the manuscript. LB contributed to the conceptualization, methodology, investigation, validation, data curation, initial drafting, and the review and editing of the manuscript.

- **Manuscript 4**: "Oligonucleotide Library Assisted Sequence Mining Reveals Promoter Sequences With Distinct Temporal Expression Dynamics For Applications In *Curvibacter* sp. AEP1-3". By Mager and Becker et al., has been submitted, reviewed and accepted by SynBio - Synthetic Biology at the time of submission of this thesis. DOI: `https://doi.org/10.1093/synbio/ysaf001`

  Contributions: SF made contributions to writing, reviewing, editing and funding acquisition. NS made contributions to methodology and resources. IA contributed to supervision, writing, reviewing, editing and funding acquisition. MM contributed to conceptualization, methodology, investigation, writing, reviewing, editing and visualization. LB made contributions to conceptualization, software, formal analysis, investigation, writing, reviewing, editing and visualization. LB and MM contributed equally to this article.

4 Manuscript 1

Extracellular Polymeric Substances Drive *Curvibacter* Host Adaptation and Co-Speciation with *Hydra*

Lukas Becker[1,2], Liam Fürbach[1], Timo Minten[1], Jay Bathia[1], Marius Karbach[3], Sylvain Foret[†,4], Nicolas M. Winterfeldt[2], Markus Pauly[3],Ilka M. Axmann[2], Sebastian Fraune[1*]

[1]Institute for Zoology and Organismic Interactions, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[2]Institute for Synthetic Microbiology, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[3]Institute for Plant Cell Biology and Biotechnology, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany
[4]Research School of Biology, Australian National University, Canberra, ACT, Australia
[†]Deceased

[*]Corresponding author: Sebastian Fraune (fraune@hhu.de)

Keywords: Phylosymbiosis, comparative genomics, symbiosis, exopolysaccharides, EPS operon, glycosyltransferase, microbiome, biofilm

4.1 Abstract

The interactions between hosts and their microbial symbionts play a crucial role in shaping biological diversity and ecosystem function. Bacteria can adapt to specific host environments over evolutionary timescales, leading to co-speciation and the formation of specialized host-microbe relationships. Understanding the molecular mechanisms underlying these adaptations provides key insights into the evolution of symbiosis and the stability of microbial communities. This study investigates the co-speciation and molecular adaptations of *Curvibacter* to its host *Hydra*, a well-established model for the study of host-microbe interactions. We provide strong evidence of co-speciation, as demonstrated by phylogenetic congruence between different *Hydra* species and their corresponding *Curvibacter* symbionts, along with preferential recolonization of germ-free *Hydra* by their native *Curvibacter* strains. Comparative genomic analyses reveal that host-associated *Curvibacter* strains exhibit distinct metabolic and biosynthetic adaptations compared to their free-living relatives. Specifically, the enrichment of proteins involved in sugar metabolism and transport, as well as the selective purification of proteins linked to macromolecule biosynthesis, highlights the specialization of *Curvibacter* symbionts within the *Hydra* glycocalyx. Functional experiments identify a symbiont-specific extracellular polymeric substances (EPS) operon as key factor for microbial adhesion and host colonization, underscoring its role in facilitating symbiont specificity and stability. These findings provide insights into the molecular mechanisms driving host-microbe co-evolution and highlight the evolutionary forces shaping microbial specialization within host-symbiont relationships.

4.2 Introduction

Microbial communities that colonize host organisms enhance the fitness of the host by supporting a range of functions, including the provision of nutrients, the maturation of the immune system and the development of resistance to pathogens (Fraune and Bosch 2010; Mcfall-Ngai et al. 2013). In return, microbial communities benefit from the ecological niche provided by the host (Kopac and Klassen 2016; Obeng et al. 2021). Consequently, the host and its associated microbiota constitute a complex, integrated biological entity, designated as 'holobiont' or 'metaorganism' (Bosch and McFall-Ngai 2011; Rosenberg and Zilber-Rosenberg 2018; Zilber-Rosenberg and Rosenberg 2008), which is shaped by evolutionary pressures that drive mutual adaptation and specialization over time. Studying the underlying evolutionary processes

is critical for understanding of how bacteria affect metaorganism maintenance and fitness. The phenomenon of phylosymbiosis, whereby the similarity of associated microbial communities correlates with the evolutionary history of their hosts, has been documented across diverse host clades (Brooks et al. 2016; Franzenburg et al. 2013; Hayward et al. 2021; Lim and Bordenstein 2020). This observation implies that host evolution exerts a long-term influence on the composition of their symbiont communities. However, the extent to which shared environments or co-evolutionary processes between host and symbiont determine the observed microbial patterns remains unclear in most cases (Moran and Sloan 2015).

Due to its simplicity, we use the freshwater polyp *Hydra* and its symbiont *Curvibacter* as host-microbe model to study mechanisms of symbiont adaptation and specificity (Minten-Lange and Fraune 2020). *Hydra* is a small freshwater polyp belonging to the phylum Cnidaria, a sister group to all bilaterians. *Hydra* features a relatively simple body plan and lifestyle, facilitating efficient experimental analysis. Comparing the bacterial communities of different *Hydra* species maintained in the lab revealed a high degree of pyhlosymbiosis (Franzenburg et al. 2013). In addition, *Hydra* polyps sampled from the field are associated with similar bacterial communities as polyps from the laboratory (Fraune and Bosch 2007; Taubenheim et al. 2022). These associated microbial species are located either epibiotic within the glycocalyx of *Hydra* or endosymbiotic within the epithelial cells (Fraune et al. 2015; Fraune and Bosch 2007). The innate immune system of *Hydra* relies on a rich repertoire of antimicrobial peptides and an evolutionary conserved set of pattern recognition receptors (Augustin et al. 2009, 2017; Bosch et al. 2009; Franzenburg et al. 2012, 2013; Fraune et al. 2010; Jung et al. 2009; Klimovich and Bosch 2024). This innate immune repertoire protects the metaorganism *Hydra* against pathogens and maintains a homeostasis with beneficial microbes. Conversely, the *Hydra* microbiota has multiple effects on the host. Polyps with their microbiota experimentally removed (germ-free animals) exhibit behavioral changes (Murillo-Rincon et al. 2017), and are prone to fungal infection (Fraune et al. 2015). In addition, members of the microbiome can induce tumor development (Boutry et al. 2023; Rathje et al. 2020), asexual reproduction (Rahat and Dimentman 1982) and pattern formation which is mediated via the activation of host peptides antagonizing the Wnt signaling pathway (Taubenheim et al. 2020). The most abundant bacterial colonizer of *Hydra vulgaris* is *Curvibacter*, a Betaproteobacterium belonging to the family of *Comamonadaceae* (Fraune et al. 2015). *Curvibacter* colonizes the mucus-like layer of the ectodermal epithelial cells of *Hydra* together with several other microbial colonizers (Fraune

et al. 2015). *Curvibacter* is involved in antifungal activity (Fraune et al. 2015) and induces the production of the Eco host peptides, which act as Wnt antagonists and thereby influence the pattern formation and behavior of stem cells (Taubenheim et al. 2020). To sustain long-term bacterial functions vertical transmission of beneficial bacterial cells is crucial. In *Hydra*, *Curvibacter* is transferred during asexual reproduction by budding, whereby the bacterial cells passively migrate from the mother polyp to the bud (Franzenburg et al. 2013; Fraune et al. 2010). During embryogenesis, however, *Curvibacter* is absent in early stages, due to the antimicrobial peptide Periculin but colonizes the embryo in later stages, likely from the mother's tissue to the embryo's cuticle (Fraune et al. 2010). After hatching, *Curvibacter* attaches from the eggshell to newly hatched polyps, ensuring transmission to the next sexual generation (Franzenburg et al. 2013; Minten-Lange and Fraune 2020). In addition, *Curvibacter* can effectively colonize germ-free *Hydra* polyps when added to the surrounding medium (Fraune et al. 2015; Wein et al. 2018). This demonstrates that *Curvibacter* has maintained the ability to have a biphasic life cycle in which it can alternate between a host-associated and a free-living phase.

In this study, we address the specificity of different *Curvibacter* strains for their respective *Hydra* hosts. We observed strong congruent phylogenetic patterns between *Curvibacter* and its *Hydra* hosts, supporting the presence of co-speciation (Kawaida et al. 2013). Reciprocal recolonization experiments with three different *Hydra* species and their associated *Curvibacter* strains showed that all strains performed best in recolonizing their native *Hydra* hosts, emphasizing the evidence for co-speciation. To further investigate host adaptation within the *Curvibacter* genus, a comparative genomic analysis of free-living and host-associated species was performed, which revealed greater genetic similarity among host-associated *Curvibacter* strains compared to their free-living relatives. This analysis also identified an extracellular polymeric substances (EPS) operon more prevalent in host-associated *Curvibacter* strains. Knockout mutants lacking components of the EPS operon ($\Delta epsH$ and $\Delta wcaJ$) exhibited altered levels of three EPS-associated monosaccharides and reduced ability to recolonize host tissue. The findings indicate that *Curvibacter* has evolved specific adaptations to thrive within the *Hydra* host species, facilitated by exopolysaccharides that reinforce the symbiotic relationship with the host.

4.3 Material & Methods

4.3.1 Animal culture

Experiments were carried out using *Hydra vulgaris* (AEP), *Hydra oligactis* (strain 10/02), *Hydra viridissima* (strain A99), *Hydra magnipapillata* (strain 105), *Hydra vulgaris* (strain Basel), and *Hydra circumcincta* (strain M7). All animals were cultured under constant, identical environmental conditions including culture medium, food (first-instar larvae of Artemia salina, fed three times per week) and temperature according to standard procedures (Lenhoff and Brown 1970). For all experiments, adult polyps without buds or gonads were used.

4.3.2 Isolation of *Curvibacter* strains

Single *Hydra* polyps from each species were placed in a 1.5-ml reaction tube and washed three times with 1 ml sterile filtered *Hydra* medium. After homogenization with a pestle, 100 µl (equates to 1/10 of a polyp) was plated on Reasoner's 2A (R2A) agar plates (Sigma-Aldrich). After incubation at 18 °C for 5 days, single colony-forming units (CFUs) were isolated and cultivated in liquid R2A medium. The bacteria were identified by Sanger sequencing of the 16S rRNA gene using the universal primers Eub-27F and Eub-1492R (Weisburg et al. 1991) and stocks were stored in Roti-Store cryo vials (Carl Roth, Karlsruhe, Germany) at −80 °C.

4.3.3 Phylogenetic

For *Hydra*, COI sequences of the desired lineages were acquired from the NCBI database: *H. vulgaris* (AEP) (EF059935), *H. carnea* (EF059940), *H. magnipapillata* (EF059934), *H. vulgaris* (EF059936), *H. oligactis* (EF059937), *H. circumcincta* (EF059938) and *H. viridissima* (EF059941). Sequence alignment for the cytochrome oxidase genes was generated using Clustal W incorporated in MEGA11 sequence analysis software package (Tamura et al. 2021). A model test was used to estimate the best-fit substitution models for phylogenetic analyses. For the maximum-likelihood analyses, genes were tested using the General Time Reversible (GTR + I) model. A bootstrap test with 1,000 replicates for maximum likelihood using a random seed was conducted.

For *Curvibacter*, sequences for 16S rDNA were acquired of all isolates via Sanger-sequencing using the universal primers Eub-27F and Eub-1492R (Weisburg et al. 1991). Evolutionary analysis was conducted in MEGA11 (Tamura et al. 2021). All

sequences were aligned using the integrated Clustal W alignment option with default parameters. The evolutionary history was inferred by using the Maximum Likelihood method and Hasegawa-Kishino-Yano model (Hasegawa et al. 1985). A discrete Gamma distribution was used to model evolutionary rate differences among sites. Bootstrap values were calculated based on 100 replicates.

4.3.4 Generation of germfree *Hydra*

Polyps were transferred to a sterile beaker containing 30 ml of sterile S-medium supplemented with antibiotic solutions (50 µg/ml each of ampicillin, rifampicin, spectinomycin, streptomycin, and neomycin). To assess any potential impact of DMSO, which was used as a solvent for rifampicin, control polyps were treated with DMSO (1 µl/ml). The beaker was sealed airtight and stored at 18 °C in the dark. Over ten days, the medium and antibiotic solution were replaced every two days under sterile conditions as previously described (Franzenburg et al. 2012). Following antibiotic treatment, the GF polyps were washed with sterile S-medium, transferred to a fresh beaker, and incubated for an additional two days. As a contamination control, two polyps from each batch were homogenized in 100 µl of sterile S-medium using zirconia beads in a screw-cap tube and plated on R2A-Agar (ROTH) plates. After 3 days of incubation at room temperature, absence of CFUs indicated successful antibiotic treatment (Franzenburg et al. 2012). For culture-independent analysis, total DNA was extracted from single polyps using the DNeasy Blood & Tissue Kit (Qiagen). The 16S rRNA genes were amplified using the universal primers Eub-27F and Eub-1492R (Weisburg et al. 1991) in a 30-cycle PCR. Sterility was verified by the absence of a PCR-product.

4.3.5 Recolonization experiments

Curvibacter isolates were cultured in liquid R2A medium for 3 days at 18 °C. Following centrifugation at 1380 × g for 10 min, the bacterial pellet was resuspended in sterile *Hydra* medium and optical density (OD_{600}) was measured. For mono-association experiments, polyps were recolonized in *Hydra*-Medium with 5,000 *Curvibacter* cells per ml. For the recolonization experiments with the *Curvibacter* AEP1.3 mutants, polyps were recolonized with 50,000 *Curvibacter* cells per ml. Non-associated bacteria were removed by washing with sterile *Hydra* medium after 24 hours. After 7 days of recolonization, the S-medium was removed, and the *Hydra* polyps were carefully washed with sterile S-medium to eliminate residual bacteria. Each polyp was then transferred to an individual 1.5 ml screw-cap tube containing

100 µl of sterile S-medium. Subsequently, zirconia beads (1 mm) were added to the tubes, and the polyps were homogenized using a shaking homogenizer (BeadBug™). Homogenates were serially diluted (1:20 and 1:40), and 100 µl of each dilution was plated onto R2A-Agar plates. The plates were incubated at room temperature for 72 hours, after which colony-forming units (CFUs) were counted and adjusted for the respective dilution factors. For the mono-association experiments, statistical analysis of the bacterial load was conducted using one-way analysis of variance (ANOVA). Dunnett's test was used as a post hoc test to compare treatment with control samples. For the recolonization experiments with the *Curvibacter* AEP1.3 mutants, statistical analysis was performed using one-way ANOVA to compare group means, followed by pairwise t-tests with Bonferroni correction.

4.3.6 Genome sequencing

For sequencing the genomes of the two *Curvibacter* isolates *Curvibacter* Hmag1.1 and *Curvibacter* Hvul1 paired-end libraries were prepared using an Illumina TruSeq LT kit with a median fragment size of 402 bp. Mate-pair libraries were prepared using an Illumina Nextera mate-pair kit with an insert size of 7.3 kb. The libraries were sequenced on a MiSeq instrument at the Biomolecular Resource Facility, The Australian National University, Canberra, Australia. The reads were quality trimmed and adaptors were clipped using libngs `https://github.com/sylvainforet/libngs`. Mate-pair libraries were processed using NextClip (Leggett et al. 2014), keeping only read pairs in which the Nextera adaptor was found. The processed reads were assembled with SPAdes v3.5.0 (Bankevich et al. 2012). Gaps were filled using GapFiller v1-11 (Nadalin et al. 2012). Genes were predicted using Prokka (Seemann 2014). Genome sequences were uploaded to NCBI as nucleotide FASTA files and are deposited under the BioProject identifier PRJNA1232435.

4.3.7 Comparative genomic analysis

The E-Direct (22.1) software from NCBI was used to search and retrieve information on available *Curvibacter* assemblies. In total there were 48 assemblies. These assemblies were then filtered by the assembly completeness level, all assemblies annotated as "Contigs" were removed, resulting in a set of 18 *Curvibacter* assemblies. The *Curvibacter* assemblies (Supplementary Table S1, S8) were downloaded from NCBI. The downloaded nucleotide FASTA files were used as input for Prokka (1.14.6) to generate GenBank and proteome FASTA files (Seemann 2014). The resulting proteomes together with the two proteome FASTA files of *Curvibacter* Hmag1.1

and *Curvibacter* Hvul1 were used as input for CATHI (Becker et al. 2023) and OrthoFinder (2.5.5) (Emms and Kelly 2015, 2019). The synteny analysis was performed using clinker (0.0.27) (Gilchrist and Chooi 2021). Clinker requires GenBank files that exclusively contain the gene regions of interest. To meet this requirement, we developed a custom Python script designed to extract relevant gene regions from the GenBank files generated by Prokka. The script identifies these regions based on reciprocal best hits (RBHs) obtained through the CATHI pipeline, particularly we checked the synteny status of the genetic loci around RBHs of EpsE (WP_087496560.1), EpsF (WP_087496559.1) and EpsG (WP_087496558.1). Gene sequences within the synteny, that do not share identities above 25% with *Curvibacter* AEP1.3 were subjected to an additional BLAST analysis. This BLAST analysis was conducted using CATHI with a subset of the Reference Proteins database, containing only high-quality bacterial proteomes (assembly status "Complete Genome" or "Chromosome"), and an e-value cut-off of 0.05. The species tree was generated based on a set of orthologous gene trees by OrthoFinder, which is using the STAG algorithm (Emms and Kelly 2018). OrthoFinder results were further used for an orthogroup (OG) overlap analysis, to check the evolutionary conservatism among the tested *Curvibacter* strains. To minimize bias arising from differences in total OG count, the number of shared OGs between each pair of strains was normalized by dividing it with the total number of OGs in each strain, reciprocally. The inference of the symbiont specific OG set was done using a custom Python script using the OG table provided by OrthoFinder. Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations were retrieved using the BlastKOALA (Kanehisa et al. 2016) webservice with the *Curvibacter* AEP1.3 proteome as input file. Gene Ontology (GO) terms were retrieved by parsing the public available Reference Sequence database GFF annotation file of *Curvibacter* AEP1.3 (GCF_002163715.1). The KEGG enrichment analysis was performed using the clusterProfiler package (3.20) (Yu et al. 2012) of R, the GO analysis was performed using the goatools (Klopfenstein et al. 2018) Python package.

4.3.8 Preparation of defined medium

The defined medium was formulated using the M9 recipe as a template, with salt concentrations adjusted to match those of R2A. To support *Curvibacter* growth, nine amino acids and ammonium were added. The medium was buffered with HEPES to counteract acidification caused by *Curvibacter* metabolism. The M9 salt, trace elements, biotin, thiamine, NH_4Cl, HEPES, $MgSO_4$, $CaCl_2$ solutions, along with distilled water, were autoclaved. The 40% glucose and L-amino acid solutions were

filter-sterilized using a 0.22 µm filter. All chemicals were stored at 4 °C prior to use. The resulting solution was sterilized by filtration using a 0.22 µm bottle-top filter to remove any potential microbial contaminants.

Table 4.1: **Components of the defined medium**

Chemical	Concentration in media
Na_2HPO_4	337 µM
KH_2PO_4	220 µM
NaCl	880 µM
NH_4Cl	935 µM
Glucose	~22,2 µM
$CaCl_2$	300 µM
$MgSO_4$	1 mM
HEPES ($C_8H_{18}N_2O_4S$)	10 mM
NH_4Cl	10 mM
Histidine	100 mg/L
Methionine	100 mg/L
Asparagine	100 mg/L
Arginine	100 mg/L
Tryptophane	100 mg/L
Isoleucine	100 mg/L
Phenylalanine	100 mg/L
Threonine	100 mg/L
Aspartate	100 mg/L
Biotin	1 mg/L
Thiamin	1 mg/L
EDTA	134 µM
$FeCl_3$-$6H_2O$	31 µM
$ZnCl_2$	6,2 µM
$CuCl_2$-$2H_2O$	0,76 µM
$CoCl_2$-$2H_2O$	0,42 µM
H_3BO_3	1,62 µM
$MnCl_2$-$4H_2O$	0,081 µM
H_2O	-

4.3.9 Generation of *Curvibacter* AEP1.3 knockout strains

Pre-cultures of *Curvibacter* AEP1.3 for genomic knockout experiments were grown in liquid R2A medium at 30 °C and 170 RPM in a New Brunswick Innova 42 incubator for approximately two days before reaching stationary phase. *Escherichia coli* (*E. coli*) DH5α and *E. coli* DH5α WM3064 (*E. coli* ΔDAP) strains were revived from cryo-cultures onto LB-Agar plates. For the *E. coli* ΔDAP strain, 0.1 mM Diaminopimelic acid (DAP) was additionally included to support growth. For plasmid isolation, conjugation, and knockout experiments, *E. coli* cultures were inoculated from LB-Agar plates into liquid LB medium. The medium was supplemented with the appropriate antibiotics and/or DAP. *E. coli* cultures grew overnight at 30 °C with continuous shaking at 170 RPM.

The deletion of the *epsH* and *wcaJ* gene regions was performed using a homologous recombination strategy as described in Wein et al. (2018). Flanking regions for the knockouts of the *epsH* and *wcaJ* genes were designed using approximately 1000 base pairs of adjacent genomic sequences. To construct the desired plasmid, the first flanking region was introduced into the vector pGT42. The insertion was performed by restriction digestion with AleI followed by ligation, designed to restore the stop codon of the ampicillin resistance gene. The resulting plasmid solution was transformed into chemically competent *E. coli* DH5α using the heat-shock method. Transformants were screened on LB-agar plates supplemented with 500 µg/ml ampicillin and 25 µg/ml chloramphenicol. Colony PCR was performed to confirm successful integration of the first flanking region. Positive clones were cultured overnight in liquid LB medium supplemented with ampicillin. Plasmid DNA was isolated using the NucleoSpin Plasmid Quick Pure Kit (Macherey-Nagel, Germany) and verified by sequencing. Next, the second flanking region was introduced into the validated pGT42 plasmid via a similar restriction-ligation process using HpaI. Cultures were treated with kanamycin for selection, and the resulting plasmid sequence was confirmed via colony PCR, bacterial growth of positive clones in liquid LB medium supplemented with 25 µg/ml KanR, plasmid isolation, and sequencing as described above. The validated pGT42 vectors, containing both flanking regions, were introduced into *E. coli* ΔDAP cells using the heat-shock method. Transformants were screened via colony PCR, and positive clones were subsequently used for biparental mating with *Curvibacter* AEP1.3. The *epsH* and *wcaJ* gene targeting vectors are provided as GenBank files in the GitHub repository of this project. For conjugation, *Curvibacter* AEP1.3 (5 ml) and the *E. coli* ΔDAP donor strain (3 ml) were grown in liquid R2A medium to stationary phase. The *E. coli* ΔDAP culture was centrifuged at 3,000 g for 3 minutes, washed with liquid R2A, and centrifuged again.

The *Curvibacter* AEP1.3 culture was then added, and the combined suspension was centrifuged at 3,000 g for 3 minutes. The pellet was resuspended in 700 µl R2A supplemented with 0.1 mM DAP and incubated for 3–6 hours at 30 °C without shaking. A 100 µl aliquot of the conjugation suspension was spotted onto R2A-Agar plates with 0.1 mM DAP and incubated for 16–24 hours at 30 °C. Cells were then scraped from the plate, resuspended in 600 µl R2A (without DAP), and 100 µl of the suspension was spread on R2A-Agar containing 5 µg/ml kanamycin. After 48 hours, *Curvibacter* AEP1.3 clones were isolated and screened for integration of the kanamycin and SacB cassettes from pGT42, as well as for successful removal of the targeted gene regions by colony PCR.

4.3.10 CLARIOstar growth experiments

Pre-cultures of *Curvibacter* AEP1.3 for growth assays were grown in defined medium at 18 °C and 170 RPM in a New Brunswick Innova 42 incubator before reaching stationary phase. Growth experiments were performed using the CLARIOstar plate reader (BMG Labtech). Pre-cultures for growth assays were diluted to an OD_{600} of 0.05 in 800 µl defined medium per well and inoculated into 48-well plates. The plate reader was programmed to measure OD_{600} values every 10 minutes throughout the experiment. For each growth experiment, at least three wells were inoculated with 800 µl of defined medium alone to serve as controls for contamination assessment and to measure blank values. The experiments were conducted at room temperature (~21 °C – 24 °C – due to shaking) with continuous shaking of 500 RPM between measurements. The generation time g was calculated based on the standard exponential bacterial growth model, assuming bifurcations.

$$n = \frac{log(OD_{600}0.1) - log(OD_{600}0.05)}{log(2)} \tag{4.1}$$

$$g = \frac{time}{n} \tag{4.2}$$

The number of generations during the time of bacterial growth is denoted by the number n, time of bacterial growth is based on the duration of how long the bacteria grow from an OD_{600} 0.05 to 0.1. Statistical analysis included the Shapiro test to assess normal distribution, the Levene test to evaluate variance differences, and one-way ANOVA to determine differences in mean values. Pairwise comparisons were conducted using Tukey's HSD (honestly significant difference) test.

4.3.11 Scanning Electron Microscopy

Curvibacter AEP1.3 wt and the mutant strains Δ*epsH* and Δ*wcaJ* were grown in liquid defined medium at 18 °C for 72 hours. Before Scanning Electron Microscopy (SEM) sample preparation, bacterial cultures were diluted to an OD_{600} of 1.6. For bacterial cell fixation, cover glasses (1 cm diameter) were coated with Poly-L-Lysine. Prior to coating, the cover glasses were washed with 70% ethanol and dried using compressed air. Cover glasses were then placed into 24-well plates and incubated for 5 minutes in a 0.1% (w/v) Poly-L-Lysine solution prepared in distilled water. The coated cover glasses were dried with compressed air and transferred to a new 24-well plate with the coated side facing upwards. Each well was filled with 1 ml of defined medium. Subsequently, 20 µL of bacterial cultures or defined medium (as a control) was added to the wells containing the coated cover glasses. The solutions were gently mixed by pipetting. The 24-well plate was then centrifuged at 1,500 g for 15 minutes at 4 °C.

The medium was removed, and samples were incubated for 1 hour in 1 mL of fixation solution (0.1 M sodium cacodylate buffer, 2.5% glutaraldehyde, 2% formaldehyde in distilled water). Samples were then washed four times with washing solution (0.1 M sodium cacodylate buffer in distilled water). To ensure lipid fixation, an additional step was performed using 1 mL of osmium tetroxide fixation solution (1% OsO_4, 0.1 M sodium cacodylate buffer in distilled water) per sample, incubating for 1 hour at 4 °C in the dark. Afterward, samples were washed four more times with washing solution and dehydrated through a graded ethanol series (30%, 50%, 70%, 80%, 90%, 96%, and 100%). Each ethanol solution was applied twice for 15 minutes, starting with 30%. After dehydration, samples were immersed in 100% ethanol and subjected to critical point drying. Therefore, samples were washed six times with liquid CO_2 and the remaining CO_2 was slowly evaporated in a pressurized chamber to remove all traces of liquid EtOH.

Next, samples were transferred to a sample holder with an adhesive pad using tweezers. The samples were subjected to sputter coating, during which a gold layer was deposited using plasma at a pressure of 0.08–0.09 mbar in an argon atmosphere. Finally, the gold coated samples were transferred to the SEM Zeiss SUPRA 55VP machine for imaging.

4.3.12 Isolation, separation and staining of bacterial lipopolysaccharides

Lipopolysaccharides (LPS) from *Curvibacter* sp. JS11-12 and *Curvibacter* AEP1.3 were isolated and separated using the method described in (Michael R. Davis and Goldberg 2012). The bacterial cultures were incubated for 72 hours. The culture was diluted to OD_{600} of 0.5 and an aliquot of 1.5 ml was centrifuged at 10,600 g for 10 minutes and the pelleted bacteria were subjected to the LPS extraction protocol. Bacterial pellets were resuspended in 200 µl of 1×SDS lysis buffer (2% β-mercaptoethanol (BME), 2% SDS, 10 % glycerol in 50 mM Tris-HCl (pH 6.8) and bromophenol blue dye) by gentle pipetting to ensure complete resuspension. The samples were boiled in a water bath at 100 °C for 15 minutes and cooled to room temperature. To each sample, 5 µl each of DNase I and RNase solutions (238 µg/ml each) were added, followed by incubation at 37 °C for 30 minutes. Proteinase K solution (455 µg/ml) was then added and incubated at 59 °C overnight. Ice-cold Tris-saturated phenol (200 µl) was added to each sample. Tubes were tightly capped and vortexed for 5–10 seconds. The samples were incubated at 65 °C for 15 minutes with occasional vortexing, cooled to room temperature, and mixed with 1 ml of room-temperature diethyl ether. The samples were vortexed for 5–10 seconds and centrifuged at 20,600 g for 10 minutes. The bottom blue layer was carefully extracted, avoiding contamination from the upper clear layer. A small amount of the blue layer was intentionally left behind to minimize contamination risk. This extraction process was repeated two times. If samples remained cloudy, additional extractions were performed as needed. Following extraction, 200 µl of 2×SDS lysis buffer was added to each sample and stored at 4 °C. Samples were separated on 15% SDS-polyacrylamide gel, using 12 µl of prepared LPS per lane for visualization.
LPS fractions were stained using the Pro-Q™ Emerald 300 Lipopolysaccharide Gel Stain Kit (Thermo Fisher Scientific). The *E. coli* O55 LPS standard (provided within the Gel Stain Kit) was diluted 10-fold (to 250 µg/ml) in 2×SDS. The 15% SDS-polyacrylamide gel was immersed in a fixation solution (50% methanol and 5% acetic acid in distilled water) with gentle agitation on an orbital shaker for 45 minutes to fixate LPS. The fixation solution was carefully removed, and the process was repeated. The gel was incubated in the wash solution (3% glacial acetic acid in distilled water) for 15 minutes with gentle agitation on an orbital shaker. The washing solution was then removed, and the step was repeated. To oxidize LPS carbohydrates, the gel was immersed in the oxidizing solution (periodic acid as provided in the Kit and 250 ml of 3% acetic acid) by gentle agitation for 30 minutes.

Following oxidation, the gel was washed twice as described earlier. A staining buffer was prepared by adding 500 µL of Pro-Q® Emerald 300 stock solution to 25 ml of staining solution. The buffer was applied to the gel and incubated for 120 minutes with gentle agitation in the dark. The gel was then washed twice with the washing solution, following the previously described procedure. The LPS fractions were visualized using the Image Lab (Bio-Rad) software and the ChemiDoc XRS+ (Bio-Rad) imaging system.

4.3.13 EPS extraction

The growth of *Curvibacter* AEP1.3 for EPS extraction for carbohydrate analysis were carried out in a 50 ml culture in Erlenmeyer flasks at 18 °C, with continuous shaking at 170 RPM in a New Brunswick Innova 42 incubator for 72 hours. For the monosaccharide analysis, *Curvibacter* AEP1.3 growth for EPS extraction was conducted using the Multi-Cultivator (MC) MC-1000-OD (PSI). Each vial was inoculated with 60 ml of *Curvibacter* AEP1.3 pre-culture at an OD_{600} of 0.1. For each experiment, at least one vial was filled with defined medium as a control. Growth in the MC device was carried out in the dark at 18 °C. Cultures grew for nine days to ensure higher concentrations of EPS. After growth, each sample was divided into two 50 ml falcon tubes and centrifuged at 4,500 g for 10 minutes. Next, the supernatant was filtered through a 0.22 µm filter and transferred to a fresh 50 ml falcon. The samples were stored at −20 °C until further use. The cell-free supernatant was freeze-dried. Each lyophilized sample was then redissolved in 700 µl of autoclaved, distilled H_2O before combining pairs of samples. The resulting 1.4 mL solution was subjected to size exclusion chromatography (SEC) by applying the sample to a Superose 12 10/300 GL (Cytiva) column, separating small molecules from the defined medium and bacterial macromolecules. Fractions containing macromolecular components were pooled and vacuum-dried for carbohydrate analyses.

4.3.14 Analysis of total sugar abundance

Vacuum-dried samples were redissolved in 1 ml of autoclaved distilled water. Total sugar abundance was quantified using the Phenol-Sulfuric Acid method, as described by (Masuko et al. 2005) using a glucose standard curve. In a 96-well plate, 50 µl of each sample or standard was added in triplicate. To each well, 150 µl of concentrated sulfuric acid was added, and the solution was mixed by pipetting. Subsequently, 30 µl of a 5% phenol solution was added to each well, and the mixture was again pipetted to ensure homogeneity. The plate was incubated at 90 °C for 10 minutes

to facilitate the reaction and then cooled to room temperature for approximately 30 minutes. Absorbance values for the samples and the standard series were measured across a wavelength range of 400–550 nm using a CLARIOstar plate reader to determine total sugar concentrations. Glucose concentrations in the samples were calculated based on the absorbance at 490 nm.

4.3.15 Monosaccharide analysis

Monosaccharide composition analysis was conducted using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Pooled macromolecular fractions containing the EPS were subjected to trifluoroacetic acid (TFA) hydrolysis. For the hydrolysis the samples were dissolved in 2 M aqueous TFA and heated at 121 °C for 90 minutes, after which they were cooled on ice and centrifuged at 12,000 RPM for 5 minutes. The remaining acid was evaporated under a constant airflow at 40 °C for approximately 80 minutes. Following this, 300 µl of isopropanol was added, the solutions were vortexed and dried again under a constant airflow at 40 °C for 15 minutes. This step was repeated a second time. The samples were then dissolved in 500 µl water and subjected to anion-exchange chromatography utilizing a Knauer Azura HPAEC system (Knauer, Berlin). The CarboPac PA20 3×30 mm column was used as pre-column, while the CarboPac PA20 3×150 mm column was used as main column (Thermo Fisher). As a flow rate 0,4 ml/min was used. The gradient profile for eluents is described in the following table.

Table 4.2: **Gradient profile of the HPAEC-PAD analysis**

Time in minutes	2 mM NaOH in H_2O as %	150 mM NaOH in H_2O as %
Initial	100	0
21	100	0
22	0	100
25	0	100
26	100	0
35	100	0

Data acquisition and integration were performed using the ClarityChrom software (7.4.2.107, Knauer). The obtained data was subjected to a two-way ANOVA statistical test followed by a Bonferroni post-test.

4.3.16 Computational scripts

All custom mathematical and plotting operations were performed with Python 3.8.16. Scripts and additional information can be found on the GitHub repository of this project: `https://github.com/Kanomble/eps_project`. Software versions and all utilized packages and third-party tools are implemented in a public Docker image (kanomble/eps_project:1.2).

4.3.17 Supplementary data

The supplementary data for this manuscript are available as a compressed .zip archive in the data folder of the project's GitHub repository: `https://github.com/Kanomble/eps_project`. Tables S1, S2, and S8, along with all supplementary figures, are provided in a separate PDF file, while the remaining supplementary tables are included in a Microsoft Excel spreadsheet.

4.4 Results

4.4.1 Co-speciation of *Curvibacter* and *Hydra*

Recognizing that *Curvibacter* is a consistent part of the microbiome of many *Hydra* species (Franzenburg et al. 2013), we performed a cultivation effort to isolate *Curvibacter* strains from all species available in the laboratory. In total, this isolation procedure yielded 16 additional *Curvibacter* strains from six different *Hydra* species (Figure 4.1A). To investigate potential co-speciation between *Curvibacter* and *Hydra*, we compared the phylogeny of the six *Hydra* species (Figure 4.1A, left) with those of 16 corresponding *Curvibacter* strains and the already described strain *Curvibacter* AEP 1.3 (Pietschke et al. 2017) as well as the *Curvibacter* strain co-sequenced with the *Hydra magnipapillata* genome (NCBI:txid667019) (Figure 4.1A, right). Phylogenetic analysis of the 16S rRNA nucleotide sequences from *Curvibacter* revealed six distinct clusters, each of which aligns congruently with its respective *Hydra* host, as reflected in the phylogeny of the cytochrome C oxidase subunit I gene of the six *Hydra* species. The closely related *Curvibacter* species within each of the six clusters originate from the same *Hydra* species. This high congruency suggests a co-speciation of *Curvibacter* and *Hydra* since early *Hydra* evolution.

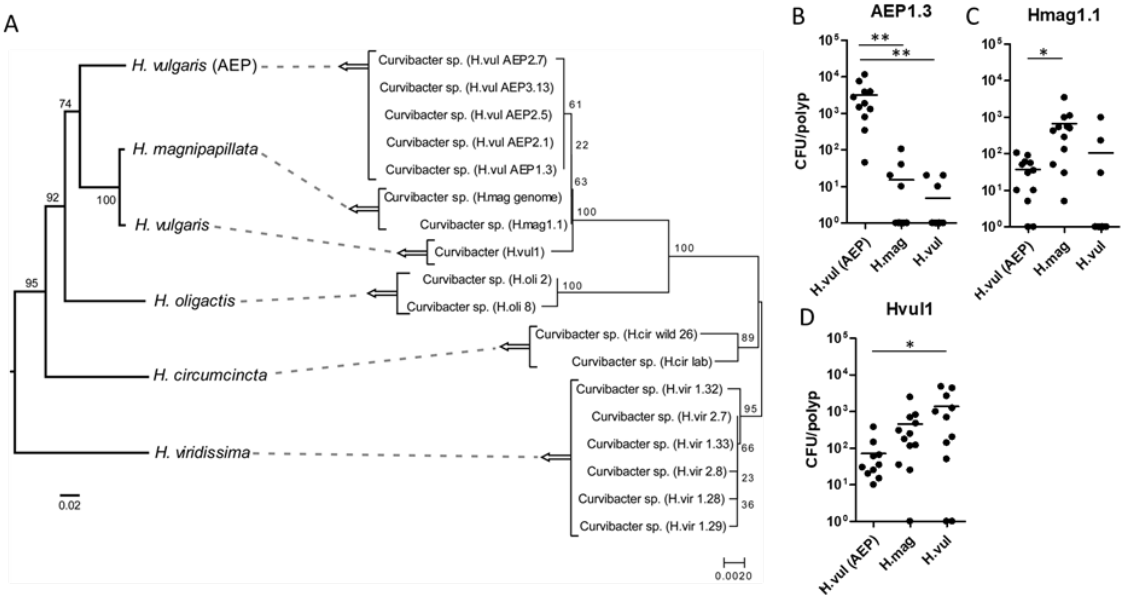

Figure 4.1: **Co-speciation of *Hydra* and their associated *Curvibacter* strains.** (**A**) Comparison of the phylogenetic trees of *Hydra* (left) and the associated *Curvibacter* species (right). (Left) Phylogenetic tree of *Hydra* species based on cytochrome oxidase genes (maximum likelihood, general time reversible (GTR+I)). Bootstrap values are shown at the corresponding nodes. The branch-length indicator displays 0.02 substitutions per site. (Right) Phylogenetic tree of *Curvibacter* isolates based on the 16S rRNA genes (maximum likelihood, Hasegawa-Kishino-Yano model). Bootstrap values are shown at the corresponding nodes. The branch-length indicator displays 0.002 substitutions per site. Note: the phylogenies of the *Hydra* hosts and their associated *Curvibacter* strains are congruent, indicating co-speciation. (**B-D**) Mono-colonization rates, measured as CFU per polyp, for three *Curvibacter* strains across three different *Hydra* species after 7 days. (n=12). Statistical significance (ANOVA, followed by a Bonferroni post-test) is indicated by asterisks, *$p < 0.05$, **$p < 0.01$.

To test whether the investigated *Curvibacter* strains have co-adapted to their *Hydra* hosts we performed reciprocal recolonization experiments with three different *Hydra* species and their corresponding *Curvibacter* isolates (Figure 4.1B-D). All tested *Curvibacter* strains recolonized their native *Hydra* host best (Figure 4.1B-D). The highest differences in colonization rate were evident for the recolonization with *Curvibacter* AEP1.3. Interestingly, the isolates Hmag1.1 and Hvul1 establish similar colonization rates on *H. magnipapillata* and *H. vulgaris* (Figure 4.1C, D). This observation is consistent with the close relationship of both *Hydra* species (Figure 4.1A).

4.4.2 Comparative genomics of free-living and *Hydra* associated *Curvibacter* strains

In the next step we aimed at analyzing the co-speciation of *Curvibacter* to the *Hydra* host on the genomic level. For this purpose, we sequenced the genomes from two isolates of *Hydra* associated *Curvibacter* strains (Hmag1.1, Hvul1) in addition to the already available genome from the strain AEP1.3 (Pietschke et al. 2017). In public databases several free-living *Curvibacter* species are reported and their genomes are available (Ding and Yokota 2004; Hahn et al. 2010; Lyu et al. 2024; Ma et al. 2016). Based on the genomes of three free-living strains (*Curvibacter delicatus* (GCA_041639495), *Curvibacter lanceolatus* (ATCC 14669) (GCF_000381265), *Curvibacter gracilis* (ATCC BAA-807) (GCF_000518645) and our three *Hydra*-associated strains (*Curvibacter* AEP1.3 (GCF_002163715), *Curvibacter* Hmag1.1, and *Curvibacter* Hvul1) a comparative genomic analysis was conducted (Table S1). To assess genetic differences among all selected *Curvibacter* strains, an all-vs-all orthogroup (OG) overlap analysis was performed (Figure 4.2A). The three symbiotic *Curvibacter* strains exhibit the highest similarities to each other with 84 to 99% of common OGs. This pattern is also evident for the two free-living *Curvibacter* strains, *C. lanceolatus* and *C. gracilis*, which both possess 97% of common OGs. Both strains show less similarities when compared to the symbiotic species with 51 to 52% of shared OGs. The *C. delicatus* genome contains the fewest protein coding genes, resulting in the lowest number of OGs (Table S2). *C. lanceolatus* and *C. gracilis*, which have the highest number of OGs, share the fewest OGs with *C. delicatus*. In contrast, the OG distribution of *C. delicatus* exhibits the highest similarity compared to the other two free-living species, sharing 90% and 91% of its OGs, respectively. This result is potentially driven by the reduced protein-coding sequences of *C. delicatus*. In comparison, it shares 82 to 83% of its OGs with host-associated *Curvibacter* strains.

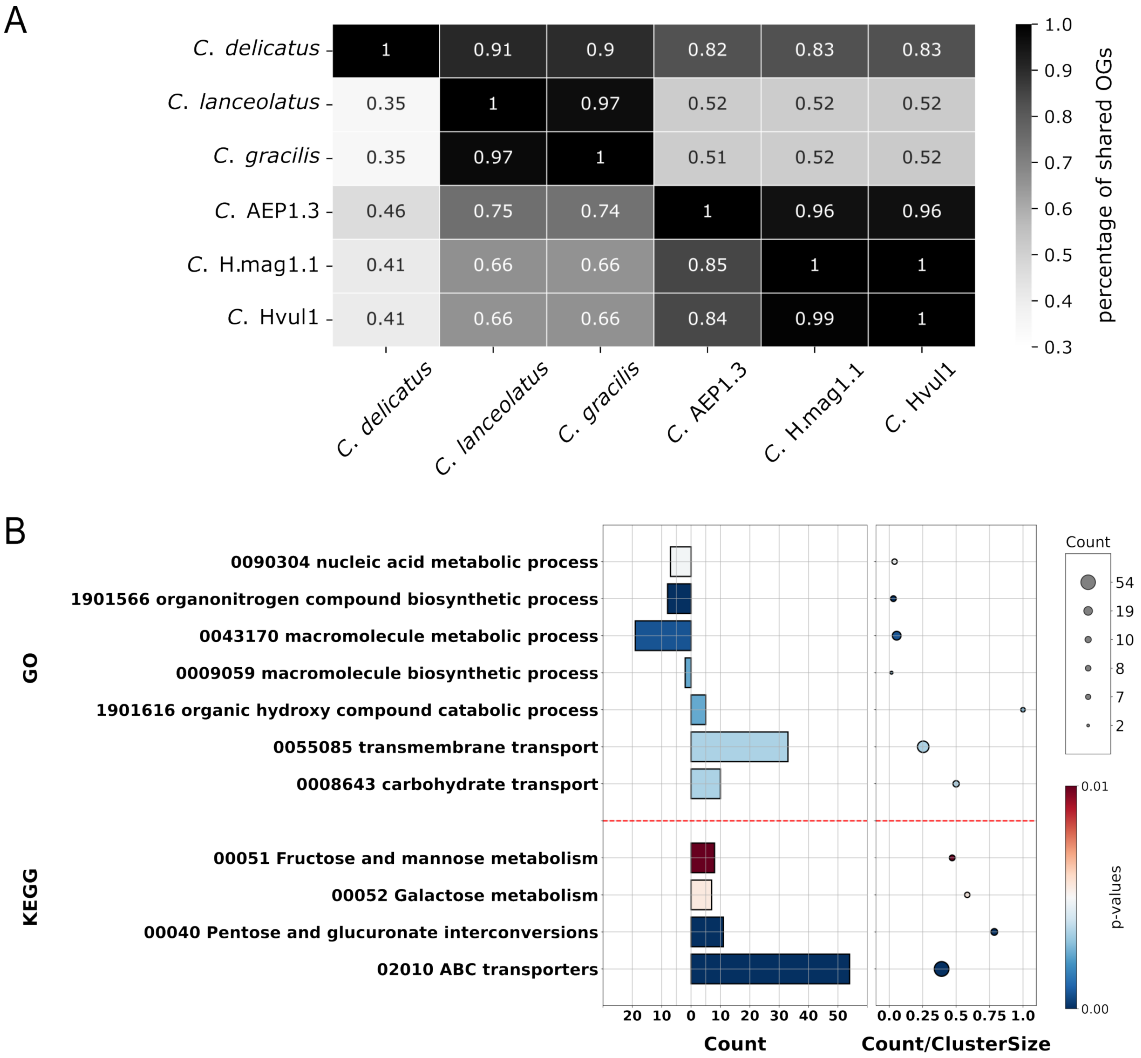

Figure 4.2: **Ortholog and KEGG/GO analysis of *Curvibacter* proteomes.** (**A**) Heatmap of orthogroup (OG) overlaps expressed as a percentage relative to the total OG count. The overlaps were normalized by dividing the OG overlap between two species by the total OG count of the species on the left axis. Highest similarities are observed within host-associated strains (C. AEP1.3, C. H.mag1.1 and C. Hvul1) and among the two free-living species *C. lanceolatus* and *C. gracilis*. (**B**) Selection of KEGG and GO enrichment analyses with the set of 693 protein sequences of *Curvibacter* AEP1.3 obtained from the symbiont specific set of OGs. The figure highlights a selection of the most notable categories, wherein the top four represent purified GO terms, and the remaining three correspond to enriched categories. The analyses reveal a significant enrichment of transporter-specific proteins (KEGG: ABC transporters; GO: carbohydrate transport and transmembrane transport). While three pathways related to sugar metabolism are enriched among the KEGG terms, the GO analysis highlights the enrichment of the organic hydroxy compound catabolic process, along with a purification of GO terms associated with the metabolic and biosynthetic pathways of macromolecules.

Subsequently, the OG data was filtered to identify OGs exclusive to the symbiotic strains. This analysis revealed 647 symbiont specific OGs (Table S3), with 693 of these genes present in *Curvibacter* AEP1.3. To assess functional enrichment within the symbiont specific OGs, a gene enrichment analysis was performed using the symbiont-specific protein sequence set of *Curvibacter* AEP1.3, incorporating

both GO terms and KEGG Orthology (KO) (Figure 4.2B). Overall, four KEGG categories show significant enrichment, while 15 and 20 GO categories were identified as enriched and purified, respectively (Table S4 and S5). Some GO terms form parent-child relationships within the GO hierarchy. Both enrichment analyses revealed a significant enrichment of transporter-specific proteins, particularly in the categories related to ABC transporters (KEGG), as well as carbohydrate and transmembrane transport (GO) (Figure 4.2B). Nearly 50% of transporters associated with carbohydrate transport (GO:0008643) are enriched, along with 45-75% of proteins involved in three KEGG pathways for sugar metabolism (ko00040, ko00052, and ko00051) (Figure 4.2B). Remarkably, all proteins within *Curvibacter* AEP1.3 associated with the organic hydroxy compound catabolic process category (GO:1901616) reside within the symbiont specific dataset. Additionally, proteins associated with macromolecule metabolism and biosynthesis (GO:0043170 and GO:0009059) exhibit significant purification. Taken together GO and KEGG terms associated with metabolic and transport pathways are significantly enriched and/or purified within the symbiont-specific gene set of 693 *Curvibacter* AEP1.3 genes, indicating that *Curvibacter* has undergone specific metabolic adaptations to thrive within its ecological niche in the glycocalyx of *Hydra*.

4.4.3 Symbiont specific genes reveal associations with an extracellular polymeric substance (EPS) operon

Within the symbiont-specific OGs we identified one gene annotated as epsH/exosortase B (WP_087496557.1) among the purified GO category 'macromolecule metabolic process' that is part of an operon potentially linked to the biosynthesis and transport of extracellular polymeric substances (EPS) (Haft et al. 2006; Yoshida et al. 2003). The complete operon consists of eleven genes (Figure 4.3A, Table S6). Among those eleven genes, five genes reside in the symbiont-specific gene set (Table S3, EpsA, EpsL, EpsD, EpsH and EpsI).

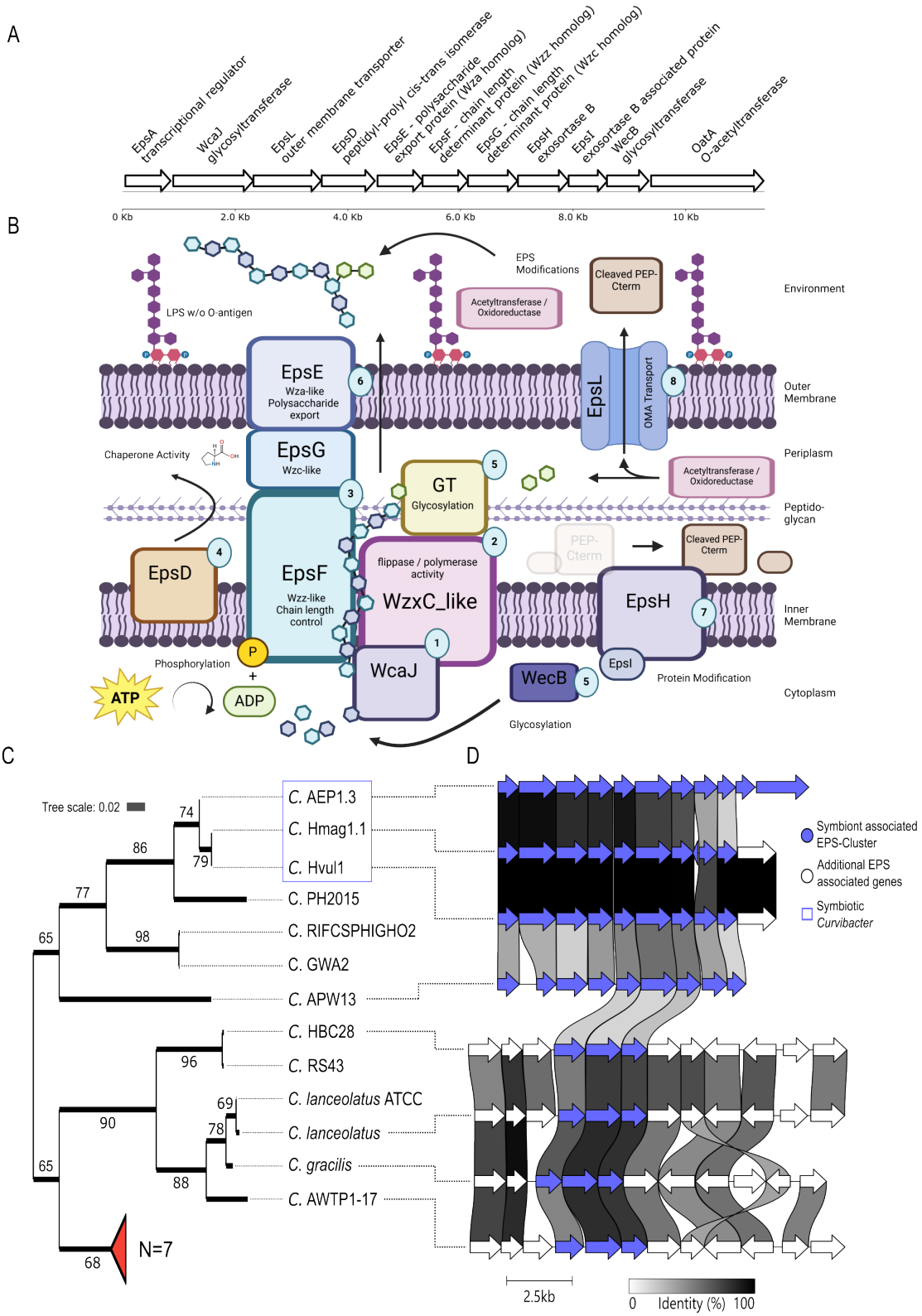

Figure 4.3: **Putative functional roles of the EPS operon enzymes and a species-level phylogeny of *Curvibacter* reflecting the syntenic conservation of this operon.**

The following section describes the content of Figure 4.3: (**A**) Locus of the EPS operon of *Curvibacter* AEP1.3. (**B**) Putative functions of the EPS operon and neighboring genes in *Curvibacter* AEP1.3 suggest involvement in exopolysaccharide production and biofilm formation, with components such as EpsF, EpsG, EpsE, and WcaJ sharing homology with known Wzy-polymerase biosynthesis pathway proteins (Cuthbertson et al. 2009; Yoshida et al. 2003). The figure was created with `https://BioRender.com`. (**C**) Orthofinder species tree inferred by the STAG algorithm from sets of gene trees. The values at each bipartition represent the percentage of occurrences of that specific bipartition across the set of inferred species trees. The phylogeny is in direct comparison to a syntenic EPS region (**D**). The synteny graph displays the conservation status of the eleven EPS operon genes of the three host-associated *Curvibacter* species and five representatives of the free-living *Curvibacter* species. The percentage identity of these genes decreases in free-living species. The red marked branch of the phylogeny shows a cluster of *Curvibacter* species without putative orthologous genes of the EPS operon. (1) WcaJ may act as a priming glycosyltransferase and performs initial glycosylation (Pal et al. 2019; Patel et al. 2012). (2) In the Wzy-Polymerase biosynthesis pathway, the growing glycan is transported across the inner membrane into the periplasm by the Wzx flippase. There is no homologous protein within the EPS operon, but a neighboring enzyme, located nine genes downstream of the O-acetyltransferase of the EPS operon, has a WzxC domain (WP_087496544.1) and may act as a flippase. (3) The enzymes EpsF and EpsG share structural similarities with polysaccharide co-polymerase (PCP) proteins. EpsF contains a Wzz domain, while EpsG features a Wzc domain. The Wzz protein functions as a regulator of glycan chain length during polysaccharide synthesis, whereas the Wzc protein may modulates the overall activity and export readiness of the polysaccharide synthesis machinery through changes in its phosphorylation state (Kintz et al. 2008; Larue et al. 2009; Reid and Whitfield 2005; Wugeditsch et al. 2001). (4) The EpsD enzyme is annotated as peptidyl-prolyl cis-trans isomerase, and it likely assists in the proper folding and stabilization of polysaccharide synthesis machinery proteins. (5) The EPS operon contains two glycosyltransferases, WcaJ and WecB, alongside four additional glycosyltransferases located in proximity (Figure S1 and Table S7). These enzymes are probably responsible for the sequential glycosylation of the polysaccharide, with each glycosyltransferase typically catalyzing the addition of a specific monosaccharide to the growing polymer chain, thereby contributing to the diversity and complexity of the polysaccharide structure. (6) Export of the polymer is likely mediated by the activity of EpsG and EpsE, which possesses a Wza domain. Wza acts as an outer

membrane polysaccharide exporter, forming a channel for the translocation of the completed polysaccharide chain (Nesper et al. 2003). (7) The EpsH exosortase is proposed to cleave the N-terminal PEP-C-term motif of proteins. These proteins are typically associated with clusters of EPS synthesis (Haft et al. 2006). Along with other enzymes such as acetyltransferases or oxidoreductases, it is hypothesized that they act as EPS modifying enzymes. (8) EpsL is an outer membrane transporter with a beta barrel domain and a signal peptide. It is likely located at the outer membrane of *Curvibacter* AEP1.3 where it forms a pore for diffusion of other enzymes or molecules. The Wzy-polymerase itself is not present within the EPS operon, however, in the genome of *Curvibacter* AEP1.3 two genes possess Wzy_C_2 domains, the corresponding enzymes may act similar to Wzy. However, the exact polymerization process of the EPS produced by the operon remains unknown.

The first gene in the operon, *epsA*, encodes a transcriptional regulator that contains an internal LuxR domain at its C-terminus but lacks an autoinducer-binding domain (Figure 4.3A). The second gene is homologous to *wcaJ*, a glycosyltransferase from *E. coli*, which functions as the priming glycosyltransferase in colanic acid polysaccharide synthesis (Pal et al. 2019; Patel et al. 2012). Many of these enzymes share similar domains with known components of the Wzy-polymerase biosynthetic machinery (Cuthbertson et al. 2009; Whitfield 2006), including those responsible for regulation (EpsA), transport (EpsL and EpsE), chain length control (EpsF and EpsG), protein modification (EpsD and EpsH/EpsI), glycosylation (WcaJ and WecB), and acetylation (OatA) (Figure 4.3B).

To assess the presence of this potential EPS synthesis and transport operon within different *Curvibacter* strains in a higher resolution, we retrieved 14 additional assemblies from free-living *Curvibacter* strains (Table S8). Using the CATHI software tool (Becker et al. 2023), a reciprocal BLAST analysis of the eleven EPS operon protein sequences was conducted against a database consisting of the 20 *Curvibacter* protein coding genes. The analysis revealed that nine of the eleven protein sequences have reciprocal best hits (RBH) within all three *Hydra*-associated *Curvibacter* strains. Furthermore, only the minority of free-living species possessed RBHs to the eleven protein sequences (Table S9).

Combining the OrthoFinder and synteny analysis we inferred a genome level phylogeny (Emms and Kelly 2018) (Figure 4.3C) as well as the conservation of the synteny of the EPS operon of all 20 *Curvibacter* strains (Figure 4.3D). All three *Hy-*

dra-associated strains form a monophyletic cluster and exhibit a high synteny conservation of the EPS operon (Figure 4.3C, D). Together with four free-living strains the *Hydra*-associated strains form a sister group to all other *Curvibacter* strains (Figure 4.3C). Interestingly, these four free-living species, similar to the symbiotic species, show homology to nine genes within the operon. However, the percentage identity of corresponding enzymes is highly reduced compared to identities observed in *Hydra*-associated strains. The syntenic status of the EPS operon is reduced to three genes within six other free-living *Curvibacter*, while seven *Curvibacter* species within this cluster lack the entire operon (Figure 4.3C, D).

Taken together, most enzymes in the EPS operon, except for EpsE, EpsF and EpsG, lack syntenic conservation in free living *Curvibacter* strains. Furthermore, conserved proteins exhibit lower sequence identity in free-living species compared to symbiotic strains. These results suggest potential functional divergence driven by different evolutionary pressures in free-living and symbiotic environments. This conclusion is supported by the fact that many protein sequences which are syntenically conserved exclusively within the free-living *Curvibacter* strains (Figure 4.3D, white arrows) also shared homologies with enzymes associated with EPS production (Table S10). In addition, in symbiotic strains several neighboring genes are predicted to be involved in EPS production, primarily comprising glycosyltransferases, oxidases, and acetylases, which are essential for determining the final EPS structure (Table S10).

4.4.4 *Curvibacter* AEP1.3 is producing extracellular polymeric substances

Based on the predicted functions (Figure 4.3B), it can be hypothesized that this operon is involved in the production of exopolysaccharide structures, such as capsular polysaccharides, glycosylated proteins and/or lipopolysaccharides (LPS). However, the exact functions of these genes remain unknown. Notably, two neighboring enzymes in the cluster are annotated as LPS biosynthesis protein (WP_087496544) and O-antigen ligase (WP_087496545). Consequently, an isolation of LPS was performed using the free-living *Curvibacter* strain JS11-12 and the symbiotic *Curvibacter* AEP1.3 (Figure 4.4A). The LPS was isolated from cultures grown in liquid R2A at 18 °C, reaching the stationary phase. While *Curvibacter* AEP1.3 exhibits distinct bands corresponding to lipid A and core oligosaccharides in *E. coli* O55, bands corresponding to the long chain O-antigen are absent (Figure 4.4A). Interestingly, the LPS extraction of the free-living species, *Curvibacter* sp. JS11-12, shows distinct bands corresponding to the long chain O-antigen polysaccharide in *E. coli*.

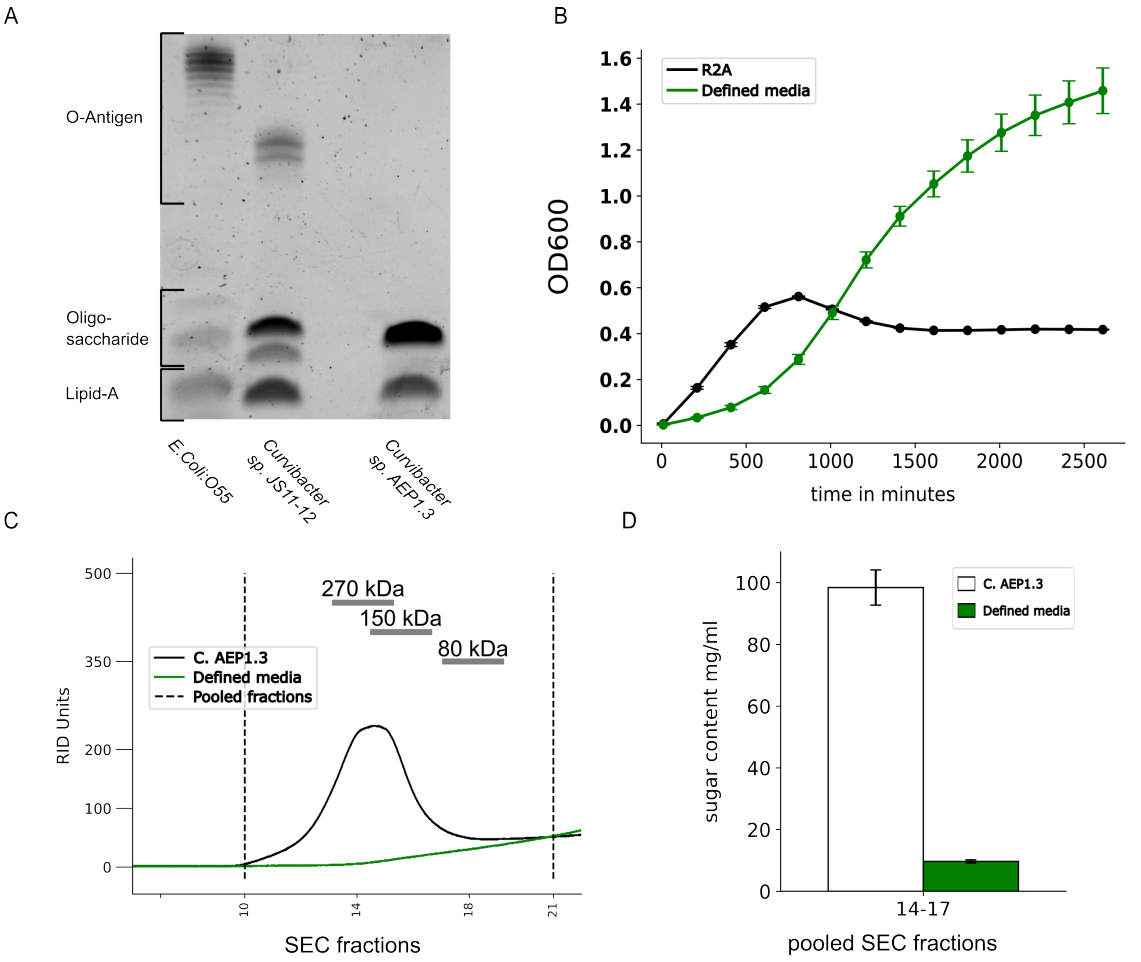

Figure 4.4: **Analysis of EPS obtained from *Curvibacter* AEP1.3.**

The following section describes the content of Figure 4.4: (**A**) 15%-SDS gel electrophoresis of Hot-Phenol-Ether isolated lipopolysaccharide (LPS) fractions. *E. coli* LPS O55, provided by the gel staining kit, was used as a standard (left lane), while the two additional LPS samples were extracted from the free-living *Curvibacter* sp. JS11-12 strain (obtained from the DSMZ) and *Curvibacter* AEP1.3. The upper bands of *E. coli* O55 and *Curvibacter* sp. JS11-12 represent long chain O-antigen regions, corresponding to varying numbers of O-antigen repeat units. The lower bands indicate intermediate and short-chain O-antigen or oligosaccharides, while the lowest bands correspond to the lipid A core region of LPS (Jacobson et al. 2018; Wang et al. 2002). Separation of LPS from *Curvibacter* AEP1.3 does not result in visible bands for the long-chain O-antigen region. (**B**) Growth curve of *Curvibacter* AEP1.3 cultured in the standard bacterial-freshwater medium R2A (black curve) and in a defined media (green curve) at room temperature. Although the generation times are longer in the defined medium (Figure S2), *Curvibacter* AEP1.3 achieves higher

OD values in this medium compared to R2A. (**C**) Size-Exclusion-Chromatographic analysis of the lyophilized and re-dissolved cell-free supernatant from *Curvibacter* AEP1.3 cultures (black), compared to lyophilized and re-dissolved defined medium as a negative control (green). In contrast to the negative control, both the Refractive Index Detector (RID) and the UV 280 nm detector (Figure S3) detected a signal in the higher molecular weight fractions of the SEC for the *Curvibacter* AEP1.3 wt supernatant. (**D**) Sugar content of *Curvibacter* AEP1.3 wt supernatant (white bars) and the defined media (green bars) as control of the higher molecular weight fractions obtained from the SEC measured with the Phenol-Sulfuric-Acid assay with glucose as standard (n=3). The detection of sugars in these fractions suggests that the high-molecular-weight molecules present in the supernatant are composed of polysaccharides.

Therefore, it is unlikely that the identified operon is responsible for the production of the O-antigen of LPS. In addition, the annotation and similarity of the potential EPS operon to the enzymes described by Yoshida et al. (2003) suggest the production of an exopolysaccharide that is not necessarily associated with the membrane. Therefore, we tested whether *Curvibacter* AEP1.3 secretes any sugar containing polymeric substances into the media. The growth medium R2A is not ideal for this purpose due to the presence of yeast extract and other polymeric substances, which influence the downstream analysis of the secreted substances. Therefore, we developed a defined growth medium for *Curvibacter* AEP1.3. Surprisingly, while exhibiting slightly reduced growth rates compared to R2A during logarithmic growth, the defined medium outcompetes R2A in terms of maximum OD_{600} values (Figure 4.4B).

In the next step, the defined medium was used to cultivate *Curvibacter* AEP1.3 at 18 °C for 76 hours until the stationary phase was reached. Subsequently, all secreted polymeric substances were isolated through centrifugation, filtration, freeze-drying, and size-exclusion chromatography (SEC). SEC revealed two distinct peaks measured by the refractive index detector (RID). The first, smaller peak appeared at the onset of the exclusion at fraction 10 and extended to fraction 21 (Figure 4.4C, black curve). As this peak is not present in the media control (Figure 4.4C), it is likely that molecules eluting at this retention time correspond to high-molecular-weight molecules secreted specifically by *Curvibacter* AEP1.3. The second, broader and larger peak appeared also in the medium control and may correspond to monosaccharides and amino acids present in the medium. In addition, there were also two peaks detected with the UV detector in the *Curvibacter* AEP1.3 sample (Figure S3). The

first peak, similar to the RID peak, indicates the presence of high-molecular-weight molecules that can absorb light at 280 nm, such as proteins or protein aggregates. The analysis of the carbohydrate concentration revealed a significantly higher sugar concentration in the *Curvibacter* AEP1.3 samples compared to the media control (Figure 4.4D). To summarize, while *Curvibacter* AEP1.3 has no distinct bands for long chain O-antigen polysaccharides, it produces and secretes polysaccharide based molecules into the medium.

4.4.5 EPS gene knockouts impact growth behavior, monosaccharide composition, and recolonization efficiency

To investigate the role of the EPS operon in the interactions with *Hydra*, knockout mutants of *Curvibacter* AEP1.3 were generated for the *epsH* and *wcaJ* genes (Figure 4.5A, B). The *epsH* gene was selected due to its presence in the enrichment analysis and because it is present within the other symbiotic strains and similarly has a very low number of RBHs in the reciprocal BLAST analysis (Figure S4 and Table S9). The *wcaJ* gene knockout was chosen due to its annotated function as a priming glycosyltransferase for the synthesis of the exopolysaccharide colanic acid (Pal et al. 2019; Patel et al. 2012).

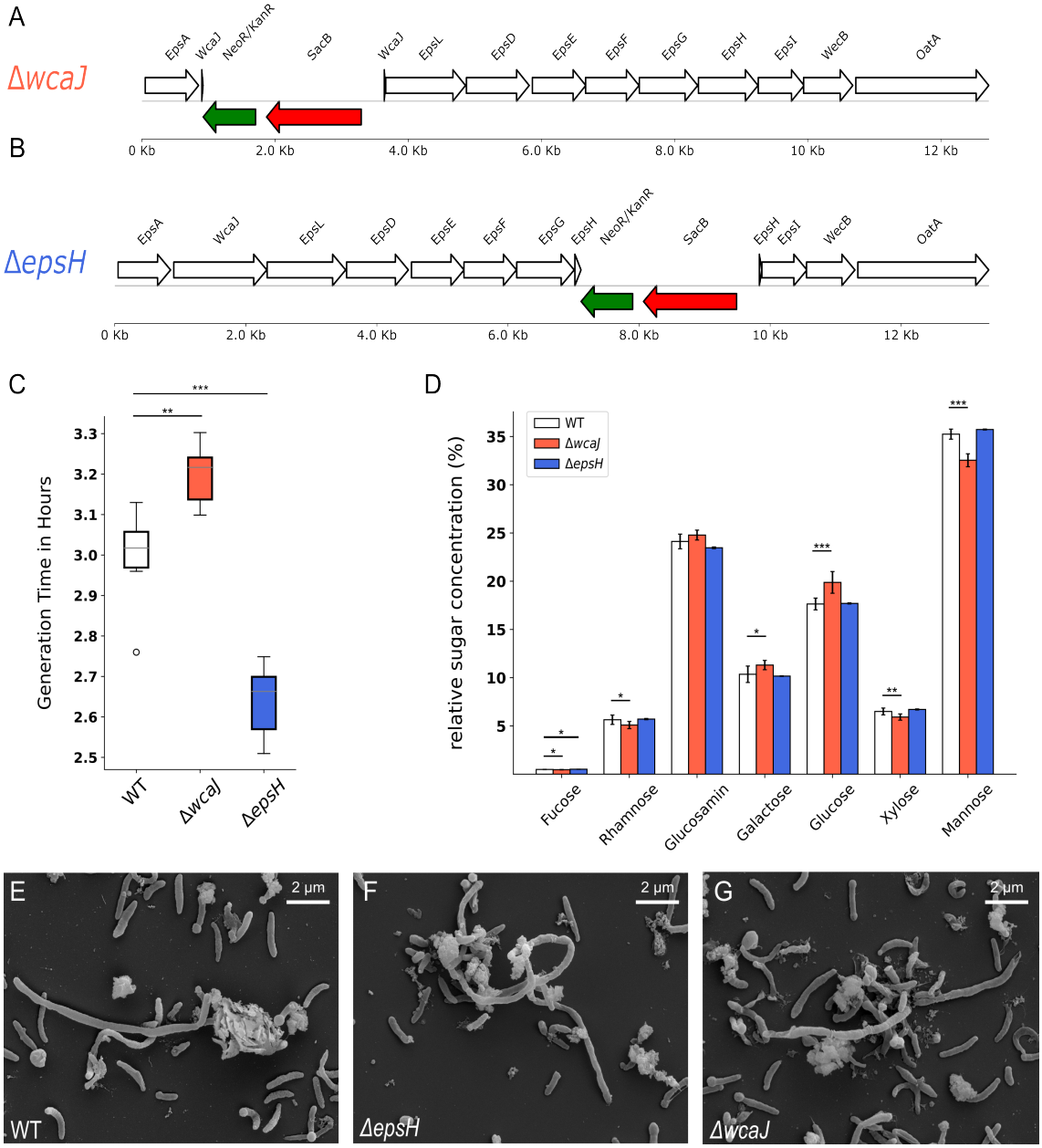

Figure 4.5: **Comparison of morphology, growth and monosaccharide abundance of *Curvibacter* AEP1.3 wt and two mutant lines.**

The following section describes the content of Figure 4.5: (**A**, **B**) Visualization of gene knockouts within the EPS operon of *Curvibacter* AEP1.3 mutant lines Δ*wcaJ* (**A**) and Δ*epsH* (**B**). (**C**) Generation times for the wt, Δ*wcaJ*, and Δ*epsH* strains of *Curvibacter* AEP1.3 (n=6) were calculated using the logarithmic growth model during the exponential growth phase. Specifically, generation time in hour was assessed during the initial logarithmic growth phase, defined as the increase in optical density from OD_{600} 0.05 to OD_{600} 0.1. The mean generation time is significantly higher in Δ*epsH* mutant and significantly lower in the Δ*wcaJ* mutant compared to the wt

strain. Statistical analyses were conducted using a one-way ANOVA, pairwise comparisons were conducted with Tukey's HSD test to identify significant differences. Statistical significance is indicated by asterisks, with the following meanings: *p < 0.05, **p < 0.01, ***p < 0.001. The OD_{600} growth curve is provided within the supplementary material (Figure S5). (**D**) Relative monosaccharide abundance of isolated Δ*epsH*, Δ*wcaJ* and wt EPS. The abundances reflect the relative sugar concentrations obtained from both Δ*wcaJ* and wt cultures (n=7), along with Δ*epsH* (n=3). Significant differences can be observed in the Mannose and Glucose abundances between wt and Δ*wcaJ*. Statistical analyses were conducted using a two-way ANOVA, p-values were adjusted using the Bonferroni correction method. Statistical significance is indicated by asterisks, with the following meanings: *p < 0.05, **p < 0.01, ***p < 0.001. (**E - G**) Scanning-Electron-Microscopy (SEM) graph of *Curvibacter* AEP1.3 strains. There are no visible morphological differences among the *Curvibacter* AEP1.3 wt (**E**) and the mutant lines Δ*epsH* (**F**) and Δ*wcaJ* (**G**).

The *epsH* and *wcaJ* knockouts were generated using the pGT42 vector-mediated double crossover technique (Wein et al. 2018), resulting in mutant strains that contain a kanamycin resistance gene and a *sacB* gene inserted at the loci of the *epsH* and *wcaJ* genes (Figure 4.5A, B, green and red arrows, respectively). Interestingly, the knockout strains exhibited significant differences in growth behavior compared to the wt strain (Figure 4.5C). While the Δ*epsH* strain growth faster, the Δ*wcaJ* strain growth significantly slower compared to the wt-strain. To investigate potential alterations in EPS sugar composition in the Δ*epsH* and Δ*wcaJ* mutants, we subjected the content of the high molecular polymer peak isolated by SEC (Figure 4.4C, region within dashed lines) to HPAEC-PAD analysis, determining the EPS monosaccharide composition.

The analysis revealed significant differences in the relative sugar abundances of the Δ*wcaJ* mutant (Figure 4.5D) with a higher relative abundance of glucose and galactose and a lower relative amount of mannose, xylose and rhamnose compared to the wt. In contrast, the Δ*epsH* mutant exhibited no significant changes in sugar composition, aside from a slight reduction in fucose levels. This finding suggests that, rather than directly affecting the synthesis and structure of EPS, EpsH may play a critical role in the proper assembly or secretion of surface proteins mediated by the selective cleavage of PEPC-term containing proteins. SEM (Figure 4.5E-G) revealed no visual morphological differences in the mutant strains compared to the wt strain.

To test the hypothesis that the EPS operon is involved in adaptation of *Curvibacter* to *Hydra*, mono-colonization experiments of germfree *Hydra* AEP were conducted (Figure 4.6A). The Δ*epsH* and Δ*wcaJ* mutants recolonized the germfree *Hydra* AEP polyps significantly worse compared to the wt strain (Figure 4.6B), supporting the hypothesis that the symbiont-specific EPS operon contributes to the co-speciation of *Hydra* and *Curvibacter*.

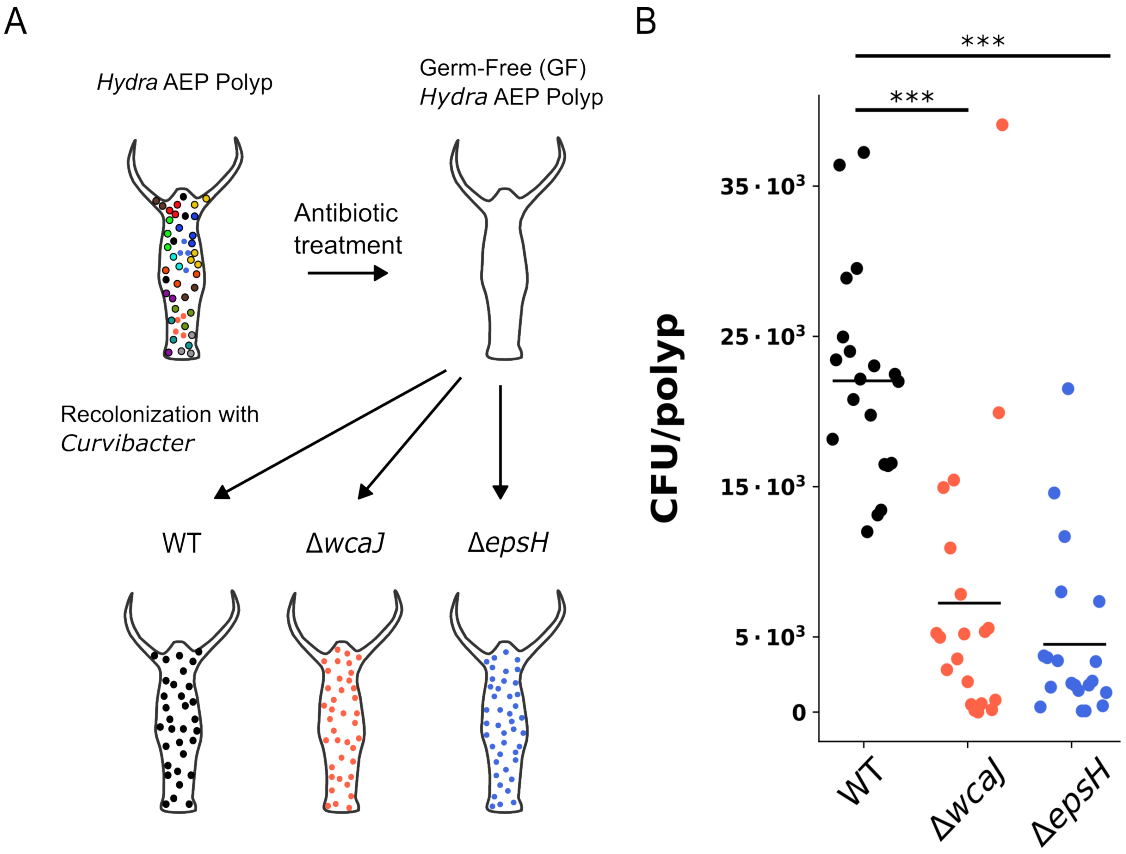

Figure 4.6: **The recolonization efficiency of *Curvibacter* AEP1.3 mutant lines is impaired compared to the wt.** (**A**) Scheme of recolonization of germ-free *Hydra* AEP polyps with *Curvibacter* AEP1.3 wt and the mutant strains Δ*wcaJ* and Δ*epsH*. (**B**) Recolonization rates after seven days recolonization, measured as CFU per polyp, for wt, Δ*wcaJ* and Δ*epsH* strains of *Curvibacter* AEP1.3. CFUs were counted for 20 polyps (n=20) per recolonization. The mutant strains exhibited significantly poorer performance compared to the wt strain; however, no significant differences were observed between the mutant lines themselves. Statistical analysis was performed using one-way ANOVA to compare group means, followed by pairwise t-tests with Bonferroni correction (***$p < 0.001$).

4.5 Discussion

4.5.1 Host-symbiont co-speciation and phylosymbiosis

The results presented in this study strongly support the concept of host-symbiont co-speciation between *Hydra* and its associated *Curvibacter* symbiont. Phylogenetic analyses revealed a congruent evolutionary pattern between the *Hydra* species and their bacterial symbionts of the genus *Curvibacter*, suggesting that evolutionary pressures have shaped the symbionts' genomes, driving them towards mutual adaptation. This result aligns with observations obtained from the *Hydra viridissima* – *Chlorella* symbiosis, in which similar congruent phylogenies were reported (Kawaida et al. 2013). Native strains consistently outperformed non-native strains in recolonization experiments, emphasizing their specialized adaptations for colonizing their respective hosts. Since *Curvibacter* colonizes the surface of the ectodermal epithelium of *Hydra* polyps (Fraune et al. 2015), while it is in direct contact with the aquatic environment, it is of great interest to understand how the symbiosis between *Curvibacter* and *Hydra* is established and maintained.

Our findings shed light on broader implications for phylosymbiosis and coevolution in host-microbe systems (Moran and Sloan 2015). Phylosymbiosis, the pattern where microbial community composition correlates with host phylogeny, has been observed across diverse host clades (Brooks et al. 2016). In mammals (Knowles et al. 2019), insects (Jackson et al. 2023; Qin et al. 2023), and invertebrates (O'Brien et al. 2020; Pollock et al. 2018), intraspecific microbiome variation is consistently lower than interspecific variation, and microbial community similarities show significant topological congruence with host phylogenies (Brooks et al. 2016). Phylosymbiosis appears to be driven by host-microbiome interactions, as evidenced by reduced survival and performance in interspecific microbiota transplants (Brooks et al. 2016). While host selection of the microbiome is often cited as a mechanism, other ecological and evolutionary processes such as dispersal, drift, and diversification may also contribute to phylosymbiotic patterns (Kohl 2020). In contrast to the community-level focus of phylosymbiosis, individual bacterial species often undergo coevolution with their hosts, resulting in highly specialized interactions. For instance, *Buchnera aphidicola* has co-evolved with aphids, providing essential nutrients that the host cannot obtain from its diet (Douglas 1998). This association, dating back 150 million years, has led to coevolution and codependence between the partners (Bennett and Moran 2015). However, this obligate symbiosis also presents potential draw-

backs, including metabolic costs, and symbiont genome degeneration, which may ultimately limit the ecological range and adaptability of aphid species (Bennett and Moran 2015). Similarly, great ape species show complex co-evolutionary dynamics with individual bacteria. Several bacterial lineages in human and ape guts have co-evolved with their hosts over the last 15 million years, with divergence times consistent with hominid evolution (Moeller et al. 2016). These distinct examples of phylosymbiosis and individual bacterial coevolution highlight the complexity and diversity of evolutionary processes shaping host-microbe relationships, from broad community-level patterns to specific, finely tuned symbiotic interactions.

In the case of *Hydra*, the innate immune system's ability to modulate microbial composition likely plays a pivotal role in maintaining phylosymbiosis (Franzenburg et al. 2013). By selectively fostering specific microbial communities by species-specific antimicrobial peptides, the host ensures compatibility and functional integration of its microbiota. On the other hand, individual bacterial symbionts, like *Curvibacter*, adapted to these host conditions ensuring the maintenance of association. By advancing our understanding of *Hydra*-symbiont co-speciation, this study contributes to the broader discourse on the evolutionary mechanisms underpinning host-microbiome dynamics. It raises compelling questions about how environmental pressures and host-specific traits influence microbial specialization, offering a framework for exploring the origins of phylosymbiosis within the *Hydra-Curvibacter* association.

4.5.2 Bacterial genomic adaptations to host organisms

The comparative genomic analysis of symbiotic and free-living *Curvibacter* strains highlights key adaptations that enable *Curvibacter* to thrive within the *Hydra* host environment. Further evidence of these adaptations is the identification of orthogroups that are unique to symbiotic strains, many of which are associated with functional categories relevant to host colonization and survival. Host-associated strains exhibit a distinct enrichment of genes involved in carbohydrate transport. These adaptations likely support the acquisition and utilization of host-derived resources, enabling the symbionts to sustain their populations within the *Hydra* environment. Furthermore, genes involved in macromolecule biosynthesis and metabolic processes exhibit significant purification, reflecting distinct evolutionary pressures acting on a specific subset of these genes. This genomic specialization may underscore the importance of metabolic plasticity in facilitating the transition from free-living to host-associated lifestyles. It is also known from other host-bacteria

models that host-associated microbes can undergo genetic diversification to exploit specific ecological niches. For instance, the coevolution of *Helicobacter pylori* with humans has resulted in lineage-specific adaptations that allow it to persist in the acidic environment of the stomach. The bacterium adapts to stomach acidity by increasing the use of positively charged amino acids in its membrane proteins (Xia and Palidwor 2005). *H. pylori* exhibit niche-specific adaptations within the stomach allowing the colonization of approximately 50% of the human population (Ailloud et al. 2019), causing overt gastric disease in only a subset of infected individuals (Salama et al. 2013).

Similar to *Curvibacter*, *Lactobacillus reuteri* strains exhibit host-specific genomic traits that enhance their ability to colonize the gastrointestinal tracts of various mammals. Comparative genomic analyses have revealed broad genetic diversity among *L. reuteri* strains from different hosts, with distinct phylogenetic clades showing host specificity (Frese et al. 2011; Yu et al. 2018). Experimental studies in gnotobiotic mice have shown that *L. reuteri* has evolved host specialization, with rodent-associated strains possessing a specific genomic content that enhances their ecological performance in the mouse gut (Frese et al. 2011). The genomic events shaping host specificity also include the acquisition of genes encoding cell surface proteins, active carbohydrate enzymes, and other functional genes that contribute to adaptation to specific intestinal habitats (Frese et al. 2013; Wegmann et al. 2015; Yu et al. 2018). In summary, the genomic adaptations observed in *Curvibacter*, in addition to the known parallels to other host-associated bacteria, underline a central role in genetic specialization for symbiotic relationships and niche adaptation within the host environment.

4.5.3 Role of exopolysaccharides in symbiont adaptations

Exopolysaccharides play a critical role in mediating the specificity and stability of the *Hydra-Curvibacter* symbiosis. The identification of a conserved EPS operon in host-associated *Curvibacter* strains underscores its importance in facilitating microbial adhesion and colonization within the host's glycocalyx. Functional analyses of this operon, particularly the *epsH* and *wcaJ* genes, revealed impaired recolonization capabilities of the corresponding mutant strains, suggesting that the production of specific glycans or glycoproteins is integral to maintaining the symbiotic relationship. These findings align with previous studies highlighting the role of EPS in microbial adhesion, biofilm formation (Vu et al. 2009), and immune evasion (Castro-Bravo et al. 2018; Fanning et al. 2012; Watters et al. 2016) in other host-microbe systems

(Gunn et al. 2016). EPS, primarily composed of polysaccharides and proteins form the matrix of microbial aggregates and biofilms (Sheng et al. 2010; Vu et al. 2009) mediate initial cell attachment to surfaces and protect bacteria from environmental stresses (Vu et al. 2009). In the context of host interactions EPS not only contribute to microbial adhesion and colonization but also play a crucial role in modulating host immune responses. Studies have shown that EPS can help symbiotic bacteria evade host immune detection by masking microbial-associated molecular patterns and inhibiting host immune receptors. For example, *Bifidobacterium breve* produces surface-associated EPS that enable evading adaptive B-cell responses and persistence in the gut (Fanning et al. 2012). Similarly, EPS from *Lactobacillus rhamnosus* forms a protective shield against host innate immune factors, enhancing its survival in the gastrointestinal tract (Lebeer et al. 2011). Furthermore, EPS produced by probiotic bacteria has been associated with various health benefits, including immune tolerance induction, anti-inflammatory effects, and protection against pathogens (Bhandary et al. 2023). Understanding the immunomodulatory properties of EPS in the *Hydra-Curvibacter* system could provide further insights into the evolution of host-microbe mutualisms and the molecular mechanisms underlying host immune tolerance.

EPS have also been shown to play pivotal roles in other host-symbiont interactions. In the squid-*Vibrio* symbiosis, for example, the EPS produced by the syp genes is essential for host colonization and biofilm formation (Shibata et al. 2012; Yip et al. 2005). Mutational studies have shown that most *syp* genes are required for successful colonization of the squid, with varying degrees of impact (Shibata et al. 2012). Similarly, in plant-*rhizobia* interactions, rhizobial EPS is essential for nodule formation and successful nitrogen fixation in leguminous plants (Acosta-jurado et al. 2021). The plant receptor Epr3 can recognize and distinguish compatible and incompatible EPS in bacterial competition studies (Kawaharada et al. 2015). The EPS produced by *Lactobacillus plantarum* play a crucial role in strain-specific probiotic effects and host interactions. *L. plantarum* strains possess multiple gene clusters responsible for EPS production, with varying impacts on surface glycan composition (Remus et al. 2012). The removal of EPS can alter bacterial surface properties, stress survival, and immunomodulatory capacities in a strain-specific manner (Lee et al. 2016). The strain-specific nature of EPS contributes to the variability in probiotic efficacy observed among different strains (Bron et al. 2013).

Similarly, the conservation of EPS-related genes in symbiotic *Curvibacter* strains (Figure 3C,D) illustrates their adaptation to the *Hydra* niche. Comparative genomic analyses revealed that these genes are absent or poorly conserved in free-living strains, indicating their evolutionarily recent acquisition or specialization in host-associated strains. This specialization likely enhances the symbiont's ability to persist in the host's dynamic environment, which is characterized by antimicrobial peptides and other selective pressures. Moreover, the distinct sugar compositions of EPS from mutant strains, as well as the significant differences in generation times compared to the wt suggest that these polysaccharides play a specific role in mediating host-symbiont interactions, potentially influencing factors such as microbial adhesion and host immune modulation. Future studies should be aimed at determining the specific structure of the *Curvibacter* EPS polymer in more detail.

4.6 Conclusion

This study elucidates the intricate dynamics of the *Hydra-Curvibacter* symbiosis, highlighting the evolutionary and molecular mechanisms that underpin this relationship. The evidence for host-symbiont co-speciation underscores the deep evolutionary connections between hosts and their symbionts, while the identification of specialized adaptations such as the EPS operon showcases the molecular innovations that enable *Curvibacter* to thrive in host environments. By integrating genomic and functional analyses, this research advances our understanding of the evolutionary processes that underpin host-microbe associations, offering valuable insights into the dynamics of microbial adaptation and the resilience of host-microbiome systems.

4.7 Appendix

4.7.1 Contributions

Contributions are mentioned in chapter 3: List of Manuscripts and Contributions.

4.7.2 Acknowledgements

We thank Katharina Grosche for excellent assistance with the carbohydrate analyses.

4.7.3 Funding

This study was supported by the DFG (Project FR 3041/3-1). MK is supported by the DFG (Project 517563163). MP acknowledges support from the Cluster of Excellence on Plant Sciences (CEPLAS) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2048/1–Project ID: 390686111.

4.7.4 Data availability

All custom scripts and data are available in the project's GitHub repository: `https://github.com/Kanomble/eps_project`. The supplementary material is also included there, as a .zip file located in the data directory.

References for Manuscript 1

Acosta-jurado, Sebastián, Francisco Fuentes-romero, Jose Enrique Ruiz-sainz, Monika Janczarek, and José María Vinardell (June 2021). "Rhizobial exopolysaccharides: Genetic regulation of their synthesis and relevance in symbiosis with legumes". In: *International Journal of Molecular Sciences* 22 (12). ISSN: 14220067. DOI: 10.3390/ijms22126233.

Ailloud, Florent et al. (May 2019). "Within-host evolution of Helicobacter pylori shaped by niche-specific adaptation, intragastric migrations and selective sweeps". In: *Nature Communications* 10 (1), p. 2273. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10050-1.

Augustin, René, Stefan Siebert, and Thomas C.G. Bosch (2009). "Identification of a kazal-type serine protease inhibitor with potent anti-staphylococcal activity as part of Hydra's innate immune system". In: *Developmental and Comparative Immunology* 33 (7), pp. 830–837. ISSN: 0145305X. DOI: 10.1016/j.dci.2009.01.009.

Augustin, René et al. (Sept. 2017). "A secreted antibacterial neuropeptide shapes the microbiome of Hydra". In: *Nature Communications* 8 (1), p. 698. ISSN: 2041-1723. DOI: 10.1038/s41467-017-00625-1.

Bankevich, Anton et al. (May 2012). "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing". In: *Journal of Computational Biology* 19 (5), pp. 455–477. ISSN: 1066-5277. DOI: 10.1089/cmb.2012.0021.

Becker, Lukas et al. (2023). "CATHI: An interactive platform for comparative genomics and homolog identification". In: *bioRxiv*. DOI: https://doi.org/10.1101/2023.09.04.556229.

Bennett, Gordon M. and Nancy A. Moran (Aug. 2015). "Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole". In: *Proceedings of the National Academy of Sciences* 112 (33), pp. 10169–10176. ISSN: 0027-8424. DOI: 10.1073/pnas.1421388112.

Bhandary, Thejaswi et al. (Mar. 2023). "Exopolysaccharides Derived from Probiotic Bacteria and their Health Benefits". In: *Journal of Pure and Applied Microbiology* 17 (1), pp. 35–50. ISSN: 09737510. DOI: 10.22207/JPAM.17.1.40.

Bosch, Thomas, René Augustin, Friederike Anton-Erxleben, Sebastian Fraune, Georg Hemmrich, et al. (Apr. 2009). "Uncovering the evolutionary history of innate immunity: The simple metazoan Hydra uses epithelial cells for host defence". In:

Developmental & Comparative Immunology 33 (4), pp. 559–569. ISSN: 0145305X. DOI: 10.1016/j.dci.2008.10.004.

Bosch, Thomas C.G. and Margaret J. McFall-Ngai (2011). "Metaorganisms as the new frontier". In: *Zoology* 114 (4), pp. 185–190. ISSN: 09442006. DOI: 10.1016/j.zool.2011.04.001.

Boutry, Justine, Marie Buysse, Sophie Tissot, Chantal Cazevielle, Rodrigo Hamede, et al. (May 2023). "Spontaneously occurring tumors in different wild-derived strains of hydra". In: *Scientific Reports* 13 (1), p. 7449. ISSN: 2045-2322. DOI: 10.1038/s41598-023-34656-0.

Bron, Peter A, Satoru Tomita, Annick Mercenier, and Michiel Kleerebezem (June 2013). "Cell surface-associated compounds of probiotic lactobacilli sustain the strain-specificity dogma". In: *Current Opinion in Microbiology* 16 (3), pp. 262–269. ISSN: 13695274. DOI: 10.1016/j.mib.2013.06.001.

Brooks, Andrew W., Kevin D. Kohl, Robert M. Brucker, Edward J. van Opstal, and Seth R. Bordenstein (2016). "Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History". In: *PLoS Biology* 15 (1), pp. 1–29. ISSN: 15457885. DOI: 10.1371/journal.pbio.1002587.

Castro-Bravo, Nuria, Jerry M. Wells, Abelardo Margolles, and Patricia Ruas-Madiedo (Oct. 2018). "Interactions of Surface Exopolysaccharides From Bifidobacterium and Lactobacillus Within the Intestinal Environment". In: *Frontiers in Microbiology* 9, p. 2426. ISSN: 1664-302X. DOI: 10.3389/fmicb.2018.02426.

Cuthbertson, Leslie, Iain L. Mainprize, James H. Naismith, and Chris Whitfield (Mar. 2009). "Pivotal Roles of the Outer Membrane Polysaccharide Export and Polysaccharide Copolymerase Protein Families in Export of Extracellular Polysaccharides in Gram-Negative Bacteria". In: *Microbiology and Molecular Biology Reviews* 73 (1), pp. 155–177. ISSN: 1092-2172. DOI: 10.1128/MMBR.00024-08.

Ding, Linxian and Akira Yokota (Nov. 2004). "Proposals of Curvibacter gracilis gen. nov., sp. nov. and Herbaspirillum putei sp. nov. for bacterial strains isolated from well water and reclassification of [Pseudomonas] huttiensis, [Pseudomonas] lanceolata, [Aquaspirillum] delicatum and [Aquaspirillum] autotrophicum as Herbaspirillum huttiense comb. nov., Curvibacter lansceolatus comb. nov., Curvibacter delicatus comb. nov." In: *International Journal of Systematic and Evolutionary Microbiology* 54 (6), pp. 2223–2230. ISSN: 14665026. DOI: 10.1099/ijs.0.02975-0.

Douglas, A. E. (Jan. 1998). "Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera". In: *Annual Review of Entomology* 43 (1), pp. 17–37. ISSN: 0066-4170. DOI: 10.1146/annurev.ento.43.1.17.

Emms, D.M. and S. Kelly (Feb. 2018). *STAG: Species Tree Inference from All Genes*. DOI: 10.1101/267914.

Emms, David M. and Steven Kelly (Aug. 2015). "OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy". In: *Genome Biology* 16 (1), p. 157. ISSN: 1474-760X. DOI: 10.1186/s13059-015-0721-2.

— (Dec. 2019). "OrthoFinder: phylogenetic orthology inference for comparative genomics". In: *Genome Biology* 20 (1), p. 238. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1832-y.

Fanning, Saranna, Lindsay J. Hall, and Douwe van Sinderen (Sept. 2012). "Bifidobacterium breve UCC2003 surface exopolysaccharide production is a beneficial trait mediating commensal-host interaction through immune modulation and pathogen protection". In: *Gut Microbes* 3 (5), pp. 420–425. ISSN: 1949-0976. DOI: 10.4161/gmic.20630.

Franzenburg, Sören, Sebastian Fraune, Sven Künzel, John F. Baines, Tomislav Domazet-Lošo, et al. (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), pp. 19374–19379. ISSN: 00278424. DOI: 10.1073/pnas.1213110109.

Franzenburg, Sören, Jonas Walter, Sven Künzel, Jun Wang, John F. Baines, et al. (Sept. 2013). "Distinct antimicrobial peptide expression determines host species-specific bacterial associations". In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (39). ISSN: 00278424. DOI: 10.1073/pnas.1304960110.

Fraune, Sebastian, Friederike Anton-Erxleben, René Augustin, Sören Franzenburg, Mirjam Knop, et al. (July 2015). "Bacteria-bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance". In: *ISME Journal* 9 (7), pp. 1543–1556. ISSN: 17517370. DOI: 10.1038/ismej.2014.239.

Fraune, Sebastian, René Augustin, Friederike Anton-Erxleben, Jörg Wittlieb, Christoph Gelhaus, et al. (Oct. 2010). "In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides". In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (42), pp. 18067–18072. ISSN: 00278424. DOI: 10.1073/pnas.1008573107.

Fraune, Sebastian and Thomas C. G. Bosch (July 2010). "Why bacteria matter in animal development and evolution". In: *BioEssays* 32 (7), pp. 571–580. ISSN: 0265-9247. DOI: 10.1002/bies.200900192.

— (2007). "Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra". In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (32), pp. 13146–13151. ISSN: 00278424. DOI: 10.1073/pnas.0703375104.

Frese, Steven A. et al. (Feb. 2011). "The Evolution of Host Specialization in the Vertebrate Gut Symbiont Lactobacillus reuteri". In: *PLoS Genetics* 7 (2). Ed. by David S. Guttman, e1001314. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1001314.

Frese, Steven A. et al. (Dec. 2013). "Molecular Characterization of Host-Specific Biofilm Formation in a Vertebrate Gut Symbiont". In: *PLoS Genetics* 9 (12). Ed. by Danielle A. Garsin, e1004057. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1004057.

Gilchrist, Cameron L M and Yit-Heng Chooi (Aug. 2021). "clinker & clustermap.js: automatic generation of gene cluster comparison figures". In: *Bioinformatics* 37 (16). Ed. by Peter Robinson, pp. 2473–2475. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btab007.

Gunn, John S., Lauren O. Bakaletz, and Daniel J. Wozniak (June 2016). "What's on the Outside Matters: The Role of the Extracellular Polymeric Substance of Gram-negative Biofilms in Evading Host Immunity and as a Target for Therapeutic Intervention". In: *Journal of Biological Chemistry* 291 (24), pp. 12538–12546. ISSN: 00219258. DOI: 10.1074/jbc.R115.707547.

Haft, Daniel H, Ian T Paulsen, Naomi Ward, and Jeremy D Selengut (Dec. 2006). "Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic". In: *BMC Biology* 4 (1), p. 29. ISSN: 1741-7007. DOI: 10.1186/1741-7007-4-29.

Hahn, Martin W. et al. (June 2010). "Limnohabitans curvus gen. nov., sp. nov., a planktonic bacterium isolated from a freshwater lake". In: *International Journal of Systematic and Evolutionary Microbiology* 60 (6), pp. 1358–1365. ISSN: 1466-5026. DOI: 10.1099/ijs.0.013292-0.

Hasegawa, Masami, Hirohisa Kishino, and Taka-Aki Yano (1985). "Journal of Molecular Evolution Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA". In: *J Mol Evol* 22, pp. 160–174. DOI: 10.1007/BF02101694.

Hayward, Alexander, Robert Poulin, and Shinichi Nakagawa (Aug. 2021). "A broad-scale analysis of host-symbiont cophylogeny reveals the drivers of phylogenetic congruence". In: *Ecology Letters* 24 (8). Ed. by Helene Morlon, pp. 1681–1696. ISSN: 1461-023X. DOI: 10.1111/ele.13757.

Jackson, Raphaella et al. (May 2023). "Evidence of phylosymbiosis in Formica ants". In: *Frontiers in Microbiology* 14. ISSN: 1664-302X. DOI: 10.3389/fmicb.2023.1044286.

Jacobson, Amy N., Biswa P. Choudhury, and Michael A. Fischbach (May 2018). "The Biosynthesis of Lipooligosaccharide from Bacteroides thetaiotaomicron". In: *mBio* 9 (2). Ed. by David A. Relman, pp. 1–14. ISSN: 2161-2129. DOI: 10.1128/mBio.02289-17.

Jung, Sascha, Andrew J. Dingley, René Augustin, Friederike Anton-Erxleben, Mareike Stanisak, et al. (2009). "Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan hydra". In: *Journal of Biological Chemistry* 284 (3), pp. 1896–1905. ISSN: 00219258. DOI: 10.1074/jbc.M804713200.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima (2016). "BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences". In: *Journal of Molecular Biology* 428 (4), pp. 726–731. ISSN: 10898638. DOI: 10.1016/j.jmb.2015.11.006.

Kawaharada, Y. et al. (July 2015). "Receptor-mediated exopolysaccharide perception controls bacterial infection". In: *Nature* 523 (7560), pp. 308–312. ISSN: 0028-0836. DOI: 10.1038/nature14611.

Kawaida, Hitomi et al. (Mar. 2013). "Symbiosis between hydra and chlorella: Molecular phylogenetic analysis and experimental study provide insight into its origin and evolution". In: *Molecular Phylogenetics and Evolution* 66 (3), pp. 906–914. ISSN: 10557903. DOI: 10.1016/j.ympev.2012.11.018.

Kintz, Erica, Jennifer M. Scarff, Antonio DiGiandomenico, and Joanna B. Goldberg (Apr. 2008). "Lipopolysaccharide O-Antigen Chain Length Regulation in Pseudomonas aeruginosa Serogroup O11 Strain PA103". In: *Journal of Bacteriology* 190 (8), pp. 2709–2716. ISSN: 0021-9193. DOI: 10.1128/JB.01646-07.

Klimovich, Alexander and Thomas Bosch (2024). "Novel technologies uncover novel 'anti'-microbial peptides in Hydra shaping the species-specific microbiome". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 379 (1901). ISSN: 14712970. DOI: 10.1098/rstb.2023.0058.

Klopfenstein, D. V., Liangsheng Zhang, Brent S. Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, et al. (2018). "GOATOOLS: A Python library for Gene

Ontology analyses". In: *Scientific Reports* 8 (1), pp. 1–17. ISSN: 20452322. DOI: 10.1038/s41598-018-28948-z.

Knowles, S. C. L., R. M. Eccles, and L. Baltrūnaitė (May 2019). "Species identity dominates over environment in shaping the microbiota of small mammals". In: *Ecology Letters* 22 (5). Ed. by James Grover, pp. 826–837. ISSN: 1461-023X. DOI: 10.1111/ele.13240.

Kohl, Kevin D. (May 2020). "Ecological and evolutionary mechanisms underlying patterns of phylosymbiosis in host-associated microbial communities". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 375 (1798), p. 20190251. ISSN: 0962-8436. DOI: 10.1098/rstb.2019.0251.

Kopac, Sarah M. and Jonathan L. Klassen (Oct. 2016). "Can they make it on their own? Hosts, microbes, and the holobiont niche". In: *Frontiers in Microbiology* 7 (OCT). ISSN: 1664302X. DOI: 10.3389/fmicb.2016.01647.

Larue, Kane, Matthew S. Kimber, Robert Ford, and Chris Whitfield (Mar. 2009). "Biochemical and Structural Analysis of Bacterial O-antigen Chain Length Regulator Proteins Reveals a Conserved Quaternary Structure". In: *Journal of Biological Chemistry* 284 (11), pp. 7395–7403. ISSN: 00219258. DOI: 10.1074/jbc.M809068200.

Lebeer, Sarah, Ingmar J. J. Claes, Tine L. A. Verhoeven, Jos Vanderleyden, and Sigrid C. J. De Keersmaecker (May 2011). "Exopolysaccharides of Lactobacillus rhamnosus GG form a protective shield against innate immune factors in the intestine". In: *Microbial Biotechnology* 4 (3), pp. 368–374. ISSN: 1751-7915. DOI: 10.1111/j.1751-7915.2010.00199.x.

Lee, I-Chiao et al. (July 2016). "Strain-Specific Features of Extracellular Polysaccharides and Their Impact on Lactobacillus plantarum-Host Interactions". In: *Applied and Environmental Microbiology* 82 (13). Ed. by D. W. Schaffner, pp. 3959–3970. ISSN: 0099-2240. DOI: 10.1128/AEM.00306-16.

Leggett, Richard M., Bernardo J. Clavijo, Leah Clissold, Matthew D. Clark, and Mario Caccamo (Feb. 2014). "NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries". In: *Bioinformatics* 30 (4), pp. 566–568. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt702.

Lenhoff, Howard M. and Ray DuBois Brown (Apr. 1970). "Mass culture of hydra: an improved method and its application to other aquatic invertebrates". In: *Laboratory Animals* 4 (1), pp. 139–154. ISSN: 0023-6772. DOI: 10.1258/002367770781036463.

Lim, Shen Jean and Seth R. Bordenstein (Mar. 2020). "An introduction to phylosymbiosis". In: *Proceedings of the Royal Society B: Biological Sciences* 287 (1922), p. 20192900. ISSN: 0962-8452. DOI: 10.1098/rspb.2019.2900.

Lyu, Jiejie et al. (Jan. 2024). "Bacterial templated carbonate mineralization: insights from concave-type crystals induced by Curvibacter lanceolatus strain HJ-1". In: *RSC Advances* 14 (1), pp. 353–363. ISSN: 2046-2069. DOI: 10.1039/D3RA06803J.

Ma, Dan, Zhenyu Hao, Rui Sun, Mark Bartlam, and Yingying Wang (2016). "Genome sequence of a typical ultramicrobacterium, Curvibacter sp. strain PAE-UM, capable of phthalate ester degradation". In: *Genome Announcements* 4 (1), pp. 1–2. ISSN: 21698287. DOI: 10.1128/genomeA.01510-15.

Masuko, Tatsuya et al. (Apr. 2005). "Carbohydrate analysis by a phenol–sulfuric acid method in microplate format". In: *Analytical Biochemistry* 339 (1), pp. 69–72. ISSN: 00032697. DOI: 10.1016/j.ab.2004.12.001.

Mcfall-Ngai, Margaret et al. (2013). "Animals in a bacterial world, a new imperative for the life sciences". In: *PNAS* 110 (9), pp. 3229–3236. DOI: https://doi.org/10.1073/pnas.1218525110.

Michael R. Davis, Jr. and Joanna B. Goldberg (May 2012). "Purification and Visualization of Lipopolysaccharide from Gram-negative Bacteria by Hot Aqueous-phenol Extraction". In: *Journal of Visualized Experiments* (63), p. 1. ISSN: 1940-087X. DOI: 10.3791/3916.

Minten-Lange, Timo and Sebastian Fraune (Jan. 2020). "Hydra and Curvibacter: An intimate crosstalk at the epithelial interface". In: *Cellular Dialogues in the Holobiont*. CRC Press, pp. 79–90. ISBN: 9781000090062. DOI: 10.1201/9780429277375-6.

Moeller, Andrew H. et al. (July 2016). "Cospeciation of gut microbiota with hominids". In: *Science* 353 (6297), pp. 380–382. ISSN: 0036-8075. DOI: 10.1126/science.aaf3951.

Moran, Nancy A. and Daniel B. Sloan (Dec. 2015). "The Hologenome Concept: Helpful or Hollow?" In: *PLoS Biology* 13 (12). ISSN: 15457885. DOI: 10.1371/journal.pbio.1002311.

Murillo-Rincon, Andrea P., Alexander Klimovich, Eileen Pemöller, Jan Taubenheim, Benedikt Mortzfeld, et al. (2017). "Spontaneous body contractions are modulated by the microbiome of Hydra". In: *Scientific Reports* 7 (1), pp. 1–9. ISSN: 20452322. DOI: 10.1038/s41598-017-16191-x.

Nadalin, Francesca, Francesco Vezzi, and Alberto Policriti (Sept. 2012). "GapFiller: a de novo assembly approach to fill the gap within paired reads". In: *BMC*

Bioinformatics 13 (S14), S8. ISSN: 1471-2105. DOI: 10.1186/1471-2105-13-S14-S8.

Nesper, Jutta et al. (Dec. 2003). "Translocation of Group 1 Capsular Polysaccharide in Escherichia coli Serotype K30". In: *Journal of Biological Chemistry* 278 (50), pp. 49763–49772. ISSN: 00219258. DOI: 10.1074/jbc.M308775200.

O'Brien, Paul A et al. (Sept. 2020). "Diverse coral reef invertebrates exhibit patterns of phylosymbiosis". In: *The ISME Journal* 14 (9), pp. 2211–2222. ISSN: 1751-7362. DOI: 10.1038/s41396-020-0671-x.

Obeng, Nancy et al. (2021). "Evolution of Microbiota-Host Associations: The Microbe's Perspective". In: *Trends in Microbiology* 29 (9), pp. 779–787. DOI: 10.1016/j.tim.2021.02.005.

Pal, Shilpa et al. (Aug. 2019). "Absence of the glycosyltransferase WcaJ in Klebsiella pneumoniae ATCC13883 affects biofilm formation, increases polymyxin resistance and reduces murine macrophage activation". In: *Microbiology* 165 (8), pp. 891–904. ISSN: 1350-0872. DOI: 10.1099/mic.0.000827.

Patel, K. B. et al. (May 2012). "Functional Characterization of UDP-Glucose:Undecaprenyl-Phosphate Glucose-1-Phosphate Transferases of Escherichia coli and Caulobacter crescentus". In: *Journal of Bacteriology* 194 (10), pp. 2646–2657. ISSN: 0021-9193. DOI: 10.1128/JB.06052-11.

Pietschke, Cleo, Christian Treitz, Sylvain Forêt, Annika Schultze, Sven Künzel, et al. (Oct. 2017). "Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts". In: *Proceedings of the National Academy of Sciences* 114 (40). ISSN: 0027-8424. DOI: 10.1073/pnas.1706879114.

Pollock, F. Joseph et al. (Nov. 2018). "Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny". In: *Nature Communications* 9 (1), p. 4921. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07275-x.

Qin, Man, Liyun Jiang, Gexia Qiao, and Jing Chen (Oct. 2023). "Phylosymbiosis: The Eco-Evolutionary Pattern of Insect–Symbiont Interactions". In: *International Journal of Molecular Sciences* 24 (21), p. 15836. ISSN: 1422-0067. DOI: 10.3390/ijms242115836.

Rahat, M. and Ch. Dimentman (Apr. 1982). "Cultivation of Bacteria-Free Hydra viridis : Missing Budding Factor in Nonsymbiotic Hydra". In: *Science* 216 (4541), pp. 67–68. ISSN: 0036-8075. DOI: 10.1126/science.7063873.

Rathje, Kai, Benedikt Mortzfeld, Marc P. Hoeppner, Jan Taubenheim, Thomas C. G. Bosch, et al. (Mar. 2020). "Dynamic interactions within the host-associated microbiota cause tumor formation in the basal metazoan Hydra". In: *PLOS*

Pathogens 16 (3). Ed. by Fanxiu Zhu, e1008375. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1008375.

Reid, Anne N. and Chris Whitfield (Aug. 2005). "Functional Analysis of Conserved Gene Products Involved in Assembly of Escherichia coli Capsules and Exopolysaccharides: Evidence for Molecular Recognition between Wza and Wzc for Colanic Acid Biosynthesis". In: *Journal of Bacteriology* 187 (15), pp. 5470–5481. ISSN: 0021-9193. DOI: 10.1128/JB.187.15.5470-5481.2005.

Remus, Daniela M et al. (Dec. 2012). "Impact of 4 Lactobacillus plantarum capsular polysaccharide clusters on surface glycan composition and host cell signaling". In: *Microbial Cell Factories* 11 (1), p. 149. ISSN: 1475-2859. DOI: 10.1186/1475-2859-11-149.

Rosenberg, Eugene and Ilana Zilber-Rosenberg (Dec. 2018). "The hologenome concept of evolution after 10 years". In: *Microbiome* 6 (1), p. 78. ISSN: 2049-2618. DOI: 10.1186/s40168-018-0457-9.

Salama, Nina R., Mara L. Hartung, and Anne Müller (June 2013). "Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori". In: *Nature Reviews Microbiology* 11 (6), pp. 385–399. ISSN: 1740-1526. DOI: 10.1038/nrmicro3016.

Seemann, Torsten (July 2014). "Prokka: rapid prokaryotic genome annotation". In: *Bioinformatics* 30 (14), pp. 2068–2069. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btu153.

Sheng, Guo-Ping, Han-Qing Yu, and Xiao-Yan Li (Nov. 2010). "Extracellular polymeric substances (EPS) of microbial aggregates in biological wastewater treatment systems: A review". In: *Biotechnology Advances* 28 (6), pp. 882–894. ISSN: 07349750. DOI: 10.1016/j.biotechadv.2010.08.001.

Shibata, Satoshi, Emily S. Yip, Kevin P. Quirke, Jakob M. Ondrey, and Karen L. Visick (2012). "Roles of the structural symbiosis polysaccharide (syp) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in Vibrio fischeri". In: *Journal of Bacteriology* 194 (24), pp. 6736–6747. ISSN: 10985530. DOI: 10.1128/JB.00707-12.

Tamura, Koichiro, Glen Stecher, and Sudhir Kumar (June 2021). "MEGA11: Molecular Evolutionary Genetics Analysis Version 11". In: *Molecular Biology and Evolution* 38 (7). Ed. by Fabia Ursula Battistuzzi, pp. 3022–3027. ISSN: 1537-1719. DOI: 10.1093/molbev/msab120.

Taubenheim, Jan, Máté Miklós, Jácint Tökölyi, and Sebastian Fraune (2022). "Population Differences and Host Species Predict Variation in the Diversity of Host-

Associated Microbes in Hydra". In: *Frontiers in Microbiology* 13 (March), pp. 1–17. ISSN: 1664302X. DOI: 10.3389/fmicb.2022.799333.

Taubenheim, Jan, Doris Willoweit-Ohl, Mirjam Knop, Sören Franzenburg, Jinru He, et al. (Sept. 2020). "Bacteria- and temperature-regulated peptides modulate β-catenin signaling in Hydra". In: *Proceedings of the National Academy of Sciences* 117 (35), pp. 21459–21468. ISSN: 0027-8424. DOI: 10.1073/pnas.2010945117.

Vu, Barbara, Miao Chen, Russell J. Crawford, and Elena P. Ivanova (July 2009). "Bacterial Extracellular Polysaccharides Involved in Biofilm Formation". In: *Molecules* 14 (7), pp. 2535–2554. ISSN: 1420-3049. DOI: 10.3390/molecules14072535.

Wang, Lei, Sandy Huskic, Adam Cisterne, Deborah Rothemund, and Peter R. Reeves (May 2002). "The O-Antigen Gene Cluster of Escherichia coli O55:H7 and Identification of a New UDP-GlcNAc C4 Epimerase Gene". In: *Journal of Bacteriology* 184 (10), pp. 2620–2625. ISSN: 0021-9193. DOI: 10.1128/JB.184.10.2620-2625.2002.

Watters, C., D. Fleming, D. Bishop, and K.P. Rumbaugh (2016). "Host Responses to Biofilm". In: *Host-Microbe Interactions*. Ed. by Michael San Francisco and Brian San Francisco. Vol. 142. Academic Press, pp. 193–239. DOI: 10.1016/bs.pmbts.2016.05.007.

Wegmann, Udo et al. (Dec. 2015). "The pan-genome of Lactobacillus reuteri strains originating from the pig gastrointestinal tract". In: *BMC Genomics* 16 (1), p. 1023. ISSN: 1471-2164. DOI: 10.1186/s12864-015-2216-7.

Wein, Tanita, Tal Dagan, Sebastian Fraune, Thomas C.G. Bosch, Thorsten B.H. Reusch, et al. (2018). "Carrying capacity and colonization dynamics of Curvibacter in the hydra host habitat". In: *Frontiers in Microbiology* 9 (MAR), pp. 1–10. ISSN: 1664302X. DOI: 10.3389/fmicb.2018.00443.

Weisburg, W G, S M Barns, D A Pelletier, and D J Lane (Jan. 1991). "16S ribosomal DNA amplification for phylogenetic study". In: *Journal of Bacteriology* 173 (2), pp. 697–703. ISSN: 0021-9193. DOI: 10.1128/jb.173.2.697-703.1991.

Whitfield, Chris (June 2006). "Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli". In: *Annual Review of Biochemistry* 75 (1), pp. 39–68. ISSN: 0066-4154. DOI: 10.1146/annurev.biochem.75.103004.142545.

Wugeditsch, Thomas et al. (Jan. 2001). "Phosphorylation of Wzc, a Tyrosine Autokinase, Is Essential for Assembly of Group 1 Capsular Polysaccharides in Escherichia coli". In: *Journal of Biological Chemistry* 276 (4), pp. 2361–2371. ISSN: 00219258. DOI: 10.1074/jbc.M009092200.

Xia, X. and G. Palidwor (Dec. 2005). "Genomic Adaptation to Acidic Environment: Evidence from Helicobacter pylori". In: *The American Naturalist* 166 (6). PMID: 16475092, pp. 776–784. ISSN: 0003-0147. DOI: 10.1086/497400.

Yip, Emily S., Brian T. Grublesky, Elizabeth A. Hussa, and Karen L. Visick (Sept. 2005). "A novel, conserved cluster of genes promotes symbiotic colonization and σ54-dependent biofilm formation by Vibrio fischeri". In: *Molecular Microbiology* 57 (5), pp. 1485–1498. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2005.04784.x.

Yoshida, Takako et al. (Feb. 2003). "Genes involved in the synthesis of the exopolysaccharide methanolan by the obligate methylotroph Methylobacillus sp. strain 12S". In: *Microbiology* 149 (2), pp. 431–444. ISSN: 1350-0872. DOI: 10.1099/mic.0.25913-0.

Yu, Guangchuang, Li Gen Wang, Yanyan Han, and Qing Yu He (2012). "ClusterProfiler: An R package for comparing biological themes among gene clusters". In: *OMICS A Journal of Integrative Biology* 16 (5), pp. 284–287. ISSN: 15362310. DOI: 10.1089/omi.2011.0118.

Yu, Jie et al. (June 2018). "Comparative Genomics of the Herbivore Gut Symbiont Lactobacillus reuteri Reveals Genetic Diversity and Lifestyle Adaptation". In: *Frontiers in Microbiology* 9 (JUN). ISSN: 1664-302X. DOI: 10.3389/fmicb.2018.01151.

Zilber-Rosenberg, Ilana and Eugene Rosenberg (Aug. 2008). "Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution". In: *FEMS Microbiology Reviews* 32 (5), pp. 723–735. ISSN: 1574-6976. DOI: 10.1111/j.1574-6976.2008.00123.x.

5 Manuscript 2

Host Colonization Triggers Amino Acid Transporter Upregulation in the Cobalamin-Dependent Methionine Auxotroph *Curvibacter*

Lukas Becker[1,2], Jay Bathia[1], Liam Fürbach[1], Said Laibacher[1,2], Ilka M. Axmann[2], Sebastian Fraune[1*]

[1]Institute for Zoology and Organismic Interactions, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[2]Institute for Synthetic Microbiology, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

[*]Corresponding author: Sebastian Fraune (fraune@hhu.de)

Keywords: Symbiosis, metabolic shift, microbial community, microbiome, nutrient exchange

5.1 Abstract

Nutrient exchange among hosts and microbes is essential for maintaining mutualistic interactions, wherein each partner provides metabolites that the other is incapable of synthesizing autonomously. While many bacteria rely on the provision of such nutrients from other members of the host microbiome, some rely directly on their hosts. These symbiotic nutrient dependencies can drive evolutionary specialization, leading to specific adaptations and interdependencies that, in turn, influence the stability and efficiency of mutualistic partnerships. This study investigates the metabolic dependencies of the Gram-negative symbiont *Curvibacter* sp. AEP1.3, which is associated with the freshwater polyp *Hydra vulgaris* AEP of the Cnidaria phylum. We provide strong evidence for a metabolic shift in *Curvibacter* towards the uptake of amino acids from the environment upon colonization of *Hydra*. Various transporter sequences, including amino acid transporters, are differentially upregulated when *Curvibacter* is associated with *Hydra*, compared to *Curvibacter* cultures in liquid media. In addition, we demonstrate that *Curvibacter* exhibits a cobalamin-dependent methionine auxotrophy, resulting from a nonfunctional cobalamin-independent methionine synthase MetE. Furthermore, we observe that upon *Hydra* mono-colonization, the cobalamin-dependent methionine synthase MetH and transporter for cobalamin are significantly downregulated. These findings suggest that methionine is supplied within the environmental niche, specifically the glycocalyx of *Hydra*, where *Curvibacter* and other members of the *Hydra* microbiome reside.

5.2 Introduction

Symbiotic interactions occur across all clades of life and are defined as close, long-term associations between taxonomically distinct organisms. These relationships shape the evolution of partners by mediating resource availability and selective pressures within a shared environment. Symbiotic partners often develop metabolic interdependencies, exchanging nutrients or compounds that one partner cannot synthesize on its own (Douglas 1998; Opatovsky et al. 2018; Ponnudurai et al. 2017; Rozmoš et al. 2022; Ryback et al. 2022). This exchange plays a crucial role in the evolution of such interactions and leads to co-adaptions of both partner (Brochet et al. 2021; Graf and Ruby 1998; Grant et al. 2014; Zhang et al. 2022). These co-adaptations can support survival, replication, and transmission, while simultaneously influencing host biology (Mandel et al. 2012; Wolschin et al. 2004; Zientz

et al. 2004). Studying these nutrient dependencies is essential for understanding the ecological and biological dynamics underlying host-microbe relationships. Studying these nutrient dependencies is essential for understanding the ecological and biological dynamics underlying host-microbe relationships.

Metabolic interdependence arises when bacteria living in nutrient-rich, stable environments, such as those provided by host organisms, lose the ability to biosynthesize certain nutrients de novo and instead rely on the direct uptake of these compounds from their surrounding environment (D'Souza et al. 2014; Ramoneda et al. 2023). This phenomenon is particularly widespread in endosymbiotic bacteria (Douglas 1998). A common feature of some of these endosymbionts is the substantial reduction of their genome size, with the retention of genes essential for survival and interaction within the host. As a result, most endosymbiotic bacteria have lost genes for certain biosynthetic pathways, rendering them unable to produce important metabolites such as specific amino acids, nucleotides, and vitamins (Grant et al. 2014; Heinken et al. 2014; Prell et al. 2010; Ramoneda et al. 2023; Shigenobu et al. 2000; Soto-Martin et al. 2020). This metabolic dependency in turn promotes specialization, which enables one symbiosis partner to concentrate on the production of the required metabolites, while the other can use its resources for complementary biological functions (Sørensen et al. 2020; Uchiumi and Sasaki 2020). Such division of labor not only stabilizes the symbiotic relationship but also facilitates the adaptation of each partner to specific ecological roles, enhancing the efficiency and resilience of the whole symbiotic system. The biosynthesis of essential metabolites through dedicated pathways is an energy-intensive process, often imposing a significant metabolic cost on the producing organism. Interestingly, it has been demonstrated that, in comparison with amino acids that are less expensive to biosynthesize, these cost-intensive amino acids tend to promote stronger cooperative interactions among a bacterial community consisting of auxotrophic *E. coli* strains (Mee et al. 2014). Consequently, the provision of nutrients in symbiotic systems represents a pivotal aspect of the evolution of mutual dependencies and fitness enhancing-strategies among all participating organisms.

In this study, we investigated the *Hydra-Curvibacter* host-microbe model system for the presence of metabolic interdependencies. *Curvibacter* is a Gram-negative bacteria belonging to the *Commamonadaceae* family. The genus *Curvibacter* comprises bacteria that inhabit diverse ecosystems, with a predominant presence in aquatic environments (Ding and Yokota 2004; Jakus et al. 2021; Ma et al. 2016; Philippi et al.

2021; Yang et al. 2021; Zhang et al. 2017). Certain species of *Curvibacter* function as ectosymbionts, forming symbiotic associations with members of the cnidarian genus *Hydra* (Fraune and Bosch 2007; Minten-Lange and Fraune 2020). *Hydra* species secrete a glycoprotein and polysaccharide network from their outer epithelial cell layer, termed as glycocalyx. This glycocalyx serves as a stable ecological niche for a microbiome that mediates fungal resistance (Fraune et al. 2015), and modulates host behavior and development (Giez et al. 2023; Murillo-Rincon et al. 2017; Nawroth et al. 2023; Rathje et al. 2020; Taubenheim et al. 2020).

The composition of *Hydra's* microbiota is influenced by host genetics and environmental factors (Taubenheim et al. 2022). Especially, nutrients available in the environment influence the composition of *Hydra's* microbiome. For instance, nutrient-rich conditions reduced the abundance of key beneficial taxa, such as *Curvibacter*, while altering the overall microbial community structure. These changes were associated with impaired host fitness, highlighting the potential for external nutrient supply to disrupt the balance of microbial symbioses in *Hydra* (Lachnit et al. 2025). However, recent research indicates that the nutrient composition of freshwater ecosystems exerts only a limited effect on microbiome stability in *Hydra* (Bathia et al. 2024). This stability is likely mediated by molecular exchanges within *Hydra's* glycocalyx, which serves as a regulatory interface facilitating both microbe-microbe and host-microbe stability (Fraune et al. 2009, 2015).

To investigate potential metabolic dependencies within the *Hydra* holobiont, we analyzed the genome of *Curvibacter* sp. AEP1.3 (hereafter referred to as *Curvibacter* AEP1.3), the primary colonizer of *Hydra vulgaris* AEP (hereafter *Hydra* AEP) (Fraune et al. 2015), focusing on its metabolic capabilities and possible auxotrophies. We demonstrate that *Curvibacter* AEP1.3 upregulates several amino acid transporter genes upon colonization of *Hydra* AEP, suggesting a metabolic adaptation to the host-provided environmental niche. Furthermore, our findings reveal that *Curvibacter* AEP1.3 exhibits a cobalamin-dependent methionine auxotrophy. Interestingly, *Curvibacter* AEP1.3 is capable of proliferating on *Hydra* AEP under laboratory conditions (Deines et al. 2020, 2017; Fraune et al. 2015; Wein et al. 2018), without external supplementation of methionine or cobalamin in the *Hydra* culture medium, indicating that methionine is likely supplied by the host. This host-derived methionine may play a regulatory role in shaping the specific abundance of *Curvibacter* AEP1.3 within the microbial community.

5.3 Material & Methods

5.3.1 Preparation of defined medium

The defined medium was formulated using the M9 recipe as a template, with salt concentrations adjusted to match those of R2A. To support *Curvibacter* growth, nine amino acids and ammonium were added. The medium was buffered with HEPES to counteract acidification caused by *Curvibacter* metabolism. The M9 salt, trace elements, biotin, thiamine, NH_4Cl, HEPES, $MgSO_4$, $CaCl_2$ solutions, along with distilled water, were autoclaved. The 40% glucose and L-amino acid solutions were filter-sterilized using a 0.22 µm filter. All chemicals were stored at 4 °C prior to use. The resulting solution was sterilized by filtration using a 0.22 µm bottle-top filter to remove any potential microbial contaminants.

Table 5.1: **Components of the defined medium**

Chemical	Concentration in media
Na_2HPO_4	337 µM
KH_2PO_4	220 µM
NaCl	880 µM
NH_4Cl	935 µM
Glucose	~22,2 µM
$CaCl_2$	300 µM
$MgSO_4$	1 mM
HEPES ($C_8H_{18}N_2O_4S$)	10 mM
NH_4Cl	10 mM
Histidine	100 mg/L
Methionine	100 mg/L
Asparagine	100 mg/L
Arginine	100 mg/L
Tryptophane	100 mg/L
Isoleucine	100 mg/L
Phenylalanine	100 mg/L
Threonine	100 mg/L
Aspartate	100 mg/L
Biotin	1 mg/L
Thiamin	1 mg/L
EDTA	134 µM
$FeCl_3$-$6H_2O$	31 µM
$ZnCl_2$	6,2 µM
$CuCl_2$-$2H_2O$	0,76 µM
$CoCl_2$-$2H_2O$	0,42 µM
H_3BO_3	1,62 µM
$MnCl_2$-$4H_2O$	0,081 µM
$C_{63}H_{88}CoN_{14}O_{14}P$	0.05 mM
H_2O	-

5.3.2 Bacterial cultivation and growth experiments

Curvibacter sp. AEP1.3 was revived from cryo-cultures on Reasoner's 2A (R2A) agar plates (Sigma-Aldrich). For growth experiments, *Curvibacter* pre-cultures were inoculated in defined media supplemented with all nine amino acids (Table 5.1) and placed in a New Brunswick Innova 42 incubator set at 18 °C and 170 RPM. Cultures grew for approximately three days until the late exponential phase, before entering stationary phase.

Prior to each experiment, fresh defined media containing the amino acids visualized in Figure 3B were prepared in 50 ml falcons. Growth experiments to assess the impact of amino acid supplementation were performed using a CLARIOstar plate reader (BMG Labtech). Pre-cultures for growth assays were inoculated to an OD_{600} of 0.05 in 800 µl of defined medium per well of a 48-well plate. The plate reader was programmed to measure OD_{600} values every 10 minutes throughout the experiment. For each growth experiment, at least two wells were inoculated with 800 µl of defined medium alone to serve as controls for contamination assessment and to measure blank values. The experiments were conducted at room temperature (~21 °C – 24 °C – due to shaking) with continuous shaking of 500 RPM between measurements.

The generation time g was calculated based on the standard exponential bacterial growth model, assuming bifurcations.

$$n = \frac{log(OD_{600}0.4) - log(OD_{600}0.1)}{log(2)} \tag{5.1}$$

$$g = \frac{time}{n} \tag{5.2}$$

N is the number of generations during the time of bacterial growth, this time is based on the duration of how long the bacteria need to grow from an OD_{600} 0.1 to 0.4. Statistical analysis included one-way ANOVA against the fully supplemented media to determine differences in mean values. Pairwise comparisons were conducted using Tukey's HSD (honestly significant difference) test.

The growth experiment with cobalamin (Cyanocobalamin, Carl Roth, s. Table 5.1) and defined media without amino acids was conducted in a Tecan Infinite M Nano+ plate reader. Briefly, pre-cultures were assessed for their optical density and an appropriate amount of pre-culture solution to reach a starting OD_{600} of 0.05 was

pipetted into 2 ml Eppendorf reaction tubes. Bacteria were then centrifuged for 4 minutes at 4,000 g. The supernatant was discarded, and the pellet was washed with defined media not containing any amino acids. This procedure was repeated two times. Finally, the *Curvibacter* AEP1.3 cell pellets were treated with defined media without amino acids and defined media without amino acids but supplemented with cobalamin. 400 µl of the bacterial cultures were then pipetted into 48-Well plates. The plate reader was programmed to measure OD_{600} values every 10 minutes throughout the experiment. For each growth experiment, at least two wells were inoculated with 400 µl of the respective media alone to serve as controls for contamination assessment and to measure blank values. Before each measurement, the plate was subjected to shaking at 500 RPM for 10 seconds. To obtain a constant temperature of 18 °C the plate-reader was placed into a refrigerator and set to hold the temperature between 17.5 and 18.5 °C.

5.3.3 Bioinformatics

For the Gene Ontoloy (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis the *Curvibacter* sp. AEP1.3 proteome (GCF_002163715.1) was obtained from NCBI. The proteome FASTA file was used as input for the PANNZER2/SPANZ tool (Törönen and Holm 2022), to infer functional annotation and GO terms. The inferred GO terms were filtered using the ARGOT_PPV value, following the authors' recommendation of applying a threshold value of 0.6. For the KEGG analysis, the proteome FASTA file was used as input for the BlastKOALA web service (Kanehisa et al. 2016). Obtained KEGG Ontology (KO) identifiers for each protein sequence were used as input for the KEGG Mapper Reconstruct tool to retrieve information on metabolic and biosynthesis pathways. The inference of the biosynthesis pathways for all proteinogenic amino acids as well as the bacterial secretion pathways was done by evaluating the Reconstruct output. The inference of the amino acid transporter sequences was done by screening the GO term dataset for GO:0006865, the term for amino acid transport.

For the transcriptome analysis, the publicly available dataset associated with the BioProject identifier PRJNA1025169 was utilized. Specifically, Supplementary File S3 was downloaded and integrated into the analysis (Giez et al. 2023). Differentially expressed genes were identified by filtering the dataset based on an adjusted p-value (≤ 0.05). For gene enrichment analysis, differentially upregulated genes were selected by applying an additional filter using only log2 fold changes ≥ 1.0. For the gene enrichment analysis, the GOATOOLS software (1.2.12) was used (Klopfenstein

et al. 2018). For Figure 2B, Transcripts-Per-Million (TPM) values were calculated using the raw reads and the gene lengths provided within the supplementary file S3 of Giez et al. (2023). Briefly, gene lengths were converted to kilobases by dividing their lengths by 1,000. Next, the raw read counts for each gene were divided by the corresponding gene length in kilobases to obtain reads per kilobase (RPK). The sum of all RPK values was then computed. Finally, the TPM value for each gene was calculated by dividing its RPK value by the total RPK sum and multiplying the result by one million.

Protein domain analysis was performed using InterProScan from the European Bioinformatics Institute (EMBL-EBI) (Jones et al. 2014). InterProScan was assessed programmatically with custom Python scripts using the *E. coli* K12 protein sequences NP_418443 for MetH and NP_418273 for MetE, as well as the *Curvibacter* AEP1.3 protein sequences WP_087495570 for MetH, WP_087497305 for the Homocysteine S-Methyltransferase domain containing protein and WP_087494645 for MetE. The protein sequences were obtained from NCBI's reference sequence database. All custom scripts and bioinformatic analyses are available in the project's GitHub repository: `https://github.com/Kanomble/amino_acid_project`.

5.4 Results

5.4.1 Metabolic adaptations of *Curvibacter* to its host habitat

To investigate metabolic capacities and potential auxotrophies of *Curvibacter* AEP1.3 in its symbiotic relationship with *Hydra* AEP, a KEGG and GO annotation of its genome was performed. This resulted in the identification of 1848 KO and 3359 GO categories. Next, we concentrated on the examination of amino acid synthesis pathways to identify potential auxotrophies. This analysis revealed that *Curvibacter* AEP1.3 has the biosynthetic potential to synthesize all proteinogenic amino acids (Figure 5.1, Synthesis of amino acids).

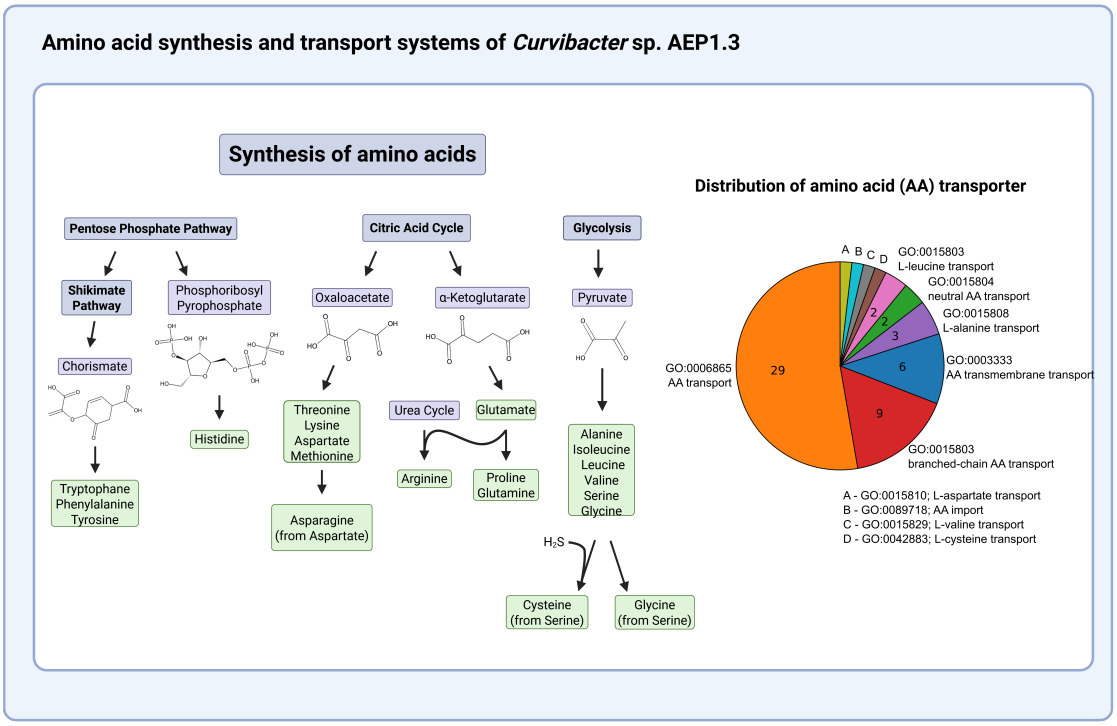

Figure 5.1: **Amino acid synthesis and transport systems of *Curvibacter* AEP1.3.** *Curvibacter* AEP1.3 has proteins for the synthesis of all proteinogenic amino acids (light green boxes). Besides the presence of biosynthetic pathways for amino acids, *Curvibacter* AEP1.3 possesses 55 sequences annotated with the Gene Ontology (GO) term of amino acid transport (GO:0006865). By traversing child GO terms associated with amino acid transport, as well as for branched-chain amino acid transport (GO:0015803) and branched-chain transmembrane amino acid transport (GO:0015658), a comprehensive set of GO terms related to amino acid transport was compiled. Using this dataset, we identified groups of the 55 transporter genes, sharing a similar distribution of GO terms. The groups and the number of sequences within those groups are visualized as a pie chart (Distribution of amino acid (AA) transporter).

To further investigate amino acid acquisition and molecular exchange capacity of *Curvibacter* AEP1.3 with its environment, GO and KO categories related to amino acid transporters were analyzed. In total, we identified 55 protein sequences of *Curvibacter* AEP1.3 associated with amino acid transporters (Figure 5.1, Distribution of amino acid transporter). Of these transporter sequences, 29 were exclusively associated with the GO term for amino acid transport (GO:0006865), with no further child transporter GO terms that may enable more precise annotation. Six transporters were associated with amino acid transmembrane transport, while nine sequences were specifically linked to branched-chain amino acid transport. The remaining eleven transporter sequences possessed additional amino acid transport GO child terms, enabling a more precise functional classification of their specific transport roles. These transporter sequences include three sequences associated with L-alanine transport, two neutral amino acid transporters, two L-leucine transporters, and one transporter sequence each for cysteine, valine, aspartate, and the import of amino acids.

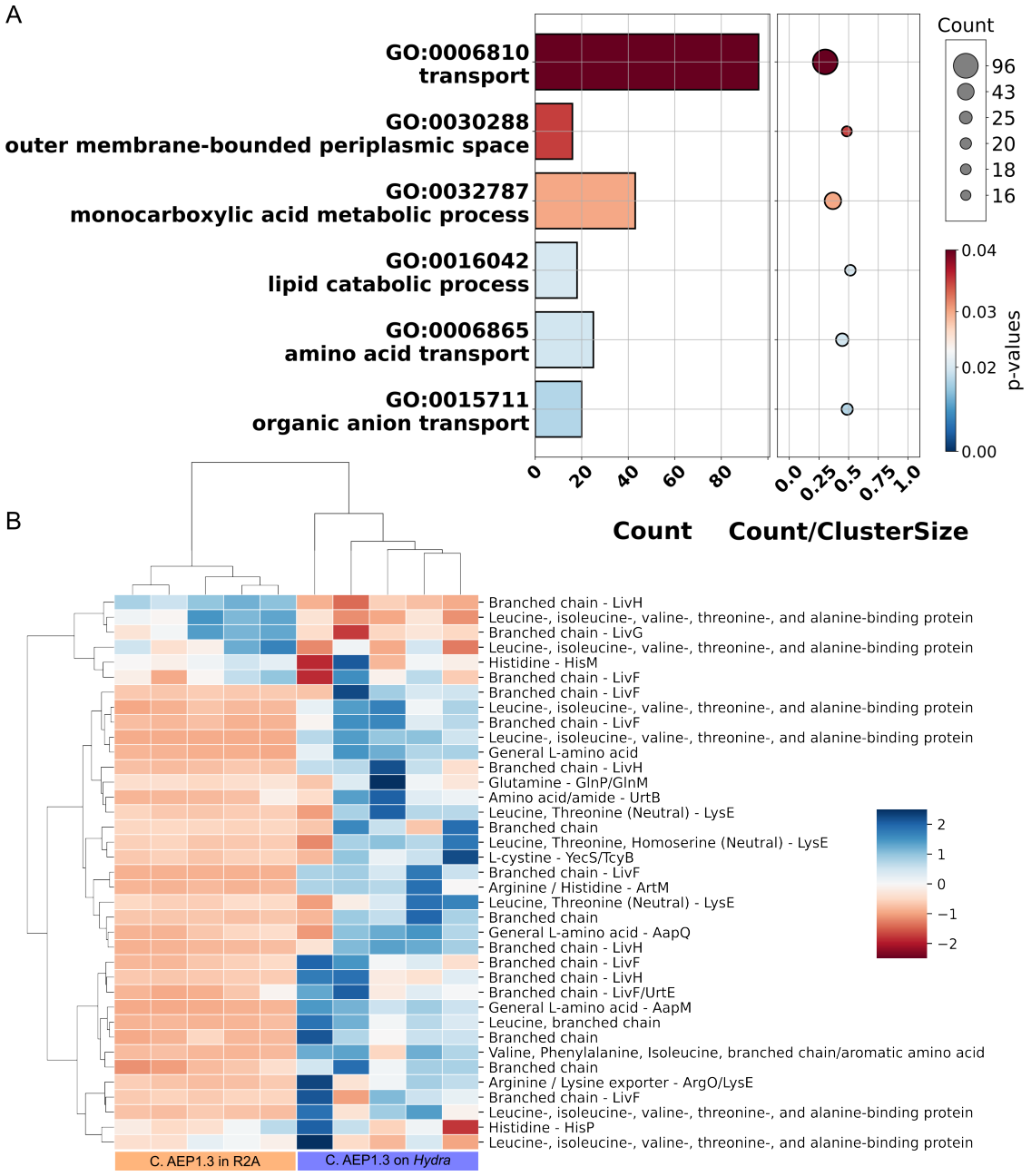

Figure 5.2: **Transcriptomic analysis of *Curvibacter* AEP1.3 on mono-colonized *Hydra* AEP compared with cultures grown in liquid R2A media.** (**A**) GO-term enrichment analysis of significantly upregulated genes of *Curvibacter* AEP1.3 monocolonizing *Hydra* AEP. The displayed GO terms represent a subset of 10 enriched categories, selected to reduce redundancy due to overlapping gene sets. (**B**) Heatmap of TPM (Transcripts Per Million) values of 37 significantly differential expressed amino acid transporter from *Curvibacter* AEP1.3. The left section of the heatmap represents the TPM values for *Curvibacter* AEP1.3 cultured in liquid R2A medium at 18 °C, while the right section displays the TPM values for *Curvibacter* AEP1.3 colonizing *Hydra* AEP. TPM values are normalized using the Z-score transformation implemented in the seaborn clustermap function, standardizing each gene across samples. The values are scaled to range between -2.5 and 2.5, with the color gradient transitioning from red (-2.5) to blue (2.5).

To further analyze the role of transporter systems and amino acid biosynthesis of *Curvibacter* AEP1.3 in relation to its host organism *Hydra* AEP, we analyzed a publicly available transcriptome dataset containing *Curvibacter* AEP1.3 samples grown in liquid R2A media and monocolonized on *Hydra* AEP (Giez et al. 2023). A GO-term enrichment analysis was performed on the transcripts from the host-associated *Curvibacter* AEP1.3 strain that had been overexpressed in comparison to liquid R2A cultures (Figure 5.2A). Among others, this analysis revealed the enrichment of transporter (GO:0006810) genes, with a significant enrichment of amino acid (GO:0006865) and organic anion (GO:0015711) transport genes (Figure 5.2A). Among the 55 amino acid transporter genes (Figure 5.1, Distribution of amino acid (AA) transporters), 37 transporter sequences are differential expressed (Figure 5.2B). 25 transporters are significantly upregulated in response to the host environment, while three transporter genes are significantly downregulated. Nine transporters exhibit less (log2 fold change $>$ -1 and $<$ 1) but significant (p-value $<$ 0.05) differential regulation. This result indicates an alteration in the amino acid transport of *Curvibacter* AEP1.3 in the glycocalyx of *Hydra* compared to cultures grown in liquid medium. In addition, genes associated with the GO-term "outer membrane-bounded periplasmic space" (GO:0030288) are significantly enriched. This GO-term is associated with proteins acting in the periplasm, such as TRAP (tripartite ATP-independent periplasmic) transporters and solute binding proteins. These proteins could facilitate the removal of host-derived toxic compounds such as antimicrobial peptides or metabolic waste produced by *Hydra* AEP as well as enhancing nutrient uptake (Davies et al. 2021; Joo et al. 2016). Additionally, genes associated with the monocarboxylic acid metabolic process (GO:0032787) and lipid catabolic process (GO:0016042) are significantly enriched among the upregulated genes, suggesting an increased utilization of host-derived lipids and organic acids. In summary, the results indicate a metabolic adaptation process of *Curvibacter* AEP1.3 within the host microenvironment.

5.4.2 Methionine auxotrophy in *Curvibacter*

To investigate potential amino acid dependencies, growth experiments with *Curvibacter* AEP1.3 were performed using a previously developed defined medium. The defined medium was supplemented with different combinations of amino acids (Figure 5.3) to assess their impact on generation time. While *Curvibacter* AEP1.3 could grow in media supplemented with nine amino acids, *Curvibacter* AEP1.3 showed no sign of growth in media lacking methionine (Figure 5.3). In addition, defined media lacking isoleucine exhibited significantly reduced generation times, while media without tryptophan resulted in significantly prolonged generation times compared to media supplemented with all nine selected amino acids (Figure 5.3).

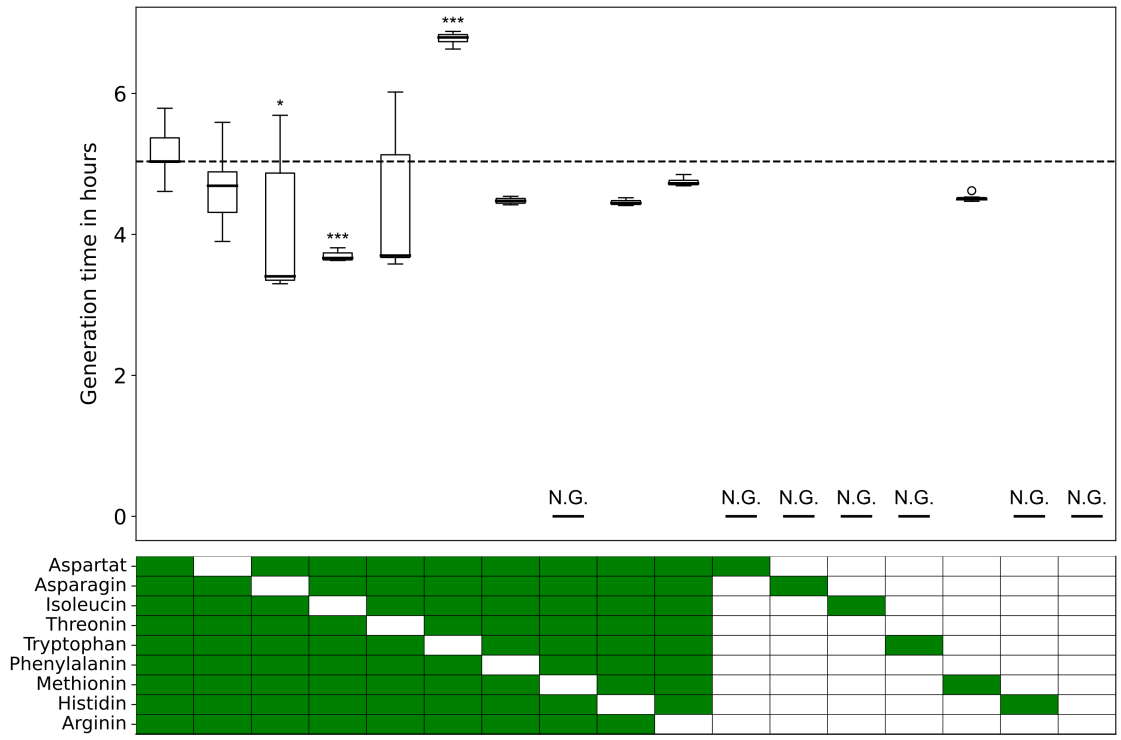

Figure 5.3: **Generation times of *Curvibacter* AEP1.3 grown in defined media supplemented with various amino acid compositions.** Box-plots showing generation times of *Curvibacter* AEP1.3 cultures grown in liquid defined media at room temperature (n=4-12). Generation time was assessed in logarithmic phase from OD_{600} 0.1 to OD_{600} 0.4 in a defined media containing selected amino acids. The dashed line represents the median generation time of *Curvibacter* AEP1.3 grown in a defined medium supplemented with nine amino acids (leftmost box-plot: fully supplemented media). To evaluate significant differences, a one-way ANOVA for each group vs. the defined medium containing all nine amino acids was performed, followed by Tukey's HSD test to identify significant differences. Statistical significance is indicated by asterisks, with the following meanings: *p $<$ 0.05, **p $<$ 0.01, ***p $<$ 0.001. The abbreviation N.G. indicates no growth of *Curvibacter* AEP1.3. The matrix displays the amino acid composition within the defined media. Green boxes indicate the presence of a specific amino acid, while white boxes denote its absence. Each column corresponds to the amino acids in the respective liquid medium for which the generation times were calculated.

5.4.3 Methionine anabolism in *Curvibacter*

To identify the cause of methionine auxotrophy in *Curvibacter* AEP1.3 we analyzed the methionine anabolism in detail at the genomic level. In bacteria methionine biosynthesis typically starts with the uptake of inorganic sulfur-containing molecules, such as sulfate (SO_4^{2-}) or thiosulfate ($S_2O_3^{2-}$), from the surrounding environment (Figure 5.4A). This process is mediated by ABC transporters or proton/sulfate symporters. In *Curvibacter* AEP1.3, multiple homologous proteins facilitate sulfur uptake, including CysW and CysT, which form a tandem permease system within the sulfate ABC transporter complex. In *E. coli*, this system is capable of transporting sulfate, thiosulfate, selenate, and selenite (Seiflein and Lawrence 2001), suggesting a similar function in *Curvibacter* AEP1.3. Once inside the cell, the inorganic sulfur source undergoes a series of reduction reactions, ultimately producing hydrogen sulfide (H_2S) (Figure 5.4A) (Seiflein and Lawrence 2001). Analysis of the *Curvibacter* AEP1.3 genome revealed KEGG homologs for all essential enzymes involved in sulfur assimilation and activation. Hydrogen sulfide serves as the biologically active form of sulfur and can be incorporated into sulfur containing molecules such as cysteine, which in turn acts as a precursor for methionine and other sulfur-containing biomolecules (Figure 5.4B).

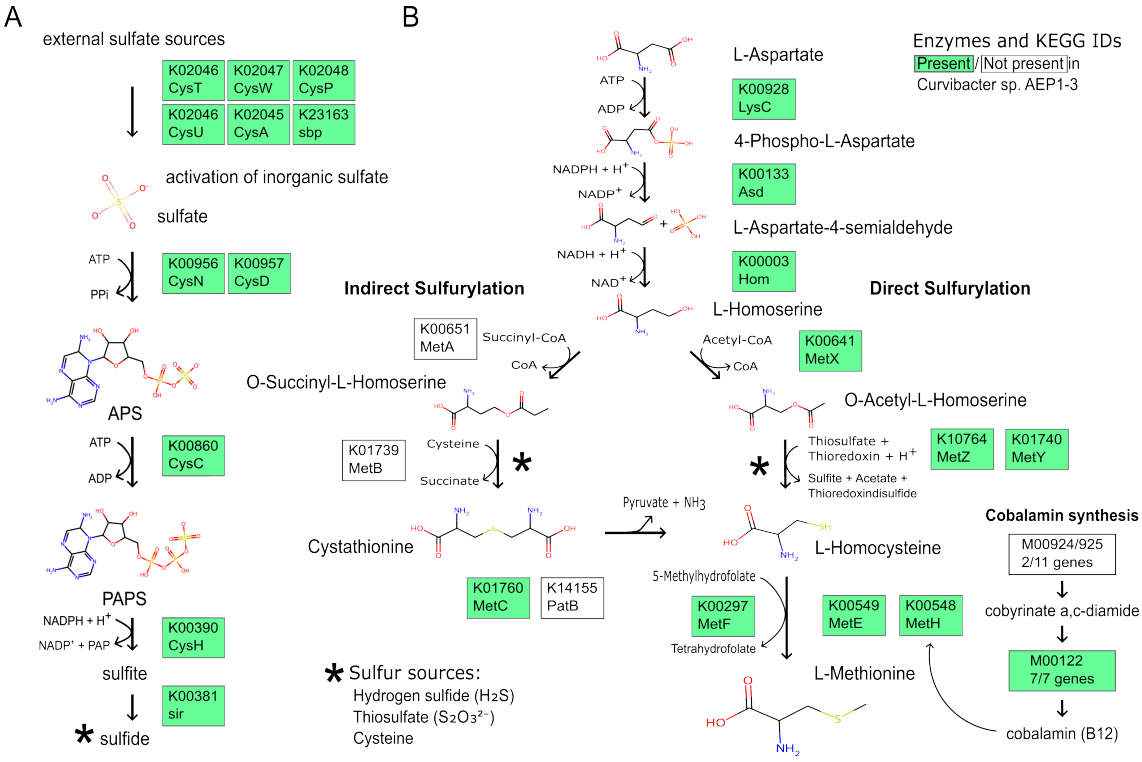

Figure 5.4: **Inorganic sulfur uptake/activation and methionine synthesis pathway in *Curvibacter* AEP1.3.** (**A**) *Curvibacter* AEP1.3 possesses the complete enzymatic repertoire required for the assimilation of external inorganic sulfur sources, such as $MgSO_4$, which is supplied in the defined medium. Inorganic sulfur is transported into the bacterial cell via specialized sulfate/thiosulfate ABC-transporter systems (CysW, CysP, CysU, CysA and sbp). Before sulfate can be utilized and incorporated into other biomolecules, it must undergo activation through a series of enzymatic steps. Initially, sulfate is transferred to an ATP molecule by the sulfate adenylyltransferase (CysN and CysD), forming adenosine 5'-phosphosulfate (APS). Subsequently, APS is phosphorylated by adenylylsulfate kinase (CysC) to generate 3'-phosphoadenosine 5'-phosphosulfate (PAPS). PAPS is then reduced to sulfite by CysH, which is further reduced to sulfide by the ferredoxin containing CysI. The resulting sulfide can then be assimilated into various biomolecules. (**B**) Methionine biosynthesis in bacteria begins with L-aspartate, which is phosphorylated by aspartokinase (LysC) to form L-aspartyl-4-phosphate. This intermediate is then reduced by aspartate semialdehyde dehydrogenase (Asd) to L-aspartate-4-semialdehyde, which is subsequently converted to L-homoserine by homoserine dehydrogenase (Hom). From this point, methionine biosynthesis diverges into two distinct pathways: the direct sulfurylation pathway and the transsulfuration (indirect sulfurylation) pathway. In *Curvibacter* AEP1.3, enzymes involved in the direct sulfurylation pathway are present, whereas MetC (cystathionine γ-synthase) is the only identified homolog associated with the transsulfuration pathway. In the direct sulfurylation route, L-homoserine is first acetylated by O-acetylhomoserine acetyltransferase (MetX) to generate O-acetylhomoserine. This intermediate then undergoes sulfur incorporation, catalyzed by MetY or MetZ, leading to the formation of L-homocysteine. Subsequently, 5,10-methylenetetrahydrofolate is reduced to 5-methyl-tetrahydrofolate by 5,10-methylene-tetrahydrofolate reductase (MetF). This methyl donor is then utilized by either the cobalamin-dependent methionine synthase (MetH) or the cobalamin-independent methionine synthase (MetE) to catalyze the remethylation of L-homocysteine to L-methionine. The pathway for de novo cobalamin (Vitamin B12) synthesis based on sirohydrochlorin (KEGG Module: M00294) or precorrin 2 (M00295) is absent in *Curvibacter* AEP1.3. Interestingly, genes for the synthesis of cobalamin from the precursor molecule cobyrinate a,c-diamide are present (M00122). However, for the MetH enzyme to function, *Curvibacter* AEP1.3 must acquire cobalamin from external sources. The illustrated pathways are based on information from Deng et al. (2018), Seiflein and Lawrence (2001), Ferla and Patrick (2014) and Rodionov et al. (2003).

The backbone of methionine is synthesized from L-homoserine, which is derived from L-aspartate through three enzymatic steps involving homologs of aspartokinase (LysC), aspartate semialdehyde dehydrogenase (Asd), and homoserine dehydrogenase (Hom). All of these enzymes are present in the *Curvibacter* AEP1.3 genome (Figure 5.4B). From L-homoserine, the pathway diverges into two alternative routes: the indirect (trans-sulfurylation) pathway and the direct sulfurylation pathway (Ferla and Patrick 2014). Most Betaproteobacteria predominantly utilize the direct sulfurylation pathway, and the analysis predicts that *Curvibacter* AEP1.3 follows this route, as it possesses homologs of all the enzymes necessary for this pathway. In this process, L-homoserine undergoes succinylation or acetylation by homoserine O-acetyltransferase/O-succinyltransferase (MetX), producing either O-succinyl-L-homoserine or O-acetyl-L-homoserine (Ferla and Patrick 2014). These intermediates are then directly sulfurylated by either MetZ (O-succinylhomoserine sulfhydrylase) or MetY (O-acetylhomoserine (thiol)-lyase), leading to the formation of L-homocysteine (Ferla and Patrick 2014). In the final step, L-homocysteine is methylated to L-methionine using methyl donor molecules, such as 5-methyltetrahydrofolate. This reaction is catalyzed by either of two methionine synthases: the cobalamin-dependent enzyme MetH or the cobalamin-independent enzyme MetE (Ferla and Patrick 2014). *Curvibacter* AEP1.3 possesses homologs of both methionine synthases. While MetE has no need for cobalamin, it is an important co-factor for the MetH methionine synthase. Interestingly, the genome of *Curvibacter* AEP1.3 lacks a complete set of coding sequences for the initial step of de novo cobalamin biosynthesis but contains the necessary sequences for synthesizing cobalamin from the precursor cobyrinate a,c-diamide. Therefore, it is likely that for the synthesis of methionine facilitated by MetH, *Curvibacter* AEP1.3 must acquire cobalamin or cobyrinate a,c-diamide from the surrounding environment. However, *Curvibacter* AEP1.3 should be able to assimilate the inorganic sulfur sources present in the defined media (Table 5.1) and synthesize L-methionine with the cobalamin-independent methionine synthase MetE, but this contradicts the observed methionine auxotrophy in the growth experiments.

5.4.4 Genetic variations and domain loss in methionine synthase loci of *Curvibacter*

A closer inspection of *Curvibacter* AEP1.3 proteins involved in methionine synthesis revealed differences in the domain structure of the methionine synthase proteins MetH and MetE in comparison to the functional sequences from *E. coli* (Figure 5.5). The homocysteine S-methyltransferase domain (PF02574) of MetH, which is catalyzing the methylation of L-Homocysteine, is missing within the enzyme of *Curvibacter* AEP1.3 (Figure 5.5A). Interestingly, a protein sequence containing the PF02574 domain is located two genes in the 5'direction of MetH (Figure 5.5B). This result suggests that the insertion of the genes may have caused a disruption to the MetH activity, which may have resulted in the production of an enzyme that is non-functional. In addition, the *Curvibacter* AEP1.3 MetE enzyme lacks the N-terminal cobalamin-independent synthase domain (Figure 5.5C, PF08267). Parallel to PF02574, PF08267 mediates the methyl transfer by binding the cofactor. The loss of this domain is likely resulting in a nonfunctional MetE. In contrast to MetH, the missing PF08267 domain of MetE is absent in *Curvibacter* AEP1.3.

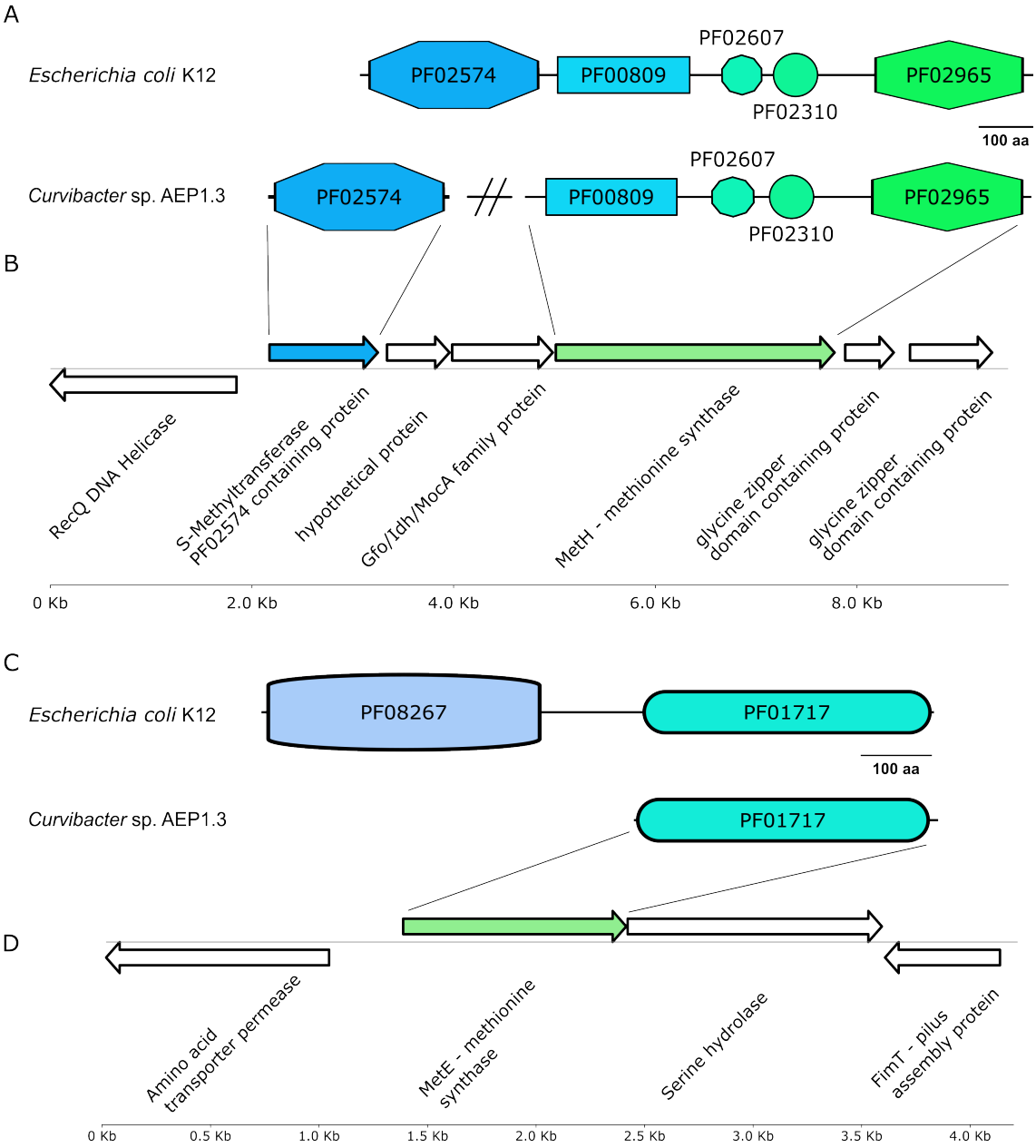

Figure 5.5: **Domain structures of the methionine synthases MetH and MetE in *Curvibacter* AEP1.3 compared to *E. coli*, along with their genomic loci in *Curvibacter* AEP1.3.** (**A**) Domain structure of the cobalamin-dependent methionine synthase enzyme MetH of *E. coli* (upper domain architecture) and *Curvibacter* AEP1.3 (lower domain architecture). The protein family (Pfam) domain identifier descriptions are as follows: PF02574 corresponds to the Homocysteine S-methyltransferase domain, PF00809 represents the Pterin-binding enzyme domain, PF02607 and PF02310 are associated with Vitamin B12-binding domains, and PF02965 is identified as the MetH activation domain. (**B**) Genomic locus of *metH* in *Curvibacter* AEP1.3. The homocysteine S-methyltransferase domain is separated from *metH* through an insertion of two other genes encoding a hypothetical protein and a Gfo/Idh/MocA family protein. (**C**) Domain structure of the cobalamin-independent methionine synthase enzyme MetE in *E. coli* (upper domain architecture) and in *Curvibacter* AEP1.3 (lower domain architecture). Pfam domain identifier descriptions are as follows: PF08267 is the MetE N-terminal domain and PF01717 is the MetE catalytic domain. (**D**) Genomic locus of *metE* in *Curvibacter* AEP1.3. The missing N-terminal domain (PF08267) in *metE* is absent within the genome of *Curvibacter* AEP1.3.

5.4.5 Vitamin B12 stimulates *Curvibacter* growth

In the next step, MetH activity was assessed through growth experiments in defined media supplemented with cobalamin. Interestingly, cobalamin supplementation stimulated the growth of *Curvibacter* AEP1.3 (Figure 5.6A). This result indicates that the methionine synthase MetH is functional despite the aforementioned gene insertion. Supplementation with methionine, cobalamin, or both had no measurable impact on the generation time.

To evaluate which scenario of methionine synthesis is potentially realized during growth on the host tissue, we analyzed the gene expression of genes related to cobalamin transport and both genes associated with MetH function during host colonization (Figure 5.6B). The transcript abundances of MetH and the cobalamin transport system BtuB in *Curvibacter* AEP1.3 were significantly lower on the *Hydra* polyp compared to liquid cultures grown in R2A (Figure 5.6B). In combination with the upregulation of several amino acid transporter genes (Figure 5.2A,B), this result suggests that active transport of methionine from the glycocalyx environment into the bacterial cell is more likely rather than a cobalamin uptake. In contrast, if cobalamin is present in the environment, *Curvibacter* AEP1.3. can grow without methionine supplementation.

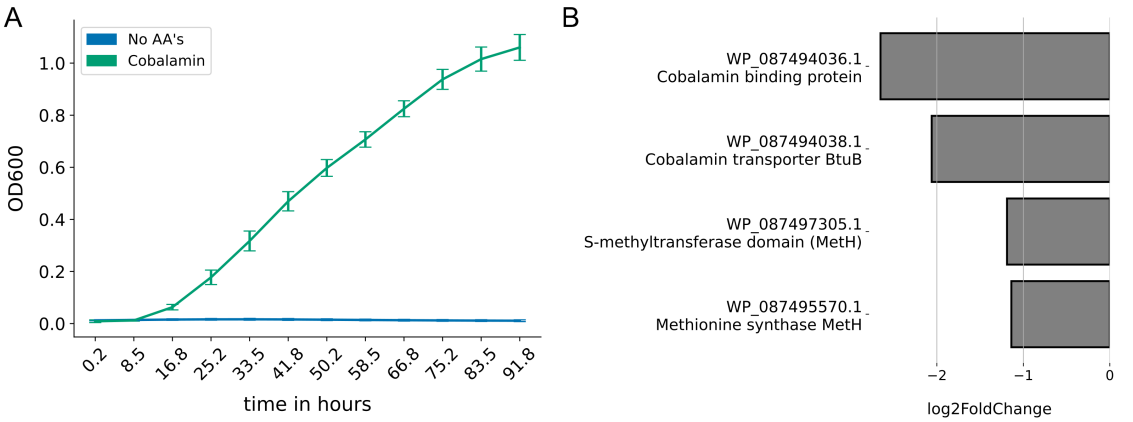

Figure 5.6: **Cobalamin induced growth and differential regulation of two cobalamin transporters and the methionine synthase MetH in *Curvibacter* AEP1.3 upon host colonization.** (**A**) OD_{600} growth curve of *Curvibacter* AEP1.3 at 18 °C in a Tecan Infinite M Nano+ plate reader (n=5). Growth of *Curvibacter* AEP1.3 within the defined media can be induced by supplementation of cobalamin. (**B**) The transcript abundance of the cobalamin transporter system BtuB and the methionine synthase MetH in *Curvibacter* AEP1.3 is significantly reduced upon host colonization.

5.5 Discussion

5.5.1 Metabolic adaptation in *Curvibacter* upon colonization of *Hydra*

Here, we demonstrate that *Curvibacter* AEP1.3 possesses a diverse repertoire of amino acid transporter genes (Figure 5.1) that are differentially regulated upon host colonization (Figure 5.2B). GO terms associated with these transporter genes are enriched (Figure 5.2A) among the upregulated genes in the host organism *Hydra* AEP, indicating metabolic adaptation from the free-living state to the host-associated state of *Curvibacter* AEP1.3. Previous studies have provided evidence for nutrient exchange in the *Hydra* holobiont (Giez et al. 2023; Hamada et al. 2018). In the *Hydra viridissima* and *Chlorella* symbiosis, *Hydra* upregulates phosphonate transport and glutamine synthetase genes, suggesting that the polyp supplies its symbiotic partner with essential nutrients, while *Chlorella* provides nutrients derived from its photosynthetic activity (Hamada et al. 2018). In the *Hydra* AEP symbiotic model, *Curvibacter* AEP1.3 enhances glutamate synthesis and glutamine uptake during host colonization (Giez et al. 2023). However, excessive glutamate negatively affects *Hydra's* mouth-opening behavior, a detrimental effect that is mitigated by glutamate-scavenging commensals such as *Duganella* and *Undibacterium*, highlighting a dynamic exchange of nutrients within the host-provided microenvironment (Giez et al. 2023). The enrichment of upregulated genes in the host organism associated with the GO term "outer membrane-bounded periplasmic space" - which includes several virulence-related transporter proteins, such as SiaP and other substrate-binding Tripartite ATP-independent Periplasmic (TRAP) transporters - suggests a significant shift in protein secretion and transport dynamics in response to the host environment.

Taken together, upon host colonization *Curvibacter* AEP1.3 undergoes a transcriptional shift, resulting in the upregulation of genes responsible for amino acid uptake and protein secretion. Similar patterns of upregulation have been observed in other symbiotic systems, including the upregulation of genes involved in amino acid metabolism and transport in *Buchnera* endosymbionts of aphids (Nakabachi et al. 2005). Similarly, *Burkholderia insecticola*, a symbiont in the midgut of the bean bug, upregulates transporters for several nutrients to ensure optimal nutrient exchange (Ohbayashi et al. 2019). The results of this study indicate that *Curvibacter* AEP1.3 is shifting its amino acid metabolism by preferentially acquiring amino acids from its

microenvironment, the glycocalyx of *Hydra*, rather than relying on de-novo biosynthesis. This finding likely reduces the metabolic burden on *Curvibacter* AEP1.3, as it minimizes the energy-intensive process of amino acid biosynthesis, enabling the bacterium to thrive on its host organism and even outcompete other members of *Hydras* microbiota (Deines et al. 2020).

5.5.2 Cobalamin dependent methionine auxotrophy in *Curvibacter* as adaptation to its host environment

Domain structure analysis revealed inconsistencies in the MetH and MetE methionine synthases of *Curvibacter* AEP1.3 compared to functional orthologs in *E. coli* (Figure 5.5). While the cobalamin independent methionine synthase MetE lost its N-terminal domain, the missing S-methyltransferase domain of MetH is encoded within another protein located two sequences downstream. The loss of the N-terminal MetE domain is probably the reason for the auxotrophy of methionine, as we could show that growth occurs if the medium is supplemented with cobalamin (Figure 5.6A). In recolonization experiments using *Curvibacter* AEP1.3 in germ-free *Hydra* polyps, *Curvibacter* AEP1.3 can successfully proliferate on the host (Deines et al. 2020; Fraune et al. 2015; Wein et al. 2018). Notably, the *Hydra* culture medium lacks supplementation with methionine or cobalamin. Given that eukaryotic organisms are incapable of synthesizing cobalamin de novo, the ability of *Curvibacter* AEP1.3 to grow under these conditions strongly suggests that methionine is supplemented within the glycocalyx of *Hydra*. In *Vibrio fischeri*, it has been shown that the production of quorum sensing molecules from the acyl-homoserine lactone (AHL) family is regulated by the availability of S-adenosylmethionine (SAM), a key methyl donor in AHL biosynthesis (Hanzelka and Greenberg 1996; Parsek et al. 1999). Since SAM synthesis depends on intracellular methionine levels, quorum sensing in *V. fischeri* is ultimately linked to methionine availability. Interestingly, Pietschke et al. (2017) demonstrated that communication between *Curvibacter* AEP1.3 and *Hydra* AEP involves quorum quenching mechanisms targeting bacterial AHL molecules. It is hypothesized that the recognition of *Curvibacter*-derived AHLs by *Hydra* triggers the expression of *eco* genes, ultimately modulating the Wnt signaling pathway through repression of β-catenin expression (Taubenheim et al. 2020). These findings connect the host-mediated provision of methionine to the production of AHL molecules by *Curvibacter* AEP1.3, which in turn appear to influence key developmental processes in *Hydra*.

In general, vitamin or amino acid auxotrophies within aquatic environments have been extensively reported for various microbial organisms (Johnson et al. 2020). In the case of vitamin B12, a significant proportion of human gut-associated bacteria have been shown to be auxotrophic for this vitamin. These bacteria depend on other microbial members of the gut microbiome or dietary intake as their primary sources of vitamin B12 (Starke et al. 2023). In the *Buchnera*-aphid symbiosis, the endosymbiotic bacterium *Buchnera aphidicola* lacks the ability to synthesize methionine and relies on other members of the microbiome or the host itself to provide this amino acid (Hansen and Moran 2011). A comparable dependency is observed in domestic ruminants such as cows and sheep, where certain symbiotic, rumen-associated bacteria are cobalamin auxotrophs, relying on other microbiome members for their vitamin B12 requirements (Franco-Lopez et al. 2020; Strobel 1992; Wang et al. 2022). In the squid-*Vibrio* symbiosis, the host provides at least nine amino acids that are critical for growth and bioluminescence activity of the symbiotic bacteria in the squid's light organ (Graf and Ruby 1998). By fine-tuning and balancing nutrient availability for microbial species in these systems, methionine and vitamin B12 auxotrophies are likely to serve as mechanisms for regulating sulfur and nitrogen cycling (Elling et al. 2020; Macdonald et al. 2012; Russell et al. 2013). Furthermore, it plays a crucial role in shaping the composition and stability of microbial communities by promoting mutualistic relationships and facilitating nutrient exchange (Rodionov et al. 2019; Starke et al. 2023). Interestingly, in *Curvibacter* AEP1.3, the transcript abundance of the cobalamin-dependent methionine synthase MetH, as well as the cobalamin transport system BtuB, is lower during *Hydra* mono-colonization compared to cultures grown in liquid R2A medium (Figure 5.6B). In conjunction with the upregulation of several amino acid transporters (Figure 5.6), we hypothesize that *Curvibacter* AEP1.3 actively imports methionine from the glycocalyx.

5.5.3 Domain structure disruption does not render MetH nonfunctional

The cobalamin-dependent methionine synthase MetH is a protein with four major domains (Figure 5.5A). Individual domains can catalyze partial reactions when provided with external methyl donors or cofactors, but a full domain ensemble is required for methionine synthesis (Ferla and Patrick 2014; Goulding et al. 1997). In *Curvibacter* AEP1.3 the S-methyltransferase domain PF02574 is separated from the other domains by two gene insertions (Figure 5.5B). Despite extensive research on the MetH methionine synthase, to our knowledge, no studies have identified domain separations with preservation of MetH function. However, a similar phe-

nomenon can be observed in the case of Toll-like receptors (TLRs) in *Hydra*, where they serve as key receptors for activating the innate immune response (Bosch et al. 2009; Franzenburg et al. 2012; Klimovich and Bosch 2024). Typically, TLRs are encoded by large protein sequences comprising an extracellular Leucine-Rich Repeat (LRR) domain and an intracellular Toll/interleukin-1 receptor (TIR) domain. The LRR domain is responsible for pattern recognition, while the TIR domain facilitates signal transduction by recruiting downstream proteins such as MyD88. In *Hydra*, these domains are separated and encoded by distinct coding sequences (Hemmrich et al. 2007), this likely contributes to an increased flexibility in the regulation and expression of TLRs, thereby enabling the polyp to precisely modulate its immune responses (Franzenburg et al. 2012). In the case of MetH, the PF02574 domain is responsible for the methylation of homocysteine. Therefore, the separation of this domain may enable *Curvibacter* AEP1.3 to finely regulate the methylation process, preventing excessive methylation of homocysteine and ensuring controlled synthesis of methionine.

5.6 Conclusion

In this study, we have shown that *Curvibacter* AEP1.3 undergoes a metabolic shift upon colonization of *Hydra* AEP, leading to an increase in the abundance of amino acid transporter genes. This upregulation can facilitate the direct uptake of amino acids from the *Hydra* glycocalyx. Moreover, *Curvibacter* AEP1.3 is auxotroph for the amino acid methionine if cobalamin is not present in the media, as the cobalamin independent methionine synthase MetE is nonfunctional, and the genes required for de novo cobalamin synthesis are incomplete. Among various sources of methionine, such as glycoproteins or metabolic byproducts present within the *Hydra* glycocalyx, active provisioning of methionine may be mediated by the organism itself. Previous fluid dynamics experiments have shown that *Hydra's* body contractions enhance the transport of chemical compounds toward its epithelial surface, potentially facilitating nutrient delivery (Nawroth et al. 2023). These chemical compounds can then serve as nutrients for the resident microbiome. Furthermore, external nutrient supply in the *Hydra* medium alters bacterial abundance, resulting in a dramatic decrease in *Curvibacter* (Lachnit et al. 2025). Similarly, nutrient-dependent shifts in microbiome composition are evident in mammals such as humans and mice, where dietary changes have been associated with increases in specific bacterial clades (Beam et al. 2021; Turnbaugh et al. 2008). Interestingly, in the case of the *Hydra-Curvibacter* relationship, the availability of the single amino acid methionine appears to be a

critical and highly favorable factor for the fitness of *Curvibacter*. This is further supported by the observation that *Curvibacter* achieves the highest relative abundance among the bacterial taxa inhabiting *Hydra* (Fraune et al. 2015; Lachnit et al. 2025). Thus, nutrient availability has a profound effect on the composition and stability of *Hydra's* microbiome. The findings of this study suggest that the glycocalyx of *Hydra* serves as a site of nutrient exchange, where the nutrient composition is determined by the secretion of chemical compounds from the host and the metabolic activity of a diverse microbiome, ultimately facilitating species-specific, long-term associations (Franzenburg et al. 2013; Fraune and Bosch 2007).

References for Manuscript 2

Bathia, Jay, Máté Miklós, István Gyulai, Sebastian Fraune, and Jácint Tökölyi (2024). "Environmental microbial reservoir influences the bacterial communities associated with Hydra oligactis". In: *Scientific Reports* 14 (1), pp. 1–15. ISSN: 20452322. DOI: 10.1038/s41598-024-82944-0.

Beam, Ashley, Elizabeth Clinger, and Lei Hao (Aug. 2021). "Effect of Diet and Dietary Components on the Composition of the Gut Microbiota". In: *Nutrients* 13 (8), p. 2795. ISSN: 2072-6643. DOI: 10.3390/nu13082795.

Bosch, Thomas, René Augustin, Friederike Anton-Erxleben, Sebastian Fraune, Georg Hemmrich, et al. (Apr. 2009). "Uncovering the evolutionary history of innate immunity: The simple metazoan Hydra uses epithelial cells for host defence". In: *Developmental & Comparative Immunology* 33 (4), pp. 559–569. ISSN: 0145305X. DOI: 10.1016/j.dci.2008.10.004.

Brochet, Silvia et al. (July 2021). "Niche partitioning facilitates coexistence of closely related honey bee gut bacteria". In: *eLife* 10. ISSN: 2050-084X. DOI: 10.7554/eLife.68583.

D'Souza, Glen, Silvio Waschina, Samay Pande, Katrin Bohl, Christoph Kaleta, et al. (Sept. 2014). "Less Is More: Selective Advantages Can Explain The Prevalent Loss Of Biosynthetic Genes In Bacteria". In: *Evolution* 68 (9), pp. 2559–2570. ISSN: 00143820. DOI: 10.1111/evo.12468.

Davies, James S. et al. (June 2021). "Selective Nutrient Transport in Bacteria: Multicomponent Transporter Systems Reign Supreme". In: *Frontiers in Molecular Biosciences* 8. ISSN: 2296889X. DOI: https://doi.org/10.3389/fmolb.2021.699222.

Deines, Peter, Katrin Hammerschmidt, and Thomas C G Bosch (2020). "Microbial Species Coexistence Depends on the Host Environment". In: DOI: 10.1128/mbio.00807-20.

Deines, Peter, Tim Lachnit, and Thomas C. G. Bosch (Sept. 2017). "Competing forces maintain the Hydra metaorganism". In: *Immunological Reviews* 279 (1), pp. 123–136. ISSN: 0105-2896. DOI: 10.1111/imr.12564.

Deng, Q, X Wu, Y Wang, and M Liu (2018). "Activity characteristics of sulfate reducing bacteria and formation mechanism of hydrogen sulfide". In: *Applied Ecology and Environmental Research* 16 (5), pp. 6369–6383. ISSN: 15891623. DOI: http://dx.doi.org/10.15666/aeer/1605_63696383.

Ding, Linxian and Akira Yokota (Nov. 2004). "Proposals of Curvibacter gracilis gen. nov., sp. nov. and Herbaspirillum putei sp. nov. for bacterial strains isolated from well water and reclassification of [Pseudomonas] huttiensis, [Pseudomonas] lanceolata, [Aquaspirillum] delicatum and [Aquaspirillum] autotrophicum as Herbaspirillum huttiense comb. nov., Curvibacter lansceolatus comb. nov., Curvibacter delicatus comb. nov." In: *International Journal of Systematic and Evolutionary Microbiology* 54 (6), pp. 2223–2230. ISSN: 14665026. DOI: 10.1099/ijs.0.02975-0.

Douglas, A. E. (Jan. 1998). "Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera". In: *Annual Review of Entomology* 43 (1), pp. 17–37. ISSN: 0066-4170. DOI: 10.1146/annurev.ento.43.1.17.

Elling, Felix J. et al. (Dec. 2020). "Vitamin B12-dependent biosynthesis ties amplified 2-methylhopanoid production during oceanic anoxic events to nitrification". In: *Proceedings of the National Academy of Sciences* 117 (52), pp. 32996–33004. ISSN: 0027-8424. DOI: 10.1073/pnas.2012357117.

Ferla, Matteo P. and Wayne M. Patrick (Aug. 2014). "Bacterial methionine biosynthesis". In: *Microbiology* 160 (8), pp. 1571–1584. ISSN: 1350-0872. DOI: 10.1099/mic.0.077826-0.

Franco-Lopez, Julian et al. (Apr. 2020). "Correlations between the Composition of the Bovine Microbiota and Vitamin B 12 Abundance". In: *mSystems* 5 (2). ISSN: 23795077. DOI: 10.1128/msystems.00107-20.

Franzenburg, Sören, Sebastian Fraune, Sven Künzel, John F. Baines, Tomislav Domazet-Lošo, et al. (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), pp. 19374–19379. ISSN: 00278424. DOI: 10.1073/pnas.1213110109.

Franzenburg, Sören, Jonas Walter, Sven Künzel, Jun Wang, John F. Baines, et al. (Sept. 2013). "Distinct antimicrobial peptide expression determines host species-specific bacterial associations". In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (39). ISSN: 00278424. DOI: 10.1073/pnas.1304960110.

Fraune, Sebastian, Yuichi Abe, and Thomas C. G. Bosch (Sept. 2009). "Disturbing epithelial homeostasis in the metazoan Hydra leads to drastic changes in associated microbiota". In: *Environmental Microbiology* 11 (9), pp. 2361–2369. ISSN: 1462-2912. DOI: 10.1111/j.1462-2920.2009.01963.x.

Fraune, Sebastian, Friederike Anton-Erxleben, René Augustin, Sören Franzenburg, Mirjam Knop, et al. (July 2015). "Bacteria-bacteria interactions within the mi-

crobiota of the ancestral metazoan Hydra contribute to fungal resistance". In: *ISME Journal* 9 (7), pp. 1543–1556. ISSN: 17517370. DOI: 10.1038/ismej.2014.239.

Fraune, Sebastian and Thomas C.G. Bosch (2007). "Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra". In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (32), pp. 13146–13151. ISSN: 00278424. DOI: 10.1073/pnas.0703375104.

Giez, Christoph, Denis Pinkle, Yan Giencke, Jörg Wittlieb, Eva Herbst, et al. (Dec. 2023). "Multiple neuronal populations control the eating behavior in Hydra and are responsive to microbial signals". In: *Current Biology* 33 (24), 5288–5303.e6. ISSN: 18790445. DOI: 10.1016/j.cub.2023.10.038.

Goulding, Celia W, Dellma Postigo, and Rowena G Matthews (1997). "Cobalamin-Dependent Methionine Synthase Is a Modular Protein with Distinct Regions for Binding Homocysteine, Methyltetrahydrofolate, Cobalamin, and Adenosylmethionine". In: *Biochemistry* 36 (26). PMID: 9201956, pp. 8082–8091. DOI: 10.1021/bi9705164.

Graf, Joerg and E. G. Ruby (Feb. 1998). "Host-derived amino acids support the proliferation of symbiotic bacteria". In: *Proceedings of the National Academy of Sciences* 95 (4), pp. 1818–1822. ISSN: 0027-8424. DOI: 10.1073/pnas.95.4.1818.

Grant, Matthew A A, Elena Kazamia, Pietro Cicuta, and Alison G Smith (July 2014). "Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal–bacterial cocultures". In: *The ISME Journal* 8 (7), pp. 1418–1427. ISSN: 1751-7362. DOI: 10.1038/ismej.2014.9.

Hamada, Mayuko, Katja Schröder, Jay Bathia, Ulrich Kürn, Sebastian Fraune, et al. (May 2018). "Metabolic co-dependence drives the evolutionarily ancient Hydra–Chlorella symbiosis". In: *eLife* 7, pp. 1–37. ISSN: 2050-084X. DOI: 10.7554/eLife.35122.

Hansen, Allison K. and Nancy A. Moran (Feb. 2011). "Aphid genome expression reveals host-symbiont cooperation in the production of amino acids". In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (7), pp. 2849–2854. ISSN: 00278424. DOI: 10.1073/pnas.1013465108.

Hanzelka, B L and E P Greenberg (Sept. 1996). "Quorum sensing in Vibrio fischeri: evidence that S-adenosylmethionine is the amino acid substrate for autoinducer synthesis". In: *Journal of Bacteriology* 178 (17), pp. 5291–5294. ISSN: 0021-9193. DOI: 10.1128/jb.178.17.5291-5294.1996.

Heinken, Almut et al. (Sept. 2014). "Functional Metabolic Map of Faecalibacterium prausnitzii, a Beneficial Human Gut Microbe". In: *Journal of Bacteriology* 196 (18), pp. 3289–3302. ISSN: 0021-9193. DOI: 10.1128/JB.01780-14.

Hemmrich, Georg, David J. Miller, and Thomas Bosch (2007). "The evolution of immunity: a low-life perspective". In: *Trends in Immunology* 28 (10), pp. 449–454. ISSN: 14714906. DOI: 10.1016/j.it.2007.08.003.

Jakus, Natalia, Nia Blackwell, Daniel Straub, Andreas Kappler, and Sara Kleindienst (Nov. 2021). "Presence of Fe(II) and nitrate shapes aquifer-originating communities leading to an autotrophic enrichment dominated by an Fe(II)-oxidizing Gallionellaceae sp". In: *FEMS Microbiology Ecology* 97 (11). ISSN: 1574-6941. DOI: 10.1093/femsec/fiab145.

Johnson, Winifred M, Harriet Alexander, Raven L Bier, Dan R Miller, Mario E Muscarella, et al. (Nov. 2020). "Auxotrophic interactions: a stabilizing attribute of aquatic microbial communities?" In: *FEMS Microbiology Ecology* 96 (11). ISSN: 1574-6941. DOI: 10.1093/femsec/fiaa115.

Jones, Philip et al. (May 2014). "InterProScan 5: genome-scale protein function classification". In: *Bioinformatics* 30 (9), pp. 1236–1240. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btu031.

Joo, Hwang-Soo, Chih-Iung Fu, and Michael Otto (May 2016). "Bacterial strategies of resistance to antimicrobial peptides". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1695), p. 20150292. ISSN: 0962-8436. DOI: https://doi.org/10.1098/rstb.2015.0292.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima (2016). "BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences". In: *Journal of Molecular Biology* 428 (4), pp. 726–731. ISSN: 10898638. DOI: 10.1016/j.jmb.2015.11.006.

Klimovich, Alexander and Thomas Bosch (2024). "Novel technologies uncover novel 'anti'-microbial peptides in Hydra shaping the species-specific microbiome". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 379 (1901). ISSN: 14712970. DOI: 10.1098/rstb.2023.0058.

Klopfenstein, D. V., Liangsheng Zhang, Brent S. Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, et al. (2018). "GOATOOLS: A Python library for Gene Ontology analyses". In: *Scientific Reports* 8 (1), pp. 1–17. ISSN: 20452322. DOI: 10.1038/s41598-018-28948-z.

Lachnit, Tim, Laura Ulrich, Fiete M. Willmer, Tim Hasenbein, Leon X. Steiner, et al. (Feb. 2025). "Nutrition-induced changes in the microbiota can cause dysbiosis

and disease development". In: *mBio*. Ed. by Joerg Graf. ISSN: 2150-7511. DOI: 10.1128/mbio.03843-24.

Ma, Dan, Zhenyu Hao, Rui Sun, Mark Bartlam, and Yingying Wang (2016). "Genome sequence of a typical ultramicrobacterium, Curvibacter sp. strain PAE-UM, capable of phthalate ester degradation". In: *Genome Announcements* 4 (1), pp. 1–2. ISSN: 21698287. DOI: 10.1128/genomeA.01510-15.

Macdonald, Sandy J., George G. Lin, Calum W. Russell, Gavin H. Thomas, and Angela E. Douglas (Aug. 2012). "The central role of the host cell in symbiotic nitrogen metabolism". In: *Proceedings of the Royal Society B: Biological Sciences* 279 (1740), pp. 2965–2973. ISSN: 0962-8452. DOI: 10.1098/rspb.2012.0414.

Mandel, Mark J, Amy L Schaefer, Caitlin A Brennan, Elizabeth A C Heath-Heckman, Cindy R Deloney-Marino, et al. (2012). "Squid-Derived Chitin Oligosaccharides Are a Chemotactic Signal during Colonization by Vibrio fischeri". In: *Applied and Environmental Microbiology* 78 (13), pp. 4620–4626. DOI: 10.1128/AEM.00377-12.

Mee, Michael T., James J. Collins, George M. Church, and Harris H. Wang (May 2014). "Syntrophic exchange in synthetic microbial communities". In: *Proceedings of the National Academy of Sciences* 111 (20). ISSN: 0027-8424. DOI: 10.1073/pnas.1405641111.

Minten-Lange, Timo and Sebastian Fraune (Jan. 2020). "Hydra and Curvibacter: An intimate crosstalk at the epithelial interface". In: *Cellular Dialogues in the Holobiont*. CRC Press, pp. 79–90. ISBN: 9781000090062. DOI: 10.1201/9780429277375-6.

Murillo-Rincon, Andrea P., Alexander Klimovich, Eileen Pemöller, Jan Taubenheim, Benedikt Mortzfeld, et al. (2017). "Spontaneous body contractions are modulated by the microbiome of Hydra". In: *Scientific Reports* 7 (1), pp. 1–9. ISSN: 20452322. DOI: 10.1038/s41598-017-16191-x.

Nakabachi, Atsushi et al. (2005). "Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera". In: DOI: https://doi.org/10.1073/pnas.0409034102.

Nawroth, Janna C., Christoph Giez, Alexander Klimovich, Eva Kanso, and Thomas C.G. Bosch (2023). "Spontaneous body wall contractions stabilize the fluid microenvironment thashapes host–microbe associations". In: *eLife* 12, pp. 1–23. ISSN: 2050084X. DOI: 10.7554/eLife.83637.

Ohbayashi, Tsubasa et al. (June 2019). "Comparative cytology, physiology and transcriptomics of Burkholderia insecticola in symbiosis with the bean bug Riptor-

tus pedestris and in culture". In: *ISME Journal* 13 (6), pp. 1469–1483. ISSN: 17517370. DOI: 10.1038/s41396-019-0361-8.

Opatovsky, Itai et al. (Dec. 2018). "Modeling trophic dependencies and exchanges among insects' bacterial symbionts in a host-simulated environment". In: *BMC Genomics* 19 (1), p. 402. ISSN: 1471-2164. DOI: 10.1186/s12864-018-4786-7.

Parsek, Matthew R., Dale L. Val, Brian L. Hanzelka, John E. Cronan, and E. P. Greenberg (Apr. 1999). "Acyl homoserine-lactone quorum-sensing signal generation". In: *Proceedings of the National Academy of Sciences* 96 (8), pp. 4360–4365. ISSN: 0027-8424. DOI: 10.1073/pnas.96.8.4360.

Philippi, Miriam et al. (Aug. 2021). "Purple sulfur bacteria fix N2 via molybdenum-nitrogenase in a low molybdenum Proterozoic ocean analogue". In: *Nature Communications* 12 (1), p. 4774. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25000-z.

Pietschke, Cleo, Christian Treitz, Sylvain Forêt, Annika Schultze, Sven Künzel, et al. (Oct. 2017). "Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts". In: *Proceedings of the National Academy of Sciences* 114 (40). ISSN: 0027-8424. DOI: 10.1073/pnas.1706879114.

Ponnudurai, Ruby et al. (Feb. 2017). "Metabolic and physiological interdependencies in the Bathymodiolus azoricus symbiosis". In: *The ISME Journal* 11 (2), pp. 463–477. ISSN: 1751-7362. DOI: 10.1038/ismej.2016.124.

Prell, Jurgen et al. (Nov. 2010). "Role of Symbiotic Auxotrophy in the Rhizobium-Legume Symbioses". In: *PLoS ONE* 5 (11). Ed. by Julian Rutherford, e13933. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0013933.

Ramoneda, Josep, Thomas B. N. Jensen, Morgan N. Price, Emilio O. Casamayor, and Noah Fierer (Nov. 2023). "Taxonomic and environmental distribution of bacterial amino acid auxotrophies". In: *Nature Communications* 14 (1), p. 7608. ISSN: 2041-1723. DOI: 10.1038/s41467-023-43435-4.

Rathje, Kai, Benedikt Mortzfeld, Marc P. Hoeppner, Jan Taubenheim, Thomas C. G. Bosch, et al. (Mar. 2020). "Dynamic interactions within the host-associated microbiota cause tumor formation in the basal metazoan Hydra". In: *PLOS Pathogens* 16 (3). Ed. by Fanxiu Zhu, e1008375. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1008375.

Rodionov, Dmitry A., Alexey G. Vitreschak, Andrey A. Mironov, and Mikhail S. Gelfand (Oct. 2003). "Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes". In: *Journal of Biological Chemistry* 278 (42), pp. 41148–41159. ISSN: 00219258. DOI: 10.1074/jbc.M305837200.

Rodionov, Dmitry A. et al. (2019). "Micronutrient requirements and sharing capabilities of the human gut microbiome". In: *Frontiers in Microbiology* 10 (JUN). ISSN: 1664302X. DOI: 10.3389/fmicb.2019.01316.

Rozmoš, Martin et al. (Mar. 2022). "Organic nitrogen utilisation by an arbuscular mycorrhizal fungus is mediated by specific soil bacteria and a protist". In: *The ISME Journal* 16 (3), pp. 676–685. ISSN: 1751-7362. DOI: 10.1038/s41396-021-01112-8.

Russell, Calum W., Sophie Bouvaine, Peter D. Newell, and Angela E. Douglas (Oct. 2013). "Shared Metabolic Pathways in a Coevolved Insect-Bacterial Symbiosis". In: *Applied and Environmental Microbiology* 79 (19), pp. 6117–6123. ISSN: 0099-2240. DOI: 10.1128/AEM.01543-13.

Ryback, Birgitta, Miriam Bortfeld-Miller, and Julia A Vorholt (Dec. 2022). "Metabolic adaptation to vitamin auxotrophy by leaf-associated bacteria". In: *The ISME Journal* 16 (12), pp. 2712–2724. ISSN: 1751-7362. DOI: 10.1038/s41396-022-01303-x.

Seiflein, Thomas A and Jeffrey G Lawrence (2001). "Methionine-to-Cysteine Recycling in Klebsiella aerogenes". In: *JOURNAL OF BACTERIOLOGY* 183 (1), pp. 336–346. DOI: 10.1128/JB.183.1.336-346.2001.

Shigenobu, Shuji, Hidemi Watanabe, Masahira Hattori, Yoshiyuki Sakaki, and Hajime Ishikawa (Sept. 2000). "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS". In: *Nature* 407 (6800), pp. 81–86. ISSN: 0028-0836. DOI: 10.1038/35024074.

Sørensen, Megan E.S. et al. (Jan. 2020). "Comparison of Independent Evolutionary Origins Reveals Both Convergence and Divergence in the Metabolic Mechanisms of Symbiosis". In: *Current Biology* 30 (2), 328–334.e4. ISSN: 09609822. DOI: 10.1016/j.cub.2019.11.053.

Soto-Martin, Eva C. et al. (Aug. 2020). "Vitamin Biosynthesis by Human Gut Butyrate-Producing Bacteria and Cross-Feeding in Synthetic Microbial Communities". In: *mBio* 11 (4). Ed. by David A. Relman, pp. 1–18. ISSN: 2161-2129. DOI: 10.1128/mBio.00886-20.

Starke, Svenja et al. (2023). "Amino acid auxotrophies in human gut bacteria are linked to higher microbiome diversity and long-term stability". In: *ISME Journal* 17 (12), pp. 2370–2380. ISSN: 17517370. DOI: 10.1038/s41396-023-01537-3.

Strobel, H J (July 1992). "Vitamin B12-dependent propionate production by the ruminal bacterium Prevotella ruminicola 23". In: *Applied and Environmental Microbiology* 58 (7), pp. 2331–2333. ISSN: 0099-2240. DOI: 10.1128/aem.58.7.2331-2333.1992.

Taubenheim, Jan, Máté Miklós, Jácint Tökölyi, and Sebastian Fraune (2022). "Population Differences and Host Species Predict Variation in the Diversity of Host-Associated Microbes in Hydra". In: *Frontiers in Microbiology* 13 (March), pp. 1–17. ISSN: 1664302X. DOI: 10.3389/fmicb.2022.799333.

Taubenheim, Jan, Doris Willoweit-Ohl, Mirjam Knop, Sören Franzenburg, Jinru He, et al. (Sept. 2020). "Bacteria- and temperature-regulated peptides modulate β-catenin signaling in Hydra". In: *Proceedings of the National Academy of Sciences* 117 (35), pp. 21459–21468. ISSN: 0027-8424. DOI: 10.1073/pnas.2010945117.

Törönen, Petri and Liisa Holm (Jan. 2022). "PANNZER –A practical tool for protein function prediction". In: *Protein Science* 31 (1), pp. 118–128. ISSN: 0961-8368. DOI: 10.1002/pro.4193.

Turnbaugh, Peter J., Fredrik Bäckhed, Lucinda Fulton, and Jeffrey I. Gordon (Apr. 2008). "Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome". In: *Cell Host & Microbe* 3 (4), pp. 213–223. ISSN: 19313128. DOI: 10.1016/j.chom.2008.02.015.

Uchiumi, Yu and Akira Sasaki (July 2020). "Evolution of division of labour in mutualistic symbiosis". In: *Proceedings of the Royal Society B: Biological Sciences* 287 (1930), p. 20200669. ISSN: 0962-8452. DOI: 10.1098/rspb.2020.0669.

Wang, Kun, Zihao Liu, Chunmei Du, Benhai Xiong, and Liang Yang (Aug. 2022). "Responses of Fermentation Characteristics and Microbial Communities to Vitamin B12 Supplementation in In Vitro Ruminal Cultures". In: *Fermentation* 8 (8), p. 406. ISSN: 2311-5637. DOI: 10.3390/fermentation8080406.

Wein, Tanita, Tal Dagan, Sebastian Fraune, Thomas C.G. Bosch, Thorsten B.H. Reusch, et al. (2018). "Carrying capacity and colonization dynamics of Curvibacter in the hydra host habitat". In: *Frontiers in Microbiology* 9 (MAR), pp. 1–10. ISSN: 1664302X. DOI: 10.3389/fmicb.2018.00443.

Wolschin, Florian, Bert Hölldobler, Roy Gross, and Evelyn Zientz (July 2004). "Replication of the Endosymbiotic Bacterium Blochmannia floridanus Is Correlated with the Developmental and Reproductive Stages of Its Ant Host". In: *Applied and Environmental Microbiology* 70 (7), pp. 4096–4102. ISSN: 0099-2240. DOI: 10.1128/AEM.70.7.4096-4102.2004.

Yang, Guoguo, Lei Li, Fuchun Li, Chonghong Zhang, and JieJie Lyu (Jan. 2021). "Mechanism of carbonate mineralization induced by microbes: Taking Curvibacter lanceolatus strain HJ-1 as an example". In: *Micron* 140 (November 2020), p. 102980. ISSN: 09684328. DOI: 10.1016/j.micron.2020.102980.

Zhang, Chonghong, Jiejie Lv, Fuchun Li, and Xuelin Li (Dec. 2017). "Nucleation and Growth of Mg-Calcite Spherulites Induced by the Bacterium Curvibacter

lanceolatus Strain HJ-1". In: *Microscopy and Microanalysis* 23 (6), pp. 1189–1196. ISSN: 1431-9276. DOI: 10.1017/S1431927617012715.

Zhang, Lin, Jiachao Zhou, Timothy S. George, Erik Limpens, and Gu Feng (Apr. 2022). "Arbuscular mycorrhizal fungi conducting the hyphosphere bacterial orchestra". In: *Trends in Plant Science* 27 (4), pp. 402–411. ISSN: 13601385. DOI: 10.1016/j.tplants.2021.10.008.

Zientz, Evelyn, Thomas Dandekar, and Roy Gross (Dec. 2004). "Metabolic Interdependence of Obligate Intracellular Bacteria and Their Insect Hosts". In: *Microbiology and Molecular Biology Reviews* 68 (4), pp. 745–770. ISSN: 1092-2172. DOI: 10.1128/MMBR.68.4.745-770.2004.

6 Manuscript 3

CATHI: An interactive platform for comparative genomics and homolog identification

Lukas Becker[1,2], Philipp Spohr[3], Gunnar W. Klau[3], Ilka M. Axmann[1], Sebastian Fraune[2], Nicolas M. Schmelling[4*]

[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
[3]Algorithmic Bioinformatics, Department of Computer Science, Heinrich-Heine-University Düsseldorf, 40225 Düsseldorf, Germany
[4]Krauts & Sprouts, 40237 Düsseldorf, Germany

[*]Corresponding author: Nicolas M. Schmelling

Keywords: Homolog Identification, Comparative Genomics, No code platform/solution/tool, Interactive and intuitive web interface, Pipeline, Snakemake, Docker, Synteny, Phylogeny, Target sequence search

6.1 Abstract

Bioinformatics has established itself as a central pillar of modern biology. Specifically, comparative genomics enables scientists to study a vast number of genomes efficiently. These comparative analyses shed light on the evolution and potential function of genomes and genes, but are also increasingly used as a key tool for metabolic engineering and synthetic biology by identifying appropriate targets for modification. While numerous sophisticated tools for comparative genomics and homolog identification exist, those tools predominantly target highly skilled bioinformatics users. Consequently, many biologists either defer such analyses to their more versed bioinformatic collaborators or resort to suboptimal tools.
Here, we present an intuitive solution available on all major operating systems, easily accessed through common web browsers. CATHI – Comparative Analysis Tool for Homolog Identification – integrates a suite of best-practice bioinformatic tools, encompassing BLAST for homology searches, MAFFT for multiple sequence alignment, FastTree2 for phylogeny reconstruction, and clinker for synteny analysis. Specifically tailored to biologists, CATHI orchestrates predefined settings and automated pipelines, obviating the need for programming expertise. This platform empowers researchers to confidently engage in detailed comparative genomics studies by streamlining the analytical process. The interactive framework provides users with a plethora of options. This includes real-time execution and progress monitoring, facilitates dynamic result tracking, and a set of search functions across NCBI databases like CDD or ProtFam. Users can interactively engage in data exploration, filtering, and visualization through CATHI's intuitive interface. Furthermore, the seamless export of project data in standard formats (FASTA, Newick, CSV, and HTML) facilitates the integration with further third-party tools such as TreeViewer and Jalview. To benchmark CATHI, we revisited the comparative analysis of cyanobacterial circadian clock proteins conducted by Schmelling et al. in 2017, revealing consistent global patterns among identified homologs, while also highlighting individual variations attributed to the expansion of available databases.

6.2 Introduction

Bioinformatics has emerged as a critical cornerstone of modern biological research, reshaping the landscape of biological exploration and transforming it from traditional laboratory confines into a dynamic, data-driven, and multidisciplinary field. Comparative genomics allows researchers to compare and contrast genomes from

different species. Those analyses provide insights into evolutionary relationships, identifying conserved genes and regions, and understanding genetic variations that underlie phenotypic diversity. A key concept of evolutionary and comparative genomics is the dichotomous differentiation of homology into orthologous and paralogous sequences (Koonin 2005; Kuzniar et al. 2008). Paralogs are genes related by an ancestral gene duplication event, while orthologs are homologous genes in different species descended from a single ancestral gene of a last common ancestor through vertical inheritance and subsequent speciation (Fitch 1970). Sets of orthologous genes are used to obtain information regarding gene function and phylogenetic relationships. Thus, a clear distinction between orthologous and paralogous genes is crucial for reliable and robust genome annotation and phylogenetic inference (Kuzniar et al. 2008; Tekaia 2016). However, those definitions were proposed before the advent of the genomic era. Assigning orthology or paralogy for genomic sequences is much more complex as sufficient information is still missing to "determine the timing of many of the speciation and gene duplication events" (Jensen 2001). Thus, it is suggested to classify homologs based on sequence-structure-function relationships as "isofunctional" or "heterofunctional" and "isospecic" or "heterospecic" homologs (Jensen 2001).

Typically, orthologous and paralogous gene relationships are disentangled by analyzing sequence similarities and their distribution within phylogenies (Koonin 2005; Kristensen et al. 2011). Different computational solutions are required depending on the biological question (Descorps-Declère et al. 2008; Dunne and Kelly 2017). Thus, various tools have been developed to infer orthologs (Nichio et al. 2017). Some of these tools, like OrFin (Midha et al. 2012), OrthoMCL (Li et al. 2003), and OrthoFinder (Emms and Kelly 2019), use a combination of existing bioinformatic software, such as BLAST (Basic Local Alignment Search Tool; (Altschul et al. 1990, 1997; Camacho et al. 2009)), DIAMOND (Buchfink et al. 2014), MMseqs2 (Steinegger and Söding 2017), Markov Clustering (MCL) (Enright et al. 2002), or phylogenetic reconstructions. Other tools are based on different approaches to identify orthologs, like SPOCS (Curtis et al. 2013), OrthoInspector (Linard et al. 2011) or JustOrthologs (Miller et al. 2019). SPOCS combines the results of pairwise species comparisons from the InParanoid (Remm et al. 2001) algorithm with graph-based predictions of orthologs and paralogs (Curtis et al. 2013), and aims to identify orthologs within species groups. OrthoInspector (Linard et al. 2011) validates the output of sequence similarity searches and creates groups of in-paralogs that are compared pairwise to assign potential orthologs and/or in-paralogs, while

JustOrthologs (Miller et al. 2019) uses coding sequence lengths and dinucleotide percentages to assign orthologs. Moreover, the focus of these tools or their respective online databases are different. While morFeus (Wagner et al. 2014) is specifically designed for identifying remotely conserved orthologs, SPOCS, on the other hand, is aimed at identifying orthologs within closely related species (Curtis et al. 2013). These tools differ in their installation, operating system compatibility, and usage (Table 6.1).

In addition, various biological databases like the Cluster of Orthologous Groups (COG) database (Tatusov et al. 2000), OrthoDB (Zdobnov et al. 2021) or InParanoidDB9 (Persson and Sonnhammer 2023) have been developed to store groups of orthologous and/or paralogous sequences, allowing for rapid identifications of homologous relationships among genes and species. EukProt (Richter et al. 2022) and PanOCT (Inman et al. 2019) are databases that primarily focus on identifying orthologs among eukaryotic and prokaryotic genomes, respectively. In contrast, more generalized approaches like the triangular Reciprocal Best Hit (RBH) method used in the COG database (Tatusov et al. 2000) take a broader approach to unravel the homologous relationships among genes.

Still, in most of these tools or their respective databases, homologs are (at least partly) identified through RBHs (Table 6.1). RBHs can be assigned by performing a reciprocal or symmetric BLAST search (Fig. 6.1). BLAST rapidly searches large sequence databases by finding local similarities. Local alignments identify regions of similarity between two sequences, highlighting regions that might share a common function. BLAST starts with a short sequence, searches for sequences sharing similarities, and extends those alignments as best as possible by applying cut-off values based on sequence substitution matrices (Altschul et al. 1990). In contrast, global alignments use the complete sequences and try to best align them with one another (Smith and Waterman 1981).

Table 6.1: **Brief comparison of a selection of analytical tools for the identification of homologs and orthologs.**

Tool	**Advantages**	**Disadvantages**	**DOI**
CATHI	Reciprocal BLAST technique, platform independent, no requirements except Docker, web-interface, interactive result analysis, combination of different bioinformatic tools for post-processing (selection constrained phylogeny, synteny etc.), customized databases, API to NCBI via E-Direct: interactive sequence/domain search	Reciprocal BLAST algorithm: large query sets will require a long run time, post-processing is user dependent, projects limited by one organism	This publication
morFeus	Reciprocal BLAST technique, Orthology network scoring, detection of distantly conserved orthologs, command line execution with one command	Web-server is offline, a lot of additional software requirements (e.g. NetworkX and BLAST+), RefSeq database must be downloaded and formatted before use outside the tool, UNIX only, additional post-processing steps outside morFeus	`https://doi.org/10.1186/1471-2105-15-263`
OrthoMCL	Grouping of orthologs using reciprocal BLAST technique with a subsequent Markov Clustering (MCL) algorithm, the OrthoMCL web-server enables simple comparison of those groups according to certain filter criteria	Many requirements (e.g. BLAST+, Relational Database, Perl, MCL), UNIX only, no pipeline, 13 steps need to be performed on the command line (`https://orthomcl.org/orthomcl/app/downloads/software/v2.0/UserGuide.txt`), the tool outputs three text files with three groups of homologous sequences (orthologs, co-orthologs, paralogs), which will require additional post-processing steps outside OrthoMCL	`https://doi.org/10.1093/nar/gkj123`

OrthoFinder	Several workflow steps: assigns genes to orthogroups (1) and infers rooted orthogroup trees by utilizing DendroBLAST (Kelly and Maini 2013) or MSA and a combination of STAG (Emms and Kelly 2018) and STRIDE (Emms and Kelly 2017) (2), identification of orthologs and paralogs by applying a hybrid algorithm based on the species-overlap method (Huerta-Cepas et al. 2007) and the duplication-loss-coalescent model (Wu et al. 2014) (3) No requirements except Docker, command line execution with one command, produces rooted species-trees and identifies gene duplication events, a lot of customizable options	Only command-line version available, no database, inference of orthologs between provided FASTA files (smaller search space but more accurate), additional post-processing steps (for visualization, extraction of results) must be performed outside OrthoFinder	https://doi.org/10.1186/s13059-019-1832-y
SPOCS	Reciprocal BLAST technique with subsequent graph-based ortholog prediction method, that combines pairwise predictions from InParanoid (Remm et al. 2001), MaxCliqueDyn for the identification of ortholog cliques (Konc and Janežič 2007), web-server and command line tool	Boost C++ and BLAST+ requirements, list of FASTA protein sequence files as input, no database, additional post-processing steps must be performed outside SPOCS, local installation is UNIX only, web-server offline	https://doi.org/10.1093/bioinformatics/btt454

Ortho-Inspector	Three steps pipeline: Reciprocal BLAST (1), comparison of in-paralog groups and assignment to potential orthologs/in-paralogs (2), best hits between organisms that contradict orthology are detected (3) Graphical interface and command line option, web-server option available, phylogenetic profiling	Installation requires configuration of a MySQL database, Java JRE and BLAST+ requirements, creation of custom databases requires some additional work (e.g. download of sequences), web-server option with a limited number of organisms (e.g. 4132 prokaryotic species)	https://doi.org/10.1186/1471-2105-12-11
JustOrthologs	Lengths of coding sequence regions and dinucleotide percentages are used for the assignment of orthologs, fast algorithm with comparable precision	Command line tool, input are altered FASTA files based on CDS regions, no database but two input FASTA files, additional post-processing steps must be performed outside JustOrthologs	https://doi.org/10.1093/bioinformatics/bty669
InParanoid (DB)	Orthologs are inferred based on pairwise similarity scores identified using DIAMOND (1), those orthologs are grouped (seed orthologs) (2) and additional sequences are added, based on similarities to those seed orthologs (3) Database of orthologous groups, accessible through a dedicated web-server, originally designed for eukaryotic orthologs (O'Brien et al. 2005; Remm et al. 2001), nowadays also listing diverse groups of organisms (Persson and Sonnhammer 2023)	Ortholog database consist of 640 organisms, for some genes orthogroups may not exist, additional post-processing steps must be performed outside InParanoid	https://doi.org/10.1093/nar/gki107
OrFin	Reciprocal BLAST technique, web-server tool	Web-server is offline, no command line tool available, no sequence database, comparison of two uploaded FASTA files, not suitable for large datasets, additional post-processing steps outside OrFin	https://doi.org/10.6026%2F97320630008738

PARIGA	Reciprocal BLAST technique, web-server tool, interactive post-processing through logical constraints	Web-server is offline, no command-line tool available, no sequence database, comparison of two uploaded FASTA files, not suitable for large datasets	https://doi.org/10.1371/journal.pone.0062224

Within a reciprocal BLAST approach, the sequence(s) of interest (SOI) from a specific species are used as input queries for the initial sequence similarity search, also known as the forward BLAST (Fig. 6.1, left). The generated output is then analyzed, and the sequences of all hits from the forward BLAST are extracted from the underlying database. These extracted sequences then serve as queries for the second sequence similarity search, also known as the backward BLAST (Fig. 6.1, right). In this step, the BLAST is constrained by explicitly searching against all sequences of the initial organism, also known as the query genome. Additionally, the output is limited to reporting only the best hit or alignment. RBHs are assigned if the query and subject sequences from the forward BLAST match against each other in the backward BLAST (Fig. 6.1, right).

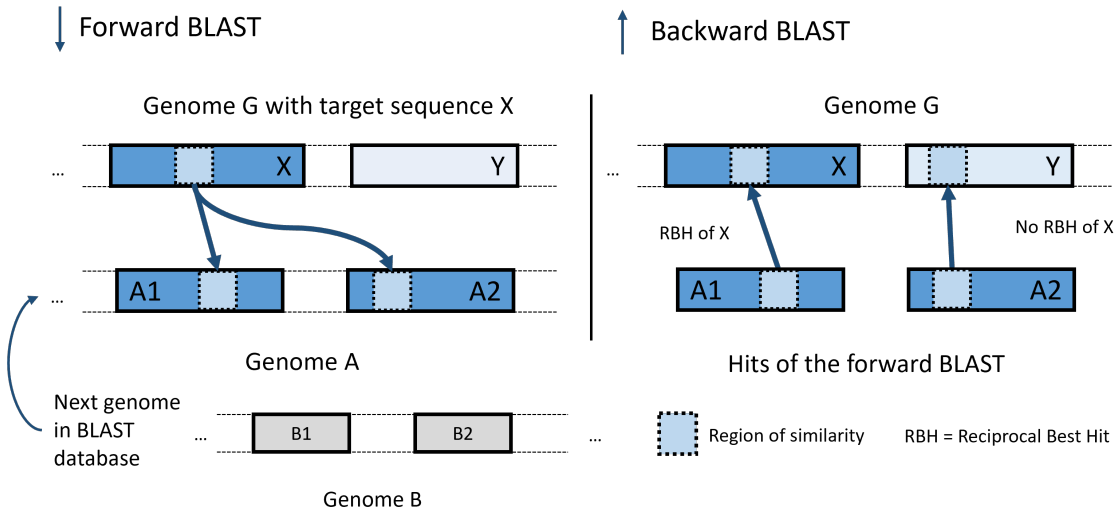

Figure 6.1: **Schematic overview of reciprocal BLAST.** Genome G is the target organism from which the query sequence X originates. X is the input (query sequence) for the first BLAST, the forward BLAST, left. The results of the forward BLAST (sequences A1 and A2) are the inputs to a second BLAST, the backward BLAST, right, which uses the query genome G as a database. A1 finds X as its best match, while A2 aligns best with Y. Thus, only A1 is a reciprocal best hit (RBH) to X.

Although sophisticated tools are available for identifying homologs (Table 1), to our experience, many biologists still do not use them. Those tools predominantly target bioinformaticians as their potential users. Thus, biologists leave those analyses to their bioinformatic collaboration partners or use less suitable tools, like the online BLAST tool available on NCBI, to identify homologs. While the online BLAST tool on NCBI is tremendously helpful to biologists, it offers limited options for correctly identifying homologs. Thus, it should instead only be used for the initial screening of candidates. In our experience, biologists tend to use tools that (1) have a graphical interface, (2) do not require any complicated installation process or command line experience, (3) can be used without programming skills, and (4) already offer many useful default settings. To overcome this discrepancy and also provide biologists with sophisticated tools to perform high-quality comparative genomic analyses and homolog identifications confidently, we combined a set of state-of-the-art tools and best-practice analyses (see Results - Identification of Homologous Sequences; Multiple Sequence Alignments and Phylogenetic Reconstruction; Genomic Structure and Synteny Analysis; Conserved Domain Database (CDD) Analysis) into an easy-to-use interactive tool with an intuitive graphical interface (see Results - Interactive Filtering).

Here, we introduce an accessible solution that can be effortlessly installed on all major operating systems, enabling straightforward usage through an interactive and intuitive web interface. The default settings are selected to provide a starting point for any homology search. However, more sophisticated users can change all settings to their personal preferences. Further, we illustrate CATHI's functionality and reliability through a reproduction and benchmarking analysis of cyanobacterial circadian clock proteins based on Schmelling et al., 2017 (Schmelling et al. 2017). This analysis revealed that overall trends in homology detection are consistent even with substantially increased database size. However, individual variations have been identified highlighting the need to thoroughly interpret results of comparative genomics analysis, especially when using different databases.

6.3 Results

CATHI combines a suite of powerful bioinformatic tools, including BLAST+ (Camacho et al. 2009) for homology searches, MAFFT (Katoh and Standley 2013) for multiple sequence alignments (MSA), FastTree2 (Price et al. 2010) for phylogeny reconstruction, and clinker (Gilchrist and Chooi 2021) for synteny analyses into

an interactive platform for comparative genomics specifically tailored to biologists. The predefined settings and pipeline automation streamline the process and enable researchers to perform sophisticated comparative genomics analyses regardless of their proficiency in programming or bioinformatics. The interactive and intuitive web interface provides users with multifaceted capabilities, including execution and monitoring, real-time result tracking, search function across diverse NCBI databases (Protein, CDD, ProtFam), and dynamic data exploration, filtering, visualization, and post-processing. CATHI runs on all major operating systems supporting Docker (Merkel 2014) and is accessible via typical browser applications (e.g., Firefox or Chrome). Through individual accounts, users have the option to create their custom projects and retrieve their comprehensive project data in standard formats (FASTA, Newick, CSV, HTML) suitable for further analysis with other third-party tools such as TreeViewer (Bianchini and Sánchez-Baracaldo 2023), Jalview (Waterhouse et al. 2009), or vector design programs. In the following, we highlight the main features of this platform and provide examples of potential use cases, illustrated through the replication of the prior comparative study on cyanobacterial circadian clock proteins by Schmelling et al., 2017 (Schmelling et al. 2017).

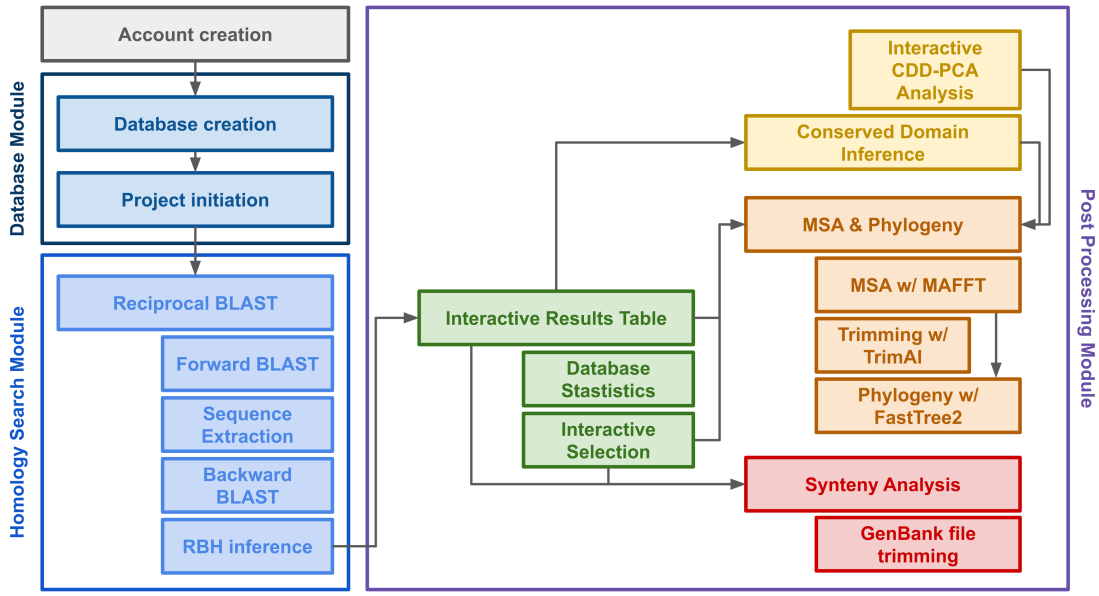

Figure 6.2: **Flowchart of CATHI's pipeline, modules, and post-processing options.** First, users have to create an account to login to the platform. They can use the **Database Module** to create a database from locally uploaded protein FASTA files or remotely available proteomes of the RefSeq or GenBank. Users can initiate a new project by selecting or creating appropriate databases, providing SOIs from one organism of interest per project and optionally customizing the default settings. After database and project creation, users perform the reciprocal BLAST analysis using the **Homology Search Module**. The underlying pipeline within the module will then automatically perform the forward and backward BLAST, sequence extraction, and reciprocal best hit (RBH) inference. The output of this module is directly forwarded to the **Post Processing Module**, where the users have different options to interactively analyze and filter their RBH data. Within the Post Processing Module, interactive results tables and statistics are available, synteny analyses and conserved domain interference can be performed, as well as multiple sequence alignments (MSA) and phylogenetic reconstruction. Furthermore, those different analyses can be combined into interactive pipelines in which users can first filter their RBH data by interactive graphs and tables before constructing MSAs and phylogenies.

6.3.1 Identification of Homologous Sequences

The homology search module within CATHI represents the core of each project. Users start each comparative genomics analysis by creating a new project through the web browser dashboard (Fig. 6.2). Before project initiation, users either choose from already formatted BLAST databases or create a new database using CATHI's database creation module (Fig. 6.2). Databases can range from the entire GenBank or RefSeq database from NCBI, over an individually filtered subset of those databases, to directly uploading annotated genome files. The latter option allows users to incorporate unpublished genome projects into their custom database. By entering the SOI(s) of one organism per project, users start the reciprocal BLAST analysis with either default (see Methods - BLAST and Identification of Reciprocal Best Hits) or custom settings. The automated pipeline then performs all steps of forward and backward BLAST as well as filtering without any further input from the user (Fig. 6.2 - Homology Search Module). Afterward, CATHI outputs a set of analysis statistics. Furthermore, the BioPython (Cock et al. 2009) library extracts taxonomic information of the organisms listed in the RBH results table from NCBI. Subsequently, these taxonomic data are integrated into the results table and written to the project directory (Fig. 6.2 - Interactive Results Table). All data tables can be downloaded as CSV files or are conveniently accessible within CATHI, facilitated by integrating the client-side JavaScript library DataTables (`http://datatables.net`). This versatile library enriches user experience with its suite of advanced functionalities, enabling seamless searching, sorting, paging, and dynamic interaction with extensive HTML tables. Through the homology search module, users can perform reliable reciprocal BLAST analysis with just a few clicks and inputs (Fig. 6.2).

6.3.2 Multiple Sequence Alignments and Phylogenetic Reconstruction

To identify conserved regions and patterns of similarity/difference within the identified RBHs and to elucidate the evolutionary relationship between them, post-processing involves the construction of an MSA and phylogenetic tree for each set of RBHs (Fig. 6.2 - MSA & Phylogeny). MAFFT (Multiple Alignment using Fast Fourier Transform) is used to conduct the MSA (Katoh and Standley (2013); Fig. 6.2). Its primary purpose is to align multiple sequences with homologous regions to identify conserved regions and evolutionary relationships, which makes it perfectly suitable for CATHI. Subsequently, the alignment undergoes a refining pro-

cess using trimAl (Capella-Gutiérrez et al. 2009), a specialized tool that automates the removal of inadequately aligned segments (Fig. 6.2). The trimming not only enhances the precision of the alignment but also contributes to the refinement of phylogenetic inference, ensuring a more robust and accurate analysis. To highlight conserved regions within the MSA and create an interface that allows interactive visualization, an HTML document is created using the MView Brown et al. (1998) tool. Phylogenetic inference is performed with FastTree2 (Price et al. (2010); Fig. 6.2). It leverages sophisticated approaches to approximate maximum-likelihood phylogenies. Subsequently, resulting phylogenetic trees are processed by shiptv (`https://github.com/peterk87/shiptv`) to generate interactive HTML visualizations (Fig. 6.3). This visualization integrates the RBH results table by aligning the data to the phylogenetic tree leaf labels, facilitating a comprehensive and insightful depiction of sequence relationships. Through the integration of the data tables, users can interactively filter and adjust the phylogenetic tree based on selection criteria such as taxonomic groups (Fig. 6.3).

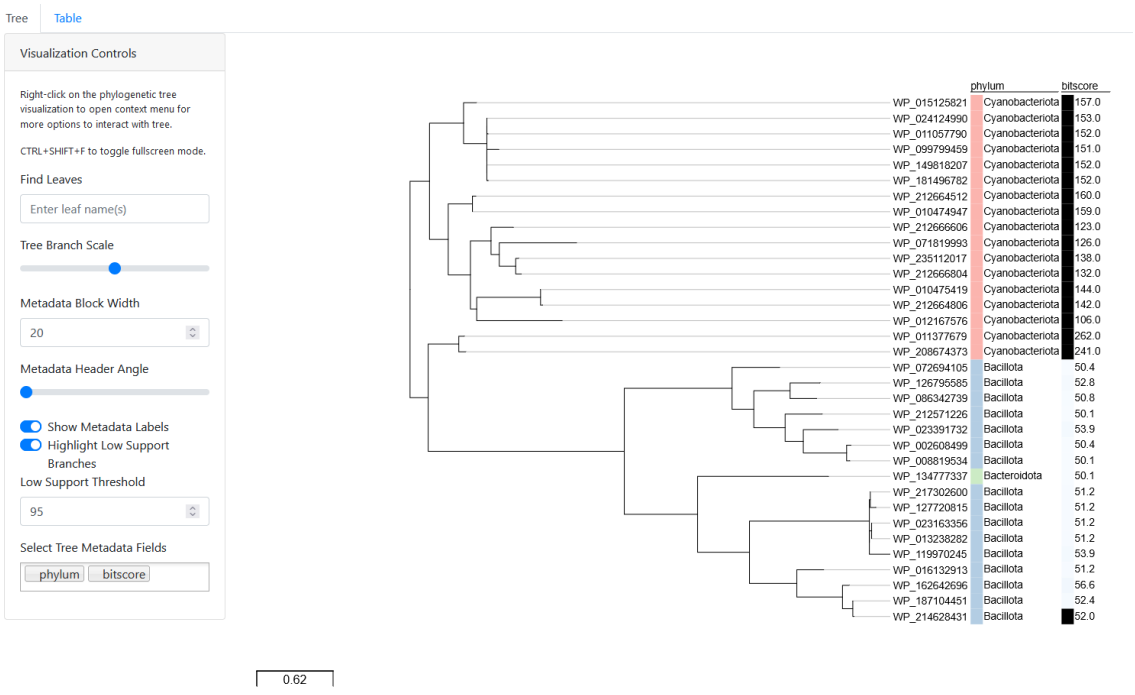

Figure 6.3: **Interactive Phylogeny Visualization.** The interactive phylogeny visualization is presented here as an exemplary illustration. This dynamic visualization was generated using a set of bioinformatic tools: MAFFT provides accurate multiple sequence alignments, FastTree2 enables rapid and efficient phylogeny reconstruction, and shiptv facilitates the creation of an interactive dashboard view. The tree's branches represent evolutionary relationships, while the distribution of a selection of RBHs of the cyanobacterial circadian clock input factor Pex is highlighted across three bacterial clades, clearly distinguishing between *Cyanobacteria* and *Bacillota*.

6.3.3 Genomic Structure and Synteny Analysis

Beyond RBHs identification, MSA, and phylogenetic reconstruction, users can analyze the genomic structure and synteny surrounding the RBHs (Fig. 6.2 - Synteny). Therefore, CATHI integrates the clinker tool, which allows researchers to compare and visualize syntenic regions across different genomes (Gilchrist and Chooi 2021). To perform this analysis, users can select up to ten RBHs of interest, and CATHI will download the corresponding GenBank files from NCBI and extract gene loci around each RBH. These sliced GenBank files are automatically passed to clinker as input, enabling users to easily identify and analyze conserved gene clusters and gene order between genomes directly from their RBH data (Fig. 6.4).

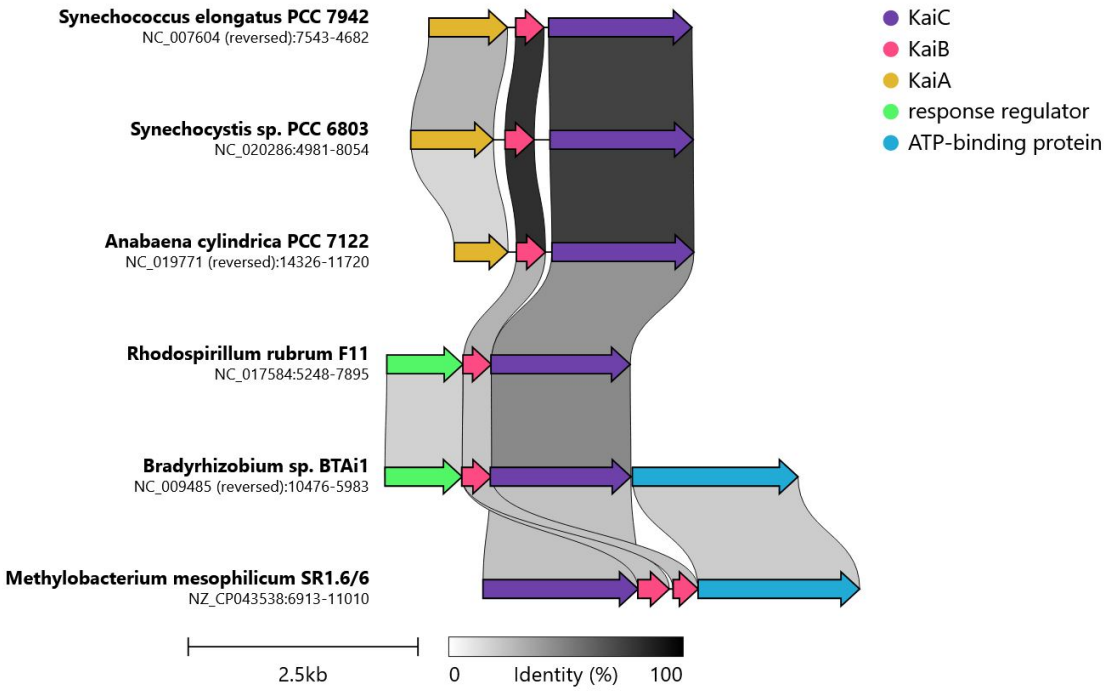

Figure 6.4: **Exemplary illustration of the synteny analysis results using clinker as part of CATHI's post-processing module.** In this analysis, the core circadian clock protein KaiC was selected. Clinker performed the synteny analysis. The graph shows the conserved genomic elements up- and downstream of KaiC that are represented in the selected genomes. The threshold for sequence identity is set to 0.25 (25%). Interestingly, the synteny analysis reveals some intricate information about the KaiABC cluster without any additional input that can easily be missed in standard RBH analyses. First, the *Anabaena/Nostoc* KaiA has a smaller sequence length compared to the KaiA sequences of *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803. Additionally, the synteny reveals the presence of a conserved response regulator/transcription factor in the genomes of the *Bradyrhizobium* and *Rhodospirillium rubrum* strains at a position where normally KaiA can be found. This protein was recently identified as a novel KaiA homolog and was named KaiA3 due to its functional similarity (Köbler et al. 2023).

6.3.4 Interactive Filtering

CATHI offers a set of additional features to analyze the RBHs based on taxonomic information and database entries (Fig. 6.2 - Interactive Selection). The number of available genomes of an organism (e.g., *E. coli*) in the BLAST database may vary, affecting the analysis of RBHs due to high data redundancy. Therefore, project-specific database statistics are calculated based on the taxonomic information within the RBH results tables and the underlying BLAST database. Each RBH is derived from a genome file of one organism. CATHI calculates database-normalized percentages of taxonomic units within identified RBHs. The resulting database statistics are integrated into the RBH results table and visualized by an interactive chart that combines RBH characteristics and a table of database entries. The table and the chart are linked, and the table data is updated according to the specified user selection (Fig. 6.5). This comprehensive graph empowers users to filter RBHs based on various features, including bitscores, RBH sequence lengths, sequence identity, e-value, and taxonomic nodes (Fig. 6.5).

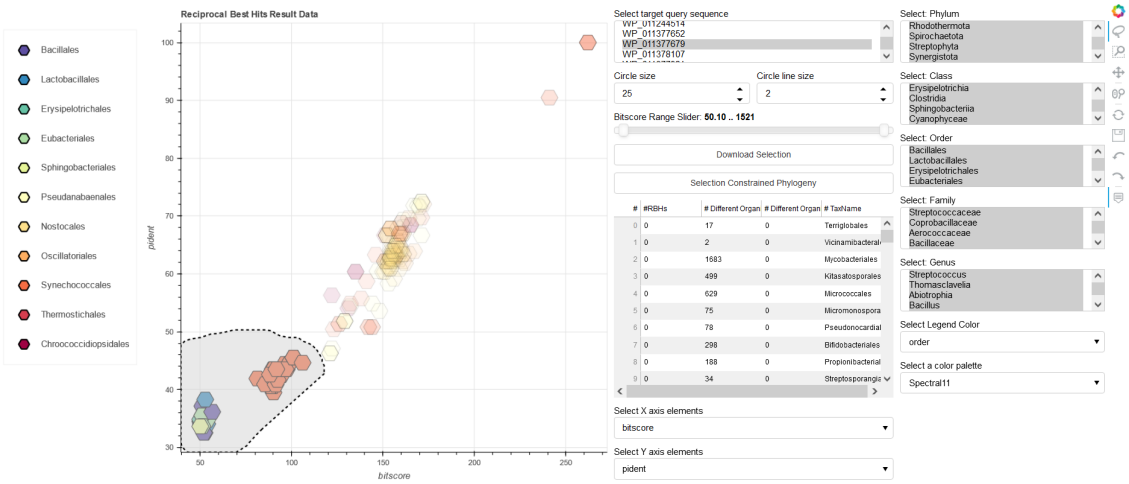

Figure 6.5: **Interactive reciprocal best hits analysis and filtering.** The interactive scatter plot leverages the reciprocal best hit (RBH) results dataset, exemplified here with the Pex protein dataset. The X and Y axes of the scatter plot are parameters of the dataset, such as bitscore, e-value, sequence identity (pident), and sequence lengths, enabling the exploration of relationships among these RBHs. The interactivity of the scatter plot is facilitated by Bokeh, a powerful visualization library (Bokeh Development Team 2018). Users can dynamically manipulate the dataset through filtering options informed by the taxonomic information associated with each RBH. In addition, Bokeh provides a lasso tool to select specific RBHs within the graph. The lasso tool can be selected from the Bokeh tool-panel displayed at the right side of the figure. This interactive feature empowers researchers to dissect real-time taxonomic trends and relationships among RBHs, unveiling underlying patterns and insights. In this example featuring the Pex protein dataset, the scatter plot offers a multidimensional view of the RBHs landscape. As users navigate through the data points, taxonomic affiliations emerge, shedding light on the distribution and relationships of Pex protein homologs across various taxa. The provided interactivity not only enhances the visualization experience but also grants researchers the capability to refine their analyses, making the scatter plot an indispensable tool for exploring the taxonomic diversity of RBHs.

To facilitate user interaction, a lasso tool is provided, enabling the direct selection of RBHs within the graph (Fig. 6.5). The database entry table is thereby filtered by the selected RBHs. The selected RBHs can be downloaded or utilized for the creation of multiple sequence alignments and phylogenetic inferences (Fig. 6.2). By leveraging this interactive graph, users can effectively explore and analyze the RBHs, gaining valuable insights and facilitating the identification of relevant sequences. The integration of RBH features and database statistics within the visual representation enhances the usability and interpretability of the results. The number of representative proteins can be examined and compared in other organisms and across phylae, classes, orders, families, and genera.

6.3.5 Conserved Domain Database Analysis

To gain further knowledge about SOI(s) CATHI offers the possibility to infer protein domains for the SOI and the associated RBHs (Fig. 6.2 - Conserved Domain Inter-

ference). The domains are inferred from a local copy of NCBI's Conserved Domain Database (CDD) (Marchler-Bauer et al. 2014). CATHI generates result tables of protein domains that are transferred into interactive HTML tables for visualization and analysis. Additionally, CATHI writes a FASTA file containing the concatenated domain sequences of the RBHs. This FASTA file is then used for an additional MSA and refined phylogenetic reconstruction, as highlighted before. Furthermore, a principal component analysis (PCA) is conducted on a percent identity (RBH vs. CDD entry) table of the conserved domains (Fig. 6.2). A graph of the first two principal components of the PCA in combination with a domain table is visualized in an interactive chart, with the possibility to filter RBHs based on taxonomy, similar to the interactive database statistic graph (Fig. 6.6). The selected RBHs can be downloaded for studies outside CATHI or used for a selection constrained phylogenetic inference based on the identified protein domains (Fig. 6.2).

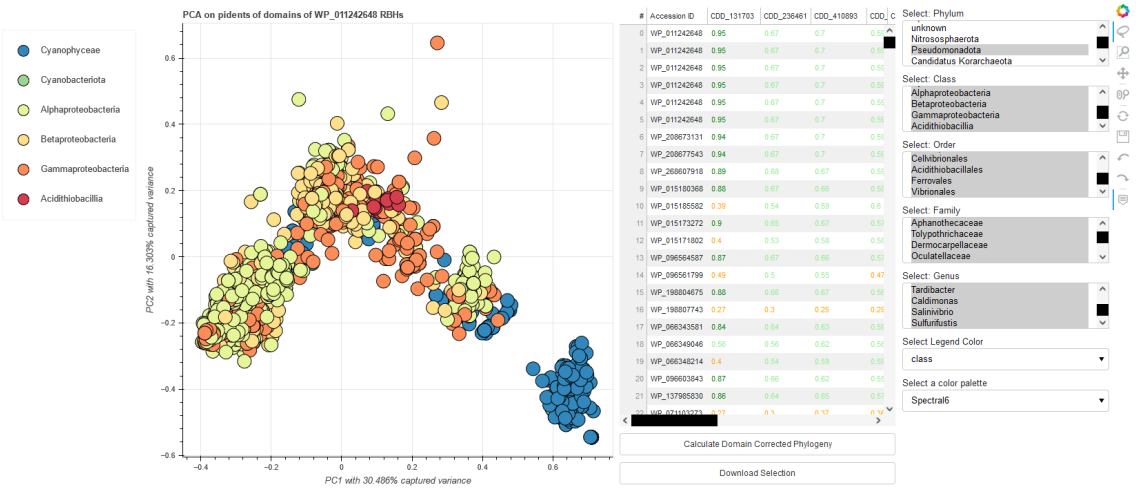

Figure 6.6: **Interactive graph of inferred CDDs of *Synechococcus* KaiC RBHs.** The figure displays the first two principal components of a principal component analysis (PCA) calculated on a percent identity table (RBH vs. CDD entry) of protein domains (displayed in the table next to the graph) inferred from RBHs of the *Synechococcus* KaiC protein. The graph and table are linked, selection of specific scatter-points will alter the percent identity table. Coloring within the graph is based on the associated taxonomic class, coloring within the table is based on the percent identity value. The figure displays a selection of organisms from the *Cyanobacteria* and *Pseudomonadota* phylum. Some of the RBHs are known KaiC2/C3 cyanobacterial proteins, paralogs of the query sequence KaiC. Interestingly, the cyanobacterial RBHs most similar to the *bona fide* KaiC form a cluster in the lower right corner of the diagram. They exhibit differences in their domain structure compared to the other RBHs.

6.3.6 Reproducibility

The amount of biological data has increased dramatically and will continue to grow, rendering data comparisons and analysis more complicated. Consequently, scalable software tools and reproducible bioinformatic analysis are essential for modern comparative genomics (Wratten et al. 2021). Within CATHI we implemented a set of measures to ensure and enhance the reproducibility and integrity of each analysis. First, deploying individualized local accounts prevents the accidental merger of projects and ensures a clear dashboard for each user. The platform enforces unique project names by design to prevent inadvertent data overwrite. Further, each BLAST database is created as a discrete and unmodifiable database, thereby preserving the content of the database at the time of initiation and insulating it from subsequent updates. This approach allows users to re-execute analyses at any time, building a dynamic environment that promotes the iterative validation and refinement of results. CATHI's architecture is grounded in utilizing a PostgreSQL database equipped with the inherent advantage of data persistence. This resilience against local file deletion safeguards against inadvertent data loss, ensuring the persistence of each dataset. Through containerization with Docker, the specific version of the underlying tools is preserved and recorded, simplifying the reporting of software versions. Furthermore, the orchestration with Snakemake (Köster and Rahmann 2012) amplifies the automation of the pipeline and ensures the correct execution, reducing potential errors from manual data manipulation and execution of tools. To further increase reproducibility, Snakemake configuration files are written into each underlying project folder, and corresponding Snakefiles and scripts can even be used outside the tool.

Each created database and analysis is assigned a timestamp. In addition, log files are generated, recording the exact steps and execution of each analysis. The resulting raw and processed data of each analysis are preserved in their entirety, while any filtering or transformation generates discrete datasets that augment interactive exploration and evaluation. This approach circumvents the overwriting of original data, encouraging users to explore and manipulate data without fear of irrevocable data loss. Incorporating interactive HTML tables allows each data table to be accessed through the platform. Further, each analysis result can be downloaded in a standardized file format, such as CSV for data tables, FASTA for sequence and alignment files, and Newick for phylogenetic trees. Conforming to standardized file formats further enhances the reproducibility of findings across disparate computational pipelines, ensuring interoperability and knowledge exchange.

6.3.7 Benchmarking

To assess the performance of CATHI's reciprocal BLAST pipeline, we conducted a comparative analysis of circadian clock proteins, as previously described in Schmelling et al., 2017, Wiegard et al., 2013 and Axmann et al., 2014 (Axmann et al. 2014; Schmelling et al. 2017; Wiegard et al. 2013). The cyanobacterial circadian clock primarily consists of three key proteins: KaiA, KaiB, and KaiC. KaiC, in particular, plays a crucial role and exhibits ATPase, auto-phosphatase, and auto-kinase activities. Activation of KaiC occurs through KaiA-mediated auto-phosphorylation at specific amino acid residues, while KaiB acts as an antagonist by occupying potential KaiA binding sites (Swan et al. 2018).

We performed the analysis on the same set of protein sequences used in the study by Schmelling et al., 2017 (Schmelling et al. 2017) (Table S1). To achieve this, we utilized CATHI's database features to download and format RefSeq proteomes into BLAST databases. The entries were filtered automatically by CATHI based on the completeness levels "Complete Genome" and "Chromosome", resulting in a final database containing 53,657 proteome entries with 28,326 unique taxonomic identifiers from 94 different phylae and 168,314,551 protein sequences. Within this database, there are 280 proteomes from 265 different cyanobacterial species. The database used in the analysis conducted by Schmelling et al., 2017 comprised 11,300 proteomes, including 94 cyanobacterial proteomes from 64 distinct cyanobacterial organisms. Each of these proteomes was used as a single BLAST database, which minimizes the e-value for possible homologous sequences, whereas CATHI uses a single BLAST database consisting of all downloaded proteomes. Following CATHI's best practices, we created two reciprocal BLAST projects: one for 16 *Synechococcus* query proteins and another for six *Synechocystis* query proteins (Table S1). The proteins WP_011378436 and WP_041425845 used in Schmelling et al., 2017 have been suppressed from NCBI as they are no longer annotated on any genome and are excluded from this study. Instead of the suppressed WP_041425845, the updated version under the accession number WP_010873802 was used, while WP_011378436 was excluded from the analysis. All remaining query proteins were extracted from the result directories provided in Schmelling et al., 2017, and removed sequence identifiers have been replaced with new identifiers. The reciprocal BLAST analysis yielded 73,023 RBHs for the *Synechococcus* queries and 5,410 RBHs for the *Synechocystis* queries. In comparison, the analysis by Schmelling et al., 2017 reported 24,027 RBHs for *Synechococcus* and 1,576 RBHs for *Synechocystis*. However, among all RBHs identified in this study, only 251 *Synechocystis* RBHs were not included

in the previous analysis, while 4,907 *Synechococcus* RBHs were missed. Of the 251 *Synechocystis* RBHs, 28 have been removed from NCBI, 196 have been suppressed by NCBI, and 25 sequences do not reside in the database, leaving two sequences not detected in this analysis. Those two sequences can be found in the results of the forward BLAST, though. Of the 4,907 *Synechococcus* RBHs, 424 have been removed from NCBI, 2,548 have been suppressed by NCBI, and 327 sequences do not reside in the database, leaving 1,608 sequences not detected in this analysis. 118 of those sequences have been detected within the forward BLAST, while 1,490 sequences do not reside in the forward BLAST table.

The circadian clock proteins KaiA, Pex, LdpA, and CdpA have been exclusively identified in *Cyanobacteria* (Schmelling et al. 2017). In the analysis conducted using CATHI, hits are observed outside the Cyanobacteria phylum for Pex and LdpA. Pex functions as a transcriptional repressor of kaiA in *Synechococcus elongatus* PCC 7942, featuring a relatively concise sequence length of 126 amino acids. For Pex, the CATHI pipeline identified 18 RBHs from 16 organisms beyond the *Cyanobacteria* phylum, displaying a mean bitscore value of 52 (Fig. 6.5). Moreover, none of those organisms possess RBHs of the central cyanobacterial circadian clock genes KaiABC. These results support the conclusion that these RBHs serve as indicators of distant homologous sequences, providing a compelling illustration of the necessity for user-driven post-processing — an aspect effortlessly facilitated through the interactive capabilities offered by CATHI. The phylogenetic tree derived from the analysis of those 18 non-cyanobacteria RBHs of Pex, in conjunction with a curated set of cyanobacterial Pex sequences, effectively illuminates the more distant relationship that exists among these sequences (Fig. 6.3). Notably, the redox-sensing protein LdpA demonstrates 268 non-cyanobacteria RBHs, characterized by a bitscore mean value of 218, signifying close homologous relationships. An intriguing pattern emerges within the 268 non-cyanobacteria hits for LdpA. A total of 248 hits are found within the *Streptophyta* phylum, originating from a diverse array of 68 distinct genera (88.46% of all *Streptophyta* possess RBHs of LdpA). Similar to the study conducted by Schmelling et al., 2017, RBHs of KaiA have been exclusively detected within the *Cyanobacteria* phylum in CATHI. Furthermore, an analysis of the KaiA protein sequence's length distribution unveiled the same three distinct subtypes of KaiA (Fig. 6.7). Consistent with the analysis by Schmelling et al., 2017, the majority of KaiA RBHs exhibit a sequence length of approximately 300 amino acids, while the remaining RBHs vary between 100 and 200 amino acids in length. Our benchmarking analysis focused on the robustness and reproducibility of results

when contrasted with previous analyses. While minor differences surfaced due to updates within NCBI's database, these deviations did not impede the overarching tendencies (Fig. 6.7, Fig. S1). Notably, we noticed a large consistency within the analysis results even as the dataset increased by a factor of five compared to the previous study. This resilience underscores the enduring reliability of the methodology, firmly affirming its capability to withstand augmented data volumes without compromising the integrity and consistency of results.

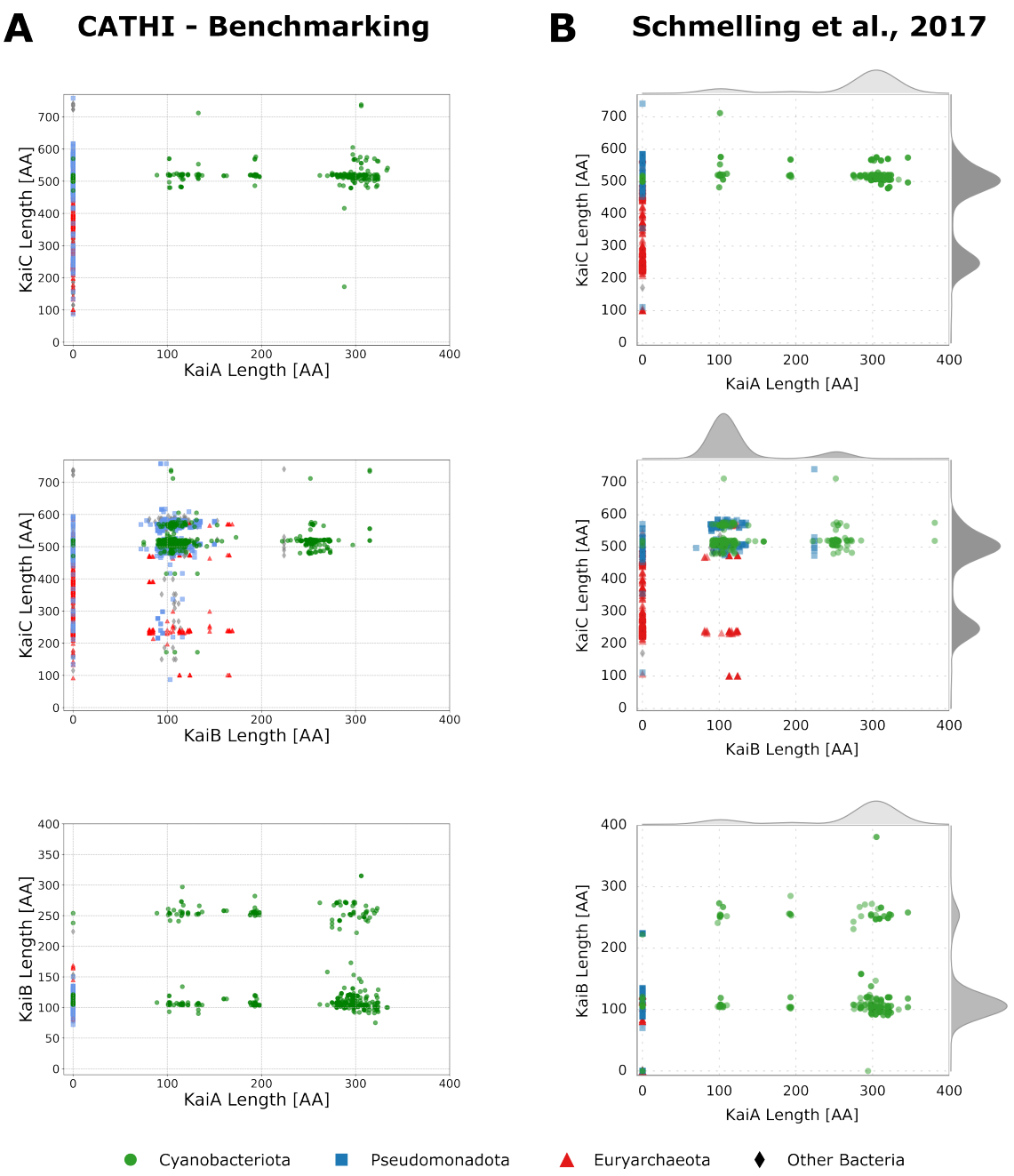

Figure 6.7: **Comparison between the benchmarking analysis and Schmelling et al., 2017 of the sequence length distribution for the core circadian clock proteins KaiA, KaiB, and KaiC.** This visual depiction offers an insightful comparison of sequence lengths for two of the three core circadian clock proteins using scatter plots. Four distinct organismic groups - *Cyanobacteriota*, *Pseudomonadota*, *Euryarchaeota*, and other bacteria — are the focal points of analysis. The overlay of the two datasets serves to highlight overarching patterns, underscoring the methodology's robustness in elucidating sequence length dynamics across different organismic groups. Notably, this comparison not only reaffirms established trends but also emphasizes the enduring reliability of the analytical framework. This analysis showcases the consistency and robustness of the methodology and its capacity to yield insights that withstand scrutiny and validation. A) Results of the benchmarking analysis. B) Previous results from the reciprocal BLAST analysis of (Schmelling et al. 2017).

6.4 Methods

During the development of the application, all changes were consistently tracked with the version control system git. The code, its developmental history, and additional material can be found under the following link: `https://github.com/Kanomble/celery_blast`.

6.4.1 Docker-Based Software Deployment and Container Orchestration

The Docker container virtualization system was utilized to package the underlying software components into distinct images. Based on these images, seven containers were established, each designated for distinct tasks within CATHI. The images utilized for container creation are available on DockerHub under the following link: `https://hub.docker.com/repositories/kanomble`.

The base image of CATHI is constructed upon the Ubuntu Focal distribution and includes essential components such as the Django web framework (Version 4.2.4, employed for constructing the web server), Miniconda (Version 4.9.2, for Python package administration), the BLAST+ (Version 2.11.0) and the E-Direct software suite (Version 20.3.20230829) from NCBI, Snakemake workflow engine (Version 7.25.0), MAFFT (Version 7.453), FastTree2 (Version 2.1.11-1), and other critical software packages for running the web server and executing the pipeline. A comprehensive table of all utilized software tools and Python packages can be found in Supplementary Table S2. Four Docker containers are derived from this foundational base image:

- The **web** container orchestrates the startup of the Django application (including database migrations) and the Python WSGI server, Gunicorn (Version 20.1.0).
- The **celery_worker** container utilizes a local celery installation (Version 5.3.1), which handles long-running background tasks, such as the Snakemake pipeline.
- The **flower** container is a monitoring interface for celery tasks and workers, utilizing the Flower (Version 2.0.1) web server.
- The **jupyter_notebook** container launches the web-based interactive computing platform Jupyter (Version based on IPython 8.12.0).

The foundational database of the web server is encapsulated within a PostgreSQL container. During the initial startup of the web container, database tables, and base models are written into the database, a process facilitated through the use of the Django web framework's makemigration and migrate commands. These programs translate Django's model classes into SQL tables (known as Object Relational Mapping, ORM). The rabbitmq container is dedicated to launching the message broker RabbitMQ (Version 3.9.4), which takes user-generated task instructions and conveys them to the celery framework. CATHI employs Nginx (Version 1.21.6) as a reverse proxy, directing client requests to a Gunicorn backend server and relaying the responses to clients. This pivotal role distributes incoming traffic, enhancing the system's overall performance. The initiation of Nginx is executed through the dedicated Nginx Docker container. To ensure proper container initialization, a specific startup sequence is vital. The web container necessitates the database container's availability; similarly, the celery container relies on the functioning rabbitmq container. The wait-for script (Version 2.1.0), reliant on netcat (Version 1.206-1ubuntu1), is employed to manage this sequence. Written in BASH, this script accepts a port number as input and continuously checks for the presence of an application listening on that port. Once a port is occupied, the script ceases its blocking function, allowing the container to initiate its operations. Once launched, users can access the application via common browser applications using this address: http://127.0.0.1:1337/blast_project/.

6.4.2 Django Web Framework

For an efficient, clean, and pragmatic design, the high-level Python web framework Django (Version 4.2.4) was used for the development of this web application. Django, a popular high-level web framework written in Python, accelerates web application development through its Model-View-Controller (MVC) architecture, integrated ORM, and dynamic template engine. Utilizing Django, the integration of additional Django-based applications is achievable by including the respective software packages within the INSTALLED_APPS environment variable, as specified in the settings.py file of the project. The following third-party applications have been incorporated into CATHI: django_extensions (Version 3.2.3), django_celery_results (Version 2.5.1), and celery_progress (Version 0.3). The django_extensions application facilitates the execution of the startup.py script, which introduces database models into the PostgreSQL (Version 13.4) database and procures the taxonomy database from NCBI during the initial launch of the web container. In parallel, the django_celery_results application extends functionality by enabling the storage of

celery task outcomes through the utilization of the Django ORM framework.

6.4.3 Web Development Tools

CATHI uses Django's built-in template language for HTML documents and employs cutting-edge web technologies to create interactive and visually appealing websites. The use of JavaScript and AJAX guarantees seamless web interactivity, complemented by Bootstrap 5.0 for bolstering aesthetics through predefined Cascading Style Sheet (CSS) components. Additionally, the HTML table plugin for JavaScript, DataTables optimizes the presentation and manipulation of HTML tables, resulting in a comprehensive toolkit specifically designed to create dynamic and user-friendly web experiences. All third-party web libraries are integrated via their respective content delivery networks (CDNs). Custom CSS and JavaScript code are located within the static directory of CATHI.

6.4.4 Project Creation

Project creation is facilitated by Django's built-in form validation. User input is validated before a project is manifested into the PostgreSQL database. There are certain input requirements for the reciprocal BLAST projects. The SOIs have to come from only one organism per project and they have to reside in the backward database. The best option is to upload the genome from which the user has obtained the sequences with CATHI's database creation module. The input is validated by Django's internal form validation functionality and accurate failure messages are displayed if form validation fails. In addition to the SOIs and the species name of the target organism, users have to enter a project title, which has to be unique, no other project should use this title. In addition, users can adjust BLAST settings (for forward and backward BLAST separately), apply a bitscore filter and limit the maximum number of RBHs used to create a phylogeny, as well as change settings of the post-processing programs trimAl (Capella-Gutiérrez et al. 2009) and MView (Brown et al. 1998).

6.4.5 Pipeline Execution with the Snakemake Workflow Engine

The CATHI pipeline process leverages the Snakemake (Köster and Rahmann 2012) workflow engine to streamline the entire process of RBH analysis. This integration not only increases the reproducibility of results but also reduces the burden on the

user by automating many of the necessary steps in the detection of putative homologs. The developed Snakemake pipelines integrate a diverse array of algorithms and software tools. Three Snakefiles have been developed: one for the core reciprocal BLAST pipeline, another for local BLAST operations, and a third for conducting remote BLASTs on NCBI servers. Alongside supplementary scripts, these Snakefiles reside within dedicated static subdirectories.
Snakemake is executed within a celery task encapsulated in a Python subprocess.Popen call. The core Snakemake pipeline for detecting RBHs comprises the following steps:

1. Forward BLAST
2. Backward BLAST preparation
3. Backward BLAST
4. Extraction of RBHs (this is done via pandas merging tools)
5. Post-processing of RBHs (inference of taxonomic information, statistics, HTML and CSV tables, basic result plots)
6. Extraction of RBH-sequences separated by query sequences
7. MSA of each set of RBHs with MAFFT
8. Phylogenetic inference of each set of RBHs with FastTree2
9. Post-processing of the phylogenetic tree and the MSA with shiptv, trimAI and MView
10. CDD domain search of target sequences

6.4.6 Database Creation

CATHI enables the creation of local BLAST databases based on publicly available or uploaded protein sequences. For the publicly available sequences, CATHI utilizes the RefSeq and GenBank assembly summary files, which inherit FTP paths to the corresponding protein assemblies. The assembly summary files can be filtered at two different levels. The first level includes the completeness status of the dedicated assembly, which corresponds to the genome coverage and completeness of the genome assembly. This is summarized by an assembly level of the corresponding protein genome file. The second filtering step involves organism-specific filtering.

For each organism and clade of the taxonomic classification (e.g., for genus, family, or order), a taxonomic node is stored in the taxonomy database (Schoch et al. 2020). These taxonomic nodes can be used to filter the assembly summary file to create BLAST databases containing only organisms with the provided taxonomic nodes. Taxonomic node files can be uploaded directly as part of the database formatting process or created beforehand by using CATHI, which provides a dashboard for translating higher taxonomic nodes to the underlying species-specific nodes. This translation procedure is realized by executing the Perl script get_species_taxids.pl provided by the BLAST+ command line tool.
Sometimes, it is necessary to work with unpublished genomes or sequences. CATHI offers two options for uploading your own protein FASTA files, which are then formatted into BLAST databases. The user can upload a concatenated protein FASTA file consisting of multiple genomes. This approach requires additional metadata; the assembly identifier (combination of characters or numbers), organism names (with valid taxonomic nodes), assembly levels (one out of four; "Contig," "Scaffold," "Chromosome," or "Complete" Genome), and a file that contains a mapping of the provided sequence identifiers to the relevant taxonomic nodes (taxmap file). This approach is error-prone and may lead to problems for inexperienced users. Therefore, a second approach was developed, which allows the uploading of multiple single protein FASTA files in combination with valid species names. Uploaded files are then used to build BLAST databases with the makeblastdb program.
Prior to the formatting procedure using makeblastdb, protein FASTA files undergo parsing, and the header of each sequence is slightly modified to incorporate the genome assembly name, thus assigning a unique identifier to each sequence. This approach mitigates the issue of multiple identical sequence identifiers within the database, which could otherwise trigger an error during the makeblastdb database formatting process. BLAST databases are divided into chunks, with each chunk comprising 500 genome files. These individual database chunks are subsequently compiled into a single .pal alias file, consolidating all database chunks into a unified BLAST database. This method prevents excessive RAM consumption for extensive databases and facilitates subsequent database updates.

6.4.7 BLAST and Identification of Reciprocal Best Hits

The forward and backward BLAST (Fig. 6.1) analysis of the user-provided SOI is performed using the blastp program within the BLAST+ software suite (Version 2.11.0) (Camacho et al. 2009). The default settings for the forward BLAST are as follows: e_value=0.001, word_size=3, num_alignments=10.000, max_hsps=500.

The default settings for the backward BLAST are similar to the forward BLAST settings, except for the num_alignment option, which is set to num_alignments=1. Default settings can be customized by the user. Other tools within the BLAST+ software suite utilized in CATHI include the makeblastdb program, employed for formatting protein BLAST databases; the blastdbcmd program, which facilitates the retrieval of protein sequences identified during the initial BLAST analysis; and the rpsblast program, which is used for the inference of CDDs among the RBHs. RBH inference is executed using the Python library Pandas, which involves a comparison between the outcomes of forward and backward BLAST analyses. Additionally, the BioPython library is harnessed to deduce the underlying taxonomic details of the organisms of the resulting RBH sequences, which are then documented in a CSV file. All tables generated are conveniently accessible within CATHI, facilitated by integrating the client-side JavaScript library DataTables (see Methods - Web Development Tools).

6.4.8 Post-Processing Software

Python's Biopython (Version 1.78) (Cock et al. 2009) library is used to extract taxonomic information from database entries. Moreover, Biopython is used to slice protein GenBank files acquired through synteny analysis conducted by CATHI. These segmented GenBank files are subsequently used as input for the clinker tool (Version 0.0.27). Matplotlib (Version 3.7.2) serves to generate fundamental result graphs, pandas (Version 1.2.4) facilitates the manipulation of tabular data and the inference of RBHs, while scipy (Version 1.10.1) and scikit-learn (Version 1.3.0) are utilized for Principal Component Analysis (PCA) within CATHI's CDD detection module. Bokeh (Version 2.4.3) is used to create interactive visualizations (based on the database and the RBH result tables or the CDD result table). Third-party tools MAFFT (Version 7.453) and FastTree2 (Version 2.1.11-1) are employed for MSAs and the inference of corresponding phylogenetic trees. TrimAl (Version v1.4.rev22) is used to trim the MSA by removing long gaps and uninformative segments from the alignment. MView (Version 1.67) and shiptv (Version 0.4.1) transform the output of these tools to generate interactive HTML documents for MSAs and phylogenies. A comprehensive table of all employed software tools and Python packages can be found in Supplementary Table S2.

6.5 Discussion

Identifying closely and distantly related homologous sequences is an essential task in the field of comparative genomics (Anisimova 2019; Koonin 2005; Kuzniar et al. 2008). Typically, orthologous and paralogous gene relationships are differentiated by evaluating and contrasting sequence similarities and their distribution within phylogenetic contexts. However, diverse biological questions may necessitate distinct computational approaches, leading to the development of various programs (Table 6.1) for homology inference (Kristensen et al. 2011; Nichio et al. 2017). Furthermore, over the past decades, biological databases with collections of orthologous sequences have emerged, expediting the swift identification of orthologous relationships within established sequences (Altenhoff et al. 2021; Li et al. 2006; Persson and Sonnhammer 2023; Richter et al. 2022; Tatusov et al. 2000; Zdobnov et al. 2021). The ever-increasing biological data demands flexible approaches to disentangle homologous relationships among genes, within newly sequenced genomes, which are not part of these databases. CATHI is a platform for the interactive exploration of the output of sequence similarity searches in custom databases that can be easily created via dedicated web-interfaces. The additional post-processing modules, such as phylogeny, synteny, and inference of conserved domains among identified homologous sequences, make CATHI a highly flexible tool that enables rapid and interactive analysis of results without the need for additional bioinformatics tools. In this chapter, we discuss CATHI's usability and highlight advantages and disadvantages in comparison to other, similar computational resources.

6.5.1 Usability

There are versatile tools for inferring homologous sequences (Table 6.1). While tools like PARIGA (Orsini et al. 2013) and orFin (Midha et al. 2012) are no longer accessible, other tools like morFeus, JustOrthologs, and orthoFinder offer different strategies for the identification of orthologs. The later tools are either accessible through the command line or hosted on dedicated web-servers. In contrast, CATHI introduces a unique paradigm. It presents an innovative approach by providing the flexibility of both local and server site installations, coupled with an interactive interface accessible via standard web browsers. Most tools provide comprehensive tables of results listing the putative orthologous sequences identified in different organisms (Table 6.1). However, a thorough analysis of these results often requires additional software tools and further post-processing steps. Tasks such as MSAs and phylogenetic tree inference frequently entail programming expertise, posing challenges for

many biologists attempting to address these complexities.

Drawing from our experience, a significant portion of biologists exhibit concerns regarding coding prerequisites. Consequently, a notable demand exists for software tools that provide graphical user interfaces, facilitating result analysis through user-friendly cursor interactions and drag-and-drop functions. Furthermore, manual intervention and examination are often necessary to analyze homologous sequences, such as distinguishing distantly from closely related homologs. Hence, an intuitive interface that streamlines result analysis and integrates crucial post-processing steps is essential for homologous sequence analysis. CATHI offers such interfaces by using e.g. the capabilities of the interactive plotting library Bokeh. Users can employ these interfaces to filter results based on taxonomy and BLAST statistics, as well as interactively select RBHs for various subsequent analyses (Fig. 6.5). Examining the RBHs of the *Synechococcus* Pex protein within the *Bacillota* and *Bacteriodota* phylum, a total of 18 RBHs from 16 organisms of the Pex protein have been found, characterized by relatively modest bitscores (mean of 52). While these identified RBHs might represent homologous sequences, their lower bitscores and reduced sequence identity could doubt their status as orthologous sequences. This underscores the need for user intervention and careful consideration through evaluating outcomes. CATHI enables such manual post-processing with a comprehensive graphical display that allows effortless filtering and editing of results (Fig. 6.5). In addition, CATHI uses standard bioinformatics algorithms whose software versions are embedded in the Docker image. This arrangement increases the stability of the pipeline in executing consistently, even if software versions change, and if a database update is required, users can use CATHI's database creation module and rerun the pipeline, ensuring robust and reproducible gene relationship detection. Nonetheless, bioinformatics software tools designed for orthologous sequence detection can assist in estimating the potential presence of homologous sequences. However, to validate a closely related homolog as a *bona fide* ortholog, it is essential to conduct wet lab experiments focusing on protein functionality.

6.5.2 Strength and Drawbacks in Homolog Identification in CATHI

The method of deducing RBHs as potential homologs may not invariably yield accurate results (Nevers et al. 2022), particularly in cases involving ancient gene duplications followed by selective paralog loss (Sjölander et al. 2011), which can partly be resolved by synteny analyses. Consequently, the extant paralog might erroneously

be categorized as an ortholog. In a broader context, if the organisms of interest harbor only a single homolog of the query sequence, and a prospective organism possesses a solitary copy of a homologous gene, that single copy could manifest as an RBH in the results, even if it represents a closely related homolog rather than a genuine ortholog. There are other tools that implement more sophisticated approaches to detect the orthologous and paralogous relationship among genes (Table 6.1) (Nevers et al. 2022). In the upcoming section, we will contrast OrthoFinder (Emms and Kelly 2019), one of the most advanced tools for ortholog detection, with CATHI. This will allow us to highlight the strengths and limitations of these two tools, emphasizing the key distinctions that prompted the development of CATHI.

OrthoFinder assigns genes to orthogroups using an orthogroup inference algorithm, infers rooted trees using the STAG and STRIDE programs (Emms and Kelly 2018; Emms and Kelly 2017), and then applies a hybrid algorithm based on the species-overlap method (Huerta-Cepas et al. 2007) and the duplication-loss-coalescent model (Wu et al. 2014) to identify orthologs and paralogs. In contrast to the RBH approach, this method leads to sophisticated ortholog and paralog assignments. However, the output of OrthoFinder are text files that need to be analyzed in additional post-processing steps outside of OrthoFinder. Furthermore, OrthoFinder's search space is constrained by the number of input FASTA files, and it focuses on whole genomes rather than individual sequences. It does not infer orthologs in comprehensive databases, which limits the potential to unveil certain features of the SOIs when searching for homologs in extensive databases.

If the number of gene duplications is relatively high, such as in angiosperms (Magadum et al. 2013), the RBH approach may not detect all orthologs, only those with the best alignment scores are reported (Hulsen et al. 2006). In this case, OrthoFinder and other tools that make use of calculating orthogroups yield more accurate results compared to the RBH approach implemented in CATHI (Koski and Golding 2001; Nevers et al. 2022). However, at least for prokaryotic genomes the RBH method can still serve as a strong indication for gene orthology (Wolf and Koonin 2012). In addition, CATHI's custom databases can expand the search scope for identifying novel orthologous or paralogous genes, particularly by allowing rapid, visual filtering of results obtained from the analysis, whereas OrthoFinder is limited to the number of FASTA files entered. CATHI also provides options for simple one-way BLAST analyses that can help identify extended duplication events when there are multiple homologous sequences in an organism with relatively high statistical significance

(e.g., bitscores > 50 or e-values < 0.001) .

However, the exact classification of genes as orthologs or paralogs is difficult to implement even for tools like OrthoFinder. To safely predict orthology, we would need a complete history of all ancestral genomes. To date, our genomic information is still insufficient and might be forever as genetic information throughout the history of life has been irrevocably lost. Thus, predictions will only yield reliable results to a certain extent. Even though it is an interesting and important question, for most laboratory biologists those distinctions are not their main focus. They rather use homology searches to better understand their protein/gene of interest before examining it in more detail in the laboratory. Here, indications about potential functions, involvement in metabolic or regulatory pathways, or location on the genome within clusters or in proximity to other well characterized genes is far more useful. In general, with CATHI it is possible for users inexperienced with command line tools or programming languages to gain insights about a gene of unknown function by identifying homologs and analyzing the synteny, phylogeny, and domain structure in an automated workflow, while exploring and filtering the data in an interactive interface. That information will then guide laboratory experiments to correctly describe the function of the gene.

CATHI utilizes some generalizations that introduce certain limitations, such as using the blastp program for conducting the sequence similarity search algorithm, thus, restricting the acceptable query sequences to proteins. Furthermore, it uses a reciprocal BLAST to search for close homologous sequences. While other tools may be more accurate at identifying orthologs (Hulsen et al. 2006), CATHI provides intuitive results analysis suitable for both experienced and less experienced users. The automated pipeline creates a generalized workflow for homolog identification. Nevertheless, CATHI may not be suitable for all SOIs due to the underlying reciprocal BLAST technique. However, it provides a great starting point for most projects, thereby extending the accessibility of homologous sequence analysis to a broader spectrum of users. Through interactive filtering, more experienced users can reduce the initial BLAST cut-offs to allow for a broader homology search and later filter those results using the extensive options in the post-processing module.

6.5.3 Database Content and Customization

Certain tools possess a restricted search space attributed to static databases (e.g., InParanoiDB9 or COG) and/or the necessity for pre-formatted databases (e.g., Or-

thoInspector, morFeus or JustOrthologs), which subsequently remain "static" in a manner that necessitates users to download and format each database in advance.

Due to the rapid growth of biological data, a flexible tool is needed that can automate rapid configuration of biological databases. CATHI provides the capability to generate customized BLAST databases utilizing publicly accessible and locally available protein sequences. Upon completing the database creation process, a dedicated webpage exhibits a table showcasing the database's contents. This comprehensive table enumerates all organisms and genome entries encompassed within the database, facilitating researchers in obtaining an overview of its contents. Furthermore, databases can be customized based on criteria such as completeness level, taxonomy, and the option to download either from RefSeq or GenBank, consequently facilitating more nuanced scrutiny of sequence similarity searches, which is vital in detecting specific homologous sequences. Once the database is downloaded, its entries remain unaltered, simplifying the monitoring of modifications in publicly accessible databases. This feature serves to alleviate potential confusion stemming from omitted or deleted sequence entries and to overcome limitations within the provided search scope.

6.5.4 Modular Expandability

The use of Docker, Django, and Snakemake provides flexibility in extending the pipeline to include additional functionality or analytics. The modular nature of these technologies allows for seamless integration of new functionality and increases the adaptability and future-proofing of the pipeline. Typically, bioinformatic tools are bundled into software packages that are executed on the UNIX command line (e.g. OrthoFinder). These packages often work independently and perform the desired task with certain inputs provided by the user. However, customizing these programs to meet specific requirements demands users to possess experience with the underlying code of these tools. On the other hand, extending CATHI's Snakemake pipeline is relatively easy for sophisticated users, giving rise to the opportunity to create novel bioinformatic workflows targeting specific research questions. Snakemake defines a workflow using a directed-acyclic-graph (DAG) builded through a set of rules whose execution is based on certain input files and some additional parameters. These rules generate output files, which in turn can serve as input files for other rules. Users can extend the provided Snakemake workflow by adding additional rules that can make use of the infrastructure generated by CATHI.

6.6 Outlook

In its current version, CATHI provides a simple entry point into comparative genomic analysis for biologists. However, thanks to its modular structure, the platform can easily be extended. The platform's homolog search module could evolve to allow users to opt for different search tools like BLAST or DIAMOND (Buchfink et al. 2014), improving the efficiency of homology searches. The integration of advanced techniques such as Markov clustering, hidden Markov models (HMMs) of sequence alignments that can be used to detect distant homologs (Chen et al. 2018), and bootstrapping into the homolog search process has the potential to enhance the precision of homolog identification, advancing the accuracy of comparative genomics analyses. Furthermore, the incorporation of iterative homology searches introduces a refinement process. Users can initiate searches, refine them through filtering, and adapt parameters based on evolving insights, ensuring a dynamic and adaptive analytical workflow. Future iterations of CATHI can encompass sophisticated co-occurrence analyses of RBHs, unveiling intricate patterns of protein relationships across diverse organisms. In addition, the inclusion of additional tools such as OrthoFinder and TreeViewer holds the potential to enrich the analytical spectrum, broadening the toolbox available to researchers. Last, the untapped potential of emerging AI tools and capabilities holds the promise to further augment this platform's functionalities, potentially unearthing novel and unforeseen avenues for analytical insights. Incorporating these enhancements will bolster the platform's versatility, ensuring its alignment with the evolving landscape of bioinformatics tools and methodologies.

6.7 Supplementary Data

Table 6.S1: **Protein input queries for the reciprocal BLAST benchmarking analysis.** Cyanobacterial circadian clock proteins were selected from *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, based on the previous analysis by Schmelling et al. (2017).

Organism Name	Protein Name	RefSeq ID
Synechococcus	KaiA	WP_011377921.1
	KaiB	WP_011242647.1
	KaiC	WP_011242648.1
	Pex	WP_011377679.1
	LdpA	WP_011377652.1
	NhtA	WP_011378346.1
	PrkE	WP_011243235.1
	CdpA	WP_011378107.1
	CikA	WP_011243194.1
	SasA	WP_011378322.1
	LabA	WP_011244514.1
	LalA	WP_011242719.1
	Crm	WP_011243720.1
	RpaA	WP_011377437.1
	RpaB	WP_011378039.1
	CpmA	WP_011377895.1
Synechocystis	KaiB1	WP_010874242.1
	KaiC1	WP_010874243.1
	KaiB2	WP_010872548.1
	KaiC2	WP_010872549.1
	KaiB3	WP_010874242.1
	KaiC3	WP_010873229.1
	LarB	WP_041426075.1
	RpaA	WP_010873880.1
	PIN/TRAM domain-containing protein	WP_010872826.1
	LdpA	WP_010874023.1
	LabA	WP_010874237.1
	DUF1269 domain-containing protein	WP_010874319.1
	DUF3370 domain-containing protein	WP_010872486.1
	histidine-kinase	WP_010873372.1
	ATP-binding protein	WP_010872820.1
	response-regulator	WP_010873029.1
	KaiB	WP_010873802.1

Table 6.S2: **Selection of CATHIs software tools and version numbers.**

Tool	**Version**
BLAST+	2.11.0
EDirect	20.3.20230829
MAFFT	7.453
FastTree2	2.1.11-1
PostgreSQL	13.4
RabbitMQ	3.9.4
nginx	1.21.6
MView	1.67
trimAl	v1.4.rev22
netcat	1.206-1ubuntu1
wait-for	2.1.0
clinker	0.0.27
shiptv	0.4.1
Snakemake	7.25.0
bokeh	2.4.3
scikit-learn	1.3.0
scipy	1.10.1
pandas	1.2.4
matplotlib	3.7.2
biopython	1.78
IPython	8.12.0
Miniconda	4.9.2
conda	23.7.3
Django	4.2.4
django-celery-results	2.5.1
django-extensions	3.2.3
flower	2.0.1
gunicorn	20.1.0
celery	5.3.1
celery-progress	0.3

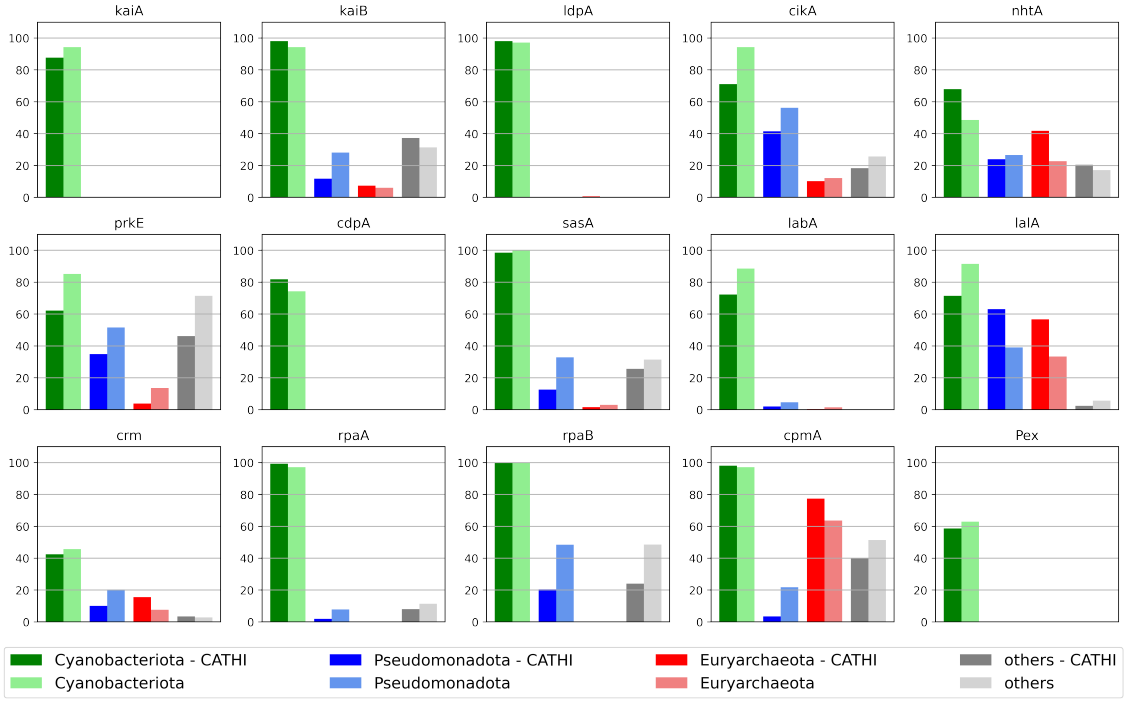

Figure 6.S1: **Comparison between CATHI and Schmelling et al., 2017 of the abundance within the RBHs of the circadian clock core-, input- and output-factors of organisms that harbor a KaiC homolog.** For each circadian clock protein, the abundance value is calculated as the ratio of the number of RBHs for the circadian clock protein divided by the number of organisms with RBHs for KaiC. While there are some differences between the studies, the overall trend remains similar. Darker colors refer to the results of this benchmark analysis, while lighter colors refer to the results of the initial study by Schmelling et al., 2017. Green: *Cyanobacteriota*; Blue: *Pseudomonadota*; Red: *Euryarchaeota*; Other Bacteria.

References for Manuscript 3

Altenhoff, Adrian M. et al. (Jan. 2021). "OMA orthology in 2021: Website overhaul, conserved isoforms, ancestral gene order and more". In: *Nucleic Acids Research* 49 (D1), pp. D373–D379. ISSN: 13624962. DOI: 10.1093/nar/gkaa1007.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). "Basic local alignment search tool". In: *Journal of Molecular Biology* 215 (3), pp. 403–410. ISSN: 00222836. DOI: 10.1016/S0022-2836(05)80360-2.

Altschul, Stephen F. et al. (Sept. 1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". In: *Nucleic Acids Research* 25.17, pp. 3389–3402. ISSN: 0305-1048. DOI: 10.1093/nar/25.17.3389.

Anisimova, Maria (2019). *Evolutionary Genomics Statistical and Computational Methods Second Edition Methods*. ISBN: 9781493990733.

Axmann, Ilka M., Stefanie Hertel, Anika Wiegard, Anja K. Dörrich, and Annegret Wilde (2014). "Diversity of KaiC-based timing systems in marine Cyanobacteria". In: *Marine Genomics* 14, pp. 3–16. ISSN: 18747787. DOI: 10.1016/j.margen.2013.12.006.

Bianchini, Giorgio and Patricia Sánchez-Baracaldo (Mar. 2023). "TreeViewer Version 2.1.0". Version v2.1.0. In: DOI: 10.5281/zenodo.7768344.

Bokeh Development Team (2018). *Bokeh: Python library for interactive visualization*.

Brown, N P, C Leroy, and C Sander (Jan. 1998). "MView: a web-compatible database search or multiple alignment viewer." In: *Bioinformatics* 14.4, pp. 380–381. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/14.4.380.

Buchfink, Benjamin, Chao Xie, and Daniel H. Huson (Jan. 2014). "Fast and sensitive protein alignment using DIAMOND". In: *Nature Methods* 12 (1), pp. 59–60. ISSN: 15487105. DOI: 10.1038/nmeth.3176.

Camacho, Christiam et al. (2009). "BLAST+: architecture and applications". In: *BMC Bioinformatics* 10 (1), p. 421. ISSN: 1471-2105. DOI: 10.1186/1471-2105-10-421.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón (Aug. 2009). "trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses". In: *Bioinformatics* 25 (15), pp. 1972–1973. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp348.

Chen, Junjie, Mingyue Guo, Xiaolong Wang, and Bin Liu (Mar. 2018). "A comprehensive review and comparison of different computational methods for protein remote homology detection". In: *Briefings in Bioinformatics* 19 (2), pp. 231–244. ISSN: 14774054. DOI: 10.1093/bib/bbw108.

Cock, Peter J.A. et al. (June 2009). "Biopython: Freely available Python tools for computational molecular biology and bioinformatics". In: *Bioinformatics* 25 (11), pp. 1422–1423. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp163.

Curtis, Darren S., Aaron R. Phillips, Stephen J. Callister, Sean Conlan, and Lee Ann McCue (Oct. 2013). "SPOCS: Software for predicting and visualizing orthology/paralogy relationships among genomes". In: *Bioinformatics* 29 (20), pp. 2641–2642. ISSN: 13674803. DOI: 10.1093/bioinformatics/btt454.

Descorps-Declère, Stéphane, Frédéric Lemoine, Quentin Sculo, Olivier Lespinet, and Bernard Labedan (2008). "The multiple facets of homology and their use in comparative genomics to study the evolution of genes, genomes, and species". In: *Biochimie* 90 (4), pp. 595–608. ISSN: 61831638. DOI: 10.1016/j.biochi.2007.09.010.

Dunne, Michael P. and Steven Kelly (May 2017). "OrthoFiller: Utilising data from multiple species to improve the completeness of genome annotations". In: *BMC Genomics* 18 (1). ISSN: 14712164. DOI: 10.1186/s12864-017-3771-x.

Emms, D.M. and S. Kelly (Feb. 2018). *STAG: Species Tree Inference from All Genes*. DOI: 10.1101/267914.

Emms, David M. and Steven Kelly (Dec. 2017). "STRIDE: Species tree root inference from gene duplication events". In: *Molecular Biology and Evolution* 34 (12), pp. 3267–3278. ISSN: 15371719. DOI: 10.1093/molbev/msx259.

— (Dec. 2019). "OrthoFinder: phylogenetic orthology inference for comparative genomics". In: *Genome Biology* 20 (1), p. 238. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1832-y.

Enright, A J, S Van Dongen, and C A Ouzounis (2002). "An efficient algorithm for large-scale detection of protein families". In: *Nucleic Acids Research* 30 (7), pp. 1575–1584.

Fitch, Walter M. (June 1970). "Distinguishing Homologous from Analogous Proteins". In: *Systematic Zoology* 19 (2), p. 99. ISSN: 00397989. DOI: 10.2307/2412448.

Gilchrist, Cameron L M and Yit-Heng Chooi (Aug. 2021). "clinker & clustermap.js: automatic generation of gene cluster comparison figures". In: *Bioinformatics* 37 (16). Ed. by Peter Robinson, pp. 2473–2475. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btab007.

Huerta-Cepas, Jaime, Hernán Dopazo, Joaquín Dopazo, and Toni Gabaldón (June 2007). "The human phylome". In: *Genome Biology* 8 (6). ISSN: 14747596. DOI: 10.1186/gb-2007-8-6-r109.

Hulsen, Tim, Martijn A. Huynen, Jacob de Vlieg, and Peter M.A. Groenen (Apr. 2006). "Benchmarking ortholog identification methods using functional genomics data". In: *Genome Biology* 7 (4). ISSN: 1474760X. DOI: 10.1186/gb-2006-7-4-r31.

Inman, Jason M. et al. (Mar. 2019). "Large-scale comparative analysis of microbial pan-genomes using PanOCT". In: *Bioinformatics* 35 (6), pp. 1049–1050. ISSN: 14602059. DOI: 10.1093/bioinformatics/bty744.

Jensen, Roy A (2001). "Orthologs and paralogs-we need to get it right". In: ISSN: 1465-6914.

Katoh, K. and D. M. Standley (Apr. 2013). "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability". In: *Molecular Biology and Evolution* 30 (4), pp. 772–780. ISSN: 0737-4038. DOI: 10.1093/molbev/mst010.

Kelly, Steven and Philip K. Maini (Mar. 2013). "DendroBLAST: Approximate Phylogenetic Trees in the Absence of Multiple Sequence Alignments". In: *PLoS ONE* 8 (3). ISSN: 19326203. DOI: 10.1371/journal.pone.0058537.

Köbler, Christin et al. (2023). "Two circadian oscillators in one cyanobacterium". In: *bioRxiv*. DOI: 10.1101/2021.07.20.453058.

Konc, Janez and Dušanka Janežič (2007). "An improved branch and bound algorithm for the maximum clique problem". In: *MATCH Communications in Mathematical and in Computer Chemistry MATCH Commun. Math. Comput. Chem* 58, pp. 569–590.

Koonin, Eugene V. (Dec. 2005). "Orthologs, Paralogs, and Evolutionary Genomics". In: *Annual Review of Genetics* 39 (1), pp. 309–338. ISSN: 0066-4197. DOI: 10.1146/annurev.genet.39.073003.114725.

Koski, Liisa B and G Brian Golding (2001). "The Closest BLAST Hit Is Often Not the Nearest Neighbor". In: *Journal of Molecular Evolution* 52 (6), pp. 540–542. ISSN: 1432-1432. DOI: 10.1007/s002390010184.

Köster, Johannes and Sven Rahmann (2012). "Snakemake-a scalable bioinformatics workflow engine". In: *Bioinformatics* 28 (19), pp. 2520–2522. ISSN: 14602059. DOI: 10.1093/bioinformatics/bts480.

Kristensen, D. M., Y. I. Wolf, A. R. Mushegian, and E. V. Koonin (Sept. 2011). "Computational methods for Gene Orthology inference". In: *Briefings in Bioinformatics* 12 (5), pp. 379–391. ISSN: 1467-5463. DOI: 10.1093/bib/bbr030.

Kuzniar, Arnold, Roeland C.H.J. van Ham, Sándor Pongor, and Jack A.M. Leunissen (Nov. 2008). "The quest for orthologs: finding the corresponding gene across genomes". In: *Trends in Genetics* 24 (11), pp. 539–551. ISSN: 01689525. DOI: 10.1016/j.tig.2008.08.009.

Li, Heng et al. (2006). "TreeFam: a curated database of phylogenetic trees of animal gene families." In: *Nucleic acids research* 34 (Database issue), pp. 10–16. ISSN: 13624962. DOI: 10.1093/nar/gkj118.

Li, L., Christian J. Stoeckert, and David S. Roos (Sept. 2003). "OrthoMCL: Identification of ortholog groups for eukaryotic genomes". In: *Genome Research* 13 (9), pp. 2178–2189. ISSN: 10889051. DOI: 10.1101/gr.1224503.

Linard, Benjamin, Julie D. Thompson, Olivier Poch, and Odile Lecompte (Jan. 2011). "OrthoInspector: Comprehensive orthology analysis and visual exploration". In: *BMC Bioinformatics* 12. ISSN: 14712105. DOI: 10.1186/1471-2105-12-11.

Magadum, S et al. (2013). "Gene duplication as a major force in evolution". In: *J. Genet* 92, pp. 155–161.

Marchler-Bauer, Aron et al. (Nov. 2014). "CDD: NCBI's conserved domain database". In: *Nucleic Acids Research* 43.D1, pp. D222–D226. ISSN: 0305-1048. DOI: 10.1093/nar/gku1221.

Merkel, Dirk (2014). "Docker: lightweight linux containers for consistent development and deployment". In: *Linux journal* 2014.239, p. 2.

Midha, Mohit et al. (2012). "OrFin: A web tool for detection of putative orthologs". In: *print) Bioinformation* 8 (15), p. 738. ISSN: 0973-8894.

Miller, Justin B., Brandon D. Pickett, and Perry G. Ridge (2019). "JustOrthologs: A fast, accurate and user-friendly ortholog identification algorithm". In: *Bioinformatics* 35 (4), pp. 546–552. ISSN: 14602059. DOI: 10.1093/bioinformatics/bty669.

Nevers, Yannis et al. (July 2022). "The Quest for Orthologs orthology benchmark service in 2022". In: *Nucleic Acids Research* 50 (W1), W623–W632. ISSN: 13624962. DOI: 10.1093/nar/gkac330.

Nichio, Bruno T. L., Jeroniza Nunes Marchaukoski, and Roberto Tadeu Raittz (Oct. 2017). "New Tools in Orthology Analysis: A Brief Review of Promising Perspectives". In: *Frontiers in Genetics* 8. ISSN: 1664-8021. DOI: 10.3389/fgene.2017.00165.

O'Brien, Kevin P., Maido Remm, and Erik L.L. Sonnhammer (Jan. 2005). "Inparanoid: A comprehensive database of eukaryotic orthologs". In: *Nucleic Acids Research* 33 (DATABASE ISS.). ISSN: 03051048. DOI: 10.1093/nar/gki107.

Orsini, M, S Carcangiu, G Cuccuru, P Uva, and A Tramontano (2013). "The PARIGA Server for Real Time Filtering and Analysis of Reciprocal BLAST Results". In: *PLoS ONE* 8 (5), p. 62224. DOI: 10.1371/journal.pone.0062224.

Persson, Emma and Erik L.L. Sonnhammer (July 2023). "InParanoiDB 9: Ortholog Groups for Protein Domains and Full-Length Proteins". In: *Journal of Molecular Biology* 435 (14). ISSN: 10898638. DOI: 10.1016/j.jmb.2023.168001.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin (2010). "FastTree 2 - Approximately maximum-likelihood trees for large alignments". In: *PLoS ONE* 5 (3). ISSN: 19326203. DOI: 10.1371/journal.pone.0009490.

Remm, Maido, Christian E.V. Storm, and Erik L.L. Sonnhammer (2001). "Automatic clustering of orthologs and in-paralogs from pairwise species comparisons". In: *Journal of Molecular Biology* 314 (5), pp. 1041–1052. ISSN: 00222836. DOI: 10.1006/jmbi.2000.5197.

Richter, Daniel J. et al. (2022). "EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes". In: *Peer Community Journal* 2. DOI: 10.24072/pcjournal.173.

Schmelling, Nicolas M. et al. (2017). "Minimal tool set for a prokaryotic circadian clock". In: *BMC Evolutionary Biology* 17 (1). ISSN: 14712148. DOI: 10.1186/s12862-017-0999-7.

Schoch, Conrad L. et al. (2020). "NCBI Taxonomy: A comprehensive update on curation, resources and tools". In: *Database* 2020. ISSN: 17580463. DOI: 10.1093/database/baaa062.

Sjölander, Kimmen, Ruchira S. Datta, Yaoqing Shen, and Grant M. Shoffner (Sept. 2011). "Ortholog identification in the presence of domain architecture rearrangement". In: *Briefings in Bioinformatics* 12 (5), pp. 413–422. ISSN: 14675463. DOI: 10.1093/bib/bbr036.

Smith, T.F. and M.S. Waterman (1981). "Identification of common molecular subsequences". In: *Journal of Molecular Biology* 147.1, pp. 195–197. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(81)90087-5.

Steinegger, Martin and Johannes Söding (Nov. 2017). "MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets". In: *Nature Biotechnology* 35 (11), pp. 1026–1028. ISSN: 1087-0156. DOI: 10.1038/nbt.3988.

Swan, Jeffrey A., Susan S. Golden, Andy LiWang, and Carrie L. Partch (Apr. 2018). "Structure, function, and mechanism of the core circadian clock in cyanobacteria". In: *Journal of Biological Chemistry* 293 (14), pp. 5026–5034. ISSN: 1083351X. DOI: 10.1074/jbc.TM117.001433.

Tatusov, Roman L., Michael Y. Galperin, Darren A. Natale, and Eugene V. Koonin (Jan. 2000). "The COG database: a tool for genome-scale analysis of protein functions and evolution". In: *Nucleic Acids Research* 28.1, pp. 33–36. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.33.

Tekaia, Fredj (Jan. 2016). "Inferring Orthologs: Open Questions and Perspectives". In: *Genomics Insights* 9, pp. 17–28. ISSN: 1178-6310. DOI: 10.4137/GEI.S37925.

Wagner, Ines et al. (Dec. 2014). "morFeus: a web-based program to detect remotely conserved orthologs using symmetrical best hits and orthology network scoring". In: *BMC Bioinformatics* 15 (1), p. 263. ISSN: 1471-2105. DOI: 10.1186/1471-2105-15-263.

Waterhouse, Andrew M., James B. Procter, David M.A. Martin, Michèle Clamp, and Geoffrey J. Barton (2009). "Jalview Version 2-A multiple sequence alignment editor and analysis workbench". In: *Bioinformatics* 25 (9), pp. 1189–1191. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp033.

Wiegard, Anika et al. (May 2013). "Biochemical analysis of three putative KaiC clock proteins from Synechocystis sp. PCC 6803 suggests their functional divergence". In: *Microbiology (United Kingdom)* 159 (PART 5), pp. 948–958. ISSN: 13500872. DOI: 10.1099/mic.0.065425-0.

Wolf, Yuri I. and Eugene V. Koonin (Dec. 2012). "A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes". In: *Genome Biology and Evolution* 4 (12), pp. 1286–1294. ISSN: 17596653. DOI: 10.1093/gbe/evs100.

Wratten, Laura, Andreas Wilm, and Jonathan Göke (Oct. 2021). "Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers". In: *Nature Methods* 18 (10), pp. 1161–1168. ISSN: 15487105. DOI: 10.1038/s41592-021-01254-9.

Wu, Yi Chieh, Matthew D. Rasmussen, Mukul S. Bansal, and Manolis Kellis (2014). "Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees". In: *Genome Research* 24 (3), pp. 475–486. ISSN: 15495469. DOI: 10.1101/gr.161968.113.

Zdobnov, Evgeny M. et al. (Jan. 2021). "OrthoDB in 2020: Evolutionary and functional annotations of orthologs". In: *Nucleic Acids Research* 49 (D1), pp. D389–D393. ISSN: 13624962. DOI: 10.1093/nar/gkaa1009.

7 Manuscript 4

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], 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
[#]contributed equally

[*]Corresponding author: Ilka M. Axmann (ilka.axmann@hhu.de)

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

7.1 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 colonizer of *Hydra vulgaris*. Yet, genetic tools for *Curvibacter* are still in an infancy; few promoters have been characterized so far. Here we employ an oligonucleotide-based strategy to develop novel expression systems *Curvibacter*. Potential promoters were systematically mined from the genome *in silico*. The sequences were cloned as a mixed library into a mCherry reporter vector and positive candidates were selected by Flow Cytometry to be further analyzed through plate reader measurements. From 500 candidate sequences, 25 were identified as active promoters of varying expression strength levels. Plate reader 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 three sequences showed higher expression levels in the exponential phase, we found 12 sequences saturating expression during stationary phase and 10 that showed little discrimination between growth phases. From our library, promoters of the genes *dnaK, rpsL* and an AHL synthase stood out as the most interesting candidates fit for a variety of applications. We identified enriched transcription factor binding motifs among the sorted 33 sequences and genes encoding for homologs of these transcription factors in close proximity to the identified motifs. In this work we show the value of employing comprehensive high-throughput strategies to establish expression systems for novel model organisms.

7.2 Introduction

Curvibacter sp. AEP1-3 (hereafter *Curvibacter*) is a rod-shaped β-proteobacterial species best known for its symbiotic interaction with *Hydra vulgaris* (hereafter *Hydra*) (Fraune et al. 2015), a freshwater polyp of the basal metazoan phylum 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 (Minten-Lange and Fraune 2020; Pietschke et al. 2017; Taubenheim et al. 2020). While the host provides an ecological niche to its colonizers, the microbiota affects mobility, asexual reproduction as well as protection against a fungus of the genus *Fusarium* (Fraune et al. 2015). 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* established through the exchange of N-acetyl homoserine lactone (Pietschke et al. 2017).
The limited genetic accessibility of *Curvibacter* restricts advancements in the genetic manipulation of its cells, thereby impeding progress in the field of interkingdom symbiosis. Genomic modifications over homologous recombination are cumbersome and have a low success rate (Holden et al. 2020). *Curvibacter* cells are amenable to transformation using RSF1010 vector (Scherzinger et al. 1984) 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 a tool kit of novel 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* (Stukenberg et al. 2021; Vasudevan et al. 2019) 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 (Akasaka et al. 2015; Berla and Pakrasi 2012; Turgeon et al. 2008). 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 choice.

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 (Cleary et al. 2004). 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 li-

braries 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 red fluorescence protein mCherry or the Green Fluorescence Protein (GFP). 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 (Oliphant et al. 1986). 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.* (Lale et al. 2022) follows exactly this approach by using long stretches of randomized nucleotides to find novel promoter sequences *de novo*.

An alternative strategy we employ here is the generation of highly curated sequence 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) (Su et al. 2021) or as in our case simply by using existing sequences harvested directly from the target species genome (Wang et al. 2020). 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.

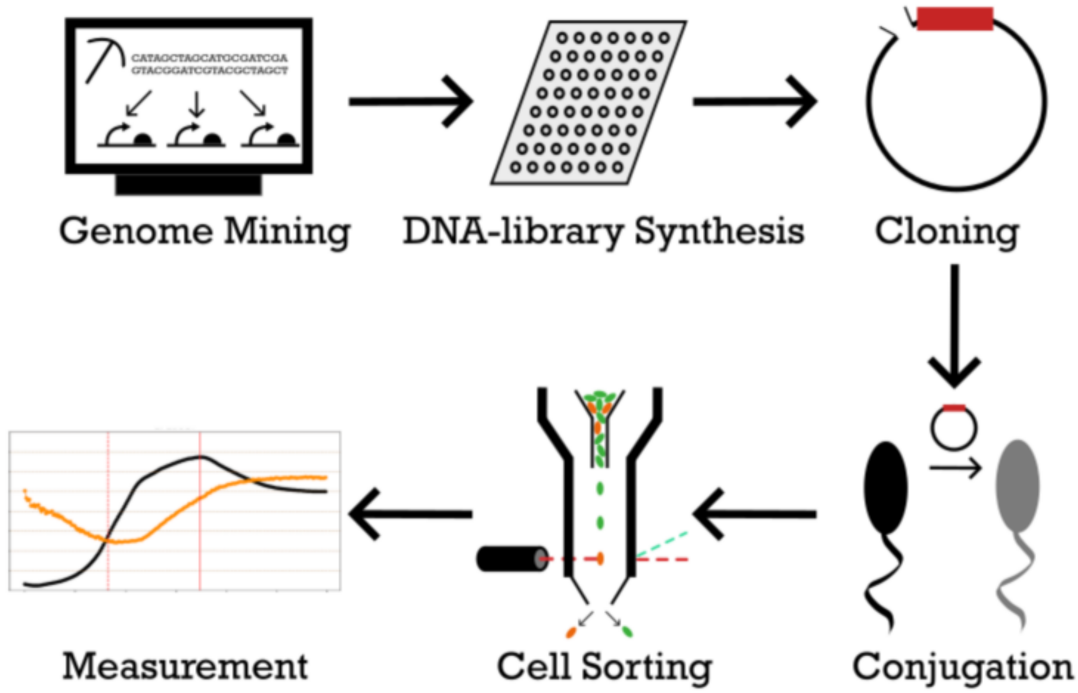

Figure 7.1: **Schematic overview of the applied workflow.** Design of the oligonucleotide library starts by mining the host genome for suitable promoter sequences. Sequences are synthesized, cloned and imported into the host species and sorted for activity by Flow Cytometry Cell Sorting. In depth assays 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 independent of growth phases in liquid media. To increase the odds of individual promoter sequences, promoter and ribosomal sequences were not further discriminated, but entire 5′ untranslated regions (5′ UTRs) were used instead. Candidate sequences for this study were directly harvested (Figure 7.1) 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 in this work as a 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′ UTRs 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 5′ UTRs showed typical increased activity in late exponential- to stationary phase, we also found some 5′ UTRs 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, high-

lighting how useful oligonucleotide library based approaches are to find promoter sequences with desired features.

7.3 Material & Methods

7.3.1 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. The scripts and all relevant data files, including supplementary material, 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 (bp) and less than 60 bp, ensuring that only appropriately sized 5′ UTRs were included in the analysis. Further filtering steps involve removing sequences with opposing orientations and overlapping segments. To remove potential tRNA and rRNA 5′ UTRs or 5′ UTRs 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)). 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 binding sites for the restriction enzymes BsaI, BsmBI and BbsI were identified among the sequences. 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 with 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, 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 bp. 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 bp.

7.3.2 Transcriptome Read Mapping

Raw RNAseq reads were obtained as described in (Pietschke et al. 2017). Briefly, RNA was isolated with the RNeasy Mini Kit (Qiagen) from *Curvibacter* grown in R2A media at 18 °C to mid-exponential growth phase. cDNA libraries were constructed using the TrueSeq Stranded mRNA LT-RiboZero Kit (Illumina) and cDNA libraries were sequenced using a NextSeq 500 machine (Illumina) in paired-ends mode. 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, (Bolger et al. 2014)). Trimmed FASTQ files were analyzed for quality using FASTQC (Version 0.11.9, (Andrews 2010)). Subsequently, the preprocessed sequences were mapped against the reference genome of *Curvibacter* utilizing the kallisto mapper (Version 0.50.0, (Bray et al. 2016)). RAW sequences are uploaded to NCBI as BioProject PRJNA1082616. The relevant trimming and mapping procedure can be found in this GitHub repository: (`https://github.com/Kanomble/curvibacter_transcriptomics`). In this work we used the results of the RNA-Seq analysis only to filter the extracted 5′ UTRs (see above section).

7.3.3 Library Golden Gate Cloning

40 ng/kb of DNA library (but a minimum of five ng per reaction) were added to 20 ng/kb of entry vector (Genbank file available on GitHub `https://github.com/Kanomble/curvibacter_promotor_studies`) to maximize yield of successful integration without compromising efficiency. Golden Gate Reaction was performed according to the LVL2 Golden Gate assembly protocol as described by Marillonet *et al.* (Weber et al. 2011), but the final digest duration was increased to one hour 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 four selection plates to reduce colony crowding. After 24 hours 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.

7.3.4 Conjugation of library vectors into *Curvibacter* sp. AEP1-3 glmS::GFP

Curvibacter with a GFP insertion in the *glmS* locus was obtained from Nawroth *et. al.* (Nawroth et al. 2023). The GFP carrying *Curvibacter* was inoculated from a fresh plate into R2A+ media and grown for 36 hours 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. Three ml of *E. coli* donor cells at OD1 and five ml of *Curvibacter* stationary phase at OD2 were mixed and the conjugation mix was centrifuged at 5000 rpm for five minutes. Cells were washed in one ml of R2A+ and centrifuged as before. Cells were resuspended in 100 µl of R2A+ and spotted on four plates of R2A media without the addition of diaminopimelic acid (DAP) or antibiotics.

Plates were incubated overnight at 30 °C and cell spots were scraped off and washed in one 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 eight 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 one ml of R2A+. Cells were diluted to OD 0.02 in R2A and sorted as described below.

7.3.5 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. The gain settings described in Table 1 were used:

Table 7.1: **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* sp. AEP1-3 *glmS*::GFP was used and the gate was set to exclude 99% of this population. To normalize the variation of fluorophore signal strength variation based on factors such as cell size, polar aggregation of fluorophores and/or cell cycles (Arnfinnsdottir et al. 2016), GFP signal strength was used to contextualize RFP signal strength as follows: the subpopulation was split into four quadrants based on their GFP and mCherry signal: green fluorescence from the genomic GFP from 100.000 to 500.000 AU was considered "high green" (HG), from 25.000 to 70.000 was considered "low green" (LG) fluorescence. Equally, red fluorescence from 8.000 to 90.000 AU was considered "high red" (HR), from 3.500 to 4.000 was considered "low red" fluorescence. The quadrant of each sorted cell is stated in the inventory list (see positive candidate table on GitHub). 2000 Cells were sorted into four 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 hours. Colonies of surviving cells were tested for successful promoter integration in the expression vector using cPCR and Sanger Sequencing. Colonies were inoculated in one ml of R2A+ containing the respective antibiotic and grown for 48 hours. One ml of 50% (v/v) glycerol in distilled water was added and cultures were frozen at −80 °C for

further use.

7.3.6 Bulk Fluorescence Intensity Measurements

Curvibacter cells containing one of the selected 5′ UTRs were inoculated from a fresh R2A+ plate into R2A+ media and grown for 36 hours at 30 °C in 24 well plates in a BMG labtech Clariostar plate reader until stationary phase. From these precultures, main cultures were inoculated to an OD of 0.05. Growth and fluorescence was monitored over 36 hours. 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. The measurements were performed in individual repetitions of three for each candidate of the *Curvibacter* mutants.

7.3.7 *Curvibacter* growth media

R2A+ media was prepared by adding additional nutrients to premixed R2A from Carl Roth according to Table 2.

Table 7.2: **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

7.3.8 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.

$$FI(fluorophore) = signal(fluorophore) - signal(media\ control) \tag{7.1}$$

The fluorescence intensity FI(fluorophore) is the value for the emission intensity of the measured fluorophore (mCherry or GFP, signal(fluorophore)) substracted by the

background emission of the media control (R2A+, signal(media control).
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:

$$RFU = FI(mCherry)/FI(GFP) \tag{7.2}$$

In order to assess changes in 5′ UTR activity across different growth phases, FI values were compared between the mid-exponential and stationary phases. This calculation was performed using the formula:

$$Differential\ activity = \frac{Normalized\ FI\ values(mid-exponential\ growth\ phase)}{Normalized\ FI\ values(stationary\ growth\ phase)} \tag{7.3}$$

Values below 0.7 indicated 5′ UTRs 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 5′ UTRs that did not show a significant preference for either growth phase. This method allowed for a comprehensive assessment of 5′ UTR behavior in relation to different growth phases.
The datasets obtained from the plate reader (BMG labtech clariostar), which included Biomass and FI data points for mCherry 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 (Savitzky and Golay 1964). 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 mid-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 mid-exponential phase, as well as the maximum OD600 value for the stationary phase.

7.3.9 Inference of transcription factor binding motifs

To identify potential transcription factor binding sites and motifs (TF-Motifs), the nucleotide sequences of the 33 5′ UTRs identified via Flow Cell Cytometry were used as input sequences for the XSTREME algorithm (Bailey 2021). 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. The proteins corresponding to the TF motifs were downloaded from UniProt and combined into a FASTA file. Using this combined FASTA file as a query, a BLAST (Basic Local Alignment Search Tool; Altschul et al. (1990, 1997)) search against the *Curvibacter* proteome was conducted with the standalone BLAST tool from NCBI (Camacho et al. 2009). The parameters were changed as follows: e_value=0.001, num_alignments=10000 and word_size=3.

7.4 Results

7.4.1 Development of a streamlined workflow for 5' UTR mining

We designed a synthetic biology approach for 5′ UTR mining in sequenced bacterial species to develop novel expression systems. The details of the workflow to filter transcriptionally active 5′ UTRs are described in the first methods section in detail. Briefly, candidate 5′ UTRs 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 regulatory sites. Filtering steps removed intergenic regions of divergent genes; with a length <60 and >170bp; 5′ UTRs 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 as candidate table together with the complete workflow 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.

7.4.2 Representativeness of extracted 5′ UTRs in *Curvibacter* sp. AEP1-3

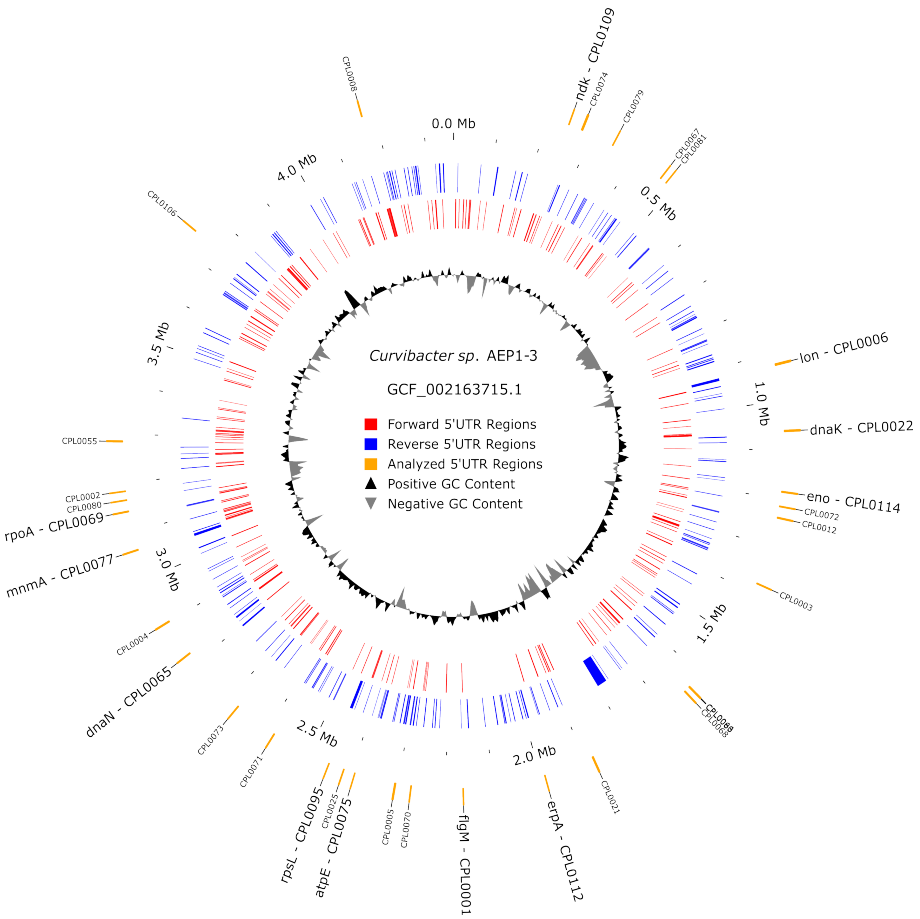

Figure 7.2: **Distribution of extracted 5′ UTRs within the genome of *Curvibacter*.** The innermost circle is composed of a density plot that showcases the GC content of the respective genome regions. Positive GC-content refers to genomic regions where the GC-content is higher than the genome-wide average, whereas negative GC-content refers to regions where it is lower. The two following red and blue circles highlight the initially extracted 500 5′ UTRs. 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′ UTRs. 5′ UTRs of CDS regions labeled as "hypothetical protein" or with protein names that are too long are not labeled.

The 33 5′ UTRs that were confirmed by Flow Cytometry (see next section) are evenly distributed throughout the entire *Curvibacter* genome (see Figure 7.2). The length of the confirmed 5′ UTRs ranges from 60 bp to 146 bp. Further, sequences of these confirmed candidate sequences were analyzed for the occurrence of common motifs (see Figure 7.3).

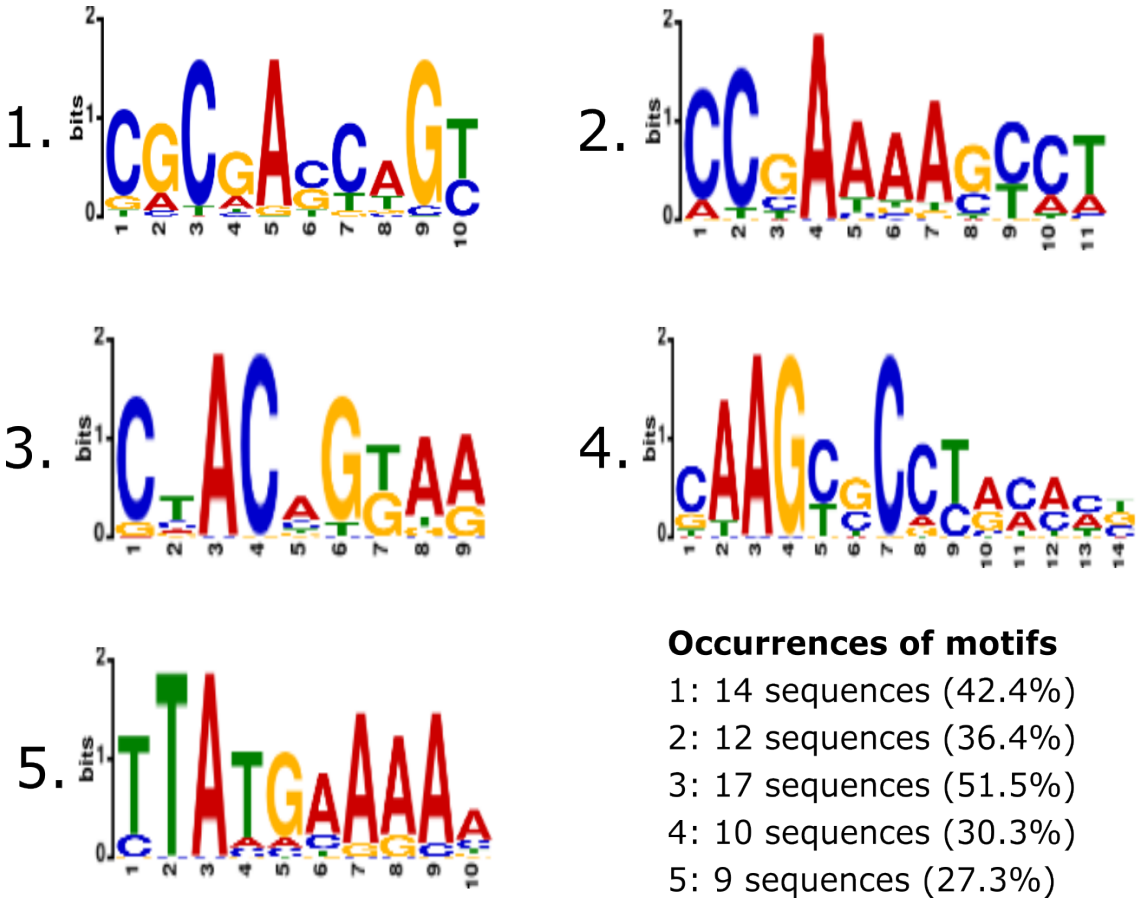

Figure 7.3: **Conserved sequence motifs within the analyzed 33 5′ UTRs of *Curvibacter*.** Sequence motifs are part of the CollectTF database. The figure highlights the 5 most prevalent motifs, additional motifs can be found in Supplementary Table S1.

Among the 33 5′ UTRs, 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 (Bailey 2021). 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 proteins of *Pseudomonas aeruginosa* (Croda-García et al. 2011; Jones et al. 2013). 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 (Gambello and Iglewski 1991) and it is considered as a transcriptional activator for virulence genes in *Pseudomonas aeruginosa* (Kiratisin et al. 2002). In 23 of 25 5′ UTRs that were further analyzed by bulk fluorescence measurement (further referred to as candidate 5′ UTRs) (see Figure 7.4a) at least one of the inferred TF-Motifs has been identified (see positive candidate table on GitHub). A detailed description of all identified motifs can be

found there as well.

BLAST analysis of the UniProt sequences obtained from the TF motifs revealed 38 unique homologous proteins from five of the eleven UniProt sequences within *Curvibacter* (see BLAST result table on GitHub). The DosR (*Mycobacterium tuberculosis*), LasR (*Pseudomonas aeruginosa*), CcPa (*Streptococcus pneumoniae*), LexA (*Vibrio parahaemolyticus*) and ExpR (*Rhizobium meliloti*) proteins, with 28, five, five, three and five homologous protein hits in *Curvibacter*, respectively. Six *Curvibacter* homologs appear multiple times in the result dataset. Transcription factors, response regulators, autoinducer-binding domain containing and chemotaxis proteins are among the homologous protein hits. None of the query sequences share a direct homology to the associated proteins of the described 33 5′ UTR regions. However, the LasR protein is a homolog to the CurR1 and CurR2 (Refseq protein identifier: WP_087495460.1 and WP_087496729.1) proteins in *Curvibacter*, which are described in Pietschke *et al.* and function directly as response regulators for the acyl-homoserine-lactone (AHL) synthases Curl1 and Curl2 (WP_232460033.1 and WP_232459811.1) in *Curvibacter* (Pietschke et al. 2017). Interestingly, with CPL0025 we identified an 5′ UTR that is associated to the AHL synthase Curl2. Among the BLAST results, the response regulator transcription factor sequence WP_157673178.1 was identified as homologous to the query protein LasR, while both WP_157673178.1 and WP_087495595.1 showed homology to the DosR query protein. They are located near CPL0077, which encodes a protein annotated as tRNA synthase MnmA, and CPL0071, which is annotated as a transferrin family substrate-binding protein (WP_087495999.1 and WP_087495596.1). Both, the response regulator WP_087495105.1 and the substrate-binding domain-containing protein WP_087495120.1 of *Curvibacter* are homologous to DosR and CcPa and are located near the 5′ UTR of CPL0112 (associated with the ErpA protein WP_087495111.1).

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

Cells carrying one of 33 unique reporter constructs sorted by flow cytometry were further analyzed for their expression level throughout different growth phases (see Figure 7.4a) using bulk fluorescence measurement in a plate reader (full list of candidates is available on GitHub as candidate table). 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. All strains are based on the same *Curvibacter* background strain containing a genomic GFP integration (Nawroth et al. 2023) in the *glmS* locus with a constant expression level relative to biomass until early stationary phase (see Supplementary Figure S2). 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 and the lack of growth phase dependant regulation 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 mCherry signal (see 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 mCherry/GFP intensity. Supplementary Figure S1-S4 contain a comparison of biomass and raw GFP expression in the background strain containing CPL0022 or CPL0017 in linear or log scale, respectively. Supplementary Figure S2 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 7.5 a-d shows a linear relation for mCherry/GFP during the stationary phase as a result of this.

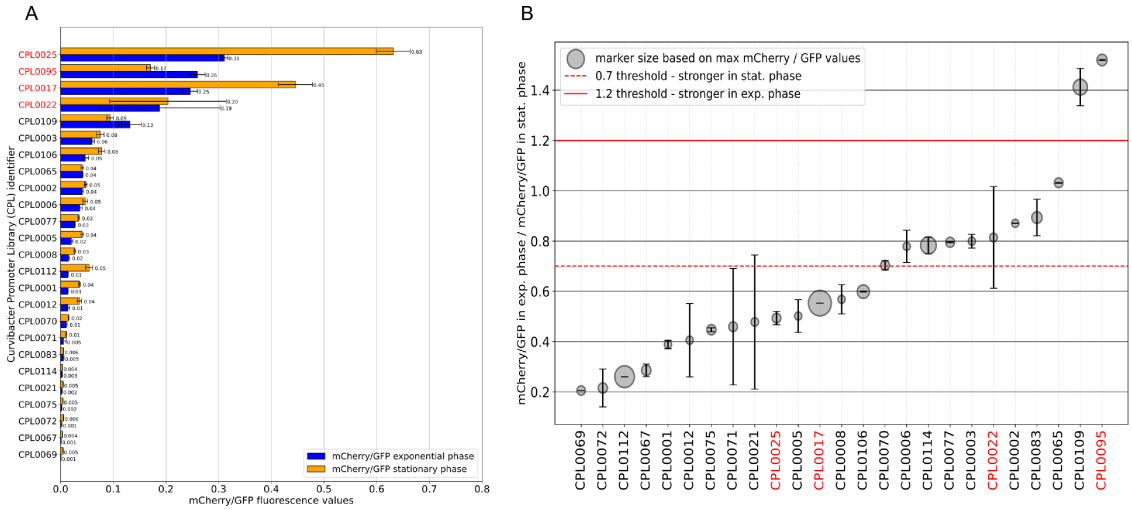

Figure 7.4: 5′ **UTR expression levels as measured by mCherry/GFP levels. A**: 5′ UTR expression levels as measured by mCherry/GFP levels within the exponential growth phase and stationary growth phase. 5′ UTR sequences are sorted in decreasing order based on the mCherry/GFP value during the exponential growth phase. The red marked CPL identifiers represent the highlighted sequences in the section below. **B**: Relative expression levels of 5′ UTRs during exponential phase and stationary phase. 5′ UTRs with values under 0.7 are more active during the stationary growth phase, while 5′ UTRs 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. Data points represent the average of three repetitions. Error bars display the standard deviation.

Figure 7.4a 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 5′ UTRs 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 7.4b). Additionally, three candidate sequences demonstrate at least 125% activity during the exponential growth phase compared to stationary phase.

7.4.4 Activity level of candidate 5’ UTRs 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 5′ UTRs of *Curvibacter*. We recommend these 5′ UTRs for further experimental use as they

cover a range of different temporal expression patterns and strengths. Additionally, we provide the graphical analysis of the 21 remaining 5′ UTRs in our GitHub repository (`https://github.com/Kanomble/curvibacter_promotor_studies`).

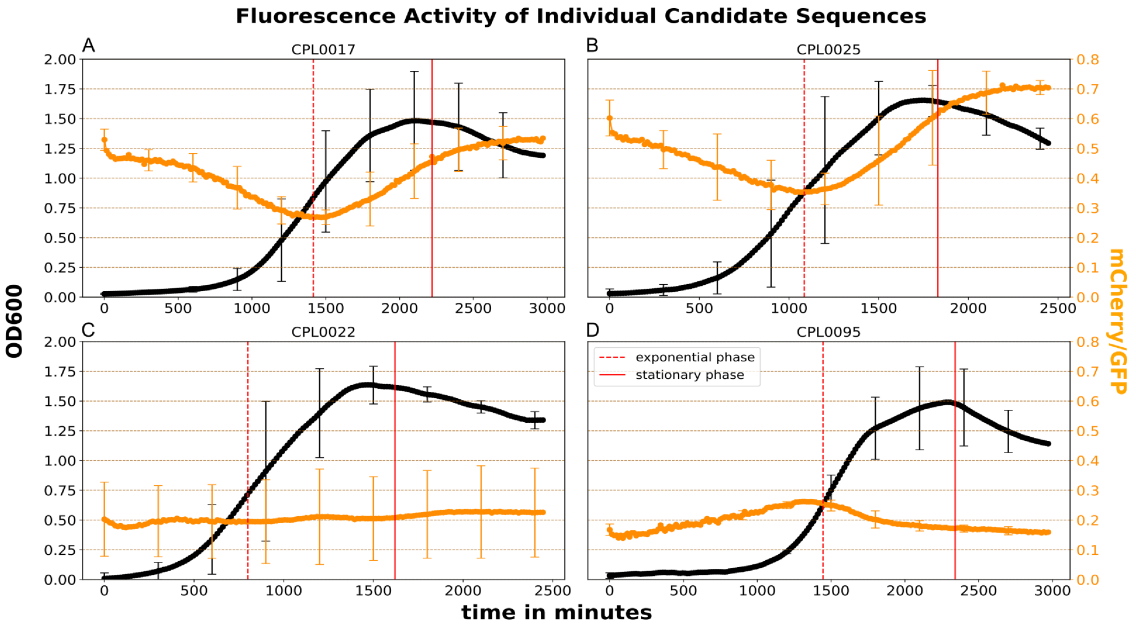

Figure 7.5: 5′ **UTR expression levels for the highlighted sequences CPL0017 (control), CPL0025, CPL0022 and CPL0095.** Full and dotted lines represent time point of exponential and stationary phase, respectively, which was used for Figure 7.4B. RFU (orange line) equals mCherry intensity (candidate 5′ UTRs) normalized to GFP intensity (which is located on the genome). Data points represent the average of three repetitions. The standard deviation of the OD and RFU values is displayed as an error bar for every 20th value.

The CPL0025 (Figure 7.5b) sequence is the 5′ UTR of the gene AEP_RS11205, expressing the AHL synthase (RefSeq protein identifier: WP_232459811), also described as Curl2 in Pietschke *et al.* (Pietschke et al. 2017). The full promoter region (519 bp) of the AHL synthase Curl2 is activated by homoserine lactones, bacterial quorum sensing molecules which play a crucial role in regulating gene expression in response to population density (Surette et al. 1999). Here we show that even the smaller 5′ UTR 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 5′ UTRs, surpassing even the expression level of the highly active J23100-RBS* 5′ UTR (CPL0017), which was used as reference (Figure 7.5a). A TF-Motif similar to the AHL activated transcriptional regulator LasR (see Supplementary Table S1) 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* (see Figure 7.3 5. and Supplementary Table S1) (Abranches et al. 2008; Ganesan and Weimer 2017; Lemos et al. 2008; Pellegrini et al. 2022).

Sequence CPL0022 (Figure 7.5c) is the 5′ UTR 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 5′ UTR shows a very constant expression level throughout all growth phases in *Curvibacter* compared to all other tested 5′ UTRs, 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, 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* (see Supplementary Table S1).

CPL0095 (Figure 7.5d) is the 5′ UTR of the gene AEP_RS11420 (RefSeq protein identifier: WP_011466063) expressing *rpsL*. RpsL is a 12S protein component of the 30S ribosomal subunit. This candidate 5′ UTR 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* (see Supplementary Table S1).

7.5 Discussion

7.5.1 25 novel promoters for the use in *Curvibacter* show distinct temporal expression dynamics

As *Curvibacter* is a promising model organism we set out in 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 5′ UTR sites in an automated fashion. Positive candidates were first picked via Flow Cy-

tometry and subsequently individual sequences 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 7.4a). We found 12 candidate 5′ UTRs which show a higher expression level in the stationary phase compared to the exponential phase. 10 candidate 5′ UTRs showed very little discrimination between growth phases, maintaining a stable level of expression throughout the observed duration and three candidate 5′ UTRs showed a higher expression level in the exponential phase compared to the stationary phase (s. Figure 7.4b).

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 5′ UTRs 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 7.4b)).

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 (Sobetzko et al. 2012). While replication induced transient changes in actual copy numbers are a factor directing genomic transcription biases along the oriC/ter axis (Teufel et al. 2023), 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 5′ UTRs of interest (as they are taken out of their natural, genomic context), they are partially affected by DNA topology (Klein et al. 2021) (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 5′ UTRs 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 tem-

poral expression dynamics. This effect is primarily attributed to the enrichment of plasmid copy numbers in the stationary phase relative to the number of cells (Akasaka et al. 2015; Berla and Pakrasi 2012; Turgeon et al. 2008). 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′ UTRs 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 (Su et al. 2021). A similar approach was recently conducted by Seo *et al.* for the cyanobacterial species *Synechocystis* sp. PCC 6803 (Seo et al. 2023). 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.

The TF-Motif discovery and enrichment analysis with the XSTREME software (s. Method section 2.9) revealed the presence of five enriched transcription factor binding motifs among the 5′ UTR sequences. This analysis identified homologous proteins in *Curvibacter* for five of the query transcription factors. Interestingly, some of these homologous transcription factors are located in close proximity to certain 5′ UTR sequences. The 5′ UTR CPL0025 contains a TF-Motif for the *Pseudomonas aeruginosa* protein LasR, among others. LasR is homologous to CurR2, which is located two genes downstream of Curl2, the protein associated with CPL0025, suggesting a potential regulatory role for CurR2 based on our dataset. This regulatory

function was indeed demonstrated by Pietschke et al. (Pietschke et al. 2017). The presence of homologous proteins and their close proximity to certain 5′ UTR sequences suggests that the motifs identified in the tested 5′ UTRs may be functionally relevant. These motifs could serve as binding sites for the homologous *Curvibacter* proteins, potentially influencing gene regulation in a manner similar to that observed in the species from which the query proteins originate.

7.5.2 Temporal expression dynamics of highlighted 5′ UTRs 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 5′ UTR, which drives expression of the *dnaK* gene in *Curvibacter* (Arsène et al. 2000). *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 7.5c). 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 5′ UTR also showed a relatively stable expression level throughout all growth phases (s. Figure 7.5c). It would be interesting to see whether this 5′ UTR 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 5′ UTR also maintains stable activity when growing on the glycocalyx of *Hydra*.

Alternatively, protein aggregation can also be prevented by using a 5′ UTR that drives lower expression levels during the stationary phase such as the CPL0095 5′ UTR. 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 (Mulder et al. 2010). Due to the high demand for protein expression during the exponential phase, genes involved in translation are upregulated during that phase, explaining the unusual temporal expression dynamic of this candidate promoter (s. Figure 7.5d). 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 7.5b). 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 7.5b) or CPL0017 (s. Figure 7.5a) 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. (Pietschke et al. 2017). 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 7.5b). 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 (Croda-García et al. 2011). 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 (Brindhadevi et al. 2020). 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 (Abranches et al. 2008; Lemos et al. 2008; Pellegrini et al. 2022). 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.

7.6 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* (see Figure 7.5) which can now be utilized in a variety of applications. It can be extended to specifically search for inducible 5′ UTRs, 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 5′ UTRs in the expression vector of both populations, the 5′ UTRs 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 find 5′ UTRs 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. However, 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 toward 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.

7.7 Appendix

7.7.1 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

7.7.2 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.

7.7.3 Material Availability

All Code used in this manuscript as well as raw data from bulk fluorescence measurements is available under (`https://github.com/Kanomble/curvibacter_promotor_studies`). A genbank file of the entry vector used to clone the promoter library is available in the GitHub data repository, including highlights for restriction sites.

7.7.4 Online Supplement

Online Supplement is provided. Supplementary Figure S1-S4 contains comparison of biomass and raw GFP expression in the background strain containing CPL0022 or CPL0017 in linear or log scale, respectively. Supplementary Table S1 contains an extended list of detected motifs in the sorted candidate sequences.

7.7.5 Funding

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

7.7.6 Conflict of interest statement

The authors declare no conflict of interest.

7.8 Supplements

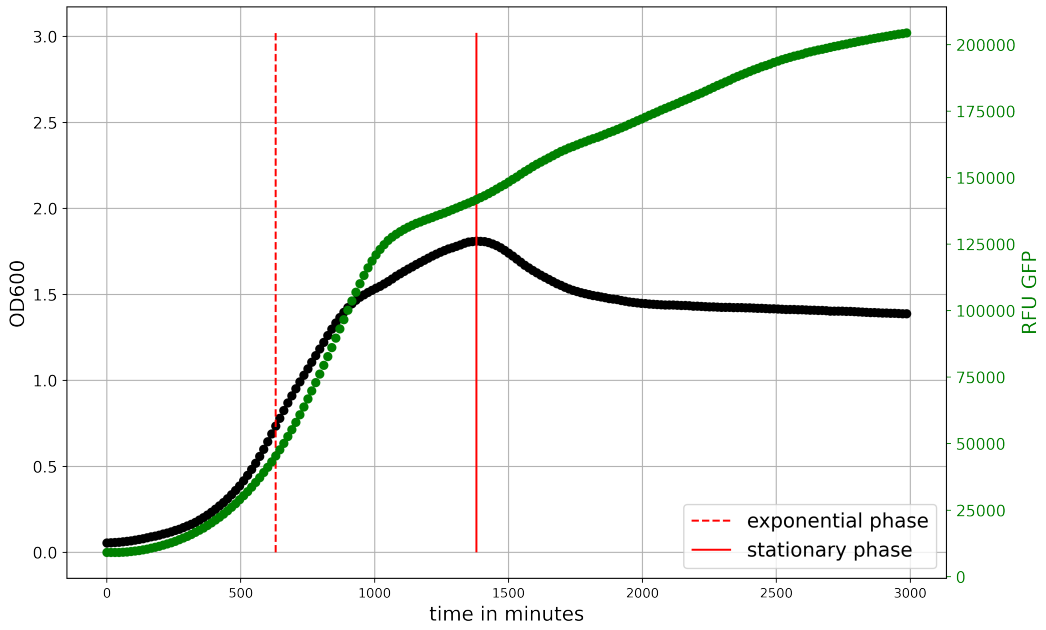

Figure 7.S1: **Comparison of change in biomass and GFP expression in *Curvibacter* AEP1-3 glmS::GFP strain carrying CPL0022 reporter plasmid.**

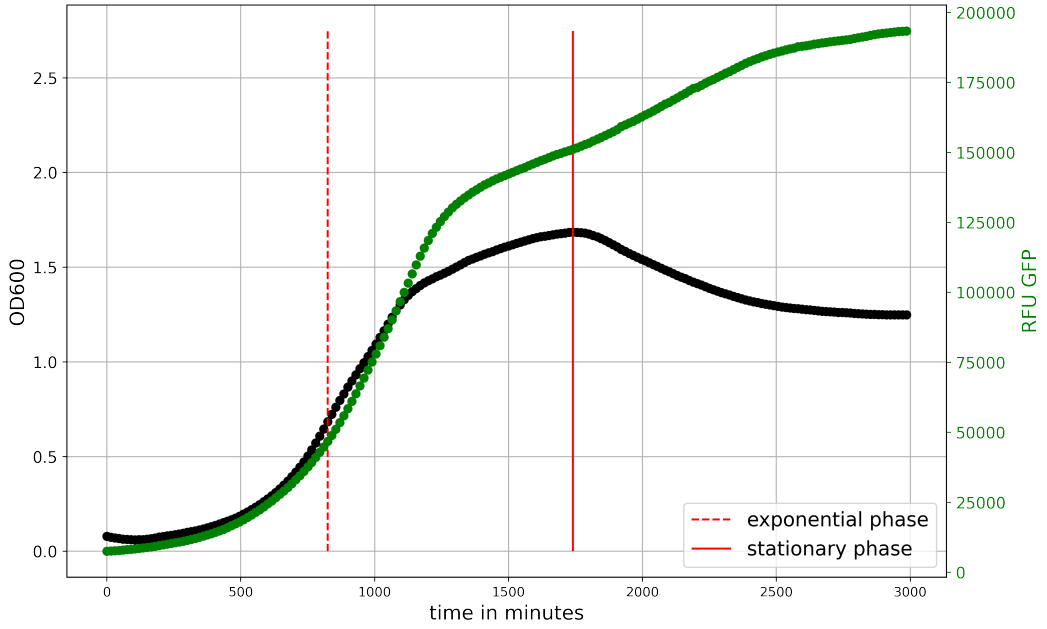

Figure 7.S2: **Comparison of change in biomass and GFP expression in *Curvibacter* AEP1-3 glmS::GFP strain carrying CPL0017 reporter plasmid.**

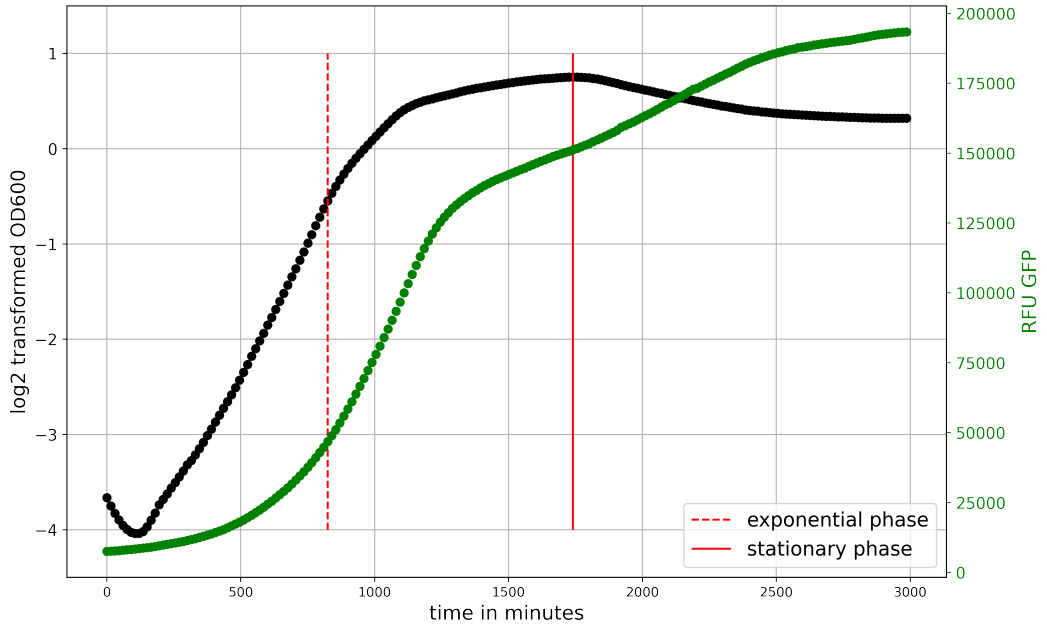

Figure 7.S3: **Comparison of change in biomass and GFP expression in *Curvibacter* AEP1-3 glmS::GFP strain carrying CPL0017 reporter plasmid, on log scale.**

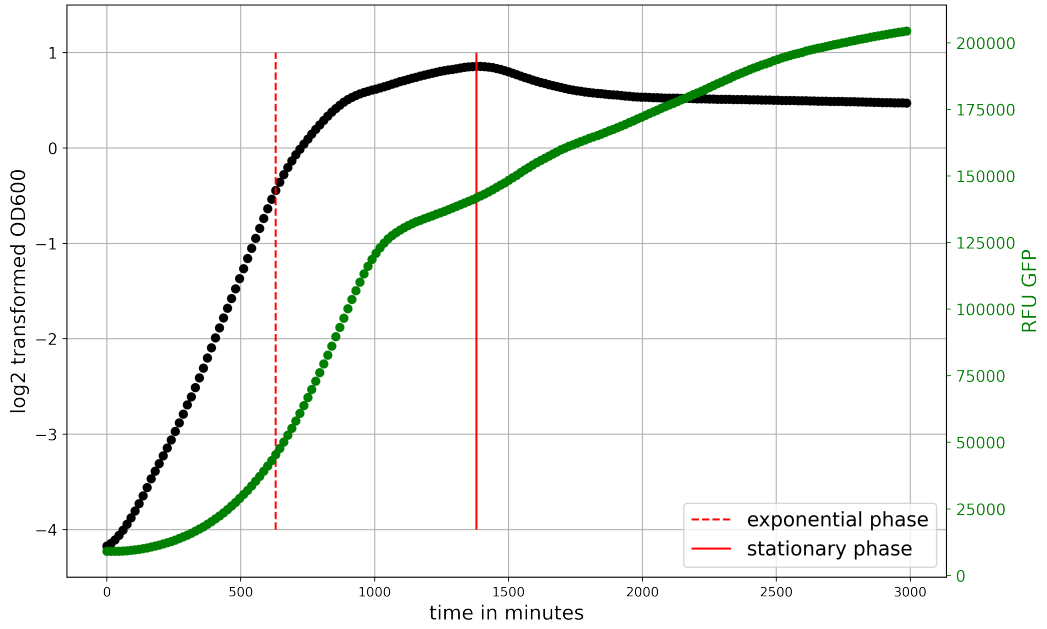

Figure 7.S4: **Comparison of change in biomass and GFP expression in *Curvibacter* AEP1-3 glmS::GFP strain carrying CPL0022 reporter plasmid, on log scale.**

References for Manuscript 4

Abranches, Jacqueline et al. (2008). "CcpA Regulates Central Metabolism and Virulence Gene Expression in Streptococcus mutans". In: *Journal of Bacteriology* 190.7, pp. 2340–2349. DOI: 10.1128/jb.01237-07.

Akasaka, Naoki et al. (2015). "Change in the plasmid copy number in acetic acid Bacteria in response to growth phase and acetic acid Concentration". In: *Journal of Bioscience and Bioengineering* 119.6, pp. 661–668. ISSN: 1389-1723. DOI: https://doi.org/10.1016/j.jbiosc.2014.11.003.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). "Basic local alignment search tool". In: *Journal of Molecular Biology* 215 (3), pp. 403–410. ISSN: 00222836. DOI: 10.1016/S0022-2836(05)80360-2.

Altschul, Stephen F. et al. (Sept. 1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". In: *Nucleic Acids Research* 25.17, pp. 3389–3402. ISSN: 0305-1048. DOI: 10.1093/nar/25.17.3389.

Andrews, Simon (2010). "FastQC: a quality control tool for high throughput sequence data." In.

Arnfinnsdottir, N. B., A. V. Bjørkøy, R. Lale, and M. Sletmoen (2016). "Heterogeneity in GFP expression in isogenic populations of: P. putida KT2440 investigated using flow cytometry and bacterial microarrays". In: *RSC Advances* 6.42, pp. 36198–36206. ISSN: 20462069. DOI: 10.1039/c5ra23757b.

Arsène, Florence, Toshifumi Tomoyasu, and Bernd Bukau (2000). "The heat shock response of Escherichia coli". In: *International Journal of Food Microbiology* 55.1, pp. 3–9. ISSN: 0168-1605. DOI: https://doi.org/10.1016/S0168-1605(00)00206-3.

Bailey, Timothy L (Mar. 2021). "STREME: accurate and versatile sequence motif discovery". In: *Bioinformatics* 37.18, pp. 2834–2840. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btab203.

Berla, Bertram M. and Himadri B. Pakrasi (2012). "Upregulation of Plasmid Genes during Stationary Phase in Synechocystis sp. Strain PCC 6803, a Cyanobacterium". In: *Applied and Environmental Microbiology* 78.15, pp. 5448–5451. DOI: 10.1128/AEM.01174-12.

Bolger, Anthony M, Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: A flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30 (15), pp. 2114–2120. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu170.

Bray, Nicolas L, Harold Pimentel, Páll Melsted, and Lior Pachter (May 2016). "Near-optimal probabilistic RNA-seq quantification". In: *Nature Biotechnology* 34 (5), pp. 525–527. ISSN: 1087-0156. DOI: 10.1038/nbt.3519.

Brindhadevi, Kathirvel et al. (2020). "Biofilm and Quorum sensing mediated pathogenicity in Pseudomonas aeruginosa". In: *Process Biochemistry* 96, pp. 49–57. ISSN: 1359-5113. DOI: https://doi.org/10.1016/j.procbio.2020.06.001.

Camacho, Christiam et al. (2009). "BLAST+: architecture and applications". In: *BMC Bioinformatics* 10 (1), p. 421. ISSN: 1471-2105. DOI: 10.1186/1471-2105-10-421.

Cleary, Michele A et al. (Dec. 2004). "Production of complex nucleic acid libraries using highly parallel in situ oligonucleotide synthesis". In: *Nature Methods* 1 (3), pp. 241–248. ISSN: 1548-7091. DOI: 10.1038/nmeth724.

Croda-García, Gerardo, Victoria Grosso-Becerra, Abigail Gonzalez-Valdez, Luis Servín-González, and Gloria Soberón-Chávez (2011). "Transcriptional regulation of Pseudomonas aeruginosa rhlR: role of the CRP orthologue Vfr (virulence factor regulator) and quorum-sensing regulators LasR and RhlR". In: *Microbiology* 157.9, pp. 2545–2555. ISSN: 1465-2080. DOI: https://doi.org/10.1099/mic.0.050161-0.

Fraune, Sebastian, Friederike Anton-Erxleben, René Augustin, Sören Franzenburg, Mirjam Knop, et al. (July 2015). "Bacteria-bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance". In: *ISME Journal* 9 (7), pp. 1543–1556. ISSN: 17517370. DOI: 10.1038/ismej.2014.239.

Gambello, M J and B H Iglewski (May 1991). "Cloning and characterization of the Pseudomonas aeruginosa lasR gene, a transcriptional activator of elastase expression". In: *Journal of Bacteriology* 173 (9), pp. 3000–3009. ISSN: 0021-9193. DOI: 10.1128/jb.173.9.3000-3009.1991.

Ganesan, Balasubramanian and Bart C. Weimer (2017). "Chapter 19 - Amino Acid Catabolism and Its Relationship to Cheese Flavor Outcomes". In: *Cheese (Fourth Edition)*. Ed. by Paul L.H. McSweeney, Patrick F. Fox, Paul D. Cotter, and David W. Everett. Fourth Edition. San Diego: Academic Press, pp. 483–516. ISBN: 978-0-12-417012-4. DOI: https://doi.org/10.1016/B978-0-12-417012-4.00019-3.

Holden, Emma R., Gregory J. Wickham, Mark A. Webber, Nicholas M. Thomson, and Eleftheria Trampari (2020). "Donor plasmids for phenotypically neutral chromosomal gene insertions in Enterobacteriaceae". In: *Microbiology* 166.12, pp. 1115–1120. ISSN: 1465-2080. DOI: https://doi.org/10.1099/mic.0.000994.

Jones, Christopher J., Cynthia R. Ryder, Ethan E. Mann, and Daniel J. Wozniak (2013). "AmrZ Modulates Pseudomonas aeruginosa Biofilm Architecture by Directly Repressing Transcription of the psl Operon". In: *Journal of Bacteriology* 195.8, pp. 1637–1644. DOI: 10.1128/jb.02190-12.

Kiratisin, Pattarachai, Kenneth D. Tucker, and Luciano Passador (2002). "LasR, a Transcriptional Activator of Pseudomonas aeruginosa Virulence Genes, Functions as a Multimer". In: *Journal of Bacteriology* 184.17, pp. 4912–4919. DOI: 10.1128/jb.184.17.4912-4919.2002.

Klein, Carlo A., Marc Teufel, Carl J. Weile, and Patrick Sobetzko (Dec. 2021). "The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity". In: *Scientific Reports* 11 (1), p. 24399. ISSN: 2045-2322. DOI: 10.1038/s41598-021-03817-4.

Lale, Rahmi et al. (Aug. 2022). "A universal approach to gene expression engineering". In: *Synthetic Biology* 7.1, ysac017. ISSN: 2397-7000. DOI: 10.1093/synbio/ysac017.

Lemos, José A., Marcelle M. Nascimento, Vanessa K. Lin, Jacqueline Abranches, and Robert A. Burne (2008). "Global Regulation by (p)ppGpp and CodY in Streptococcus mutans". In: *Journal of Bacteriology* 190.15, pp. 5291–5299. DOI: 10.1128/jb.00288-08.

Minten-Lange, Timo and Sebastian Fraune (Jan. 2020). "Hydra and Curvibacter: An intimate crosstalk at the epithelial interface". In: CRC Press, pp. 79–90. ISBN: 9781000090062. DOI: 10.1201/9780429277375-6.

Mulder, Anke M. et al. (2010). "Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30S Subunit". In: *Science* 330.6004, pp. 673–677. DOI: 10.1126/science.1193220.

Nawroth, Janna C., Christoph Giez, Alexander Klimovich, Eva Kanso, and Thomas C.G. Bosch (2023). "Spontaneous body wall contractions stabilize the fluid microenvironment thashapes host–microbe associations". In: *eLife* 12, pp. 1–23. ISSN: 2050084X. DOI: 10.7554/eLife.83637.

Oliphant, Arnold R, Alexander L. Nussbaum, and Kevin Struhl (1986). "Cloning of random-sequence oligodeoxynucleotides". In: *Gene* 44.2, pp. 177–183. ISSN: 0378-1119. DOI: https://doi.org/10.1016/0378-1119(86)90180-0.

Pellegrini, Angelica et al. (2022). "CodY Is a Global Transcriptional Regulator Required for Virulence in Group B Streptococcus". In: *Frontiers in Microbiology* 13. ISSN: 1664-302X. DOI: 10.3389/fmicb.2022.881549.

Pietschke, Cleo, Christian Treitz, Sylvain Forêt, Annika Schultze, Sven Künzel, et al. (Oct. 2017). "Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts". In: *Proceedings of the National Academy of Sciences* 114 (40). ISSN: 0027-8424. DOI: 10.1073/pnas.1706879114.

Savitzky, Abraham and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." In: *Analytical Chemistry* 36.8, pp. 1627–1639. DOI: 10.1021/ac60214a047.

Scherzinger, E et al. (1984). "Replication of the broad host range plasmid RSF1010: requirement for three plasmid-encoded proteins." In: *Proceedings of the National Academy of Sciences* 81.3, pp. 654–658. DOI: 10.1073/pnas.81.3.654.

Seo, Euijin, Yun-Nam Choi, Ye Rim Shin, Donghyuk Kim, and Jeong Wook Lee (May 2023). "Design of synthetic promoters for cyanobacteria with generative deep-learning model". In: *Nucleic Acids Research* 51.13, pp. 7071–7082. ISSN: 0305-1048. DOI: 10.1093/nar/gkad451.

Sobetzko, Patrick, Andrew Travers, and Georgi Muskhelishvili (2012). "Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle". In: *Proceedings of the National Academy of Sciences* 109.2, E42–E50. DOI: 10.1073/pnas.1108229109.

Stukenberg, Daniel et al. (2021). "Correction to "The Marburg Collection: A Golden Gate DNA Assembly Framework for Synthetic Biology Applications in Vibrio natriegens"". In: *ACS Synthetic Biology* 10.11. PMID: 34698492, pp. 3236–3236. DOI: 10.1021/acssynbio.1c00497.

Su, Wei et al. (2021). "PPD: A Manually Curated Database for Experimentally Verified Prokaryotic Promoters". In: *Journal of Molecular Biology* 433.11. Computation Resources for Molecular Biology, p. 166860. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2021.166860.

Surette, Michael G., Melissa B. Miller, and Bonnie L. Bassler (1999). "Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: A new family of genes responsible for autoinducer production". In: *Proceedings of the National Academy of Sciences* 96.4, pp. 1639–1644. DOI: 10.1073/pnas.96.4.1639.

Taubenheim, Jan, Doris Willoweit-Ohl, Mirjam Knop, Sören Franzenburg, Jinru He, et al. (Sept. 2020). "Bacteria- and temperature-regulated peptides modulate β-

catenin signaling in Hydra". In: *Proceedings of the National Academy of Sciences* 117 (35), pp. 21459–21468. ISSN: 0027-8424. DOI: 10.1073/pnas.2010945117.

Teufel, Marc, Werner Henkel, and Patrick Sobetzko (2023). "The role of replication-induced chromosomal copy numbers in spatio-temporal gene regulation and evolutionary chromosome plasticity". In: *Frontiers in Microbiology* 14. ISSN: 1664-302X. DOI: 10.3389/fmicb.2023.1119878.

Turgeon, Nathalie, Christian Laflamme, Jim Ho, and Caroline Duchaine (2008). "Evaluation of the plasmid copy number in B. cereus spores, during germination, bacterial growth and sporulation using real-time PCR". In: *Plasmid* 60.2, pp. 118–124. ISSN: 0147-619X. DOI: https://doi.org/10.1016/j.plasmid.2008.05.001.

Vasudevan, Ravendran et al. (Feb. 2019). "CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax". In: *Plant Physiology* 180.1, pp. 39–55. ISSN: 0032-0889. DOI: 10.1104/pp.18.01401.

Wang, Ye et al. (May 2020). "Synthetic promoter design in Escherichia coli based on a deep generative network". In: *Nucleic Acids Research* 48.12, pp. 6403–6412. ISSN: 0305-1048. DOI: 10.1093/nar/gkaa325.

Weber, Ernst, Carola Engler, Ramona Gruetzner, Stefan Werner, and Sylvestre Marillonnet (Feb. 2011). "A Modular Cloning System for Standardized Assembly of Multigene Constructs". In: *PLOS ONE* 6.2, pp. 1–11. DOI: 10.1371/journal.pone.0016765.

8 Discussion

8.1 General Discussion

The aim of this thesis was to investigate specific traits and adaptations of the *Curvibacter* symbiont of *Hydra* in the context of host-microbe interaction. While previous studies have demonstrated that *Curvibacter* maintains a stable relationship with *Hydra* through various modes of interaction (s. Introduction 1.6), only a limited number of published studies have explored these adaptations at both the genetic and functional level. One such study by Wein et al. (2018) showed that knockout mutants of the flagellar protein FlgC exhibit reduced colonization efficiency in *Curvibacter* AEP, indicating that motility plays a crucial role in host association. However, compared to other well-studied model microorganisms such as *P. aeruginosa* or *E. coli*, the available microbiological tools and methodologies for studying *Curvibacter* AEP are limited. This is primarily due to challenges such as low conjugation efficiency and inefficient homologous recombination. Consequently, an aspect of this thesis was the development of novel methodologies and techniques to facilitate the investigation of *Curvibacter* AEP. As part of this effort, I established a defined minimal medium, which enabled metabolic studies, leading to the identification of *Curvibacter* AEP's cobalamin-dependent methionine auxotrophy (s. Manuscript 2). Furthermore, through the screening of putative promoter sequences, we identified regulatory elements that can be utilized for gene expression studies in *Curvibacter* AEP (s. Manuscript 4). The analysis of genetic traits underlying adaptation involves genomic screening of putative target genes. To facilitate these investigations, I developed the Comparative Analysis Tool for Homolog Identification (CATHI), a software tool for the analysis of homologous protein sequences (s. Manuscript 3) (Becker et al. 2023). With the aid of CATHI and other bioinformatic tools, I identified several genes and gene clusters that are more prevalent in host-associated *Curvibacter* species, one of which is responsible for the production of extracellular polysaccharides (EPS) (s. Manuscript 1). By applying microbiological and analytical methodologies, we demonstrated that *Curvibacter* AEP secretes EPS that influence both its growth behavior and colonization efficiency. The following discussion places these new methodologies, as well as the discovered adaptations and genetic traits in a broader scientific context.

8.2 Adaptations of host-associated *Curvibacter* as response to the *Hydra* host environment

The *Hydra-Curvibacter* symbiosis is a valuable model for exploring interspecies interactions and the role of the microbiome in holobionts (Bathia and Bosch 2020; Kovačević et al. 2024; Minten-Lange and Fraune 2020). While some aspects of this relationship have been elucidated, much of its complexity remains unknown. To establish a basis for investigating the bacterial contribution in the *Hydra* holobiont, I analyzed specific genetic traits and adaptations in symbiotic *Curvibacter*.

8.2.1 Patterns of phylosymbiosis in the *Hydra-Curvibacter* model system

In general, symbiotic systems with organisms in close relationships adapt to each other, which can be measured in terms of similar and congruent phylogenetic patterns. An example provided by Liang et al. (2024) reveals significant patterns of phylogenetic congruence between endosymbiotic *Buchnera* bacteria and their aphid hosts, in which the phylogeny of the bacterial endosymbionts mirrors the phylogeny of their corresponding aphid hosts (Liang et al. 2024). This similarity is termed phylosymbiosis and can be explained by multiple ecological, evolutionary, and molecular mechanisms (Brooks et al. 2016). Patterns of phylosymbiosis are also evident within the *Hydra* holobiont (s. Figure 8.1), as distinct *Hydra* species exhibit similar microbiome compositions (Franzenburg et al. 2013).

The microbiome composition is species-specific and robust, pairwise co-culture experiments of *Hydra* species revealed that host-specific communities remain stable, with minimal convergence in microbiome composition despite close physical association (Franzenburg et al. 2013). The glycocalyx of *Hydra* serves as environmental niche for its microbiome, promoting microbial adaptation that led to long-term associations (Fraune and Bosch 2007). The longer a bacterium persists within a host environment the more likely it can become a niche-specialist, leading to reduced adaptability to alternative environments (Buckling et al. 2003; Elena and Sanjuán 2003). In manuscript 1, we demonstrated that *Curvibacter* isolates cluster congruently with their respective *Hydra* hosts, indicating not only phylosymbiosis as shown by Franzenburg et al. (2013) but also co-speciation. Moreover, we demonstrated that *Curvibacter* colonization efficiency is highest in their native *Hydra* strains, highlighting that prolonged host-specific adaptation enhances *Curvibacter's* ability

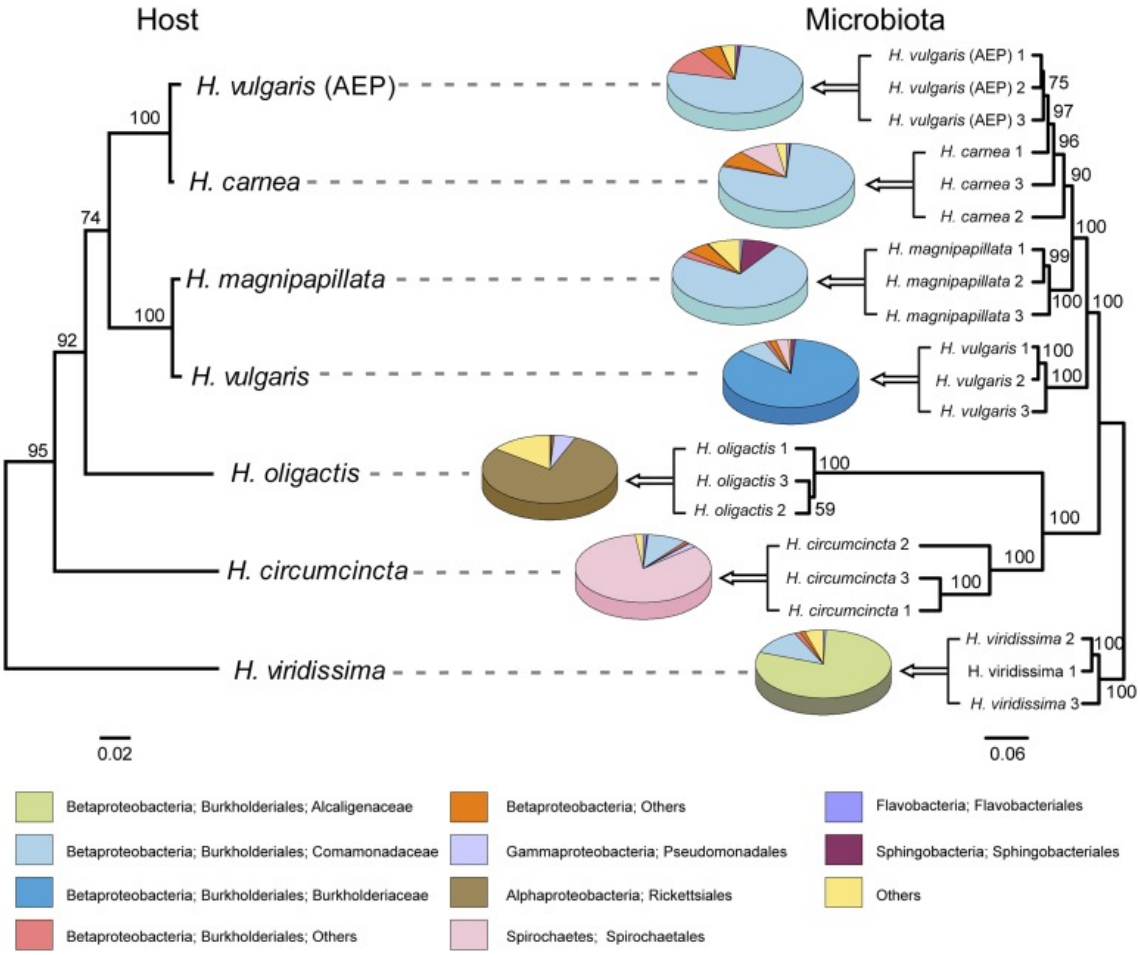

Figure 8.1: **Phylosymbiosis within the *Hydra* genus.** The phylogeny of seven different species from the *Hydra* genus remarkably mirrors a species-specific microbiome composition. The figure was taken from Franzenburg et al. (2013).

to successfully colonize its host. This long-term and intimate association between *Curvibacter* and *Hydra* may have shaped the evolution of specialized traits that enhance bacterial-host interactions and promote bacterial persistence within the host. Phylosymbiosis is also evident among Bilateria (Brooks et al. 2016). The genetic adaptations that have driven the co-speciation of symbiotic *Curvibacter* strains in the basal metazoan holobiont *Hydra* provide a framework for extrapolating to Bilaterian holobionts, offering insights into the processes underlying organismic interactions even in higher animals, such as humans or mice.

8.2.2 Conserved genomic traits in host associated *Curvibacter*

To gain insights into these adaptations I analyzed the genomes of the three *Hydra* associated *Curvibacer* strains - AEP1.3 (symbiont of *Hydra vulgaris* AEP), Hmag1.1 (symbiont of *Hydra magnipapillata*) and Hvul1 (symbiont of *Hydra vulgaris*) - and compared them to their free-living relatives (s. Manuscript 1). The genome comparison revealed a distinct dataset of 693 protein sequences in *Curvibacter* AEP, forming a unique homologous protein group conserved within the symbiotic *Curvibacter* strains. This dataset is referred to in the following sections as the symbiont specific orthogroup (OG) dataset. The observed phylosymbiosis may have shaped the genetic repertoire of symbiotic *Curvibacter* strains, leading to the retention of proteins that enhance symbiotic interactions or contribute to the fitness of *Curvibacter* within the holobiont. However, with 693 protein sequences remaining, further refinement was necessary to identify key players in the *Hydra-Curvibacter* symbiosis. A previous transcriptome analysis revealed that the plasma membrane associated protein group was the largest among differentially regulated genes in sessile *Curvibacter* AEP, compared to its planktonic-living and host-associated counterparts (Ulrich et al. 2022).

This finding indicates that the outer membrane of *Curvibacter* AEP serves as an interface for adaptive lifestyle processes. These processes typically include differential regulation of nutrient uptake, the secretion of virulence factors, and the production of extracellular polymeric substances. Consequently, I focused on proteins associated with the bacterial outer membrane. This structure serves as the primary interface with the environment (s. Figure 8.2), making it a key structural feature for mediating interactions with external conditions and other organisms (Galdiero et al. 2012; Sharma et al. 2022). Gram-negative bacteria possess two membrane structures: an inner membrane and an outer membrane, separated by the periplasm, which contains structural components such as peptidoglycan (s. Figure 8.2). The outer membrane acts as a primary barrier between the bacterium and its environment. It is composed of various proteins and other molecules that are responsible for its functional role (Hinchliffe et al. 2013; Whitfield and Trent 2014). Interestingly, annotation and gene enrichment analysis of the symbiont-specific OG dataset revealed the enrichment of transporter and other plasma membrane associated proteins, such as secretion systems and substrate-binding proteins. Thus, the persistence of these proteins in symbiotic *Curvibacter* strains may facilitate environmental sensing and the initia-

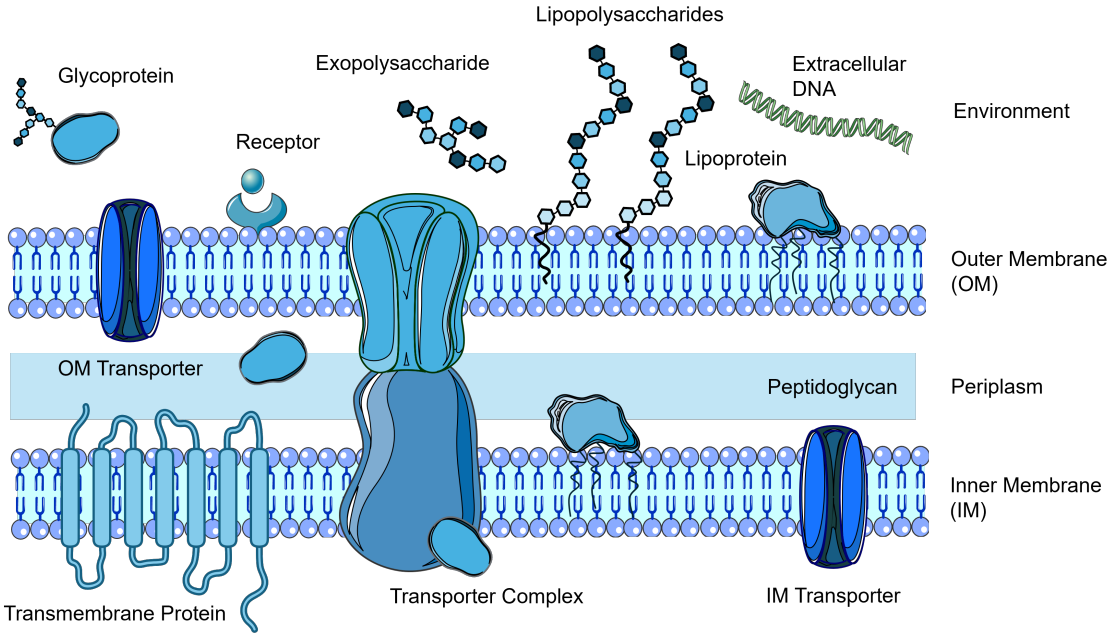

Figure 8.2: **Schematic overview of the Gram-negative bacterial membrane system.** A diverse array of proteins and protein complexes are embedded within both the outer membrane (OM) and inner membrane (IM). These membrane-associated proteins include channel and ABC transporters, porins, and specialized secretion systems (Green and Mecsas 2016; Özkan et al. 2024) such as the Wzy-dependent polymerase complex (Whitfield 2006). These enzymes can mediate the biosynthesis and secretion of various molecules such as glycoproteins, lipoproteins, extracelullar DNA, EPS and LPS (Vorkapic et al. 2016; Whitfield 2006; Zückert 2014).

tion of specialized adaptive responses, potentially mediated by host-derived cues or interactions with other members of *Hydra's* microbiome. This further strengthens the hypothesis that *Curvibacter's* plasma membrane serves as a primary interface for mediating symbiotic interactions.

8.2.3 Extracellular polymeric substances contribute to host colonization

A subset of the symbiont-specific OG dataset contains proteins associated with the biosynthesis of extracellular polymeric substances, including exopolysaccharides (EPS) and lipopolysaccharides (LPS). These molecules are either an integral part of the bacterial plasma membrane or are secreted into the environment by specialized transporter complexes (Whitfield 2006; Whitfield et al. 2020). The analysis of a gene cluster linked to EPS production highlights the role of *Curvibacter*-derived EPS in the *Hydra-Curvibacter* symbiosis (s. Manuscript 1). Using two knockout mutants of this gene cluster, we demonstrated alterations in the composition of the produced EPS, a reduced recolonization efficiency of germ-free *Hydra* AEP polyps, and significant differences in bacterial growth rates when cultured in liquid media. The

findings of this study align with observations from other symbiotic systems, where EPS plays a crucial role in colonization dynamics and host-microbe interactions, as extensively discussed in manuscript 1 (Downie 2010; Lebeer et al. 2011; Lee et al. 2016; Skorupska et al. 2006; Yip et al. 2005).

The precise role of this gene cluster in the *Hydra-Curvibacter* interaction remains unclear. The *Curvibacter*-derived EPS may contribute to bacterial biofilm formation, facilitating adhesion to the host surface, or it may function as a structural component of the glycocalyx, aiding *Curvibacter* AEP in immune evasion. In its host-associated state, *Curvibacter* AEP resides within the *Hydra* glycocalyx, a complex network of glycoproteins and glucosaminoglycans that forms five distinct cell layers (Böttger et al. 2012). A study of *Hydra's* glycocalyx identified proteins of the sweet tooth (SWT) and putative peroxidase (PPOD) family located within this extracellular layer. The PPOD proteins possess lectin like carbohydrate-binding properties, conferring the ability to bind specific sugar molecules (Böttger et al. 2012). Hypothetically, this function may serve as a mechanism for microbiome control by mediating interaction through selective binding of monosaccharides, as well as with monosaccharides in the *Curvibacter*-derived EPS. However, most genes within this EPS operon are not differentially regulated during host colonization compared to free-living conditions (unpublished data). Therefore, in conjunction with the observed alterations in growth rates, *Curvibacter* EPS may also serve as a protective molecule, shielding the bacterium from bacteriophage predation or other environmental stressors.

8.2.4 Bacterial EPS and their role in adaptive lifestyle processes

In general, extracellular polymeric substances and particularly EPS contribute to the formation of bacterial biofilms. Biofilms are dynamic and complex biological systems, with various functions in bacterial life. By establishing a self-produced, protective microenvironment, biofilms enhance survival under fluctuating environmental conditions, providing stability and resilience against external stressors (Hall-Stoodley et al. 2004). Biofilms represent an ancient and fundamental aspect of the bacterial lifestyle (Limoli et al. 2015). In particular, the secreted extracellular polymeric substances such as EPS or other glycoproteins, play a crucial role in bacterial survival on host organisms (Cohen 2015; Fanning et al. 2012). They facilitate bacterial attachment to host tissues and surfaces (Busscher and Mei 2012), aid in modulating and evading host immune responses to prevent recognition as

pathogens (Gunn et al. 2016), provide protection against environmental stressors such as desiccation (Alp and Aslim 2010; Primo et al. 2020), and influence host specificity (Fanning et al. 2012). The production of extracellular polymeric substances is tightly regulated and influenced by various environmental and genetic factors (Janczarek 2011; Mueller and Gonzaĺez 2011; Skorupska et al. 2006). It is typically triggered when bacteria encounter surfaces that necessitate attachment, prompting a phenotypic transition from a motile, free-floating planktonic state to a sessile one, which facilitates substratum colonization (Tischler and Camilli 2004).

Genes involved in the biosynthesis of these macromolecules are typically organized in operons, as exemplified by the EPS gene cluster identified in *Curvibacter* AEP. The transcription of these operon-associated genes is controlled by specific transcription factors, which abundance is, in turn, regulated by global bacterial transcriptional and translational regulatory networks (Chen et al. 2021; Netrusov et al. 2023). A key example of this regulation is cyclic diguanylate (c-di-GMP)-mediated regulation, where elevated c-di-GMP levels in response to favorable environmental conditions trigger the expression of biofilm-promoting genes, leading to reduced motility and a sessile lifestyle (Liang 2015; Tischler and Camilli 2004). Interestingly, the symbiont specific OG dataset (s. Manuscript 1) includes proteins annotated as c-di-GMP phosphodiesterases (WP_087495319 and WP_087495309), which are known to regulate intracellular c-di-GMP levels (Andersen et al. 2021). These enzymes may play a role in controlling EPS production in *Curvibacter* AEP and could serve as key regulators of symbiotic interactions by modulating c-di-GMP levels in response to the host environment. In other symbiotic systems, a regulatory role for c-di-GMP has been demonstrated. For instance, in plant-*Rhizobia* associations, elevated c-di-GMP levels enhance bacterial adhesion to plant roots, facilitating colonization and symbiosis establishment (Pérez-Mendoza et al. 2014). In the squid-*Vibrio* interaction, high levels of c-di-GMP can impair the colonization efficiency of *V. fischeri*. Furthermore, in *Pseudomonas lurida* elevated c-di-GMP levels, arising from mutations in multiple genes that collectively upregulate this secondary messenger, led to reduced motility and enhanced biofilm production, ultimately facilitating a stronger association with its host *C. elegans* (Obeng et al. 2023). These examples highlight the complex regulatory role of c-di-GMP in microbial-host interactions (Isenberg et al. 2022).

8.3 Nutrient provisioning and exchange in the *Hydra-Curvibacter* symbiosis

8.3.1 Metabolic flexibility in freshwater systems

In the context of the *Hydra-Curvibacter* relationship, the transition from free-living planktonic forms to a host-associated lifestyle necessitates transcriptional changes (Giez et al. 2023; Ulrich et al. 2022). *Curvibacter* must balance survival between its free-living state - characterized by limited nutrients and fluctuating environmental conditions - and its host-associated lifestyle, which offers a stable, nutrient-rich niche. These nutrient-rich environments often favor the loss of genes for biosynthetic functions, as corresponding essential metabolites are readily available, providing a significant fitness advantage to auxotrophic mutants (D'Souza and Kost 2016; D'Souza et al. 2014). In obligate endosymbiotic bacteria, such genome reductions are commonly observed, leading to the selective loss of nonessential genes while preserving those for maintaining the symbiotic relationship (Latorre and Manzano-Marín 2017). In the analyzed *Curvibacter* strains, genome reduction does not appear to have occurred, as their gene counts remain comparable to those of other free-living freshwater bacteria. *Curvibacter* AEP and the other symbiotic strains are facultative host-associated bacteria, meaning they can survive independently. Their ability to grow in a free-living form outside their typical niche depends on nutrient availability and the chemical composition of their environment. The retention of genes supporting free-living survival in host-associated *Curvibacter* strains mirrors similar adaptations observed in members of the human and *Drosophila* gut microbiomes, enabling them to persist both within the host and in external environments (Winans et al. 2017; Zaneveld et al. 2008). This metabolic flexibility may be essential for *Curvibacter* and other facultative symbionts during host transmission.

8.3.2 Nutrient availability affects *Hydra's* microbiome composition

External nutrient supplementation leads to a dramatic decrease in the relative abundance of *Curvibacter* within *Hydra's* microbiome (s. Figure 8.3B), from a dominant 60–80% to below 7% (Lachnit et al. 2025). While the absolute abundance of *Curvibacter* may remain relatively stable, the observed shift likely results from an increase in the proliferation of less abundant microbial taxa. This effect may reflect a disruption of *Hydra's* nutrient-mediated microbial control, potentially mediated by

host-driven mechanisms such as rhythmic body contractions (Murillo-Rincon et al. 2017; Nawroth et al. 2023) and the selective properties of the glycocalyx (Franzenburg et al. 2013; Fraune et al. 2009). Building on my first question and the previously described findings, I sought to understand whether *Curvibacter* has nutrient dependencies that influence its dynamic interactions within the *Hydra* holobiont.

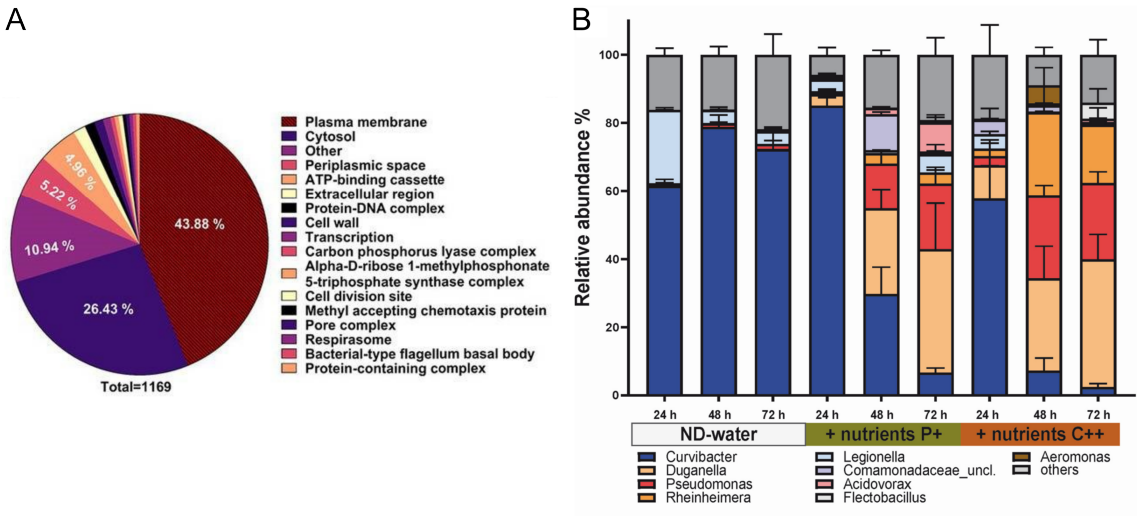

Figure 8.3: **Transcriptome analysis of *Curvibacter* AEP and the effect of nutrients on the *Hydra* microbiome. (A)** Pie diagram showcasing the percentages of differential expressed genes of sessile *Curvibacter* AEP in comparison to planktonic and host-associated *Curvibacter* AEP. In total the analysis identified 1169 differential regulated genes, with 43.88% of genes associated with the plasma membrane. The pie chart was taken from Ulrich et al. (2022). **(B)** Stacked bar plots illustrating the relative abundance of microbial taxa of laboratory-maintained *Hydra* AEP polyps exposed to varying nutrient environments over a period of three days. In nutrient-deficient water (ND-water), the relative abundance of *Curvibacter* AEP remains stable. However, in water containing a protein source (P+) or a carbon source (C++), the relative abundance of *Curvibacter* AEP declines. The stacked bar plot was taken from Lachnit et al. (2025).

The transcriptome study of *Curvibacter* AEP conducted by Ulrich et al. (2022) revealed a differential regulation of plasma membrane-associated proteins in its different lifestyles (s. Figure 8.3A). This finding was further supported by the transcriptome analysis of Giez et al. (2023), which I used to demonstrate significant upregulation of several amino acid transporter genes of *Curvibacter* AEP in mono-association with *Hydra* AEP. Furthermore, the retention and significant enrichment of transporter-related genes in the symbiont specific OG dataset (s. Manuscript 1) indicates a metabolic shift in *Curvibacter* towards the uptake of essential nutrients during its association with *Hydra*. *Curvibacter* resides in the *Hydra* glycocalyx. This extracellular layer serves as a source of metabolites and nutrients for the resident microbiome, as certain bacteria, particularly *Curvibacter*, can proliferate on *Hydra* even in the absence of external nutrient supplementation in the culture medium

(Deines et al. 2020, 2017; Fraune et al. 2015; Wein et al. 2018). Thus, *Curvibacter* must acquire these essential nutrients from the glycocalyx microenvironment in order to proliferate and establish its abundance on the *Hydra* polyp.

8.3.3 Nutrient mediated interactions in the *Hydra* holobiont

An active nutrient exchange within the *Hydra* holobiont was observed by Giez et al. (2023), who demonstrated that the availability of glutamate mediates microbe-microbe interactions, particularly between glutamate-producing bacteria like *Curvibacter* and glutamate-consuming bacteria such as *Duganella* and *Undibacterium* (Giez et al. 2023). In other symbiotic systems a similar pattern can be observed, where the synthesis of nutrients and upregulation of transporter genes facilitates a nutrient exchange. For example, in the *Aphid-Buchnera* symbiosis, the endosymbiotic bacteria supply their aphid hosts with essential amino acids, which are otherwise scarce in their phloem-based diet (Smith and Moran 2020). In the squid-*Vibrio* symbiosis, the symbiotic bacterium *V. fischeri* relies on nutrients, such as sulfur (Wasilko et al. 2019) and amino acids (Graf and Ruby 1998), that are provided within the micro-environment of the host squid species *E. scolopes*. *V. fischeri* upregulates the transcription of specific transporter genes exclusively in its symbiotic state (Jones and Nishiguchi 2006). Interestingly, one of the gene-products of these transcripts is the protein sequence BtuD, which forms a complex responsible for vitamin B12 (cobalamin) transport (Bass et al. 2003; DeVeaux and Kadner 1985). Cobalamin serves as an important cofactor for methionine synthesis, catalyzed by the methionine synthase MetH (Ferla and Patrick 2014). In addition to its role in protein synthesis, methionine availability in *V. fischeri* influences the levels of S-adenosylmethionine (SAM), a key metabolite for the bacterial quorum sensing molecules of the AHL family (Hanzelka and Greenberg 1996). In the *Hydra-Curvibacter* interaction, methionine availability appears to be a pivotal element in determining the abundance of *Curvibacter*. My findings demonstrate that *Curvibacter* AEP exhibits a cobalamin-dependent methionine auxotrophy due to a nonfunctional cobalamin-independent methionine synthase (MetE) (s. Manuscript 2). In general, vitamin or amino acid auxotrophies have been reported for various microbial organisms in aquatic environments (Johnson et al. 2020). Thus, while the cobalamin-dependent methionine auxotrophy might explain a transition of planktonic *Curvibacter* to a host-associated state, it might not be enough for establishing the long-term association that we observe today (Franzenburg et al. 2013; Fraune and Bosch 2007). Based on the identified cobalamin-dependent methionine aux-

otrophy, other interactions may have evolved that favor the host-associated status of *Curvibacter* AEP. One such interaction could be that *Curvibacter* and *Hydra* interact through the AHL signalling molecules produced by *Curvibacter* (Pietschke et al. 2017). As described in the example of *V. fischeri* (Hanzelka and Greenberg 1996), methionine availability regulates the SAM pool, which is essential for the biosynthesis of quorum-sensing signals. Given that *Curvibacter* AEP proliferates on *Hydra* AEP in the absence of methionine or cobalamin supplementation in the *Hydra*-culture medium, it is reasonable to infer that the host environment supplies methionine. This supplementation may occur either through the degradation of glycoproteins in the glycocalyx mediated by *Curvibacter*, or via active secretion of methionine by *Hydra*. Consequently, this nutrient-provisioning may modulate the synthesis of AHLs by *Curvibacter*, through the availability of methionine and SAM. This finding potentially links nutrient availability and exchange in the *Hydra* holobiont to the organismic interactions between the host and its resident *Curvibacter* population. However, most metazoans are incapable of synthesizing methionine de novo without cobalamin (Bromke and Hesse 2015; Matthews et al. 2003), which is typically supplied by associated bacteria. Since *Curvibacter* is unable to produce cobalamin in defined media (see Manuscript 2), the involvement of other microbial taxa capable of synthesizing cobalamin should be considered in explaining the methionine exchange networks within the *Hydra* holobiont.

8.4 Bacterial transporter systems facilitate lifestyle adaptations

In several sections of the preceding discussion, bacterial transporter systems were involved. The differential regulation of transporter genes, and in particular amino acid transporters and plasma membrane-associated proteins, as well as the secretion and transport of EPS, are associated with the contribution of *Curvibacter* to the *Hydra* holobiont. Differential regulation of transport proteins is also evident in other organismic interactions, such as in the opportunistic pathogen *Burkholderia cepacia* during its association with Sprague-Dawley rat hosts (O'Grady 2011). Interestingly, protein sequences associated with transport are significantly enriched within the symbiont specific OG dataset (s. Manuscript 1). Half of the carbohydrate transporter genes in *Curvibacter* AEP are found within this dataset. The retention of carbohydrate transporters in symbiotic *Curvibacter* species appears to be a distinct genetic feature, particularly when compared to the lower abundance of such transporters observed in other *Hydra*-associated symbionts (Chapman et al.

2010). Similarly, genome comparisons between plant-associated bacteria and their free-living relatives have revealed the retention of carbohydrate transporters in the symbiotic strains (Pini et al. 2011). Another example for the role of carbohydrate transporters in symbiotic interactions is provided by the facultative symbiosis between the cyanobacterium *Nostoc punctiforme* and higher plants, where the glucose transporter GlcP is essential for symbiosis establishment (Ekman et al. 2013). In combination with the previously described upregulation of amino acid transporters in *Curvibacter* AEP (s. Section 8.3.2), this finding highlights the pivotal role of bacterial transport systems in symbiotic interactions by mediating the bidirectional exchange of nutrients with the surrounding environment. Accordingly, *Curvibacter* AEP actively contributes to the shared nutrient pool and the availability of key metabolites - either by utilizing host-derived compounds such as methionine or by supplying nutrients like glutamate (Giez et al. 2023). Furthermore, the secretion of exopolysaccharides, as demonstrated in manuscript 1, along with the export of AHLs and putative virulence factors via specialized *Curvibacter* transporter systems, represent key components of the interaction within the *Hydra* holobiont.

Taken together, these results lead to the following considerations: Host-associated *Curvibacter* strains dynamically interact with their environment by (a) the ability to biosynthesize or catabolize important nutrients such as glutamate, by (b) possessing specialized transport complexes for the secretion or uptake of various molecules such as EPS or methionine and (c) adapting to the host environment through differential regulation of these transport systems.

8.5 Ortholog inference as a gateway to understanding symbiotic interactions

In the previous chapters, I examined various genetic traits of microbial species involved in holobiont interactions, with the special focus on the symbiotic relationship between *Curvibacter* and *Hydra*. In general, symbiotic interactions mediated by microbial species exhibit diverse modes of action. However, their underlying mechanisms can often be traced back to specific genetic traits and their regulation within the holobiont. Analyzing these genetic traits is challenging, as it requires programming expertise and the use of various tools. The results obtained often require post-processing despite the potential for task streamlining, as discussed in the introduction of this thesis and in manuscript 3. Thus, I asked the question: How can we bioinformatically improve the analysis of genetic traits and gene products

involved in adaptation processes?

The analysis of genetic traits and protein sequences is often the starting point for biological research. This process involves identifying orthologous protein sequences from proteins with known functions to obtain corresponding proteins in the target organisms (Koonin 2005; Tekaia 2016). This is especially important for the analysis of non-model organisms as most protein functions remain unknown and, as in the case of *Curvibacter*, research is limited to a smaller community. Thus, a detailed bioinformatic analysis of protein sequences is a crucial component of most biological research projects. Sophisticated, reliable, and fast algorithms, such as BLAST (Altschul et al. 1990, 1997) and Diamond (Buchfink et al. 2015), are already available. Furthermore, the use of these algorithms on various websites with user-friendly interfaces, along with a range of post-processing methods, further enhances the analysis of biological sequences. However, most tools are limited by (1) the availability and transparency of biological databases, (2) changing software versions that can render certain tools non-functional, (3) the number of sequences that can be analyzed simultaneously, (4) post-processing and visualization methods, (5) the customization of settings and algorithmic workflows, (6) the need for special dependencies before use, (7) programming expertise and (8) the structuring and saving of projects for future utilization. Thus, the answer to the previous question is straightforward; we need to ease the computational workflows for disentangling orthologous sequences. To address these challenges and limitations, I developed the CATHI software (s. Manuscript 3).

CATHI is based on the container-virtualization platform Docker, which enables cross-platform compatibility and makes it accessible to a wide range of users in the scientific community. Additionally, Docker's encapsulated virtualization ensures clear versioning of the software tools, allowing continued use of CATHI even if software packages are updated or changed. CATHI streamlines bioinformatics workflows using the Snakemake (Köster and Rahmann 2012) workflow management system and provides an intuitive web interface for users without programming experience. The development of bioinformatics software is ongoing, with new tools and algorithms constantly emerging to replace older programs. With the use of Snakemake, CATHI is easily modifiable, allowing for custom-tailored modifications and extensions. Its integrated database and use of the system's hard disk storage to save projects allow for the reuse of these results in future analyses, eliminating the need to repeat certain bioinformatic tasks. The previously discussed genetic traits of host-associated

Curvibacter species were initially identified through a comparative genomic approach using OrthoFinder (Emms and Kelly 2019). However, further refinement and detailed analysis of the identified orthologous protein sequences were carried out using CATHI, demonstrating its effectiveness in analyzing such sequences and untangling the complex evolutionary relationships among them.

8.6 Native promoter sequences for expression studies in *Curvibacter* AEP

Another important aspect of bacterial genetic traits and their influence within the holobiont is their differential transcriptional regulation. Transcriptome analyses by Ulrich et al. (2022) and Giez et al. (2023) have already demonstrated the adaptive transcriptional responses of *Curvibacter* AEP. To fully understand these regulatory networks, it is crucial to investigate the native promoter sequence systems. Promoter sequences are usually located adjacent to the corresponding gene, with specialized DNA motifs, such as the TATA-binding box, positioned relative to the transcriptional start site. To investigate the native promoter sequences of *Curvibacter* AEP, we isolated the native 5' untranslated regions (5'UTRs) of its genome and constructed an oligonucleotide library of these sequences to identify suitable expression systems, as extensively discussed in manuscript 4. This approach identified a set of 25 promoter sequences that exhibited significant fluorescence from the reporter gene (mCherry). Notably, one of the highest fluorescence levels, which positively correlated with bacterial growth, was observed in a *Curvibacter* AEP mutant carrying the 5'UTR of the AHL synthase LasI (WP_232459811), previously described as Curl2 by Pietschke et al. (2017). The identification of suitable expression systems is crucial for designing genetic constructs that enable controlled and predictable expression of target sequences. Therefore, this work expands the current understanding of expression systems in *Curvibacter* AEP and facilitates the study of gene regulation under controlled environmental conditions, such as varying temperatures or during host colonization. Thus, using our technique, *Curvibacter* AEP cells carrying the reporter plasmids can be cultured under various conditions and cells exhibiting mCherry activity can be screened and sorted using a cell sorter. This procedure enables the identification of active 5'UTRs that are specific to defined environmental conditions. Unlike conventional RNA-sequencing approaches, this method facilitates the *in-vivo* investigation of regulatory networks by allowing real-time measurement of reporter gene expression at defined time points, thereby providing temporal resolution of regulatory activity.

8.7 Summary

Through comparative genomics of host-associated and free-living *Curvibacter* strains, I identified several genetic traits conserved specifically within the host-associated lineage. To further investigate their functional relevance, I generated *Curvibacter* AEP knockout mutants targeting genes from a cluster implicated in EPS biosynthesis. Next, I were able to show, that these knockouts lead to alterations in the EPS monosaccharide distribution, in the bacterial growth and colonization behavior. On a theoretical level, these results provide an answer to my first question - What are genetic traits that enable the transition of *Curvibacter* AEP from a free-living state to host-association? - on a functional level they partially answer the question, as further analysis needs to be performed to effectively describe their functional role in the symbiotic interaction. However, a detailed bioinformatic analysis of the EPS operon and other symbiont-associated genetic traits was carried out using the CATHI software. CATHI thus serves as a valuable tool that streamlines and enhances the analysis of such genetic features, effectively addressing my second research question concerning methodological improvements for trait analysis. Another advancement for the microbiological analysis of *Curvibacter* AEP was achieved through the investigation of native promoter sequences. By constructing an oligonucleotide library in which 5'UTR of *Curvibacter* AEP genes were fused to a mCherry reporter and screened for activity, we enabled the identification of active promoter elements that can be leveraged for future genetic engineering, such as the investigation of regulatory networks based on these promoter sequences under diverse environmental conditions. During the analysis of the EPS I developed a defined media, thereby I could show, that *Curvibacter* AEP has a cobalamin-dependent methionine auxotrophy. This finding suggests that methionine or cobalamin is supplied within the *Hydra* holobiont, given the fact that *Curvibacter* is capable of growth on the host - even under mono-association conditions. Furthermore, nutrient exchange - particularly the transfer of amino acids - appears to play a key role in *Hydra-Curvibacter* interactions. My findings demonstrate that symbiotic *Curvibacter* species retain genes associated with transport functions and exhibit an upregulation of amino acid transporters in response to the *Hydra* host environment. These results underscore the significance of nutrient exchange in host-microbe associations. Overall, this work provides new insights into key aspects of the *Hydra-Curvibacter* symbiotic model system and introduces novel methodologies for its analysis.

9 Conclusion

Bacterial species are an intriguing part of the *Hydra* holobiont. They are responsible for diverse developmental and phenotypic effects of the *Hydra* host. Disentangling these bacterially controlled processes can contribute to the understanding of symbiotic interactions. The retention of certain genetic traits, which can be explained by a prolonged adaptation to the host environment, favors symbiotic interactions. In a similar fashion, these evolutionary processes gave rise to intricate obligate symbiotic relationships with the eukaryotic cell as an overarching example. The endosymbiosis theory posits that this evolutionary process constitutes a form of obligate mutualism, wherein the host archae-cell gets associated with a free-living bacterium, which subsequently becomes the mitochondrion and ultimately results in the evolution of a singular biological entity: the eucaryotic cell (Archibald 2015; Gray and Doolittle 1982; Raval et al. 2024). Bacterial species adapt to their environment, and the connection to the environment is facilitated by the interface of the bacterial plasma membrane. In this interface, several protein complexes, including transporter and secretory proteins, mediate the exchange of chemicals. In this thesis, I demonstrated that this interface offers multiple anchoring points for organismic interactions, by (a) secreting sugar-containing polymers and (b) absorbing key metabolites. However, these adaptations are potentially driven by a co-evolution of *Curvibacter* and *Hydra*. For instance, considering a free-living *Curvibacter* species, that is horizontally transmitted to a *Hydra* polyp. Equipped with certain favorable traits, it can adhere to the host environment and proliferate. These beneficial and other genes may evolve over the course of evolutionary time, followed by selection in the host-environment. This selection process can result in genes that exhibit altered functions, which in turn support the host-associated lifestyle, whether in a pathogenic, commensal or mutualistic manner. The ongoing evolution of *Curvibacter* on its *Hydra* host has shaped its genetic repertoire in such a way that *Curvibacter* has a beneficial effect within the *Hydra* holobiont. Considering the fact that this mutualistic relationship is a tenuous thread with the potential to undergo rapid modification, it can concomitantly result in the establishment of a persistent association, which in turn can precipitate further adaptation, thereby ultimately metamorphosing *Curvibacter* into an obligatory symbiont. The selection processes are driven by stochastic phenomena, which are in turn influenced by holistic effects such as interactions within the holobiont and the environment. It is therefore imperative to consider the genetic

traits that have been investigated as a mere glimpse in the evolutionary time of *Curvibacter*. They are subject to the ongoing evolution of *Curvibacter* and form the basis of the aforementioned selection processes. Their function, however, appears to be that of typical traits necessary for organismic interactions.

General References

Acosta-jurado, Sebastián, Francisco Fuentes-romero, Jose Enrique Ruiz-sainz, Monika Janczarek, and José María Vinardell (June 2021). "Rhizobial exopolysaccharides: Genetic regulation of their synthesis and relevance in symbiosis with legumes". In: *International Journal of Molecular Sciences* 22 (12). ISSN: 14220067. DOI: 10.3390/ijms22126233.

Aleru, Omoshola and Matthew F. Barber (Sept. 2020). "Battlefronts of evolutionary conflict between bacteria and animal hosts". In: *PLOS Pathogens* 16 (9). Ed. by Jorn Coers, e1008797. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1008797.

Alhede, Maria, Thomas Bjarnsholt, Michael Givskov, and Morten Alhede (2014). "Pseudomonas aeruginosa Biofilms". In: ed. by Sima Sariaslani and Geoffrey M Gadd. Vol. 86. Academic Press, pp. 1–40. DOI: 10.1016/B978-0-12-800262-9.00001-9.

Alp, Gulcin and Belma Aslim (Apr. 2010). "Relationship between the resistance to bile salts and low pH with exopolysaccharide (EPS) production of Bifidobacterium spp. isolated from infants feces and breast milk". In: *Anaerobe* 16 (2), pp. 101–105. ISSN: 10759964. DOI: 10.1016/j.anaerobe.2009.06.006.

Alteri, Christopher J. and Harry L. T. Mobley (June 2015). "Metabolism and Fitness of Urinary Tract Pathogens". In: *Microbiology Spectrum* 3 (3). Ed. by Tyrrell Conway and Paul Cohen. ISSN: 2165-0497. DOI: 10.1128/microbiolspec.MBP-0016-2015.

Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). "Basic local alignment search tool". In: *Journal of Molecular Biology* 215 (3), pp. 403–410. ISSN: 00222836. DOI: 10.1016/S0022-2836(05)80360-2.

Altschul, Stephen F. et al. (Sept. 1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". In: *Nucleic Acids Research* 25.17, pp. 3389–3402. ISSN: 0305-1048. DOI: 10.1093/nar/25.17.3389.

Andersen, Jens Bo, Kasper Nørskov Kragh, Louise Dahl Hultqvist, Morten Rybtke, Martin Nilsson, et al. (Mar. 2021). "Induction of Native c-di-GMP Phosphodiesterases Leads to Dispersal of Pseudomonas aeruginosa Biofilms". In: *Antimicrobial Agents and Chemotherapy* 65 (4). ISSN: 0066-4804. DOI: 10.1128/AAC.02431-20.

Archibald, John M. (Oct. 2015). "Endosymbiosis and Eukaryotic Cell Evolution". In: *Current Biology* 25 (19), R911–R921. ISSN: 09609822. DOI: 10.1016/j.cub.2015.07.055.

Augustin, René, Sebastian Fraune, and Thomas C.G. Bosch (2010). "How Hydra senses and destroys microbes". In: *Seminars in Immunology* 22 (1), pp. 54–58. ISSN: 10445323. DOI: 10.1016/j.smim.2009.11.002.

Augustin, René, Stefan Siebert, and Thomas C.G. Bosch (2009). "Identification of a kazal-type serine protease inhibitor with potent anti-staphylococcal activity as part of Hydra's innate immune system". In: *Developmental and Comparative Immunology* 33 (7), pp. 830–837. ISSN: 0145305X. DOI: 10.1016/j.dci.2009.01.009.

Bary, Anton de (1879). "Die Erscheinung der Symbiose". In: Karl J. Trübner.

Bass, Randal B, Kaspar P Locher, Elizabeth Borths, Yan Poon, Pavel Strop, et al. (Nov. 2003). "The structures of BtuCD and MscS and their implications for transporter and channel function". In: *FEBS Letters* 555 (1), pp. 111–115. ISSN: 0014-5793. DOI: 10.1016/S0014-5793(03)01126-8.

Bathia, Jay and Thomas C.G. Bosch (2020). "Symbiotic interactions in the holobiont Hydra". In: *Cellular Dialogues in the Holobiont* (Galliot), pp. 65–78. DOI: 10.1201/9780429277375-5.

Bathia, Jay, Máté Miklós, István Gyulai, Sebastian Fraune, and Jácint Tökölyi (2024). "Environmental microbial reservoir influences the bacterial communities associated with Hydra oligactis". In: *Scientific Reports* 14 (1), pp. 1–15. ISSN: 20452322. DOI: 10.1038/s41598-024-82944-0.

Bathia, Jay, Katja Schröder, Sebastian Fraune, Tim Lachnit, Philip Rosenstiel, et al. (2022). "Symbiotic Algae of Hydra viridissima Play a Key Role in Maintaining Homeostatic Bacterial Colonization". In: *Frontiers in Microbiology* 13 (June), pp. 1–11. ISSN: 1664302X. DOI: 10.3389/fmicb.2022.869666.

Becker, Anke, Hélène Bergès, Elizaveta Krol, Claude Bruand, Silvia Rüberg, et al. (Mar. 2004). "Global Changes in Gene Expression in Sinorhizobium meliloti 1021 under Microoxic and Symbiotic Conditions". In: *Molecular Plant-Microbe Interactions®* 17 (3), pp. 292–303. ISSN: 0894-0282. DOI: 10.1094/MPMI.2004.17.3.292.

Becker, Lukas et al. (2023). "CATHI: An interactive platform for comparative genomics and homolog identification". In: *bioRxiv*. DOI: https://doi.org/10.1101/2023.09.04.556229.

Bélanger, Louise, Kristin A. Dimmick, Jacquelyn S. Fleming, and Trevor C. Charles (Dec. 2009). "Null mutations in Sinorhizobium meliloti exoS and chvI demon-

strate the importance of this two-component regulatory system for symbiosis". In: *Molecular Microbiology* 74 (5), pp. 1223–1237. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.2009.06931.x.

Black, Michael, Paula Moolhuijzen, Brett Chapman, Roberto Barrero, John Howieson, et al. (Feb. 2012). "The Genetics of Symbiotic Nitrogen Fixation: Comparative Genomics of 14 Rhizobia Strains by Resolution of Protein Clusters". In: *Genes* 3 (1), pp. 138–166. ISSN: 2073-4425. DOI: 10.3390/genes3010138.

Bode, Hans, Tobias Lengfeld, Bert Hobmayer, and Thomas W. Holstein (2008). "Detection of Expression Patterns in Hydra Pattern Formation". In: *Wnt Signaling*. Ed. by Elizabeth Vincan. Humana Press, pp. 69–84. ISBN: 978-1-60327-469-2. DOI: 10.1007/978-1-60327-469-2_7.

Boehm, Anna-Marei, Konstantin Khalturin, Friederike Anton-Erxleben, Georg Hemmrich, Ulrich C. Klostermeier, et al. (Nov. 2012). "FoxO is a critical regulator of stem cell maintenance in immortal Hydra". In: *Proceedings of the National Academy of Sciences* 109 (48), pp. 19697–19702. ISSN: 0027-8424. DOI: 10.1073/pnas.1209714109.

Bosch, Thomas, René Augustin, Friederike Anton-Erxleben, Sebastian Fraune, Georg Hemmrich, et al. (Apr. 2009). "Uncovering the evolutionary history of innate immunity: The simple metazoan Hydra uses epithelial cells for host defence". In: *Developmental & Comparative Immunology* 33 (4), pp. 559–569. ISSN: 0145305X. DOI: 10.1016/j.dci.2008.10.004.

Bosch, Thomas C. G. (2019). "Hydra as Model to Determine the Role of FOXO in Longevity". In: pp. 231–238. ISBN: 93989003_19. DOI: 10.1007/978-1-4939-8900-3_19.

Bosch, Thomas C. G., Friederike Anton-Erxleben, Georg Hemmrich, and Konstantin Khalturin (Jan. 2010). "The Hydra polyp: Nothing but an active stem cell community". In: *Development, Growth & Differentiation* 52 (1), pp. 15–25. ISSN: 0012-1592. DOI: 10.1111/j.1440-169X.2009.01143.x.

Bosch, Thomas C.G. (Sept. 2013). "Cnidarian-Microbe Interactions and the Origin of Innate Immunity in Metazoans". In: *Annual Review of Microbiology* 67 (1), pp. 499–518. ISSN: 0066-4227. DOI: 10.1146/annurev-micro-092412-155626.

Bosch, Thomas C.G., Juris A. Grasis, and Tim Lachnit (Mar. 2015). "Microbial ecology in Hydra: Why viruses matter". In: *Journal of Microbiology* 53 (3), pp. 193–200. ISSN: 1225-8873. DOI: 10.1007/s12275-015-4695-2.

Böttger, Angelika, Andrew C. Doxey, Michael W. Hess, Kristian Pfaller, Willi Salvenmoser, et al. (Dec. 2012). "Horizontal Gene Transfer Contributed to the Evolution of Extracellular Surface Structures: The Freshwater Polyp Hydra Is Cov-

ered by a Complex Fibrous Cuticle Containing Glycosaminoglycans and Proteins of the PPOD and SWT (Sweet Tooth) Families". In: *PLoS ONE* 7 (12). ISSN: 19326203. DOI: 10.1371/journal.pone.0052278.

Boutry, Justine, Marie Buysse, Sophie Tissot, Chantal Cazevielle, Rodrigo Hamede, et al. (May 2023). "Spontaneously occurring tumors in different wild-derived strains of hydra". In: *Scientific Reports* 13 (1), p. 7449. ISSN: 2045-2322. DOI: 10.1038/s41598-023-34656-0.

Bright, Monika and Silvia Bulgheresi (Mar. 2010). "A complex journey: transmission of microbial symbionts". In: *Nature Reviews Microbiology* 8 (3), pp. 218–230. ISSN: 1740-1526. DOI: 10.1038/nrmicro2262.

Bromke, Mariusz A and Holger Hesse (Dec. 2015). "Phylogenetic analysis of methionine synthesis genes from Thalassiosira pseudonana". In: *SpringerPlus* 4 (1), p. 391. ISSN: 2193-1801. DOI: 10.1186/s40064-015-1163-8.

Bronstein, Judith L. (Mar. 1994). "Our Current Understanding of Mutualism". In: *The Quarterly Review of Biology* 69 (1), pp. 31–51. ISSN: 0033-5770. DOI: 10.1086/418432.

Brooks, Andrew W., Kevin D. Kohl, Robert M. Brucker, Edward J. van Opstal, and Seth R. Bordenstein (2016). "Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History". In: *PLoS Biology* 15 (1), pp. 1–29. ISSN: 15457885. DOI: 10.1371/journal.pbio.1002587.

Buchfink, Benjamin, Chao Xie, and Daniel H Huson (Jan. 2015). "Fast and sensitive protein alignment using DIAMOND". In: *Nature Methods* 12 (1), pp. 59–60. ISSN: 1548-7091. DOI: 10.1038/nmeth.3176.

Buckling, Angus, Matthew A. Wills, and Nick Colegrave (Dec. 2003). "Adaptation Limits Diversification of Experimental Bacterial Populations". In: *Science* 302 (5653), pp. 2107–2109. ISSN: 0036-8075. DOI: 10.1126/science.1088848.

Busscher, Henk J. and Henny C. van der Mei (Jan. 2012). "How Do Bacteria Know They Are on a Surface and Regulate Their Response to an Adhering State?" In: *PLoS Pathogens* 8 (1). Ed. by Virginia Miller, e1002440. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1002440.

Buzgariu, W, S Al Haddad, S Tomczyk, Y Wenger, and B Galliot (Oct. 2015). "Multi-functionality and plasticity characterize epithelial cells in Hydra". In: *Tissue Barriers* 3 (4), e1068908. ISSN: 2168-8370. DOI: 10.1080/21688370.2015.1068908.

Cai, L., D. Juedes, and E. Liakhovitch (2000). "Evolutionary computation techniques for multiple sequence alignment". In: *Proceedings of the 2000 Congress on*

Evolutionary Computation. CEC00 (Cat. No.00TH8512). Vol. 2. IEEE, pp. 829–835. ISBN: 0-7803-6375-2. DOI: 10.1109/CEC.2000.870716.

Cao, Mengyi and Heidi Goodrich-Blair (Aug. 2017). "Ready or Not: Microbial Adaptive Responses in Dynamic Symbiosis Environments". In: *Journal of Bacteriology* 199 (15). Ed. by George O'Toole. ISSN: 0021-9193. DOI: 10.1128/JB.00883-16.

Casadevall, Arturo and Liise-anne Pirofski (Jan. 2015). "What Is a Host? Incorporating the Microbiota into the Damage-Response Framework". In: *Infection and Immunity* 83 (1). Ed. by H. L. Andrews-Polymenis, pp. 2–7. ISSN: 0019-9567. DOI: 10.1128/IAI.02627-14.

Cazet, Jack F., Stefan Siebert, Hannah Morris Little, Philip Bertemes, Abby S. Primack, et al. (2023). "A chromosome-scale epigenetic map of the Hydra genome reveals conserved regulators of cell state". In: *Genome Research* 33 (2), pp. 283–298. ISSN: 15495469. DOI: 10.1101/gr.277040.122.

Chapman, Jarrod A., Ewen F. Kirkness, Oleg Simakov, Steven E. Hampson, Therese Mitros, et al. (Mar. 2010). "The dynamic genome of Hydra". In: *Nature* 464 (7288), pp. 592–596. ISSN: 0028-0836. DOI: 10.1038/nature08830.

Chen, Chengjie, Hao Chen, Yi Zhang, Hannah R Thomas, Margaret H Frank, et al. (2020). "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data". In: *Molecular Plant* 13, pp. 1194–1202. DOI: 10.1016/j.molp.2020.06.009.

Chen, J., A. Zhang, Z. Xiang, M. Lu, P. Huang, et al. (Aug. 2021). "EpsR Negatively Regulates Streptococcus mutans Exopolysaccharide Synthesis". In: *Journal of Dental Research* 100 (9), pp. 968–976. ISSN: 0022-0345. DOI: 10.1177/00220345211000668.

Chen, Jason Z., Anthony Junker, Iris Zheng, Nicole M. Gerardo, and Nic M. Vega (Oct. 2024). "A strong priority effect in the assembly of a specialized insect-microbe symbiosis". In: *Applied and Environmental Microbiology* 90 (10). Ed. by Knut Rudi. ISSN: 0099-2240. DOI: 10.1128/aem.00818-24.

Chen, Xi, Jun Yang, Paul M Evans, and Chunming Liu (July 2008). "Wnt signaling: the good and the bad". In: *Acta Biochimica et Biophysica Sinica* 40 (7), pp. 577–594. ISSN: 1672-9145. DOI: 10.1111/j.1745-7270.2008.00440.x.

Chipman, Ariel D. (May 2024). "45Diploblastic organisms: Cnidaria and Ctenophora". In: *Organismic Animal Biology: An Evolutionary Approach*. Oxford University Press. ISBN: 9780192893581. DOI: 10.1093/oso/9780192893581.003.0009.

Cohen, Miriam (Sept. 2015). "Notable Aspects of Glycan-Protein Interactions". In: *Biomolecules* 5 (3), pp. 2056–2072. ISSN: 2218-273X. DOI: 10.3390/biom5032056.

Collins, A. G. (May 2002). “Phylogeny of Medusozoa and the evolution of cnidarian life cycles”. In: *Journal of Evolutionary Biology* 15 (3), pp. 418–432. ISSN: 1010-061X. DOI: 10.1046/j.1420-9101.2002.00403.x.

Collins, A. G., Peter Schuchert, Antonio C. Marques, Thomas Jankowski, Mónica Medina, et al. (Feb. 2006). “Medusozoan Phylogeny and Character Evolution Clarified by New Large and Small Subunit rDNA Data and an Assessment of the Utility of Phylogenetic Mixture Models”. In: *Systematic Biology* 55 (1). Ed. by Tim Collins, pp. 97–115. ISSN: 1076-836X. DOI: 10.1080/10635150500433615.

Cresci, Gail A. and Emmy Bawden (Dec. 2015). “Gut Microbiome”. In: *Nutrition in Clinical Practice* 30 (6), pp. 734–746. ISSN: 0884-5336. DOI: 10.1177/0884533615609899.

Crick, Francis (1958). “On Protein Synthesis”. In: *Symposia of the Society for Experimental Biology* (12), pp. 138–163.

— (1970). “Central Dogma of Molecular Biology”. In: *NATURE* 227, p. 1970.

D’Souza, Glen and Christian Kost (Nov. 2016). “Experimental Evolution of Metabolic Dependency in Bacteria”. In: *PLOS Genetics* 12 (11). Ed. by Diarmaid Hughes, e1006364. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1006364.

D’Souza, Glen, Silvio Waschina, Samay Pande, Katrin Bohl, Christoph Kaleta, et al. (Sept. 2014). “Less Is More: Selective Advantages Can Explain The Prevalent Loss Of Biosynthetic Genes In Bacteria”. In: *Evolution* 68 (9), pp. 2559–2570. ISSN: 00143820. DOI: 10.1111/evo.12468.

Dale, Colin and Nancy A. Moran (Aug. 2006). “Molecular Interactions between Bacterial Symbionts and Their Hosts”. In: *Cell* 126 (3), pp. 453–465. ISSN: 00928674. DOI: 10.1016/j.cell.2006.07.014.

Deines, Peter, Katrin Hammerschmidt, and Thomas C G Bosch (2020). “Microbial Species Coexistence Depends on the Host Environment”. In: DOI: 10.1128/mbio.00807-20.

Deines, Peter, Tim Lachnit, and Thomas C. G. Bosch (Sept. 2017). “Competing forces maintain the Hydra metaorganism”. In: *Immunological Reviews* 279 (1), pp. 123–136. ISSN: 0105-2896. DOI: 10.1111/imr.12564.

DeVeaux, L C and R J Kadner (June 1985). “Transport of vitamin B12 in Escherichia coli: cloning of the btuCD region”. In: *Journal of Bacteriology* 162 (3), pp. 888–896. ISSN: 0021-9193. DOI: 10.1128/jb.162.3.888-896.1985.

Dey, Priyankar (May 2024). “Good girl goes bad: Understanding how gut commensals cause disease”. In: *Microbial Pathogenesis* 190, p. 106617. ISSN: 08824010. DOI: 10.1016/j.micpath.2024.106617.

Ding, Linxian and Akira Yokota (Nov. 2004). "Proposals of Curvibacter gracilis gen. nov., sp. nov. and Herbaspirillum putei sp. nov. for bacterial strains isolated from well water and reclassification of [Pseudomonas] huttiensis, [Pseudomonas] lanceolata, [Aquaspirillum] delicatum and [Aquaspirillum] autotrophicum as Herbaspirillum huttiense comb. nov., Curvibacter lansceolatus comb. nov., Curvibacter delicatus comb. nov." In: *International Journal of Systematic and Evolutionary Microbiology* 54 (6), pp. 2223–2230. ISSN: 14665026. DOI: 10.1099/ijs.0.02975-0.

Douglas, Angela E. (2010). *The Symbiotic Habit*. Princeton University Press. ISBN: 9780691134241.

— (Dec. 2019). "Simple animal models for microbiome research". In: *Nature Reviews Microbiology* 17 (12), pp. 764–775. ISSN: 1740-1526. DOI: 10.1038/s41579-019-0242-1.

Downie, J. Allan (Mar. 2010). "The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots". In: *FEMS Microbiology Reviews* 34 (2), pp. 150–170. ISSN: 1574-6976. DOI: 10.1111/j.1574-6976.2009.00205.x.

Egerton, Frank N. (Jan. 2015). "History of Ecological Sciences, Part 52: Symbiosis Studies". In: *The Bulletin of the Ecological Society of America* 96 (1), pp. 80–139. ISSN: 0012-9623. DOI: 10.1890/0012-9623-96.1.80.

Ekman, Martin, Silvia Picossi, Elsie L. Campbell, John C. Meeks, and Enrique Flores (Apr. 2013). "A Nostoc punctiforme Sugar Transporter Necessary to Establish a Cyanobacterium-Plant Symbiosis ". In: *Plant Physiology* 161 (4), pp. 1984–1992. ISSN: 1532-2548. DOI: 10.1104/pp.112.213116.

Elena, Santiago F. and Rafael Sanjuán (Dec. 2003). "Climb Every Mountain?" In: *Science* 302 (5653), pp. 2074–2075. ISSN: 0036-8075. DOI: 10.1126/science.1093165.

Elena Aledo (2025). *Glutamine Synthetase from Conifers: Unraveling the Hidden Paralogs*. Accessed: 2025-04-14. URL: https://cran.r-project.org/web/packages/orthGS/vignettes/Unraveling_the_hiden_paralogous.html.

Emms, David M. and Steven Kelly (Dec. 2019). "OrthoFinder: phylogenetic orthology inference for comparative genomics". In: *Genome Biology* 20 (1), p. 238. ISSN: 1474-760X. DOI: 10.1186/s13059-019-1832-y.

Encyclopaedia Britannica (2024). *Hydra – Greek Mythology*. Accessed: 2025-04-14. URL: https://www.britannica.com/topic/Hydra-Greek-mythology.

Fagorzi, Camilla, Alexandru Ilie, Francesca Decorosi, Lisa Cangioli, Carlo Viti, et al. (Dec. 2020). "Symbiotic and Nonsymbiotic Members of the Genus Ensifer (syn.

Sinorhizobium) Are Separated into Two Clades Based on Comparative Genomics and High-Throughput Phenotyping". In: *Genome Biology and Evolution* 12 (12). Ed. by Esperanza Martinez-Romero, pp. 2521–2534. ISSN: 1759-6653. DOI: 10.1093/gbe/evaa221.

Fanning, Saranna, Lindsay J. Hall, and Douwe van Sinderen (Sept. 2012). "Bifidobacterium breve UCC2003 surface exopolysaccharide production is a beneficial trait mediating commensal-host interaction through immune modulation and pathogen protection". In: *Gut Microbes* 3 (5), pp. 420–425. ISSN: 1949-0976. DOI: 10.4161/gmic.20630.

Feng, Da-Fei and Russell F. Doolittle (Aug. 1987). "Progressive sequence alignment as a prerequisite to correct phylogenetic trees". In: *Journal of Molecular Evolution* 25 (4), pp. 351–360. ISSN: 0022-2844. DOI: 10.1007/BF02603120.

Ferla, Matteo P. and Wayne M. Patrick (Aug. 2014). "Bacterial methionine biosynthesis". In: *Microbiology* 160 (8), pp. 1571–1584. ISSN: 1350-0872. DOI: 10.1099/mic.0.077826-0.

Finlay, B. Brett and Grant McFadden (Feb. 2006). "Anti-Immunology: Evasion of the Host Immune System by Bacterial and Viral Pathogens". In: *Cell* 124 (4), pp. 767–782. ISSN: 00928674. DOI: 10.1016/j.cell.2006.01.034.

Fisher, Roberta M., Lee M. Henry, Charlie K. Cornwallis, E. Toby Kiers, and Stuart A. West (July 2017). "The evolution of host-symbiont dependence". In: *Nature Communications* 8 (1), p. 15973. ISSN: 2041-1723. DOI: 10.1038/ncomms15973.

Fitch, Walter M. (June 1970). "Distinguishing Homologous from Analogous Proteins". In: *Systematic Zoology* 19 (2), p. 99. ISSN: 00397989. DOI: 10.2307/2412448.

— (May 2000). "Homology". In: *Trends in Genetics* 16 (5), pp. 227–231. ISSN: 01689525. DOI: 10.1016/S0168-9525(00)02005-9.

Franzenburg, Sören, Sebastian Fraune, Sven Künzel, John F. Baines, Tomislav Domazet-Lošo, et al. (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), pp. 19374–19379. ISSN: 00278424. DOI: 10.1073/pnas.1213110109.

Franzenburg, Sören, Jonas Walter, Sven Künzel, Jun Wang, John F. Baines, et al. (Sept. 2013). "Distinct antimicrobial peptide expression determines host species-specific bacterial associations". In: *Proceedings of the National Academy of Sciences of the United States of America* 110 (39). ISSN: 00278424. DOI: 10.1073/pnas.1304960110.

Fraune, Sebastian, Yuichi Abe, and Thomas C. G. Bosch (Sept. 2009). "Disturbing epithelial homeostasis in the metazoan Hydra leads to drastic changes in associated microbiota". In: *Environmental Microbiology* 11 (9), pp. 2361–2369. ISSN: 1462-2912. DOI: 10.1111/j.1462-2920.2009.01963.x.

Fraune, Sebastian, Friederike Anton-Erxleben, René Augustin, Sören Franzenburg, Mirjam Knop, et al. (July 2015). "Bacteria-bacteria interactions within the microbiota of the ancestral metazoan Hydra contribute to fungal resistance". In: *ISME Journal* 9 (7), pp. 1543–1556. ISSN: 17517370. DOI: 10.1038/ismej.2014.239.

Fraune, Sebastian, René Augustin, Friederike Anton-Erxleben, Jörg Wittlieb, Christoph Gelhaus, et al. (Oct. 2010). "In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides". In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (42), pp. 18067–18072. ISSN: 00278424. DOI: 10.1073/pnas.1008573107.

Fraune, Sebastian and Thomas C.G. Bosch (2007). "Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra". In: *Proceedings of the National Academy of Sciences of the United States of America* 104 (32), pp. 13146–13151. ISSN: 00278424. DOI: 10.1073/pnas.0703375104.

Galdiero, Stefania, Annarita Falanga, Marco Cantisani, Rossella Tarallo, Maria Elena Della Pepa, et al. (Dec. 2012). "Microbe-Host Interactions: Structure and Role of Gram-Negative Bacterial Porins". In: *Current Protein and Peptide Science* 13 (8), pp. 843–854. ISSN: 13892037. DOI: 10.2174/138920312804871120.

Gerth, Michael and Christoph Bleidorn (Dec. 2016). "Comparative genomics provides a timeframe for Wolbachia evolution and exposes a recent biotin synthesis operon transfer". In: *Nature Microbiology* 2 (3), p. 16241. ISSN: 2058-5276. DOI: 10.1038/nmicrobiol.2016.241.

Ghantasala, Swathi and Swarup Roy Choudhury (Dec. 2022). "Nod factor perception: an integrative view of molecular communication during legume symbiosis". In: *Plant Molecular Biology* 110 (6), pp. 485–509. ISSN: 0167-4412. DOI: 10.1007/s11103-022-01307-3.

Giez, Christoph, Denis Pinkle, Yan Giencke, Jörg Wittlieb, Eva Herbst, et al. (Dec. 2023). "Multiple neuronal populations control the eating behavior in Hydra and are responsive to microbial signals". In: *Current Biology* 33 (24), 5288–5303.e6. ISSN: 18790445. DOI: 10.1016/j.cub.2023.10.038.

Goodwin, Sara, John D. McPherson, and W. Richard McCombie (June 2016). "Coming of age: ten years of next-generation sequencing technologies". In: *Nature Re-*

views Genetics 17 (6), pp. 333–351. ISSN: 1471-0056. DOI: 10.1038/nrg.2016.49.

Graf, Joerg and E. G. Ruby (Feb. 1998). "Host-derived amino acids support the proliferation of symbiotic bacteria". In: *Proceedings of the National Academy of Sciences* 95 (4), pp. 1818–1822. ISSN: 0027-8424. DOI: 10.1073/pnas.95.4.1818.

Gray, M W and W F Doolittle (Mar. 1982). "Has the endosymbiont hypothesis been proven?" In: *Microbiological Reviews* 46 (1), pp. 1–42. ISSN: 0146-0749. DOI: 10.1128/mr.46.1.1-42.1982.

Green, Erin R. and Joan Mecsas (Jan. 2016). "Bacterial Secretion Systems: An Overview". In: *Microbiology Spectrum* 4 (1). Ed. by Indira T. Kudva. ISSN: 2165-0497. DOI: 10.1128/microbiolspec.VMBF-0012-2015.

Gresham, David, Maitreya J. Dunham, and David Botstein (2008). "Comparing whole genomes using DNA microarrays". In: *Nature Reviews Genetics* 9 (4), pp. 291–302. ISSN: 14710056. DOI: 10.1038/nrg2335.

Groussin, Mathieu, Florent Mazel, and Eric J. Alm (July 2020). "Co-evolution and Co-speciation of Host-Gut Bacteria Systems". In: *Cell Host & Microbe* 28 (1), pp. 12–22. ISSN: 19313128. DOI: 10.1016/j.chom.2020.06.013.

Grundmann, Oliver (June 2020). "Gastrointestinal Inflammation and the Gut Microbiome: An Evolving Conceptual Framework with Implications for Diagnosis and Therapy in Inflammatory Bowel Disorders". In: *EMJ Microbiology & Infectious Diseases*, pp. 42–50. ISSN: 2732-5326. DOI: 10.33590/emjmicrobiolinfectdis/20-00050.

Gunn, John S., Lauren O. Bakaletz, and Daniel J. Wozniak (June 2016). "What's on the Outside Matters: The Role of the Extracellular Polymeric Substance of Gram-negative Biofilms in Evading Host Immunity and as a Target for Therapeutic Intervention". In: *Journal of Biological Chemistry* 291 (24), pp. 12538–12546. ISSN: 00219258. DOI: 10.1074/jbc.R115.707547.

Gyles, C. and P. Boerlin (Mar. 2014). "Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease". In: *Veterinary Pathology* 51 (2), pp. 328–340. ISSN: 0300-9858. DOI: 10.1177/0300985813511131.

Hall-Stoodley, Luanne, J. William Costerton, and Paul Stoodley (Feb. 2004). "Bacterial biofilms: from the Natural environment to infectious diseases". In: *Nature Reviews Microbiology* 2 (2), pp. 95–108. ISSN: 1740-1526. DOI: 10.1038/nrmicro821.

Hammer, Tobin J, Jon G Sanders, and Noah Fierer (May 2019). "Not all animals need a microbiome". In: *FEMS Microbiology Letters* 366 (10). ISSN: 1574-6968. DOI: 10.1093/femsle/fnz117.

Hansen, Georg Norgaard, Michael Williamson, and Cornelis J.P. Grimmelikhuijzen (July 2000). "Two-color double-labeling in situ hybridization of whole-mount Hydra using RNA probes for five different Hydra neuropeptide preprohormones: evidence for colocalization". In: *Cell and Tissue Research* 301 (2), pp. 245–253. ISSN: 0302-766X. DOI: 10.1007/s004410000240.

Hanzelka, B L and E P Greenberg (Sept. 1996). "Quorum sensing in Vibrio fischeri: evidence that S-adenosylmethionine is the amino acid substrate for autoinducer synthesis". In: *Journal of Bacteriology* 178 (17), pp. 5291–5294. ISSN: 0021-9193. DOI: 10.1128/jb.178.17.5291-5294.1996.

Hemmrich, Georg, David J. Miller, and Thomas Bosch (2007). "The evolution of immunity: a low-life perspective". In: *Trends in Immunology* 28 (10), pp. 449–454. ISSN: 14714906. DOI: 10.1016/j.it.2007.08.003.

Henikoff, S and J G Henikoff (Nov. 1992). "Amino acid substitution matrices from protein blocks." In: *Proceedings of the National Academy of Sciences* 89 (22), pp. 10915–10919. ISSN: 0027-8424. DOI: 10.1073/pnas.89.22.10915.

Hernández-Salmerón, Julie E and Gabriel Moreno-Hagelsieb (May 2020). "Progress in quickly finding orthologs as reciprocal best hits". In: 30, pp. 1–5. ISSN: 14712164. DOI: 10.1101/2020.05.04.077222.

Hilgenboecker, Kirsten, Peter Hammerstein, Peter Schlattmann, Arndt Telschow, and John H. Werren (Apr. 2008). "How many species are infected with Wolbachia? - a statistical analysis of current data". In: *FEMS Microbiology Letters* 281 (2), pp. 215–220. ISSN: 03781097. DOI: 10.1111/j.1574-6968.2008.01110.x.

Hinchliffe, Philip, Martyn F. Symmons, Colin Hughes, and Vassilis Koronakis (Sept. 2013). "Structure and Operation of Bacterial Tripartite Pumps". In: *Annual Review of Microbiology* 67 (1), pp. 221–242. ISSN: 0066-4227. DOI: 10.1146/annurev-micro-092412-155718.

Holstein, Thomas W. (July 2022). "The role of cnidarian developmental biology in unraveling axis formation and Wnt signaling". In: *Developmental Biology* 487, pp. 74–98. ISSN: 00121606. DOI: 10.1016/j.ydbio.2022.04.005.

Hosokawa, Takahiro, Ryuichi Koga, Yoshitomo Kikuchi, Xian-Ying Meng, and Takema Fukatsu (Jan. 2010). "Wolbachia as a bacteriocyte-associated nutritional mutualist". In: *Proceedings of the National Academy of Sciences* 107 (2), pp. 769–774. ISSN: 0027-8424. DOI: 10.1073/pnas.0911476107.

Hurst, Gregory D.D. and Crystal L. Frost (May 2015). "Reproductive Parasitism: Maternally Inherited Symbionts in a Biparental World". In: *Cold Spring Harbor*

Perspectives in Biology 7 (5), a017699. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a017699.

Hussa, Elizabeth A. and Heidi Goodrich-Blair (Sept. 2013). "It Takes a Village: Ecological and Fitness Impacts of Multipartite Mutualism". In: *Annual Review of Microbiology* 67 (1), pp. 161–178. ISSN: 0066-4227. DOI: 10.1146/annurev-micro-092412-155723.

Isenberg, Ruth Y., David G. Christensen, Karen L. Visick, and Mark J. Mandel (Aug. 2022). "High Levels of Cyclic Diguanylate Interfere with Beneficial Bacterial Colonization". In: *mBio* 13 (4). Ed. by Nicole Dubilier. ISSN: 2150-7511. DOI: 10.1128/mbio.01671-22.

Itoh, Hideomi, Seonghan Jang, Kazutaka Takeshita, Tsubasa Ohbayashi, Naomi Ohnishi, et al. (Nov. 2019). "Host–symbiont specificity determined by microbe–microbe competition in an insect gut". In: *Proceedings of the National Academy of Sciences* 116 (45), pp. 22673–22682. ISSN: 0027-8424. DOI: 10.1073/pnas.1912397116.

Janczarek, Monika (Nov. 2011). "Environmental Signals and Regulatory Pathways That Influence Exopolysaccharide Production in Rhizobia". In: *International Journal of Molecular Sciences* 12 (11), pp. 7898–7933. ISSN: 1422-0067. DOI: 10.3390/ijms12117898.

Jansen, Gunther, Lena L. Crummenerl, Felix Gilbert, Timm Mohr, Roxana Pfefferkorn, et al. (Nov. 2015). "Evolutionary Transition from Pathogenicity to Commensalism: Global Regulator Mutations Mediate Fitness Gains through Virulence Attenuation". In: *Molecular Biology and Evolution* 32 (11), pp. 2883–2896. ISSN: 0737-4038. DOI: 10.1093/molbev/msv160.

Ji, Ni and Steven W. Flavell (Apr. 2017). "Hydra: Imaging Nerve Nets in Action". In: *Current Biology* 27 (8), R294–R295. ISSN: 09609822. DOI: 10.1016/j.cub.2017.03.040.

Johnson, Winifred M, Harriet Alexander, Raven L Bier, Dan R Miller, Mario E Muscarella, et al. (Nov. 2020). "Auxotrophic interactions: a stabilizing attribute of aquatic microbial communities?" In: *FEMS Microbiology Ecology* 96 (11). ISSN: 1574-6941. DOI: 10.1093/femsec/fiaa115.

Jones, B. W. and M. K. Nishiguchi (Dec. 2006). "Differentially expressed genes reveal adaptations between free-living and symbiotic niches of Vibrio fischeri in a fully established mutualism". In: *Canadian Journal of Microbiology* 52 (12), pp. 1218–1227. ISSN: 00084166. DOI: 10.1139/W06-088.

Juliano, Celina E., Haifan Lin, and Robert E. Steele (Sept. 2014). "Generation of Transgenic Hydra by Embryo Microinjection". In: *Journal of Visualized Experiments* (91). ISSN: 1940-087X. DOI: 10.3791/51888.

Jung, Sascha, Andrew J. Dingley, René Augustin, Friederike Anton-Erxleben, Mareike Stanisak, et al. (2009). "Hydramacin-1, structure and antibacterial activity of a protein from the basal metazoan hydra". In: *Journal of Biological Chemistry* 284 (3), pp. 1896–1905. ISSN: 00219258. DOI: 10.1074/jbc.M804713200.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima (2016). "BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences". In: *Journal of Molecular Biology* 428 (4), pp. 726–731. ISSN: 10898638. DOI: 10.1016/j.jmb.2015.11.006.

Kent, W. James (Apr. 2002). "BLAT –The BLAST -Like Alignment Tool". In: *Genome Research* 12 (4), pp. 656–664. ISSN: 1088-9051. DOI: 10.1101/gr.229202.

Klimovich, Alexander and Thomas Bosch (2024). "Novel technologies uncover novel 'anti'-microbial peptides in Hydra shaping the species-specific microbiome". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 379 (1901). ISSN: 14712970. DOI: 10.1098/rstb.2023.0058.

Klimovich, Alexander, Stefania Giacomello, Åsa Björklund, Louis Faure, Marketa Kaucka, et al. (July 2020). "Prototypical pacemaker neurons interact with the resident microbiota". In: *Proceedings of the National Academy of Sciences* 117 (30), pp. 17854–17863. ISSN: 0027-8424. DOI: 10.1073/pnas.1920469117.

Klopfenstein, D. V., Liangsheng Zhang, Brent S. Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, et al. (2018). "GOATOOLS: A Python library for Gene Ontology analyses". In: *Scientific Reports* 8 (1), pp. 1–17. ISSN: 20452322. DOI: 10.1038/s41598-018-28948-z.

Klug, Marianne, Jakob Weber, and Pierre Tardent (Jan. 1989). "Hemolytic and toxic properties of Hydra attenuata nematocysts". In: *Toxicon* 27 (3), pp. 325–339. ISSN: 00410101. DOI: 10.1016/0041-0101(89)90180-3.

Koh, Clara W.T., Justin S.G. Ooi, Eugenia Ziying Ong, and Kuan Rong Chan (Dec. 2023). "STAGEs: A web-based tool that integrates data visualization and pathway enrichment analysis for gene expression studies". In: *Scientific Reports* 13 (1). ISSN: 20452322. DOI: 10.1038/s41598-023-34163-2.

Koonin, Eugene V. (Dec. 2005). "Orthologs, Paralogs, and Evolutionary Genomics". In: *Annual Review of Genetics* 39 (1), pp. 309–338. ISSN: 0066-4197. DOI: 10.1146/annurev.genet.39.073003.114725.

Kopac, Sarah M. and Jonathan L. Klassen (Oct. 2016). "Can they make it on their own? Hosts, microbes, and the holobiont niche". In: *Frontiers in Microbiology* 7 (OCT). ISSN: 1664302X. DOI: 10.3389/fmicb.2016.01647.

Köster, Johannes and Sven Rahmann (Oct. 2012). "Snakemake–a scalable bioinformatics workflow engine". In: *Bioinformatics* 28 (19), pp. 2520–2522. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/bts480.

Kovačević, Goran, Petra Korać, Davor Želježić, Mirela Sertić Perić, Petra Peharec Štefanić, et al. (July 2024). "Hydra for 21st Century - A Fine Model in Freshwater Research". In: *Water* 16 (15), p. 2114. ISSN: 2073-4441. DOI: 10.3390/w16152114.

Kristensen, D. M., Y. I. Wolf, A. R. Mushegian, and E. V. Koonin (Sept. 2011). "Computational methods for Gene Orthology inference". In: *Briefings in Bioinformatics* 12 (5), pp. 379–391. ISSN: 1467-5463. DOI: 10.1093/bib/bbr030.

Kucuk, R A (July 2020). "Gut Bacteria in the Holometabola: A Review of Obligate and Facultative Symbionts". In: *Journal of Insect Science* 20 (4). Ed. by Oliver Martin, pp. 1–12. ISSN: 1536-2442. DOI: 10.1093/jisesa/ieaa084.

Lachnit, Tim, Laura Ulrich, Fiete M. Willmer, Tim Hasenbein, Leon X. Steiner, et al. (Feb. 2025). "Nutrition-induced changes in the microbiota can cause dysbiosis and disease development". In: *mBio*. Ed. by Joerg Graf. ISSN: 2150-7511. DOI: 10.1128/mbio.03843-24.

Latorre, Amparo and Alejandro Manzano-Marín (Feb. 2017). "Dissecting genome reduction and trait loss in insect endosymbionts". In: *Annals of the New York Academy of Sciences* 1389 (1), pp. 52–75. ISSN: 0077-8923. DOI: 10.1111/nyas.13222.

Lebeer, Sarah, Ingmar J. J. Claes, Tine L. A. Verhoeven, Jos Vanderleyden, and Sigrid C. J. De Keersmaecker (May 2011). "Exopolysaccharides of Lactobacillus rhamnosus GG form a protective shield against innate immune factors in the intestine". In: *Microbial Biotechnology* 4 (3), pp. 368–374. ISSN: 1751-7915. DOI: 10.1111/j.1751-7915.2010.00199.x.

Lee, I-Chiao et al. (July 2016). "Strain-Specific Features of Extracellular Polysaccharides and Their Impact on Lactobacillus plantarum-Host Interactions". In: *Applied and Environmental Microbiology* 82 (13). Ed. by D. W. Schaffner, pp. 3959–3970. ISSN: 0099-2240. DOI: 10.1128/AEM.00306-16.

Li, Xiang Yi, Tim Lachnit, Sebastian Fraune, Thomas C.G. Bosch, Arne Traulsen, et al. (2017). "Temperate phages as self-replicating weapons in bacterial competition". In: *Journal of the Royal Society Interface* 14 (137), pp. 1–7. ISSN: 17425662. DOI: 10.1098/rsif.2017.0563.

Li, Xiang Yi, Cleo Pietschke, Sebastian Fraune, Philipp M. Altrock, Thomas C.G. Bosch, et al. (July 2015). "Which games are growing bacterial populations playing?" In: *Journal of the Royal Society Interface* 12 (108). ISSN: 17425662. DOI: 10.1098/rsif.2015.0121.

Liang, Yukang, Rebecca B. Dikow, Xu Su, Jun Wen, and Zhumei Ren (June 2024). "Comparative genomics of the primary endosymbiont Buchnera aphidicola in aphid hosts and their coevolutionary relationships". In: *BMC Biology* 22 (1), p. 137. ISSN: 1741-7007. DOI: 10.1186/s12915-024-01934-w.

Liang, Zhao-Xun (2015). "The expanding roles of c-di-GMP in the biosynthesis of exopolysaccharides and secondary metabolites". In: *Natural Product Reports* 32 (5), pp. 663–683. ISSN: 0265-0568. DOI: 10.1039/C4NP00086B.

Limoli, Dominique H., Christopher J. Jones, and Daniel J. Wozniak (June 2015). "Bacterial Extracellular Polysaccharides in Biofilm Formation and Function". In: *Microbiology Spectrum* 3 (3). Ed. by Mahmoud Ghannoum, Matthew Parsek, Marvin Whiteley, and Pranab Mukherjee, pp. 1–30. ISSN: 2165-0497. DOI: 10.1128/microbiolspec.MB-0011-2014.

Lin, Baohui, Xiaoling Luo, Yumeng Liu, and Xiaopeng Jin (May 2024). "A comprehensive review and comparison of existing computational methods for protein function prediction". In: *Briefings in Bioinformatics* 25 (4). ISSN: 1467-5463. DOI: 10.1093/bib/bbae289.

Lood, Rolf, Kristofer Wollein Waldetoft, and Pontus Nordenfelt (Oct. 2015). "Localization-Triggered Bacterial Pathogenesis". In: *Future Microbiology* 10 (10), pp. 1659–1668. ISSN: 1746-0913. DOI: 10.2217/fmb.15.89.

Lousada, Marta B., Tim Lachnit, Janin Edelkamp, Ralf Paus, and Thomas C.G. Bosch (May 2022). "Hydra and the hair follicle – An unconventional comparative biology approach to exploring the human holobiont". In: *BioEssays* 44 (5). ISSN: 15211878. DOI: 10.1002/bies.202100233.

Ma, Chenchen, Hongyang Guo, Haibo Chang, Shi Huang, Shuaiming Jiang, et al. (2020). "The effects of exopolysaccharides and exopolysaccharide-producing Lactobacillus on the intestinal microbiome of zebrafish (Danio rerio)". In: *BMC Microbiology* 20 (1), pp. 1–11. ISSN: 14712180. DOI: 10.1186/s12866-020-01990-6.

Ma, Dan, Zhenyu Hao, Rui Sun, Mark Bartlam, and Yingying Wang (2016). "Genome sequence of a typical ultramicrobacterium, Curvibacter sp. strain PAE-UM, capable of phthalate ester degradation". In: *Genome Announcements* 4 (1), pp. 1–2. ISSN: 21698287. DOI: 10.1128/genomeA.01510-15.

Mandel, Mark J, Amy L Schaefer, Caitlin A Brennan, Elizabeth A C Heath-Heckman, Cindy R Deloney-Marino, et al. (2012). "Squid-Derived Chitin Oligosaccharides Are a Chemotactic Signal during Colonization by Vibrio fischeri". In: *Applied and Environmental Microbiology* 78 (13), pp. 4620–4626. DOI: 10.1128/AEM.00377-12.

Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, et al. (Sept. 2005). "Genome sequencing in microfabricated high-density picolitre reactors". In: *Nature* 437 (7057), pp. 376–380. ISSN: 0028-0836. DOI: 10.1038/nature03959.

Matthews, Rowena G. et al. (Dec. 2003). "Cobalamin-Dependent and Cobalamin-Independent Methionine Synthases: Are There Two Solutions to the Same Chemical Problem?" In: *Helvetica Chimica Acta* 86 (12), pp. 3939–3954. ISSN: 0018-019X. DOI: 10.1002/hlca.200390329.

McFall-Ngai, Margaret and Thomas C.G. Bosch (Jan. 2021). "Animal development in the microbial world: The power of experimental model systems". In: *Current Topics in Developmental Biology*. Vol. 141. Academic Press Inc., pp. 371–397. ISBN: 9780128149683. DOI: 10.1016/bs.ctdb.2020.10.002.

McKenzie, Valerie J, Robert M Bowers, Noah Fierer, Rob Knight, and Christian L Lauber (Mar. 2012). "Co-habiting amphibian species harbor unique skin bacterial communities in wild populations". In: *The ISME Journal* 6 (3), pp. 588–596. ISSN: 1751-7362. DOI: 10.1038/ismej.2011.129.

Minten-Lange, Timo and Sebastian Fraune (Jan. 2020). "Hydra and Curvibacter: An intimate crosstalk at the epithelial interface". In: *Cellular Dialogues in the Holobiont*. CRC Press, pp. 79–90. ISBN: 9781000090062. DOI: 10.1201/9780429277375-6.

Miwa, Hiroki and Shin Okazaki (Aug. 2017). "How effectors promote beneficial interactions". In: *Current Opinion in Plant Biology* 38, pp. 148–154. ISSN: 13695266. DOI: 10.1016/j.pbi.2017.05.011.

Moreno-Hagelsieb, Gabriel and Kristen Latimer (2008). "Choosing BLAST options for better detection of orthologs as reciprocal best hits". In: *Bioinformatics* 24 (3), pp. 319–324. ISSN: 13674803. DOI: 10.1093/bioinformatics/btm585.

Mortzfeld, Benedikt M., Jan Taubenheim, Sebastian Fraune, Alexander V. Klimovich, and Thomas C. G. Bosch (Apr. 2018). "Stem Cell Transcription Factor FoxO Controls Microbiome Resilience in Hydra". In: *Frontiers in Microbiology* 9. ISSN: 1664-302X. DOI: 10.3389/fmicb.2018.00629.

Mueller, Konrad and Juan E. González (Jan. 2011). "Complex Regulation of Symbiotic Functions Is Coordinated by MucR and Quorum Sensing in Sinorhizobium

meliloti". In: *Journal of Bacteriology* 193 (2), pp. 485–496. ISSN: 0021-9193. DOI: 10.1128/JB.01129-10.

Murillo-Rincon, Andrea P., Alexander Klimovich, Eileen Pemöller, Jan Taubenheim, Benedikt Mortzfeld, et al. (2017). "Spontaneous body contractions are modulated by the microbiome of Hydra". In: *Scientific Reports* 7 (1), pp. 1–9. ISSN: 20452322. DOI: 10.1038/s41598-017-16191-x.

Nawroth, Janna C., Christoph Giez, Alexander Klimovich, Eva Kanso, and Thomas C.G. Bosch (2023). "Spontaneous body wall contractions stabilize the fluid microenvironment thashapes host–microbe associations". In: *eLife* 12, pp. 1–23. ISSN: 2050084X. DOI: 10.7554/eLife.83637.

Needleman, Saul B. and Christian D. Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins". In: *Journal of Molecular Biology* 48 (3), pp. 443–453. ISSN: 00222836. DOI: 10.1016/0022-2836(70)90057-4.

Netrusov, Alexander I., Elena V. Liyaskina, Irina V. Kurgaeva, Alexandra U. Liyaskina, Guang Yang, et al. (June 2023). "Exopolysaccharides Producing Bacteria: A Review". In: *Microorganisms* 11 (6), p. 1541. ISSN: 2076-2607. DOI: 10.3390/microorganisms11061541.

Nielsen, Claus, Nikolaj Scharff, and Danny Eibye-Jacobsen (1996). "Cladistic analyses of the animal kingdom". In: *Biological Jmml ofthe Linnean Society* 57, p. 385410.

Nikoh, Naruo, Takahiro Hosokawa, Minoru Moriyama, Kenshiro Oshima, Masahira Hattori, et al. (July 2014). "Evolutionary origin of insect – Wolbachia nutritional mutualism". In: *Proceedings of the National Academy of Sciences* 111 (28), pp. 10257–10262. ISSN: 0027-8424. DOI: 10.1073/pnas.1409284111.

Nikoh, Naruo, Takahiro Hosokawa, Kenshiro Oshima, Masahira Hattori, and Takema Fukatsu (2011). "Reductive evolution of bacterial genome in insect gut environment". In: *Genome Biology and Evolution* 3 (1), pp. 702–714. ISSN: 17596653. DOI: 10.1093/gbe/evr064.

Nobrega, Marcelo A. and Len A. Pennacchio (Jan. 2004). "Comparative genomic analysis as a tool for biological discovery". In: *The Journal of Physiology* 554 (1), pp. 31–39. ISSN: 0022-3751. DOI: 10.1113/jphysiol.2003.050948.

Nourabadi, Neda and Michele K. Nishiguchi (Feb. 2021). "pH Adaptation Drives Diverse Phenotypes in a Beneficial Bacterium-Host Mutualism". In: *Frontiers in Ecology and Evolution* 9. ISSN: 2296701X. DOI: 10.3389/fevo.2021.611411.

O'Grady, Eoin (2011). "Burkholderia cenocepacia differential gene expression during host–pathogen interactions and adaptation to the host environment". In:

Frontiers in Cellular and Infection Microbiology 1, p. 15. ISSN: 22352988. DOI: 10.3389/fcimb.2011.00015.

Obeng, Nancy, Anna Czerwinski, Daniel Schutz, Jan Michels, Jan Leipert, et al. (Aug. 2023). "Bacterial c-di-GMP has a key role in establishing host–microbe symbiosis". In: *Nature Microbiology* 8 (10), pp. 1809–1819. ISSN: 2058-5276. DOI: 10.1038/s41564-023-01468-x.

Ortíz-Castro, Randy, Hexon Angel Contreras-Cornejo, Lourdes Macías-Rodríguez, and José López-Bucio (Aug. 2009). "The role of microbial signals in plant growth and development". In: *Plant Signaling & Behavior* 4 (8), pp. 701–712. ISSN: 1559-2324. DOI: 10.4161/psb.4.8.9047.

Otto, M. (2006). "Bacterial Evasion of Antimicrobial Peptides by Biofilm Formation". In: *Antimicrobial Peptides and Human Disease*. Ed. by William M Shafer. Springer Berlin Heidelberg, pp. 251–258. ISBN: 978-3-540-29916-5. DOI: 10.1007/3-540-29916-5_10.

Özkan, Melek, Hilal Yılmaz, Pınar Ergenekon, Esra Meşe Erdoğan, and Mustafa Erbakan (Feb. 2024). "Microbial membrane transport proteins and their biotechnological applications". In: *World Journal of Microbiology and Biotechnology* 40 (2), p. 71. ISSN: 0959-3993. DOI: 10.1007/s11274-024-03891-6.

Parker, Benjamin J. (Aug. 2021). "Mechanisms and Evolution of Heritable Microbial Density in Insect Hosts". In: *mSystems* 6 (4). ISSN: 2379-5077. DOI: 10.1128/msystems.00728-21.

Pearson, William R. (June 2013). "An Introduction to Sequence Similarity ("Homology") Searching". In: *Current Protocols in Bioinformatics* 42 (1). ISSN: 1934-3396. DOI: 10.1002/0471250953.bi0301s42.

Pérez-Mendoza, Daniel, Isabel M. Aragón, Harold A. Prada-Ramírez, Lorena Romero-Jiménez, Cayo Ramos, et al. (Mar. 2014). "Responses to Elevated c-di-GMP Levels in Mutualistic and Pathogenic Plant-Interacting Bacteria". In: *PLoS ONE* 9 (3). Ed. by Jesús Murillo, e91645. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0091645.

Pervez, Muhammad Tariq, Masroor Ellahi Babar, Asif Nadeem, Muhammad Aslam, Ali Raza Awan, et al. (Jan. 2014). "Evaluating the Accuracy and Efficiency of Multiple Sequence Alignment Methods". In: *Evolutionary Bioinformatics* 10, pp. 205–217. ISSN: 1176-9343. DOI: 10.4137/EBO.S19199.

Pietschke, Cleo, Christian Treitz, Sylvain Forêt, Annika Schultze, Sven Künzel, et al. (Oct. 2017). "Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts". In: *Proceedings of the National*

Academy of Sciences 114 (40). ISSN: 0027-8424. DOI: 10.1073/pnas.1706879114.

Pini, Francesco, Marco Galardini, Marco Bazzicalupo, and Alessio Mengoni (Nov. 2011). "Plant-Bacteria Association and Symbiosis: Are There Common Genomic Traits in Alphaproteobacteria?" In: *Genes* 2 (4), pp. 1017–1032. ISSN: 2073-4425. DOI: 10.3390/genes2041017.

Primo, Emiliano, Pablo Bogino, Sacha Cossovich, Emiliano Foresto, Fiorela Nievas, et al. (Oct. 2020). "Exopolysaccharide II Is Relevant for the Survival of Sinorhizobium meliloti under Water Deficiency and Salinity Stress". In: *Molecules* 25 (21), p. 4876. ISSN: 1420-3049. DOI: 10.3390/molecules25214876.

Rathje, Kai, Benedikt Mortzfeld, Marc P. Hoeppner, Jan Taubenheim, Thomas C. G. Bosch, et al. (Mar. 2020). "Dynamic interactions within the host-associated microbiota cause tumor formation in the basal metazoan Hydra". In: *PLOS Pathogens* 16 (3). Ed. by Fanxiu Zhu, e1008375. ISSN: 1553-7374. DOI: 10.1371/journal.ppat.1008375.

Raval, P.K., V. Zimorski, and W.F. Martin (Apr. 2024). "Endosymbiotic Theory". In: *Reference Module in Life Sciences*. Elsevier, pp. 511–517. ISBN: 9780128004265. DOI: 10.1016/B978-0-443-15750-9.00011-2.

Reddy, Bharath and Richard Fields (Apr. 2024). "Performance Analysis of Multiple Sequence Alignment Tools". In: *Proceedings of the 2024 ACM Southeast Conference on ZZZ*. ACM, pp. 167–174. ISBN: 9798400702372. DOI: 10.1145/3603287.3651216.

Reshef, Leah, Omry Koren, Yossi Loya, Ilana Zilber-Rosenberg, and Eugene Rosenberg (Dec. 2006). "The Coral Probiotic Hypothesis". In: *Environmental Microbiology* 8 (12), pp. 2068–2073. ISSN: 1462-2912. DOI: 10.1111/j.1462-2920.2006.01148.x.

Roberts, Larry S, John Jr. Janovy, and Steve Nadler (2009). *Foundations of Parasitology*. 8th ed. McGraw-Hill Higher Education. ISBN: 9780073028279.

Rosenberg, Eugene, Omry Koren, Leah Reshef, Rotem Efrony, and Ilana Zilber-Rosenberg (May 2007). "The role of microorganisms in coral health, disease and evolution". In: *Nature Reviews Microbiology* 5 (5), pp. 355–362. ISSN: 1740-1526. DOI: 10.1038/nrmicro1635.

Rosenberg, Eugene and Ilana Zilber-Rosenberg (Dec. 2018). "The hologenome concept of evolution after 10 years". In: *Microbiome* 6 (1), p. 78. ISSN: 2049-2618. DOI: 10.1186/s40168-018-0457-9.

Saada, Bacem, Tianchi Zhang, Estevao Siga, Jing Zhang, and Maria Malane Magalhães Muniz (June 2024). "Whole-Genome Alignment: Methods, Challenges, and

Future Directions". In: *Applied Sciences* 14 (11), p. 4837. ISSN: 2076-3417. DOI: 10.3390/app14114837.

Sackton, Timothy B, Brian P Lazzaro, Todd A Schlenke, Jay D Evans, Dan Hultmark, et al. (Dec. 2007). "Dynamic evolution of the innate immune system in Drosophila". In: *Nature Genetics* 39 (12), pp. 1461–1468. ISSN: 1061-4036. DOI: 10.1038/ng.2007.60.

Savitskaya, Anna, Jorge Masso-Silva, Imen Haddaoui, and Shymaa Enany (Nov. 2023). "Exploring the arsenal of antimicrobial peptides: Mechanisms, diversity, and applications". In: *Biochimie* 214, pp. 216–227. ISSN: 03009084. DOI: 10.1016/j.biochi.2023.07.016.

Sharma, Arpita, Shashi Prakash Yadav, Dwipjyoti Sarma, and Arunika Mukhopadhaya (2022). "Modulation of host cellular responses by gram-negative bacterial porins". In: *Membrane Proteins*. Ed. by Rossen Donev. Vol. 128. Academic Press, pp. 35–77. DOI: 10.1016/bs.apcsb.2021.09.004.

Shibata, Satoshi, Emily S. Yip, Kevin P. Quirke, Jakob M. Ondrey, and Karen L. Visick (2012). "Roles of the structural symbiosis polysaccharide (syp) genes in host colonization, biofilm formation, and polysaccharide biosynthesis in Vibrio fischeri". In: *Journal of Bacteriology* 194 (24), pp. 6736–6747. ISSN: 10985530. DOI: 10.1128/JB.00707-12.

Siebert, Stefan, Jeffrey A. Farrell, Jack F. Cazet, Yashodara Abeykoon, Abby S. Primack, et al. (July 2019). "Stem cell differentiation trajectories in Hydra resolved at single-cell resolution". In: *Science* 365 (6451). ISSN: 10959203. DOI: 10.1126/science.aav9314.

Sivashankari, Selvarajan and Piramanayagam Shanmughavel (Jan. 2007). "Comparative genomics - A perspective". In: *Bioinformation* 1 (9), pp. 376–378. ISSN: 09738894. DOI: 10.6026/97320630001376.

Skorupska, Anna, Monika Janczarek, Małgorzata Marczak, Andrzej Mazur, and Jarosław Król (Dec. 2006). "Rhizobial exopolysaccharides: genetic control and symbiotic functions". In: *Microbial Cell Factories* 5 (1), p. 7. ISSN: 1475-2859. DOI: 10.1186/1475-2859-5-7.

Smith, T.F. and M.S. Waterman (Mar. 1981). "Identification of common molecular subsequences". In: *Journal of Molecular Biology* 147 (1), pp. 195–197. ISSN: 00222836. DOI: 10.1016/0022-2836(81)90087-5.

Smith, Thomas E. and Nancy A. Moran (Jan. 2020). "Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and Buchnera". In: *Proceedings of the National Academy of Sciences* 117 (4), pp. 2113–2121. ISSN: 0027-8424. DOI: 10.1073/pnas.1916748117.

Steele, Robert E. (Aug. 2002). "Developmental Signaling in Hydra: What Does It Take to Build a "Simple" Animal?" In: *Developmental Biology* 248 (2), pp. 199–219. ISSN: 00121606. DOI: 10.1006/dbio.2002.0744.

Steinegger, Martin and Johannes Söding (Nov. 2017). "MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets". In: *Nature Biotechnology* 35 (11), pp. 1026–1028. ISSN: 1087-0156. DOI: 10.1038/nbt.3988.

Steinert, M., U. Hentschel, and J. Hacker (Jan. 2000). "Symbiosis and Pathogenesis: Evolution of the Microbe-Host Interaction". In: *Naturwissenschaften* 87 (1), pp. 1–11. ISSN: 0028-1042. DOI: 10.1007/s001140050001.

Szymanski, John R. and Rafael Yuste (June 2019). "Mapping the Whole-Body Muscle Activity of Hydra vulgaris". In: *Current Biology* 29 (11), 1807–1817.e3. ISSN: 09609822. DOI: 10.1016/j.cub.2019.05.012.

Tarracchini, Chiara, Gabriele Andrea Lugli, Leonardo Mancabelli, Douwe van Sinderen, Francesca Turroni, et al. (Oct. 2024). "Exploring the vitamin biosynthesis landscape of the human gut microbiota". In: *mSystems* 9 (10). Ed. by Jack A. Gilbert. ISSN: 2379-5077. DOI: 10.1128/msystems.00929-24.

Tatusov, Roman L., Michael Y. Galperin, Darren A. Natale, and Eugene V. Koonin (Jan. 2000). "The COG database: a tool for genome-scale analysis of protein functions and evolution". In: *Nucleic Acids Research* 28.1, pp. 33–36. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.33.

Taubenheim, Jan, Máté Miklós, Jácint Tökölyi, and Sebastian Fraune (2022). "Population Differences and Host Species Predict Variation in the Diversity of Host-Associated Microbes in Hydra". In: *Frontiers in Microbiology* 13 (March), pp. 1–17. ISSN: 1664302X. DOI: 10.3389/fmicb.2022.799333.

Taubenheim, Jan, Doris Willoweit-Ohl, Mirjam Knop, Sören Franzenburg, Jinru He, et al. (Sept. 2020). "Bacteria- and temperature-regulated peptides modulate β-catenin signaling in Hydra". In: *Proceedings of the National Academy of Sciences* 117 (35), pp. 21459–21468. ISSN: 0027-8424. DOI: 10.1073/pnas.2010945117.

Tekaia, Fredj (Jan. 2016). "Inferring Orthologs: Open Questions and Perspectives". In: *Genomics Insights* 9, pp. 17–28. ISSN: 1178-6310. DOI: 10.4137/GEI.S37925.

Theis, Kevin R., Nolwenn M. Dheilly, Jonathan L. Klassen, Robert M. Brucker, John F. Baines, et al. (Apr. 2016). "Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes". In: *mSystems* 1 (2). Ed. by Jack A. Gilbert. ISSN: 2379-5077. DOI: 10.1128/mSystems.00028-16.

Thorne, Jeffrey L (2000). "Models of protein sequence evolution and their applications". In: *Current Opinion in Genetics & Development* 10 (6), pp. 602–605. ISSN: 0959-437X. DOI: https://doi.org/10.1016/S0959-437X(00)00142-8.

Tischler, Anna D. and Andrew Camilli (Aug. 2004). "Cyclic diguanylate (c-di-GMP) regulates Vibrio cholerae biofilm formation". In: *Molecular Microbiology* 53 (3), pp. 857–869. ISSN: 0950-382X. DOI: 10.1111/j.1365-2958.2004.04155.x.

Tommaso, Paolo Di, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo, et al. (Apr. 2017). "Nextflow enables reproducible computational workflows". In: *Nature Biotechnology* 35 (4), pp. 316–319. ISSN: 1087-0156. DOI: 10.1038/nbt.3820.

Ulrich, Laura, Christoph Giez, Leon X. Steiner, Ute Hentschel, and Tim Lachnit (2022). "Adaptive lifestyle of bacteria determines phage-bacteria interaction". In: *Frontiers in Microbiology* 13 (December), pp. 1–16. ISSN: 1664302X. DOI: 10.3389/fmicb.2022.1056388.

Visick, Karen L. and Line M. Skoufos (Feb. 2001). "Two-Component Sensor Required for Normal Symbiotic Colonization of Euprymna scolopes by Vibrio fischeri". In: *Journal of Bacteriology* 183 (3), pp. 835–842. ISSN: 0021-9193. DOI: 10.1128/JB.183.3.835-842.2001.

Voolstra, Christian R., David J. Suggett, Raquel S. Peixoto, John E. Parkinson, Kate M. Quigley, et al. (Oct. 2021). "Extending the natural adaptive capacity of coral holobionts". In: *Nature Reviews Earth & Environment* 2 (11), pp. 747–762. ISSN: 2662-138X. DOI: 10.1038/s43017-021-00214-3.

Vorkapic, Dina, Katharina Pressler, and Stefan Schild (Feb. 2016). "Multifaceted roles of extracellular DNA in bacterial physiology". In: *Current Genetics* 62 (1), pp. 71–79. ISSN: 0172-8083. DOI: 10.1007/s00294-015-0514-x.

Wade, Michael J. (Mar. 2007). "The co-evolutionary genetics of ecological communities". In: *Nature Reviews Genetics* 8 (3), pp. 185–195. ISSN: 1471-0056. DOI: 10.1038/nrg2031.

Wang, Lusheng and Tao Jiang (Jan. 1994). "On the Complexity of Multiple Sequence Alignment". In: *Journal of Computational Biology* 1 (4), pp. 337–348. ISSN: 1066-5277. DOI: 10.1089/cmb.1994.1.337.

Wasilko, Nathan P. et al. (Mar. 2019). "Sulfur availability for Vibrio fischeri growth during symbiosis establishment depends on biogeography within the squid light organ". In: *Molecular Microbiology* 111 (3), pp. 621–636. ISSN: 0950-382X. DOI: 10.1111/mmi.14177.

Wein, Tanita, Tal Dagan, Sebastian Fraune, Thomas C.G. Bosch, Thorsten B.H. Reusch, et al. (2018). "Carrying capacity and colonization dynamics of Curvibacter in the hydra host habitat". In: *Frontiers in Microbiology* 9 (MAR), pp. 1–10. ISSN: 1664302X. DOI: 10.3389/fmicb.2018.00443.

Whitfield, Chris (June 2006). "Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli". In: *Annual Review of Biochemistry* 75 (1), pp. 39–68. ISSN: 0066-4154. DOI: 10.1146/annurev.biochem.75.103004.142545.

Whitfield, Chris and M. Stephen Trent (June 2014). "Biosynthesis and Export of Bacterial Lipopolysaccharides". In: *Annual Review of Biochemistry* 83 (1), pp. 99–128. ISSN: 0066-4154. DOI: 10.1146/annurev-biochem-060713-035600.

Whitfield, Chris, Samantha S. Wear, and Caitlin Sande (Sept. 2020). "Assembly of Bacterial Capsular Polysaccharides and Exopolysaccharides". In: *Annual Review of Microbiology* 74 (1), pp. 521–543. ISSN: 0066-4227. DOI: 10.1146/annurev-micro-011420-075607.

Wiesmann, Christina L, Nicole R Wang, Yue Zhang, Zhexian Liu, and Cara H Haney (Nov. 2023). "Origins of symbiosis: shared mechanisms underlying microbial pathogenesis, commensalism and mutualism of plants and animals". In: *FEMS Microbiology Reviews* 47 (6). ISSN: 1574-6976. DOI: 10.1093/femsre/fuac048.

Williams, Paul (Dec. 2007). "Quorum sensing, communication and cross-kingdom signalling in the bacterial world". In: *Microbiology* 153 (12), pp. 3923–3938. ISSN: 1350-0872. DOI: 10.1099/mic.0.2007/012856-0.

Winans, Nathan J., Alec Walter, Bessem Chouaia, John M. Chaston, Angela E. Douglas, et al. (Sept. 2017). "A genomic investigation of ecological differentiation between free-living and Drosophila -associated bacteria". In: *Molecular Ecology* 26 (17), pp. 4536–4550. ISSN: 0962-1083. DOI: 10.1111/mec.14232.

Wittlieb, Jörg, Konstantin Khalturin, Jan U. Lohmann, Friederike Anton-Erxleben, and Thomas C. G. Bosch (Apr. 2006). "Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis". In: *Proceedings of the National Academy of Sciences* 103 (16), pp. 6208–6211. ISSN: 0027-8424. DOI: 10.1073/pnas.0510163103.

Yip, Emily S., Brian T. Grublesky, Elizabeth A. Hussa, and Karen L. Visick (Sept. 2005). "A novel, conserved cluster of genes promotes symbiotic colonization and σ54-dependent biofilm formation by Vibrio fischeri". In: *Molecular Microbiology* 57 (5), pp. 1485–1498. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2005.04784.x.

Yu, Guangchuang, Li Gen Wang, Yanyan Han, and Qing Yu He (2012). "clusterProfiler: An R package for comparing biological themes among gene clusters". In: *OMICS A Journal of Integrative Biology* 16 (5), pp. 284–287. ISSN: 15362310. DOI: 10.1089/omi.2011.0118.

Zamioudis, Christos and Corné M. J. Pieterse (Feb. 2012). "Modulation of Host Immunity by Beneficial Microbes". In: *Molecular Plant-Microbe Interactions®* 25 (2), pp. 139–150. ISSN: 0894-0282. DOI: 10.1094/MPMI-06-11-0179.

Zaneveld, Jesse, Peter J Turnbaugh, Catherine Lozupone, Ruth E Ley, Micah Hamady, et al. (Feb. 2008). "Host-bacterial coevolution and the search for new drug targets". In: *Current Opinion in Chemical Biology* 12 (1), pp. 109–114. ISSN: 13675931. DOI: 10.1016/j.cbpa.2008.01.015.

Zeng, M.Y., N. Inohara, and G. Nuñez (Jan. 2017). "Mechanisms of inflammation-driven bacterial dysbiosis in the gut". In: *Mucosal Immunology* 10 (1), pp. 18–26. ISSN: 19330219. DOI: 10.1038/mi.2016.75.

Zhao, Min'an, Jiayi Chu, Shiyao Feng, Chuanhao Guo, Baigong Xue, et al. (Aug. 2023). "Immunological mechanisms of inflammatory diseases caused by gut microbiota dysbiosis: A review". In: *Biomedicine & Pharmacotherapy* 164, p. 114985. ISSN: 07533322. DOI: 10.1016/j.biopha.2023.114985.

Zilber-Rosenberg, Ilana and Eugene Rosenberg (Aug. 2008). "Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution". In: *FEMS Microbiology Reviews* 32 (5), pp. 723–735. ISSN: 1574-6976. DOI: 10.1111/j.1574-6976.2008.00123.x.

Zückert, Wolfram R. (Aug. 2014). "Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1843 (8), pp. 1509–1516. ISSN: 01674889. DOI: 10.1016/j.bbamcr.2014.04.022.

Zug, Roman and Peter Hammerstein (Feb. 2015). "Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts". In: *Biological Reviews* 90 (1), pp. 89–111. ISSN: 1464-7931. DOI: 10.1111/brv.12098.

10 Acknowledgements

Zuallererst möchte ich meinen Betreuern, Prof. Dr. Ilka M. Axmann und Prof. Dr. Sebastian Fraune, für die fachliche Unterstützung, die wertvollen Anregungen und Gespräche sowie die Möglichkeit danken, meine wissenschaftlichen Arbeiten in ihren Laboren durchzuführen. Ich bin zutiefst dankbar für die schöne Zeit, all das neu Gelernte und die unglaublich vielseitigen wissenschaftlichen Austausche und Bekanntschaften, die ich während meiner Doktorarbeit an euren Instituten erlebt habe. Ich freue mich auf weitere wissenschaftliche Zusammenarbeit.

Ich möchte mich bei den SynMibi und OrgInt Teams für die kollegiale Unterstützung und die zahlreichen Diskussionen und Gespräche bedanken. Während besonders anstrengenden und herausfordernden Phasen haben mir die offenen Ohren und aufmunternden Gespräche den Arbeitsalltag erleichtert und mich motiviert weiterzumachen. Ohne euren Input, eure geduldigen Erklärungen sowie eure Bereitschaft, mir immer wieder zuzuhören, wäre ich niemals so weit gekommen.

Zuweilen kann Forschen anstrengend und nervenaufreibend sein. Meinen liebsten Menschen, die mich auf diesem Weg begleitet, meinen Zweifeln und Sorgen zugehört und mir über alle Probleme hinweg stets das Gefühl gegeben haben, verstanden zu werden, möchte ich besonders danken. Danke, Mama, Papa und Lena.

11 Statement on AI Assistance

In the preparation of this thesis, AI-based tools were utilized to enhance language clarity, refine grammar, and improve the readability of the text. The core ideas, research, and arguments presented in this work are entirely my own, and AI assistance was limited to stylistic improvements and minor corrections. All critical thinking, analysis, and conclusions remain the product of my own academic effort.

I have ensured that the use of AI aligns with my institution's academic integrity policies and does not compromise the originality or ethical standards of this research.

12 Eidestattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Ferner erkläre ich, dass ich in keinem anderen Dissertationsverfahren mit oder ohne Erfolg versucht habe, diese Dissertation einzureichen.

Düsseldorf, den _____

Lukas Becker: _____