Faculty of Medicine

SYSTEMS BIOLOGY

Department of Human and Medical Genetics

Petr Sharygin, year 2, group 1

Master's Thesis

Metabolinių Sąveikų Modeliavimas Melsvabakterių Mikrobiome

Modelling Metabolic Interactions Within the Cyanobacteria Microbiome

Supervisor Assoc. Prof. Dr. Sigitas Šulčius

Head of the department or clinic Prof. Dr. Algirdas Utkus

Vilnius, 2026

Student's email petr.sharygin@mf.stud.vu.lt

CONTENTS

LIST OF ABBREVIATIONS **5**

INTRODUCTION **6**

LAY DESCRIPTION **8**

LITERATURE REVIEW **9**

1. Cyanobacteria and their microbiomes 9

2. *Aphanizomenon flos-aquae* 9

3. Metabolic interactions in microbial communities 10

3.1 Cross-feeding 11

3.2 Metabolic specialization and genome streamlining. 12

4. Cross-feeding in cyanobacterial microbiomes. 13

4.1 Carbon-nitrogen coupling. 14

4.2 Vitamin exchange and production. 14

5. Genome-scale metabolic modeling 15

5.1 Genome-scale metabolic reconstruction and flux balance analysis 16

5.2 Community modeling 18

6. Computational approaches for metagenome reconstruction and analysis 20

6.1 Long-read metagenomic sequencing 21

6.2 Metagenome-assembled genomes 21

6.3 Genome-based taxonomy 22

6.4 Functional annotation with Hidden Markov Models 23

6.5 Pangenome reconstruction and comparative genomics 23

7. Summary and research gap 24

METHODS **26**

1. Experimental organisms 26

2. Metagenomic assembly and MAG reconstruction 26

2.1 Read quality control 26

2.2 Long-read metagenomic assembly 26

2.3 Read mapping 26

2.4 Metagenomic genome binning 27

2.5 Bin-specific read extraction and polishing 27

2.6 Quality assessment 27
2.7 Taxonomic classification 27
3. Microbial network reconstruction and analysis 28
4. Individual genome-scale metabolic model reconstruction 28
5. Community-scale metabolic model reconstruction 30
6. *Aphanizomenon flos-aquae* pangenome analysis 31
6.1 Dataset compilation 31
6.2 Pangenome reconstruction 31
6.2.2 Functional annotation 32
6.2.3 Pangenome computation 32
6.2.4 Pangenome analysis and classification 32
6.2.5 Rarefaction curves and Heaps' Law fitting 32
7. Code availability 33
RESULTS **34**
1. Metagenomic assembly and MAG reconstruction 34
1.1 AFA_2012_KM_D3 assembly 34
2. Taxonomic classification of reconstructed MAGs 36
3. Geographic structure of strain-associated microbiomes 38
4. Individual genome-scale metabolic-model reconstruction 40
5. Community-scale metabolic interaction analysis 41
6. Aphanizomenon flos-aquae pangenome analysis 44
6.1 Pairwise ANI analysis 45
6.2 Pangenome composition 47
6.3 Rarefaction analysis and Heaps' Law fitting 49
6.4 Functional characterization of unique *Aphanizomenon flos-aquae* gene clusters 49
DISCUSSION **51**
1. Reconstruction of genome-scale metabolic models 51
2. Integration of individual genome-scale metabolic models into community-level reconstructions 52
3. Shared and strain-specific metabolic interaction patterns across microbiomes 53
4. Microbiome composition and AFA genomic variation 54
CONCLUSIONS **58**
ACKNOWLEDGEMENTS **59**

REFERENCES **60**

SUMMARY **70**

SUMMARY IN LITHUANIAN **72**

APPENDIX 1: FULL METABOLIC INTERACTIONS NETWORK OF AFA_2012_KM_D3 ..... **74**

APPENDIX 2: MICROBIOME COMPOSITION OF STRAINS AFA_2012_KM_D3, AFAB4C, AFAB4F, AND AFAB10C. **75**

LIST OF ABBREVIATIONS

AFA – *Aphanizomenon Flos-aquae*

ANI – Average Nucleotide Identity

COG – Clusters of Orthologous Groups

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

DNA - Deoxyribonucleic Acid

FBA – Flux Balance Analysis

GEM – Genome-scale Metabolic Model

GTDB – Genome Taxonomy Database

HCB - Harmful Cyanobacterial Bloom

HMM – Hidden Markov Model

KEGG - Kyoto Encyclopedia of Genes and Genomes

KO – KEGG Orthology

MAG – Metagenome Assembled Genome

MCL - Markov Cluster Algorithm

NCBI - National Center for Biotechnology Information

ONT – Oxford Nanopore Technology

PERMANOVA - Permutational Multivariate Analysis of Variance

RNA - Ribonucleic Acid

INTRODUCTION

Cyanobacteria are oxygenic photosynthetic prokaryotes with major roles in primary production and biogeochemical cycling in aquatic ecosystems (Garcia-Pichel *et al.*, 2020). In freshwater environments, bloom-forming cyanobacteria can strongly affect nutrient dynamics, food webs, and water quality, particularly under eutrophic and warming conditions (Huisman *et al.*, 2018). Many bloom-forming taxa are filamentous and heterocyst-forming, allowing them to combine photosynthesis with nitrogen fixation and to persist under nitrogen-limited conditions (Arévalo *et al.*, 2021).

Cyanobacteria do not occur alone. They are surrounded by a chemically structured microenvironment, referred to as the cyanosphere, in which released organic compounds and other metabolites support interactions with heterotrophic bacteria (Fuster *et al*., 2023). These interactions can include exchange of carbon compounds, vitamins, cofactors, and other metabolites, as well as detoxification, signaling, and nutrient remineralization (Pacheco, Moel and Segrè, 2019).

Aphanizomenon flos-aquae is a bloom-forming filamentous cyanobacterium associated with freshwater systems and is a relevant model for studying cyanobacterial microbiomes (Cirés & Ballot, 2016). Its associated bacteria are expected to contribute to nutrient cycling, metabolism of host-derived compounds, and general community functioning, but the interaction structure of these communities remains poorly resolved. Most studies of cyanobacterial microbiomes remain focused on community composition rather than on explicit predictions of metabolic exchange.

Recent advances in shotgun metagenomics allow the recovery of metagenome-assembled genomes (MAGs) from mixed microbial communities and have made it possible to investigate microbiomes at genome resolution (Grossart *et al.*, 2020). Genome-resolved metagenomics is particularly useful for host-associated microbiomes, where many community members are difficult to isolate individually.

Recovered genomes can be used to reconstruct genome-scale metabolic models (GEMs), which describe the metabolic reactions encoded by a genome and allow prediction of growth, substrate use, and metabolic capabilities under defined environmental conditions (O'Brien, Monk and Palsson, 2015). Extending these models to the community level makes it possible to predict

candidate cross-feeding interactions and shared metabolic dependencies among microbiome members (Diener, Gibbons and Resendis-Antonio, 2020). For cyanobacterial microbiomes, this is particularly relevant because many expected interactions involve metabolites that cannot be inferred from taxonomy alone, including amino acids, cofactors, vitamins, and central carbon compounds. Community metabolic modelling therefore provides a way to examine the potential interaction network underlying cyanobacterial-bacterial associations.

In addition to microbial community composition, the genotype of the cyanobacterial host may influence microbiome assembly and interaction structure. Strains within the same cyanobacterial species can differ substantially in genome content, transport systems, regulatory genes, and secondary metabolite biosynthesis potential, all of which may affect associated bacteria directly or indirectly (McInerney, McNally and O'Connell, 2017). If host strains differ in metabolic capabilities or in the compounds they release into their surrounding environment, those differences may contribute to variation in both microbiome composition and predicted cross-feeding networks.

Host genomic variation can be examined using pangenome analysis, which distinguishes between the core genome, shared by all strains, and the accessory genome, which varies among strains (Matthews *et al.*, 2024). This framework is useful because differences in gene content provide a direct basis for differences in reconstructed metabolic potential.

The main aim of this research is to characterize metabolic interactions in *Aphanizomenon flos-aquae* microbiomes and evaluate the role of cyanobacterial genotypes in structuring community composition and metabolic cross-feeding networks.
To achieve this, the following objectives were set:

1. Reconstruct genome-scale metabolic models for *Aphanizomenon flos-aquae* strains and co-occurring heterotrophic bacteria.
2. Integrate individual genome-scale metabolic models into community-level reconstructions.
3. Assess how genotypic variation among *Aphanizomenon flos-aquae* strains influences microbiome composition and predict metabolic interactions.

LAY DESCRIPTION

Cyanobacteria are tiny organisms that live in water and produce oxygen through photosynthesis, the same way that plants do. While they play an important role in nature, some species can multiply rapidly and form harmful blooms, thick green layers on the water surface that can make water unsafe for both humans and aquatic life. One such species is *Aphanizomenon flos-aquae*, which forms blooms in lakes and seas around the world, including Lithuania and the Baltic Sea.

These cyanobacteria, however, do not live alone. They are surrounded by a diverse community of other bacteria that live very close to them, in a region called the "cyanosphere." These neighboring bacteria and host cyanobacteria exchange nutrients with each other, for example, cyanobacteria produce certain sugars through photosynthesis and can provide these to other bacteria, while in return receiving nitrogen or vitamins that help them grow. Understanding these exchanges is extremely important because they can influence bloom dynamics.

However, studying these interactions in the laboratory is difficult, expensive, and time-consuming. To overcome this challenge, this research uses computer models that simulate the metabolism of each bacterium based on its genetic code. By analyzing the genomes of five different strains of *Aphanizomenon flos-aquae* and the bacteria living with them, we reconstructed virtual metabolic networks that predict which nutrients each organism can produce and which it needs from others.

This work aims to reveal how genetic differences between cyanobacterial strains influence the composition of their surrounding bacterial community and the metabolic interactions that occur. By predicting these nutrient exchanges, we can better understand what drives bloom formation and potentially develop strategies to monitor or manage harmful cyanobacterial blooms in the future. Ultimately, this research contributes to protecting water quality and public health.

LITERATURE REVIEW

1. Cyanobacteria and their microbiomes

Cyanobacteria are a vastly diverse group of photosynthetic prokaryotes, known to be the first prokaryotes to evolve oxygenic photosynthesis, using the energy obtained from light to split water molecules, and producing oxygen as a byproduct of this reaction (Planavsky *et al.*, 2014). This evolutionary event has fundamentally transformed our planet as we know it, leading to the rapid oxygenation of Earth's atmosphere during the early Proterozoic, 2.5-2.3 billion years ago, in what is known as the Great Oxidation Event (Schirrmeister, Gugger and Donoghue, 2015). In addition to currently accounting for up to 25% of all photosynthesis in marine ecosystems, cyanobacteria are a major primary producer, playing an important role in cycling of nitrogen, carbon, sulfur, iron, and redox-sensitive trace metals (Sánchez-Baracaldo *et al.*, 2022). Furthermore, while the exact mechanism is still being debated, it is undoubtedly known that the existence and ecological prevalence of cyanobacteria has directly led to the evolution of chloroplasts, and eukaryotic phototrophs in general (Sato, 2021).

Cyanobacteria inhabit a vast and diverse range of habitats, and can potentially form dense, bloom-forming populations, that greatly impact the structure and composition of aquatic ecosystems, of which they are part of (Huisman *et al.*, 2018). These blooms develop as a multi-species microbial consortium, with heterotrophic bacteria growing in close association with the cyanobacteria, as part of the so-called cyanosphere (Fuster *et al.*, 2023). This narrow zone adjacent to the cyanobacterial cells tends to exhibit great microbial diversity and abundance, which, through cross-feeding or other metabolic interactions, can either promote or inhibit the growth of the entire system (Amin *et al.*, 2015). The importance of these interactions in the case of cyanobacteria, however, remains under-investigated, and the understanding of the precise processes that occur within the cyanosphere could bring novel insights into understanding and predicting bloom occurrence and dynamics (Louati *et al.*, 2015).

2. *Aphanizomenon flos-aquae*

The genus *Aphanizomenon*, within the Nostocales order, comprises filamentous, often bloom-forming cyanobacteria that have been frequently implicated in harmful cyanobacterial blooms (HCBs), especially in eutrophic freshwater systems (Cirés & Ballot, 2016). These blooms can severely impair water quality, disrupt aquatic food webs, and limit the use of water bodies for

recreation, aquaculture, or drinking supply due to biomass accumulation, anoxia, and potential toxin production. *Aphanizomenon flos-aquae* (AFA) is a globally distributed species, found in freshwater ecosystems across Europe, Asia, and North America. It is a Gram-negative, filamentous nitrogen-fixing obligate phototroph, found in brackish and freshwater ecosystems, and often dominates bloom events (Komárek and Komárková, 2006). Morphologically, its most characteristic feature is the ability to form bundles of filaments, called fascicles, that can be up to 2 cm in length, and visible to the naked eye (Carmichael, Drapeau and Anderson, 2000). In addition to its vegetative cells, it can be form two types of specialized cells: heterocysts, which allow for nitrogen fixation, and akinetes: dormant cells that allow for AFAs continued survival under unfavorable environmental conditions (Scoglio et al., 2024). *Aphanizomenon flos-aquae* blooms display a strong seasonal pattern, typically occurring from late spring to early autumn in the Northern hemisphere, usually peaking around July (Kononen and Nômmann, 1992).

Despite its wide spread, the taxonomy of the *Aphanizomenon* genus, and AFA specifically, remains unclear, due to its polyphyletic nature (Rajaniemi *et al.*, 2005a). This taxonomic uncertainty has implications for ecological and genomic research, as strains classified as AFA may differ substantially in physiology, toxin production, and metabolic capabilities (Strunecký, Ivanova and Mareš, 2023).

3. Metabolic interactions in microbial communities

Organisms in the microbial world rarely exist in isolation, instead forming complex dynamic networks of interactions among species, forming a web of interconnected metabolisms (Faust and Raes, 2012). In fact, failure to account for the full range of metabolic interactions for any given microbe can heavily impede our understanding of not only its ecological role, but even its basic physiology (Traxler *et al.*, 2013). The access to a particular nutrient can be a determining factor in whether a certain microbe is able to persist and thrive in any given ecological niche (Johnson *et al.*, 2012). As such, metabolic cooperation between species in symbiotic relationships can provide them metabolic products that neither of them would be able to have access to otherwise, which lets the bacteria involved occupy unique niches, which they would not be able to inhabit individually (Johnson *et al.*, 2012). These nutritional interactions that define a microbe's ability to carry out certain metabolic processes can be broadly divided into three categories (Figure 1). Cross-feeding, or sharing metabolites between microbes, relationships specifically are as complex,

as they are incredibly important to multiple ecological mechanisms, such as community establishment and dynamics, resilience to external perturbation, and many more (Culp and Goodman, 2023). Despite this, the underlying mechanisms still remain understudied, specifically when it comes to highly spatially and temporally dynamic aquatic microbial communities (Ma *et al.*, 2022). As such, progress in this area will lead to important advancements in multiple areas of science, including microbiology, evolutionary biology, ecology, and even nutrition.

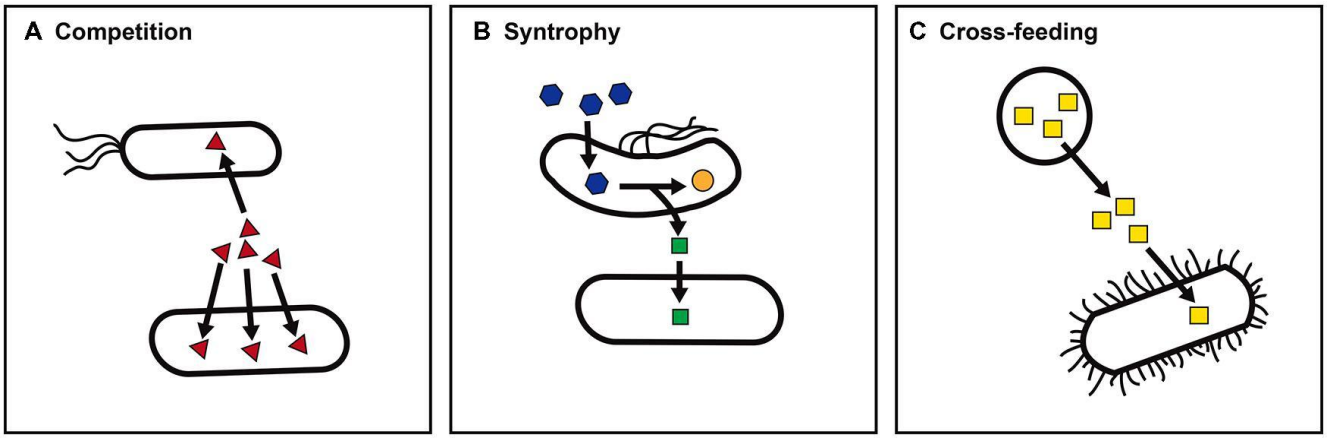


Figure 1: Nutritional interactions between microbes. Adapted from "Nutrient cross-feeding in the microbial world" by E. C. Seth & M. E. Taga (2014). *Frontiers in microbiology*, 5, 350

3.1 Cross-feeding

Even though bacterial organisms fundamentally compete for nutrients, an evolutionary trend toward metabolic specialization has generated substantial variation in feeding preferences. This diversification increases the likelihood that metabolic by-products excreted by one species or strain serve as essential nutrients for another (Pacheco, Moel and Segrè, 2019). Despite all the available information and studies conducted on microbial cross-feeding, the language surrounding it and exact terminology can remain varied and vague. One author defines it as simply "an interaction between bacterial strains in which molecules resulting from the metabolism of one strain are further metabolized by another strain" (Smith *et al.*, 2019). While useful, such a definition could be seen as overly narrow, and failing to capture the full range of interactions observed in microbial communities.

A broader conceptual framework proposed by Fritts et al. requires that cross-feeding minimally requires four conditions to be met (2021). First, material must be transferred from a producer to a recipient; this material may include not only metabolites but also subatomic particles such as electrons or protons, and potentially even shared electrochemical energy rather than discrete molecules. Second, the transferred material must be assimilated by the recipient or directly participate in energy transformation in either the producer or recipient, thereby excluding interactions such as detoxification or signal degradation that do not constitute "feeding" in a metabolic sense (Mataigne *et al.*, 2021). Third, the assimilation or energy gain must measurably alter the fitness of the producer and/or recipient, even if such effects are subtle or only detectable over long timescales. Finally, cross-feeding must occur between different species or between genotypically or phenotypically distinct populations, ensuring that the interaction reflects genuine metabolic interdependence rather than trivial exchange within a homogeneous population (Ackermann, 2015). Together, these criteria provide a more inclusive and mechanistically grounded definition of cross-feeding that better reflects the diversity and ecological significance of microbial metabolic interactions.

3.2 Metabolic specialization and genome streamlining.

Since microbial organisms generally reside in extremely diverse ecosystems and form complex communities with a large number of large species, there have been certain evolutionary adaptations to this lifestyle.

Metabolic specialization refers to a certain phenomenon, when organisms narrow their functional metabolic capabilities, focusing on a certain subset of pathways and rely on other members of the community to produce the complimentary metabolites (Kreft, Griffin and González-Cabaleiro, 2020). Such environmental adaptation through loss of function can occur through several distinct mechanisms, and result in both individual and community-level advantages (Hottes *et al.*, 2013).

One of such mechanisms responsible for the evolution of metabolic interdependence is genome streamlining (Hessen *et al.*, 2010). In many species of free-living bacteria, particularly those occupying relatively stable environments, evolutionary selection can favor extreme metabolic efficiency achieved through gene loss, since cutting down on the size of the genome reduces the cellular costs of DNA replication, and other related processes, conferring a certain advantage in oligotrophic environments (M.-Y. Chen *et al.*, 2021). An interesting variant of such loss of

function, however, can occur specifically in communities heavily relying on cross-feeding interactions between species (Morris, Lenski and Zinser, 2012). When a specific metabolic function is particularly costly to maintain for an organism, yet it provides a public good for the community as a whole, natural selection can favor mutants that lose some other metabolic pathways, which can be compensated by other members of the ecosystem (Coyte, Schluter and Foster, 2015). Over a period of time, this results in the complete or partial loss of unnecessary biosynthetic genes, leading to obligate metabolic dependencies, as these organisms become auxotrophic for the metabolites they can no longer synthesize, but, notably, not leading to a qualitive change in community-level metabolic production (Moran, McLaughlin and Sorek, 2009).

This is particularly important for highly spatially structured microenvironments, such as microbial mats and films, and, specifically, the mutualistic micro-niche found around the cyanobacterial heterocyst, where the specific nutrient and oxygen gradients especially favor phenotypic specialization (van Gestel, Vlamakis and Kolter, 2015). Studies centering on denitrifying microbial systems have shown that there is an extremely widespread division of labor within them, with each individual organism only performing a specific part of the denitrification process, which reduces the metabolic burden on any individual microbe, without losing any sort of functionality at the community level (Roothans, van Loosdrecht and Laureni, 2025). This is not universally stable, however, and many factors can influence this process, meaning that we cannot generalize this phenomenon to be occurring in any given specialized microbial community without further investigation of that particular system (Mridha and Kümmerli, 2022). As such, in what ways, and how strongly metabolic specialization and labor division influence the structures of freshwater cyanobacterial microbiomes, such as *Aphanizomenon flos-aquae,* remains an open question.

4. Cross-feeding in cyanobacterial microbiomes.

The microenvironment immediately around the cyanobacterial cells, known as the cyanosphere, tends to be a hotspot for metabolic exchange due to the production of a wide array of compounds and metabolites, which contributes to the functioning and growth of related heterotrophic bacteria, leading to an increased microbial abundance and diversity in such systems (Fuster *et al.*, 2023). Understanding the specific mechanisms and metabolites involved in cyanosphere cross-feeding is essential for predicting bloom dynamics, nutrient cycling, and community responses to environmental change.

4.1 Carbon-nitrogen coupling.

One of the most fundamental interactions which led to the evolution of such communities, is the production of photosynthetically fixed carbon in exchange for nitrogen compounds (Gobler *et al.*, 2016). As obligate photoautotrophs, cyanobacteria convert atmospheric CO_2 into organic carbon through the Calvin cycle, releasing a substantial fraction as extracellular metabolites including simple sugar, organic acids, and photorespiratory byproducts such as glycolate (Raines, 2003). Studies show, that these mutualistic "C for N" exchanges occur in a number of communities, including ones based around *M. vaginatus*, a soil cyanobacterium which is unable to fix dinitrogen, and has to receive it from the related heterotrophic bacteria (Nelson, Giraldo-Silva and Garcia-Pichel, 2021). In nitrogen-fixing cyanobacteria, which *Aphanizomenon* also is, after receiving these nitrogenous compounds from heterotrophic bacteria, specialized heterocyst cells reduce them to ammonia, which is then released into the immediate microenvironment to the benefit of both the heterotrophs, as well as the vegetative cells of the same cyanobacterium (Rolff, Almesjö and Elmgren, 2007). Notably, in some cases the introduction of externally provided nitrogen results in the loss of these heterotrophs from the community, further confirming the obligate nature of this metabolically specialized relationship (Nelson and Garcia-Pichel, 2021).

4.2 Vitamin exchange and production.

Vitamin exchange represents another critical point of interdependence in the cyanosphere. Particularly important group of cofactors is the B group vitamins, which are essential for many enzymatic reactions, including cellular energy production. However, their biosynthesis itself is immensely energetically costly, leading to the metabolic pathways implicated in its production being divided across multiple species within these communities (Santo *et al.*, 2022).

4.2.1 Cobalamin

For example, cobalamin, more commonly known as vitamin $B_{12,}$ is required by the vast majority of both prokaryotes as well as eukaryotes for key metabolic functions, however, surveys of marine phytoplankton have discovered that over half of all investigated species are auxotrophic for it, creating dependence on the few microbial species capable of its biosynthesis (Croft *et al.*, 2005).

However, different functional forms of the molecule exist. Cobalamin structurally consists of a corrin ring with a cobalt atom in the center, which can form to additional bonds, to a group below and above the ring, with these groups influencing how it functions and what enzymes it can bind

to (Markle and Greenway, 1996). Most cyanobacteria produce a structural variant known as pseudocobalamin instead, which has reduced bioavailability for most eukaryotes (Helliwell *et al.*, 2016). Studies have also shown that the production of this vitamin is highly dependent on metabolic division of labor, leading to increased stability of microbial communities capable of its production (Grant *et al.*, 2014).

4.2.2 Biotin

Another example of the importance of vitamin exchange in cyanobacterial communities is vitamin B_7. Like B_{12}, biotin is only produced de novo by a limited subset of marine microbes, despite it's universal necessity in many reactions (Sirithanakorn and Cronan, 2021).

A significant percentage of cyanosphere-associated heterotrophic bacteria are auxotrophic for biotin (Wienhausen *et al.*, 2022). Many cyanobacterial taxa are known to be primary producers of this vitamin, and it can be secreted both actively and passively by the cyanobacterium during its growth (Gornicki, Scappino and Haselkorn, 1993). The most dramatic release of biotin, however, can be observed during bloom collapses, when lysed cyanobacterial cells release intracellular vitamin pools, which is also known to sometimes trigger secondary blooms of biotin-auxotrophic bacteria or eukaryotic algae, significantly impacting the structure and dynamics of the ecosystem (Sakaki *et al.*, 2020).

However, for freshwater bloom systems biotin, and other vitamin cross-feeding dynamic remain almost entirely uncharacterized (Fuster *et al.*, 2023). While genomic data does suggest that bloom-forming cyanobacteria, including *Aphanizomenon flos-aquae,* likely possess complete, or nearly-complete, vitamin biosynthesis pathways, potentially providing them for the cyanosphere as well, the exact mechanisms and implications of this on the dynamics and behavior of the microbial system remains to be determined (Louati *et al.*, 2015).

5. Genome-scale metabolic modeling

True understanding of these cross-feeding mechanisms requires a more diligent approach than simple correlative observation and individual-species level predictions. However, traditional experimental approaches face a potentially insurmountable number of limitations when trying to analyze such systems, from practical difficulties in microbial isolation and co-culture experiments, to a lack of methods capable of capturing the complex webs of metabolic interactions in necessary

resolution (Zengler and Palsson, 2012). Genome-scale metabolic models (GEMs/GSMs) offer a systems-level approach to investigating these interactions by using genomic data to infer and predict the metabolic phenotype of an organism (Zhang and Hua, 2016). This approach, on its own, still lacks the necessary capacity to predict metabolic interactions on a community level. This, however, is being increasingly addressed in recent years, with community-scale modelling frameworks extending this predictive power from single-organism level to niches and entire ecosystems (Diener, Gibbons and Resendis-Antonio, 2020). However, the application of these methods to phototrophic microbial communities, and bloom-forming cyanobacteria specifically remains extremely limited despite their immense ecological importance.

5.1 Genome-scale metabolic reconstruction and flux balance analysis

GEMs are comprehensive mathematical abstractions, aiming to represent the entire metabolic network of an organism, usually reconstructed from genomic reads, and curated using known experimentally validated biochemical knowledge (McCloskey, Palsson and Feist, 2013). These models integrate all known metabolic pathway an organism is predicted to possess, based on its genomic sequence, and organizes them into a complete network that is able to predict metabolic phenotypes, including growth rates, nutritional requirements of an organism, as well as its secretion profile under given environmental conditions (O'Brien, Monk and Palsson, 2015).

GEM construction generally starts with genome annotation, identifying any protein-coding genes and their products, based on homology with known enzymes from curated databases (Gu *et al.*, 2019). Each detected enzyme is then associated with a biochemical reaction, which are compiled into a stochiometric matrix, used to mathematically represent the whole network. (Figure 2) (Baart and Martens, 2012).

Reaction Network

R1: 2A + B → C + 3D

R2: A + 3B → C + E

R3: A → 2B + D + E

R4: 4B → D + A

R5: D + 2B → C + 2E

R6: C + 4E → 3B + D

Stoichiometric Matrix

	R1	R2	R3	R4	R5	R6
A	-2	-1	-1	1	0	0
B	-1	-3	2	-4	-2	3
C	1	1	0	0	1	-1
D	3	0	1	1	-1	1
E	0	1	1	0	2	-4

Figure 2: Example of the stoichiometric matrix for a hypothetical reaction network. Adapted from Singh, et al. (2013). Flux-based classification of reactions reveals a functional bow-tie organization of complex metabolic networks. *Physical review*. E, Statistical, nonlinear, and soft matter physics. 87. 052708. 10.1103/PhysRevE.87.052708.

In such a matrix, rows represent the metabolites, columns the reactions, and the numerical entries are the stochiometric coefficients, or how many molecules of each metabolite are produced or consumed in any given reaction (Opdam *et al.*, 2017).

Once constructed, such a model is typically analyzed using constraint-based modeling approaches, usually some form of flux balance analysis (FBA) (Lakshmanan *et al.*, 2014). The most important assumption – and drawback – of FBA is that it operates under a metabolic steady state, which is mathematically expressed as $S \cdot v = 0$, where S is the stoichiometric matrix and v is a vector of metabolic fluxes (Kaste and Shachar-Hill, 2024). In simpler terms, this means that no metabolite is accumulated or lost over a period of time. Unfortunately, as a metabolic networks generally has more reactions than metabolites within it, such a system of equations contains a near-infinite amount of possible solutions (Antoniewicz, 2015). This is most commonly addressed by defining an objective function that is to be maximized for a system of equations, usually biomass production, which tends to provide a more biologically meaningful solution for the model, but comes with its own unique set of challenges and drawbacks (Basile *et al.*, 2020).

This method has been gaining popularity in recent years due to the recent development and advances in automated GEM reconstruction and gapfilling tools, such as CarveMe and Gapseq. (Machado *et al.*, 2018; Zimmermann, Kaleta and Waschina, 2021). These tools employ the use of curated databases, such as KEGG, BiGG, and ModelSEED to automatically identify metabolic

pathways and fill possible gaps that occur because of incomplete input genome completeness, making it possible to reconstruct individual models in the matter of minutes to hours, depending on the tools used and the desired sensitivity (King *et al.*, 2016).

Despite their apparent usefulness, these models come with a set of limitations. Firstly, as previously mentioned, they assume a metabolic steady-state, which is not necessarily characteristic or realistic for many organisms, especially in aquatic environments, which tend to be highly dynamical (Succurro and Ebenhöh, 2018). Secondly, GEMs assume an unlimited enzyme budget, meaning that an organism can produce and allocate as many proteins as necessary for a given reaction to occur at its maximal possible flux under the given environmental conditions, which may not reflect reality, especially in a resource-limited environment (Gu *et al.*, 2019). Thirdly, these models do not account for additional biological regulation, such as transcriptional control or post-translational modification, unless such omics data is obtained and provided in addition to the genome sequence (Nilsson and Nielsen, 2017). Lastly, the biomass reaction, which is perhaps the most important part of the model due to the specifics of FBA, is often generalized for large taxonomic groups, and does not necessarily represent realistic biological realities (Bernstein *et al.*, 2021).

Despite all these limitations, when using a careful approach based on your specific data and organism-specific considerations, GEMs can be a highly valuable tool for both hypothesis generation, as well as directly examining and understanding microbial metabolic phenotypes.

5.2 Community modeling

While single-organism GEMs have proven highly successful in helping understand individual metabolic capabilities and pathway completeness, microbial communities tend to exhibit system-level emergent properties, arising from cross-feeding networks, interdependent metabolisms, and metabolic labor division, which cannot be predicted from analyzing any species in isolation (Grosskopf and Soyer, 2014). Scaling individual-level modeling up to a community level, however, presents with a number of unique challenges, both conceptual and computational, as it does not simply involve extending an abstract stochiometric matrix, but considering how these organisms compete for resources, excrete and uptake metabolites from the environments, and engage in a number of other biologically meaningful processes (Wang *et al.*, 2023). Over the last

decade or so, several frameworks have been developed to address these concerns, each utilizing different conceptual strategies.

One of the earliest and most important tools developed is OptCom (Optimization of Community Metabolism), which frames a microbial community as a multi-level optimization problem (Zomorrodi and Maranas, 2012). Firstly, each individual model is optimized to achieve maximum biomass production through FBA, then, as a second-level objective, community-level biomass is optimized as well, while keeping the individual organism growth within a biologically feasible range. This allows the algorithm to predict metabolite exchanges and simulate both competitive and cooperative interactions within a relatively stable community composition. However, OptCom requires quite extensive computational resources for its use, and is generally not very practical when applied to communities larger than around 10 species (Fang, Lloyd and Palsson, 2020).

To address the problem of scalability, an extension of this framework was later developed, known as SteadyCom (Chan, Simons and Maranas, 2017). It reformulates the multi-level community problem into a linear one by assuming that all the species within the community grow at the same rate. This has been successfully applied to several microbial communities, leading to prediction of metabolic interactions and identification of key species in bacterial systems, however this assumption of steady growth does not reflect the ecological reality in many contexts, such as aquatic ecosystems, especially during a bloom collapse or establishment (Afarin and Naeimpoor, 2024).

One of the most recent and effective tools in this area is MICOM (Microbial Community Modeling), which introduces a cooperative tradeoff approach to balance community-level metabolic function with individual species fitness (Diener, Gibbons and Resendis-Antonio, 2020). The cooperative tradeoff parameter can be adjusted to either optimize for individual biomass production, as in OptCom, to model a purely competitive community structure, or maximize community-level growth, to model a fully cooperative structure, with values in-between these two being used to reflect a more intermediate ecological scenario. For studying cross-feeding interactions specifically, MICOM explicitly models an extracellular shared environment for all the members of a community, which realistically reflects how microbes exchange metabolites with each other in a real ecosystem (Dal Bello *et al.*, 2021).

All the frameworks mentioned above, however, share a critical flaw when it comes to modelling aquatic communities specifically, which is assuming a non-spatially stratified, well-mixed medium environment, which is more reflective of a laboratory environment than an ecologically realistic one. A number of tools exist to address this, modeling bacterial communities across time and space as well, such as BacArena and COMETS (Computation of Microbial Ecosystems in Time and Space) (Bauer *et al.*, 2017; Dukovski *et al.*, 2021). COMETS works by dividing the abstract environment, into which the microorganisms are populated, into a grid of spatially discrete compartments and simulates the diffusion of metabolites across grid lines, allowing the model to capture the dynamics of complex spatially structured communities, such as microbial mats, or the cyanosphere around individual cyanobacterial heterotrophs. At each time step, FBA is performed for each organism in each spatial location, and metabolite concentrations are updated based on consumption, secretion, and diffusion, which leads to the main limitation of the algorithm – its rather high computational cost and complexity.

Despite the diversity of these and the number of other existing, but unmentioned, algorithms, they tend to share the same limiting factors and common assumptions. They are generally shared with the limitations and assumptions of the individual GEMs, such as the metabolic steady state, lack of incorporation of genetic regulation, and certain ecological factors, such as viral lysis, and other environmental phenomena.

Nevertheless, especially when experimentally validated, these models provide a very powerful framework for hypothesis generation, as well as direct investigation of complex metabolic interactions shaping microbial systems (Frioux *et al.*, 2020). In the context of cyanobacterial blooms, community modeling holds particular promise for elucidating how photosynthetically derived carbon, fixed nitrogen, and essential vitamins flow through the cyanosphere, shaping the composition and dynamics of heterotrophic bacterial assemblages.

6. Computational approaches for metagenome reconstruction and analysis

This section delves deeper into the theoretical foundations and computational algorithms underlining some of the more central methods used in this work. As this study integrates several distinct approaches, such as comparative genomics, genome-scale metabolic modelling, pangenome analysis, genome assembly and more, understanding the theoretical basis of these approaches is essential for interpreting the obtained results and their biological significance.

6.1 Long-read metagenomic sequencing

Long-read sequencing technologies such as Oxford Nanopore have made a significant contribution to the field of microbial genomics and metagenomics by enabling the recovery of high-quality genome reads directly from complex environmental samples, allowing us to investigate such systems without laboratory cultivation (Sereika *et al.*, 2022). This is particularly important in the context of metagenomics, since it improves resolution of repetitive regions, mobile elements, and operons which are more often fragmented during short-read sequencing, reading to poorer assembly quality (Wick, Judd and Holt, 2023).

In microbiome studies, long-read technologies provide another advantage, allowing us to recover near-complete metagenome-assembled genomes from heterogeneous and complex systems, which is otherwise exceedingly difficult, if not impossible to achieve with short-read technologies, such as Illumina (Latorre-Pérez *et al.*, 2020). For this study, near-complete genomes are particularly important, as gaps in the genome can lead to an incomplete genomic, and therefore metabolic profile, most commonly false auxotrophies or metabolic dependencies between organisms that do not truly exist biologically.

Despite these advantages, ONT sequencing has a higher error rate than short-read approaches, which has to be mitigated by employing extensive quality control and post-assembly genome polishing during the bioinformatical analysis of raw read data (Y. Chen *et al.*, 2021).

6.2 Metagenome-assembled genomes

Metagenome-assembled genomes are reconstructed microbial genomes generated directly from environmental or culture sequencing data, eliminating the need for cultivation and isolation of every separate species within a sample (Setubal, 2021). This approach has lead to great advancements in the research on microbial diversity, allowing us to investigate complex systems, like aquatic ones, which were previously largely enigmatic due to the limitations of traditional laboratory methods (Hugerth *et al.*, 2015).

Reconstructing genomes from such sequencing data, however, poses a unique challenge from a bioinformatical perspective. Metagenomic assemblers have traditionally been developed for short-read data, and utilize probabilistic de Bruijn graph approaches, where reads are fragmented into overlapping segments – k-mers – which are then used to reconstruct genomes by using graph

traversal optimization algorithms, usually finding a Hamiltonian cycle of a given graph (Pell *et al.*, 2012). However, long-read sequencing datasets tend to have more noise and errors within them, which makes this approach less optimal for such data. Recently, unique algorithms have been developed specifically to deal with assembling genomes from metagenomic long-read data, the latest one being metaMDBG, which uses minimizer-based sparse de Bruijn graph strategies specifically designed for ONT and PacBio long-read datasets, which reduce graph complexity to improve scalability and assembly quality (Benoit *et al.*, 2026).

After assembly, the next step in the reconstruction process generally involves the grouping of assembled contigs into provisional genomes, which is known as 'binning'. These tools rely on machine learning methods, but the exact algorithms vary greatly between tools, and encompass supervised, semi-supervised, and unsupervised approaches (Yue *et al.*, 2020). Currently, the optimal strategy for this step of reconstructing MAGs is to employ several different techniques, and then additionally use a consensus refinement tool to integrate results from several different binners into non-redundant collections (Vosloo *et al.*, 2021).

Due to the relatively high rate of errors in ONT data, which persists even after assembly, genome polishing is an essential step for generating quality genomes suitable for downstream genomic and metabolic investigation. A commonly used tool for this is Racon, which utilizes iterative partial-order alignment correction based on read support (Vaser *et al.*, 2017). Sequential read polishing utilizing both Racon and Medaka, a neural-based polishing algorithm specifically trained on ONT error profiles, is currently a common strategy used in long-read microbial metagenomic workflows, and seems to meaningfully improve the quality of reconstructed MAGs, but should, however, be used sparingly, as overuse of polishing tools can lead to increased genome contamination (Lee *et al.*, 2021).

6.3 Genome-based taxonomy

Historically, microbial taxonomy relied on either morphological characteristics, or single-gene, usually 16S rRNA, analyses. This, however, is suboptimal for classification of closely related or morphologically similar taxa, as well as environmental metagenomic analysis (Parks *et al.*, 2018).

Genome-based taxonomy is currently emerging as a more robust alternative, utilizing marker genes and whole-genome comparisons for classification of prokaryotic organisms, with the

Genome Taxonomy Database (GTDB) being the most comprehensive efforts towards a centralized and standardized prokaryotic taxonomy databases (Parks *et al.*, 2022).

Such frameworks are particularly important for cyanobacterial research, due to many morphologically classified taxa exhibiting polyphyletic nature and substantial genomic diversity, which is poorly reflected by traditional morphology-based taxonomical approaches (Rajaniemi et al., 2005).

6.4 Functional annotation with Hidden Markov Models

Accurate functional annotation of our genomes is crucial for us, as it directly impacts the results of the metabolic network analysis. While traditional methods such as BLAST are useful, they are known to fail to identify distantly related homologs with low overall sequence similarity (Pearson, 2014). This is particularly important when analyzing microbiomes from understudied or complex systems, as many proteins in such organisms originate from poorly characterized or phylogenetically distant organisms, due to a high degree of horizontal gene transfer in such systems (Pérez-Carrascal *et al.*, 2020).

Hidden Markov Models (HMMs) provide an improvement over those methods in the form of a probabilistic framework for identifying homologous proteins based on conserved motifs and position-specific amino acid conservation patterns (Eddy, 1998). Multiple representatives of a protein family are used to generate a statistical profile describing expected amino acid frequencies and insertion/deletion probabilities across each alignment position.

Query sequences are then compared against these profiles to identify likely homologs. Because HMMs capture evolutionary conservation patterns across entire protein families, they are substantially more sensitive than simple pairwise alignment approaches for detecting distantly related proteins (Aramaki *et al.*, 2020).

6.5 Pangenome reconstruction and comparative genomics

The increasing amount and availability of prokaryotic whole-genome sequences is revealing that strains traditionally classified as the same species can differ greatly in gene content, genome size, metabolic capabilities, and other phenotypical characteristics, leading to an increased interest in intraspecies genomic analysis (Lan *et al.*, 2000). Due to this, the concept of a “pangenome” was recently introduced, which examines the complete set of genes present across multiple genomes

of a given species (Rouli *et al.*, 2015). The pangenome is generally divided into the core genome, containing genes shared by all strains, and the accessory genome, consisting of genes present only in subsets of strains or unique to individual genomes (Matthews *et al.*, 2024). Pangenome reconstruction itself generally begins with prediction and annotation of protein-coding genes within each genome assembly, with homologous genes identified through all-versus-all sequence similarity searches and grouped into gene clusters representing orthologous or closely related genes shared among genomes; this clustering is commonly performed using graph-based algorithms, such as the Markov Cluster Algorithm (Ding, Baumdicker and Neher, 2018). This allows for the investigation of gene presence-absence patterns, identification of strain-specific genes, and analysis of metabolic or ecological specialization within microbial populations (Dewar *et al.*, 2024).

Despite their usefulness, pangenome analyses remain sensitive to genome quality, contamination, and assembly fragmentation, which may influence gene prediction and orthology assignment accuracy (Verma, Satyanarayana and Dias, 2024). Furthermore, horizontal gene transfer and gene duplication events can complicate interpretation of accessory genome structure (Marin *et al.*, 2025). Nevertheless, when combined with metabolic modeling and comparative functional annotation, pangenome reconstruction provides a powerful framework for investigating genomic diversity and ecological adaptation within microbial species.

7. Summary and research gap

Despite increasing recognition of the ecological importance of metabolic interactions within cyanobacterial microbiomes, mechanistic understanding of these processes remains limited, particularly for bloom-forming freshwater cyanobacteria. Existing studies largely rely on correlative community analyses or focus on single host genotypes, leaving the role of intraspecific genomic variation in shaping microbiome structure and metabolic interdependencies insufficiently explored. Furthermore, genome-scale metabolic modeling approaches remain underutilized in phototrophic microbial communities, especially in strain-resolved and community-level applications.

The present study addresses these gaps by integrating strain-resolved microbiome reconstruction, genome-scale metabolic modeling, and pangenome analysis to investigate metabolic interactions within *Aphanizomenon flos-aquae* microbiomes. This work is conducted within the framework of

the “Interactome of *Aphanizomenon flos-aquae* for environmental and biotechnological advancement” **(**INTEC**)** project (P-MIP-25-48), funded by the Research Council of Lithuania under the Researcher Groups Projects initiative. By applying complementary metabolic modeling frameworks to genetically distinct *A. flos-aquae* strains and their associated microbiomes, this thesis contributes a systems-level perspective on cyanosphere metabolism and supports broader efforts to understand and mitigate harmful cyanobacterial blooms through integrative, ecosystem-scale approaches.

METHODS

1. Experimental organisms

Raw Oxford Nanopore long-read sequencing data of 28 different strains of the clonal cyanobacteria *Aphanizomenon flos-aquae* together with their associated microbiomes, was used for the purposes of this research. Local samples were acquired from field work and previous sampling in Lithuania, strains from other countries were purchased from relevant national culture collections. All these strains are kept as unialgal, but non-axenic cultures in the Laboratory of Algology and Microbial Ecology at the Nature Research Centre. Sequencing and initial read data analysis was performed by Seqvision, UAB (Vilnius, Lithuania).

2. Metagenomic assembly and MAG reconstruction

Raw Oxford Nanopore long-read sequencing data obtained from each *Aphanizomenon flos-aquae* culture, and its associated microbiome were processed using a custom Snakemake based metagenomic assembly and genome reconstruction pipeline. The workflow integrated quality filtering, long-read metagenomic assembly, consensus genome binning, polishing, genome quality assessment, and taxonomic classification.

2.1 Read quality control

Initial Nanopore reads were filtered using Chopper version 0.8 to remove low-quality and extremely short reads prior to assembly. Reads with an average quality score lower than Q10 and reads shorter than 1000 bp were excluded from the data. This was performed independently for each strain-associated read set.

2.2 Long-read metagenomic assembly

Previously filtered reads were assembled using MetaMDBG version 1.4, a metagenomic assembler specifically designed to handle ONT long-read data. Assemblies were generated independently for each microbiome sample using default assembly parameters. Following this, contigs shorter than 1500 bp were removed to improve the quality and reliability of the following downstream analyses.

2.3 Read mapping

Filtered reads were mapped back to assembled contigs using Minimap2 version 2.3 in Oxford Nanopore alignment mode. Resulting alignments were sorted and converted into BAM files using SAMtools. Coverage profiles generated from this step were used as input for binning algorithms.

2.4 Metagenomic genome binning

To maximize both the number and the quality of recovered metagenome-assembled genomes, three independent binning algorithms utilizing different computational approaches were separately applied to each assembly:

- MetaBAT2 version 2.18 was run with default settings on generated contig abundance profiles
- COMEBin version 1.0 was used with default ensemble clustering settings
- SemiBin2 version 2.2 was run in long-read self-supervised mode

The outputs of these three algorithms were then optimized and integrated using DAS Tool version 1.1.7 with default parameters. The resulting non-redundant bins were used in downstream analysis.

2.5 Bin-specific read extraction and polishing

For each consensus genome bin generated by DAS Tool, associated reads were extracted from the assembly alignment files. Then, only reads exhibiting at least 90% alignment identity and query coverage relative to bin contigs were retained, to minimize potential read misalignment. Genome polishing was performed in two consecutive stages on each bin individually. First, Racon version 1.5 was run with default parameters, and its output was processed in the second stage, using Medaka version 2.2.1 with the model trained on r1041_e82_400bps_sup_v5.2.0 Nanopore model.

2.6 Quality assessment

Quality of polished MAGs was assessed using CheckM2 version 1.1.0 using default parameters. A quality gate of above 50% completeness and below 10% contamination was established, and any genomes not meeting this standard discarded.

2.7 Taxonomic classification

Taxonomic classification of the resulting MAGs was performed using the GTDB-Tk version 2.7.1 based on the Genome Taxonomy Database (GTDB). Classification was performed using phylogenetic marker genes with ANI pre-screening disabled due to computational limitations. The resulting MAG collection and taxonomy data was analyzed using the methods described further.

3. Microbial network reconstruction and analysis

The bacterial community associated with each AFA strain was characterized at the genus level and compiled into a presence/absence matrix. Cultures were grouped a priori by geographic origin of the cyanobacterial host into four categories: Europe, Lithuania, Asia, and USA.

To visualize and analyze the overall similarities between community compositions both within and between geographical groups, the presence/absence matrix was standardized (zero mean, unit variance per genus), and reduced to two dimensions by principal component analysis (PCA) using scikit-learn version 1.8.0.

To test whether heterotrophic community composition differed significantly among cyanobacterial cultures grouped by geographic origin we used PERMANOVA (Permutational Multivariate Analysis of Variance) to test whether the centroids of the four geographic groups in multivariate space differed significantly. Additionally, pairwise PERMANOVA was performed between all pairs of geographic groups to identify which specific groups differed. For all tests, significance was declared at $\alpha = 0.05$

4. Individual genome-scale metabolic model reconstruction

Genome-scale metabolic models of *Aphanizomenon flos-aquae* as well as each individual heterotrophic bacteria were reconstructed using gapseq version 1.4 providing modified BG11 medium, to be used as gap-filling medium for the algorithm. Medium was modified to only contain the minimal possible number of compounds that support growth of all modeled organisms. This modification was implemented to minimize the availability of excess nutrients and reduce artificial cross-feeding interactions that could arise from compounds present in richer media. Exact final medium composition, including uptake rate constraints can be seen in Table 1.

Table 1. Composition of the modified BG11 medium used for genome-scale metabolic model reconstruction and gap-filling.

Compound ID	Compound Name	Maximum flux (mmol/gDW/hr)
cpd00063	Calcium ion	100
cpd00099	Chloride ion	100
cpd00137	Citrate	5

cpd00009	Phosphate	100
cpd00205	Potassium ion	100
cpd11632	Photon	100
cpd00254	Magnesium	100
cpd10516	Ferric ion	5
cpd00013	Ammonium	100
cpd00007	Oxygen	100
cpd00149	Carbon dioxide	100
cpd00058	Cupric ion	100
cpd00067	Hydrogen ion	1
cpd00030	Manganese ion	100
cpd00034	Zinc ion	100
cpd00048	Sulfate	100
cpd00209	Nitrate	100

The maximum uptake flux for most inorganic nutrients was set to 100 mmol gDW^{-1} h^{-1}, which effectively represents a non-limiting availability in flux balance analysis. This value is commonly used in genome-scale metabolic modeling to avoid artificially constraining metabolic fluxes when the environmental concentration of the compound is not intended to limit growth. Lower flux bounds were assigned to selected compounds (citrate, ferric ion, and hydrogen ion) to better approximate environmentally realistic availability and physiological constraints. Citrate was limited to avoid unrealistic heterotrophic carbon uptake that could bias cross-feeding predictions, while ferric ion uptake was constrained to reflect the typically low bioavailability of iron in aquatic environments. The hydrogen ion exchange flux was restricted to maintain physiologically reasonable proton balance and prevent excessive flux through proton-coupled transport reactions, which can otherwise lead to unrealistic metabolic solutions in constraint-based models.

Both draft and gap-filled models were initially generated and benchmarked. However, after reviewing the MEMOTE reports and the reaction content of the resulting reconstructions, the draft models were retained for further analysis. This decision was made because draft models more conservatively reflect genome-encoded metabolic potential, whereas gap-filled models may include additional reactions that improve in silico growth but are not directly supported by

genomic evidence. Since all draft models were already capable of growth simulation and showed acceptable quality scores, they were considered appropriate for subsequent community-scale modeling.

The final curated genome-scale metabolic models for each modelled *Aphanizomenon flos-aquae* strain and the associated heterotrophic bacterial genomes were exported in SBML format and used for further community metabolic interaction analyses.

5. Community-scale metabolic model reconstruction

The following tools and software packages were used to construct community-scale metabolic models of AFA strains and their related microbiomes, as well as to infer cross-feeding interactions between the cyanobacteria and the heterotrophs:

MICOM version 0.37.1 was used to build community models based on cooperative tradeoff optimization, with the following parameters: bacterial taxonomy was defined up to the genus level based on genome annotations, species abundances were set to equal values (1/n for n species) in the absence of experimental abundance data, representing a neutral assumption of community composition., tradeoff parameter was set to 0.9, to prioritize interactions between bacteria over individual growth rates.

SMETANA version 1.2.1 was used to predict metabolic dependencies and cross-feeding potential through algorithmic analysis of individual genome-scale models without requiring integrated community simulations. SMETANA was run with the --detailed flag, which outputs metabolite-level interaction predictions rather than only summary scores, enabling identification of specific compounds involved in potential cross-feeding relationships.

Metage2metabo version 1.6.2 was employed to assess metabolic complementarity and gap-filling potential within the community using a scope-expansion algorithm. This tool identifies compounds that can be produced collectively by the community but not by individual members, thereby predicting potential metabolic cooperation. Only the metacom workflow was used, which specifically analyzes community-level metabolic capabilities and identifies targets for metabolic exchanges, gap-filling requirements, and producibility of key metabolites under the given medium conditions.

For all these tools the modified BG11 medium was provided as a modeling parameter (Table 1). The Gurobi solver, used under an academic license, was used for all the calculations performed. All these tools were compared against each other based on the number of cross-feeding interactions successfully predicted, as well as the biological feasibility of those reactions. Interactions were considered to be biologically realistic based on previous literature, and presence of complete biosynthetic pathways in the bacterial genome, cross-referenced with KEGG.

6. *Aphanizomenon flos-aquae* pangenome analysis

6.1 Dataset compilation

To contextualize the 28 *Aphanizomenon flos-aquae* strains cultured at the Laboratory of Algology and Microbial Ecology within the broader genomic diversity of the species and assess how this diversity could be connected to the metabolic capabilities of the strains, sixteen publicly available AFA assemblies were retrieved from the NCBI Assembly database in September of 2025. Selection criteria for public genomes included: taxonomic identification as *Aphanizomenon flos-aquae* at the species level, assembly status of "scaffold" level or better, and availability of complete assembly files in FASTA format. The final dataset comprised 44 *A. flos-aquae* genomes (28 from this study + 16 public assemblies). An external genomes file was prepared listing all 44 genome databases with associated metadata for pangenome computation.

Pairwise Average Nucleotide Identity (ANI) values were calculated for all 44 genomes using FastANI version 1.33 with default parameters. ANI analysis was performed to verify species-level taxonomic assignment of all genomes and assess genomic similarity among strains. The ANI matrix was visualized as a heatmap using the Python seaborn library version 0.13.2.

6.2 Pangenome reconstruction

Pangenome analysis was conducted using anvi'o (analysis and visualization platform for omics data) version 9 following the pangenomics workflow. The analysis pipeline consisted of the following steps:

6.2.1. Genome database generation

Individual anvi'o genome databases were created for each of the 44 *A. flos-aqua*e assemblies using anvi-gen-contigs-database. Gene prediction was performed using Prodigal version 2.6.3 in metagenomic mode, which is standard for anvi'o pangenome workflows. Hidden Markov Model

(HMM) searches were conducted using anvi-run-hmms with the Bacteria_71 single-copy core gene collection to assess genome completeness within the anvi'o framework.

6.2.2 Functional annotation

Genes were functionally annotated using anvi-run-ncbi-cogs to assign NCBI Clusters of Orthologous Groups (COG) categories, providing functional context for pangenome gene clusters. Additional annotation was performed using anvi-run-kegg-kofams to map genes to KEGG Orthology (KO) terms and metabolic pathways.

6.2.3 Pangenome computation

The pangenome was computed using anvi-pan-genome with the following parameters:

- --use-ncbi-blast: employed NCBI BLASTP (version 2.13.0+) for all-versus-all protein similarity searches instead of the default DIAMOND aligner, prioritizing sensitivity over speed for this relatively small genome set
- --mcl-inflation 10: set the Markov Cluster Algorithm (MCL) inflation parameter to 10 (range: 1-100), which controls the granularity of gene cluster delineation. Higher values (such as 10) produce more granular, smaller gene clusters by increasing cluster tightness, reducing the likelihood of distantly related paralogs being grouped together.

6.2.4 Pangenome analysis and classification

Pangenome statistics, including total number of gene clusters, core genome size, accessory genome size, and singleton genes (unique to single genomes), were calculated using built-in anvi'o functions. The pangenome was visualized using anvi-display-pan with hierarchical clustering of both genomes (based on gene cluster presence/absence) and gene clusters (based on distribution across genomes).

Singleton gene clusters (those unique to individual genomes) were ranked by total gene length, and the three largest were selected for detailed functional characterization to identify potential strain-specific adaptations. Functional investigation was performed using COG analysis

All data visualization was performed using anvi'o interactive interface and exported figures, with additional customization using Python matplotlib version 3.5.2 and seaborn version 0.13.2

6.2.5 Rarefaction curves and Heaps' Law fitting

To assess how completely the 44-genome dataset captured the pangenome diversity of *Aphanizomenon flos-aquae* and to quantify pangenome openness, rarefaction curves were computed from the anvi'o pangenome database using anvi-compute-rarefaction-curves.

Rarefaction analysis was run for 100 iterations, and the resulting accumulation curve was exported as an SVG figure.

Heaps' Law was fitted to the rarefaction curve using the model $P(n) = K \cdot n^{\alpha}$, where $P(n)$ is the cumulative number of gene clusters observed after sampling n genomes, K is a fitted constant, and α describes the rate at which novel gene clusters continue to be discovered with increasing genome sampling. The shape of the rarefaction curve was used to evaluate whether gene cluster discovery approached saturation, while the Heaps' Law α parameter was used as a quantitative descriptor of pangenome openness.

7. Code availability

The workflows used in metagenome assembly and binning, constructing metabolic models, and creating the community models with MICOM are available on GitHub (https://github.com/caermorn/aphanizomenon_analysis)

RESULTS

1. Metagenomic assembly and MAG reconstruction

Oxford Nanopore long-read metagenomic sequencing was successfully performed for 28 non-axenic Aphanizomenon flos-aquae cultures and their associated microbiomes. After assembly of the metagenomic data, followed by genome binning, polishing, and quality filtering, we have acquired 600 metagenome-assembled genomes across all analyzed microbiomes, however only 570 passed the quality gate. The reconstructed MAG collection consisted predominantly of high-completeness genomes. Based on CheckM2 completeness and contamination estimates, 476 MAGs were classified as near-complete genomes (>90% completeness and <5% contamination), 55 MAGs were classified as high-quality genomes (>70% completeness and <10% contamination), and 39 MAGs met medium-quality criteria (>50% completeness and <10% contamination) (Table 2).

Table 2. Quality distribution of reconstructed MAGs

MAG quality category	Completeness threshold	Contamination threshold	Number of MAGs
Near-complete	>90%	<5%	476
High-quality	>70%	<10%	55
Medium-quality	>50%	<10%	39

1.1 AFA_2012_KM_D3 assembly

The MAGs presented in Table 3 originated from the microbiome associated with strain AFA_2012_KM_D3. This dataset was selected for subsequent in-depth community-scale metabolic interaction analyses because it contained multiple high-quality and near-complete MAGs with low contamination and high assembly continuity across both the cyanobacterial host and several associated heterotrophic bacteria. In particular, the microbiome included multiple genomes reconstructed as single-contig assemblies together with a highly complete *Aphanizomenon* MAG, providing a technically robust dataset for genome-scale metabolic reconstruction. In addition, the recovered MAGs represented phylogenetically distinct bacterial

taxa with varying genome sizes and metabolic potential, enabling downstream analyses of potential metabolic complementarity and cross-feeding interactions within the reconstructed microbial community.

Table 3. Assembly statistics and quality metrics of MAGs reconstructed from the microbiome associated with strain AFA_2012_KM_D3.

Completeness (%)	Contamination (%)	Size (bp)	Contigs	GC (%)	Genus	CDS	rRNAs	tRNAs	ncRNAs	CRISPR
86,71	0,43	3,38E+06	13	68,4	*Caulobacter*	4001	3	45	48	
94,33	0,48	3,16E+06	9	52,6	*Limnobacter*	3592	6	39	45	
99,53	0,93	4,96E+06	1	65,2	*Hydrogenophaga*	4698	6	49	56	1
94,09	0,49	5,37E+06	3	45,1	*Phnomibacter*	5377	6	45	52	13
90,57	0,00	3,44E+06	8	39,1	*Sediminibacterium*	4085	3	33	37	2
97,14	1,46	4,09E+06	1	41,8	*SB11*	3670	6	39	46	5
96,50	0,44	6,10E+06	1	38,3	*Aphanizomenon*	6512	14	42	57	43
63,07	2,92	2,28E+06	50	55,7	*Sphingorhabdus_B*	3817	3	31	34	1

Several bacterial genomes were recovered as single-contig assemblies, including representatives assigned to *Hydrogenophaga*, *SB11*, and *Aphanizomenon flos-aquae*, with completeness values ranging from 96.50% to 99.53%. Other MAGs, including *Caulobacter, Limnobacter, Phnomibacter,* and *Sediminibacterium*, were reconstructed in fewer than 15 contigs while maintaining completeness values above 86%.

Genome sizes among the recovered MAGs ranged from 2.28 Mbp in *Sphingorhabdus_B* to 6.10 Mbp in AFA, while GC content ranged from 38.3% in AFA to 68.4% in *Caulobacter*. The number of predicted coding sequences varied from 3,592 CDS in *Limnobacter* to 6,512 CDS in *Aphanizomenon*. Ribosomal RNA genes, transfer RNAs, and other non-coding RNAs were identified in all reported MAGs, while CRISPR arrays were detected in multiple genomes, most notably in the *Aphanizomenon* MAG, which contained 43 predicted CRISPR regions.

2. Taxonomic classification of reconstructed MAGs

Taxonomic classification of the reconstructed MAG collection using GTDB-Tk revealed a diverse assemblage of bacteria associated with *Aphanizomenon flos-aquae* cultures. In total, 190 MAGs were classified to the genus level, of which 107 could be further resolved to the species level. Across the full dataset, the recovered MAGs represented 10 phyla, 13 classes, 40 orders, 58 families, 116 genera, and 93 species-level taxa. At the genus level, individual cultures contained on average 16.7 genera, with richness ranging from 4 to 31 genera per sample (Figure 3).

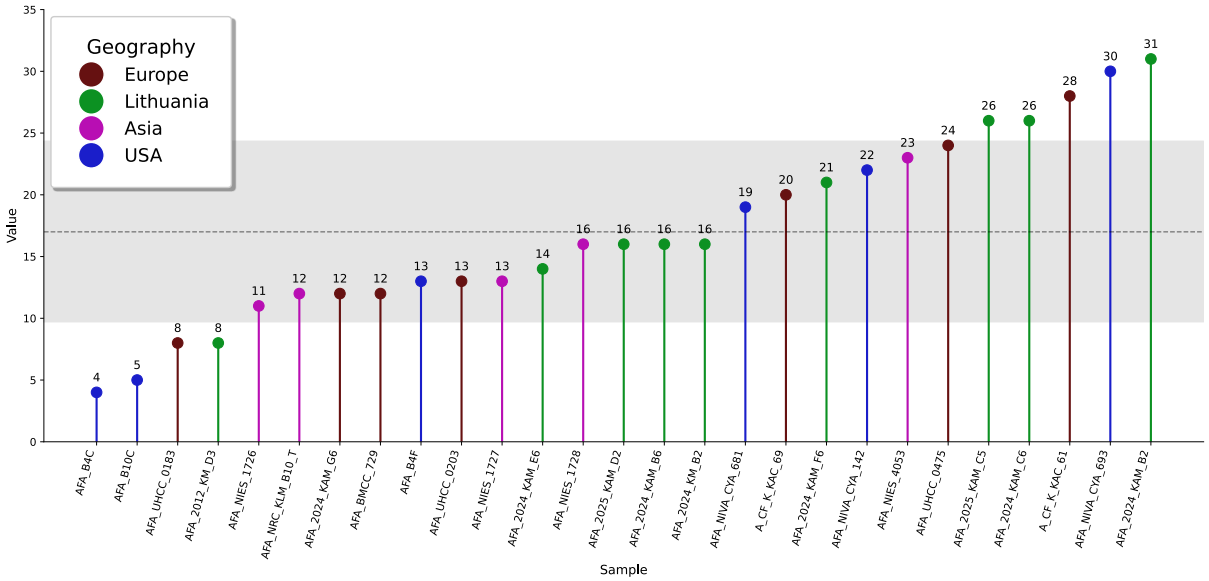


Figure 3: Distribution of heterotrophic genera across different strains of AFA.

At higher taxonomic ranks, the microbiomes showed a relatively conserved overall structure. At the phylum level, Pseudomonadota and Bacteroidota were detected in all analyzed cultures. Similarly, at the class level, Alphaproteobacteria, Bacteroidia, and Gammaproteobacteria were present in all cultures. Greater variability was observed at lower taxonomic levels. The most prevalent orders were Burkholderiales (100.0%), Rhizobiales (92.6%), Sphingomonadales (92.6%). At the family level, the most common groups were Burkholderiaceae and Sphingomonadaceae (both 92.6%), followed by Beijerinckiaceae (85.2%). At the genus level, no taxon was present in all cultures, but several genera were recurrent, including *Bosea* (77.8%), *Sphingorhabdus_B* (77.8%), *Flavobacterium* (70.4%), *Mesorhizobium* (66.7%), *Pseudomonas_E* (66.7%), *Hydrogenophaga* (66.7%), and *Variovorax* (63.0%).

The extent of taxon sharing decreased markedly with increasing taxonomic resolution. Of the 10 detected phyla, 2 were present in all cultures, whereas among the 116 genera, none formed a strict core microbiome. Overall, these results indicate that *A. flos-aquae* cultures share a broad community backbone at high taxonomic ranks, while the composition at family, genus, and species levels is substantially more variable among strains (Table 4).

Table 4: Distribution of taxon prevalence across taxonomic levels.

Taxonomic level	Number of taxa	Taxa present in >50% of cultures	Taxa present in ≥75% of cultures	Taxa present in ≥90% of cultures	Core taxa
Phylum	10	3	3	2	2
Class	13	4	4	3	3
Order	40	7	5	3	1
Family	58	8	4	2	0
Genus	116	9	3	0	0
Species	93	6	0	0	0

3. Geographic structure of strain-associated microbiomes

To assess whether the composition of Aphanizomenon flos-aquae-associated bacterial communities differed according to host geographic origin, genus-level presence/absence profiles were compared across 27 cultures assigned to four geographic groups: Lithuania (n = 9), Europe (n = 7), USA (n = 6), and Asia (n = 5). The genus-level microbiome matrix used for this analysis showed substantial variation in community size among cultures, ranging from 4 to 31 genera per sample, with an overall mean of 17.0 genera. Among the four regions, cultures from Lithuania had the largest average microbiome size (19.3 genera), whereas cultures from Asia and the USA showed somewhat lower average richness (15.0 and 15.5 genera, respectively) (Table 5).

Table 5: Distribution of microbiome sizes across geographic groups.

Geographic group	Number of cultures	Mean microbiome size (genera)	Range
Asia	5	15.0	11–23
Europe	7	16.7	8–28
Lithuania	9	19.3	8–31
USA	6	15.5	4–30
Total / overall	27	17.0	4–31

Visualization of this data by PCA indicated large overlap among the four geographic groups, with no discrete clustering of the samples (Figure 4).

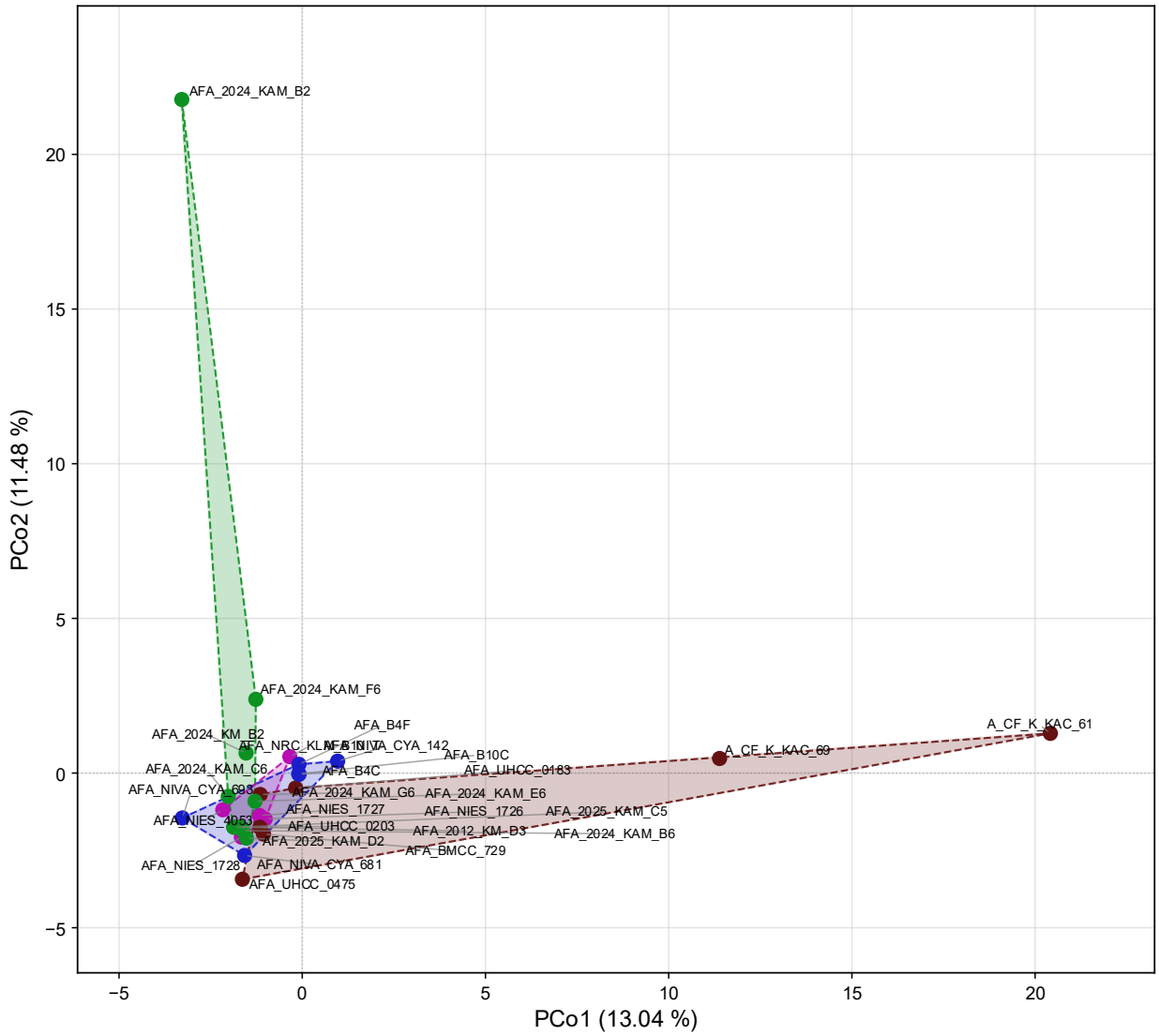


Figure 4: PCA analysis of AFA microbiome compositions.

However, global PERMANOVA analysis detected a very small, but statistically significant impact of geographic origin on microbiome composition ($R^2 = 0.149$, $p = 0.0451$), with geographical data accounting for approximately 15% of the observed variation.

Prevalence analysis further showed that the shared portion of the genus-level microbiome was small. Only 7 genera were present in more than 50% of samples, and no genus occurred in >75% of cultures (Table 6).

Table 6: Genera present in more than 50% of analyzed cultures

Genus	Number of cultures	Prevalence (%)
Sphingorhabdus_B	18	66.7
Hydrogenophaga	18	66.7
Variovorax	17	63.0
Mesorhizobium	17	63.0
Pelomonas	16	59.3
Pseudomonas_E	16	59.3
Flavobacterium	15	55.6

4. Individual genome-scale metabolic-model reconstruction

Genome-scale metabolic models were reconstructed for MAGs recovered from the analyzed Aphanizomenon flos-aquae-associated microbiomes. However, for detailed evaluation and subsequent community-scale analysis, the microbiome associated with strain AFA_2012_KM_D3 was selected as the focal dataset. Within the AFA_2012_KM_D3 microbiome, 8 genome-scale metabolic models were reconstructed.

All reconstructed draft models from the AFA_2012_KM_D3 microbiome were able to sustain simulated growth under the modified BG11 medium used during model reconstruction. Predicted individual growth rates ranged from 0.11 to 0.65, indicating substantial variation in modeled growth performance under identical environmental constraints (Table 7).

Table 7: Properties of the final genome-scale metabolic models associated with AFA_2012_KM_D3.

Bin	Taxonomy	Individual growth rate (gapseq)	Draft model quality score (MEMOTE)
1	*Caulobacter sp.*	0.29	70%
2	*Limnobacter sp002954425*	0.30	70%
3	*Hydrogenophaga sp.*	0.34	70%
4	*Phnomibacter sp.*	0.23	68%

5	*Sediminibacterium sp.*	0.12	69%
6	*SB11 sp.*	0.43	70%
7	*Aphanizomenon flos-aquae*	0.65	70%
8	*Sphingorhabdus_B* sp.	0.11	70%

Model quality assessment using MEMOTE showed that the draft reconstructions were technically consistent and of comparable quality. MEMOTE scores ranged from 68% to 70%, with six of the eight models receiving a score of 70%.

5. Community-scale metabolic interaction analysis

Following reconstruction of individual genome-scale metabolic models, community-scale metabolic interaction analysis was performed to evaluate potential cross-feeding relationships within reconstructed *Aphanizomenon flos-aquae* microbiomes.

The first objective was to compare how different community-scale modeling approaches describe interaction potential within the AFA_2012_KM_D3 microbiome under the selected medium. Among the tools that explicitly report exchange reactions, MICOM produced the largest interaction network, predicting 322 exchange reactions involving 57 unique metabolites, whereas SMETANA predicted 188 reactions involving 61 unique metabolites (Table 8). By comparison, the gapseq-based screening yielded only 8 unique candidate produced metabolites, and no exchange reactions were predicted (by design).

Table 8: Comparison of community-scale interaction predictions for the AFA_2012_KM_D3 microbiome

Tool	Predicted reactions	Unique metabolites
gapseq	N/A	8
SMETANA	188	61
MICOM	322	57

To translate the observed community composition of AFA_2012_KM_D3 into mechanistic hypotheses about metabolic interactions, the MICOM solution for this microbiome was further

visualized as a metabolite exchange network (Figure 5). In this network, outer sectors correspond to the host cyanobacterium and the associated heterotrophic genera, and inner connections correspond to predicted uptake or secretion interactions. Connection width reflects the relative magnitude of exchange in the MICOM solution, with connection colors corresponding to category of metabolite exchanged, with grey representing organic compounds, pink representing vitamins, cyan representing amino acids, etc. Full exchange network can be found in Appendix 1.

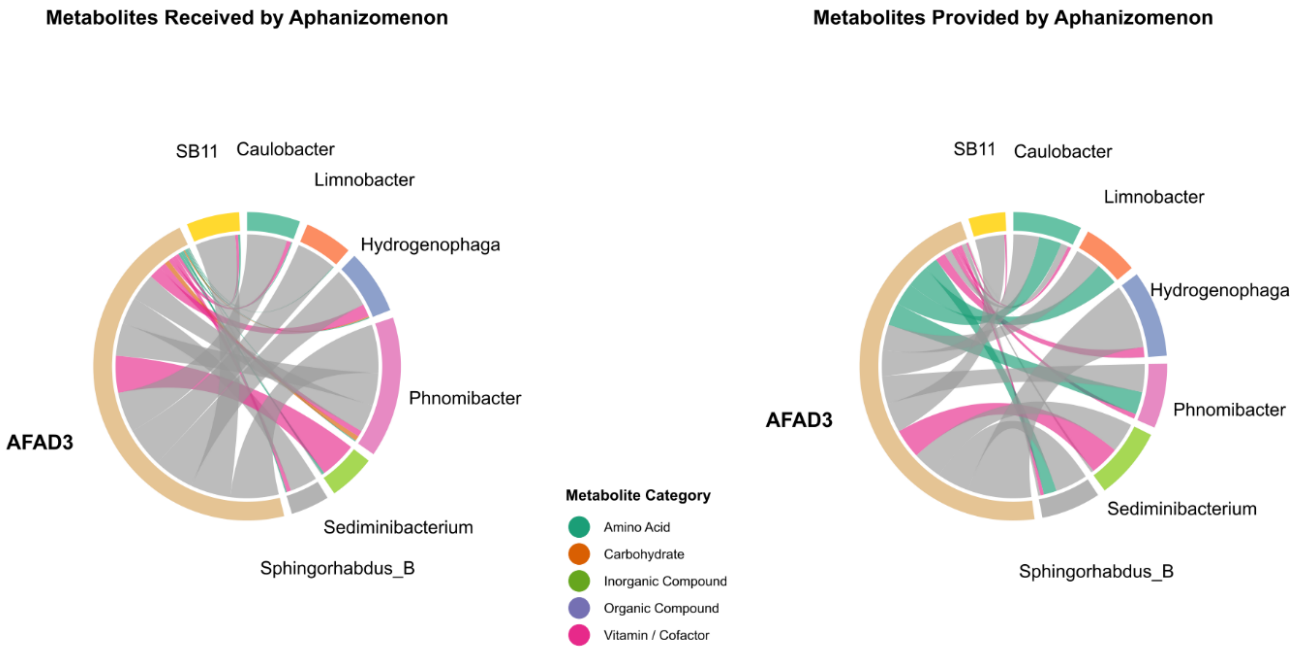


Figure 5: Metabolic interactions within the AFA_2012_KM_D3 microbiome.

A complementary perspective was provided by metage2metabo, which does not primarily quantify pairwise exchange reactions but instead evaluates how much each community member expands its metabolic scope in community context. For the AFA_2012_KM_D3 microbiome, all eight reconstructed models showed an increase in the number of metabolites predicted to be producible when analyzed as part of the community rather than individually (Table 9). The number of metabolites produced individually ranged from 188 to 837, while the corresponding community-context values ranged from 453 to 925. The largest absolute increases in metabolic scope were observed for *Sediminibacterium sp.*, *Aphanizomenon flos-aquae*, and *Limnobacter sp002954425*.

In relative terms, the largest increases were found for *Sphingorhabdus_B sp*., *Sediminibacterium sp*., and *Aphanizomenon flos-aquae*.

Table 9: Metabolic scope expansion.

Taxonomy	Produced alone (m2m)	Produced in community (m2m)	Community metabolic gain	Community metabolic gain (%)
Caulobacter sp.	533	764	231	43.3
Limnobacter sp002954425	398	754	356	89.4
Hydrogenophaga sp.	837	925	88	10.5
Phnomibacter sp.	418	742	324	77.5
Sediminibacterium sp.	291	694	403	138.5
SB11 sp.	384	637	253	65.9
Aphanizomenon flos-aquae	317	719	402	126.8
Sphingorhabdus_B sp.	188	453	265	141.0

The AFA_2012_KM_D3 microbiome was also compared with three additional reconstructed communities, AFAB4C, AFAB4F, and AFAB10C, using MICOM-derived metabolite exchange profiles. Compositions of the additional microbiomes can be found in Appendix 2. This comparison identified metabolites that were unique to a single microbiome among the four tested communities. For uniquely provided metabolites, AFAB4C had 2, AFAB4F had 6, AFAB10C had

1, and AFA_2012_KM_D3 had 8 (Table 10). For uniquely received metabolites, AFAB4C had 1, AFAB4F had 3, AFAB10C had 3, and AFAD3 had 11 (Table 11).

Table 10: Unique metabolites provided by AFA.

Strain	Unique provided metabolites
AFAB4C	Pantothenate (vitamin B5); menaquinol-7
AFAB4F	Adenosine; guanosine; L-alanine; L-proline; L-tyrosine; putrescine
AFAB10C	Glycerol
AFA_2012_KM_D3	Acetate; ethanol; fumarate; L-aspartate; NADP; nicotinamide; ornithine; xylose

Table 11: Unique metabolites received by AFA

Strain	Unique received metabolites
AFAB4C	Menaquinone-7
AFAB4F	Deoxyguanosine; glycerol; H_2
AFAB10C	D-glyceraldehyde; D-ribose; formate
AFA_2012_KM_D3	D-fructose; L-arabinose; L-lysine; L-malate; L-serine; L-tryptophan; MTTL; putrescine; spermidine; succinate; acetaldehyde

6. Aphanizomenon flos-aquae pangenome analysis

To place the genomes analyzed in this study into a broader genomic context, a comparative analysis was performed on a dataset of 44 *Aphanizomenon flos-aquae* genomes, consisting of 28 genomes reconstructed in this study and 16 publicly available assemblies. Pairwise Average Nucleotide Identity (ANI) values were calculated for all genomes using FastANI, and the resulting matrix was visualized as a heatmap.

6.1 Pairwise ANI analysis

Pairwise ANI values across the 44 genomes ranged from 74.7% to 100.0% (excluding self-comparisons, rounded at 3rd decimal point) (Figure 6, Table 12). Several genomes were nearly identical. For example, AFAB10C, AFAB10F, AFAB4C, and AFAB4F showed pairwise ANI values close to 100%.

High pairwise ANI values were also observed among several public genomes, including FACHB_1171, FACHB_1249, FACHB_1265, FACHB_1287, FACHB_1290, and FACHB_1416, which formed a high-identity group with pairwise values above 99.9% in multiple comparisons.

The ANI matrix also contained genomes with lower identity relative to the main high-identity groups. AFAD3 showed 98.4% ANI with AFA_UHCC_0183, 96.6% with FACHB_1040, and 96.0% with AFA_NIES_1728_A, but approximately 86.7–87.3% ANI to many genomes in the larger high-identity block. Additional genomes with comparatively low ANI to the main block included A_CF_K_KAC_69 and AFA_UHCC_0475. The lowest overall ANI values were observed for AFA_UHCC_0344, which showed values of approximately 74.7–75.9% relative to the remaining genomes.

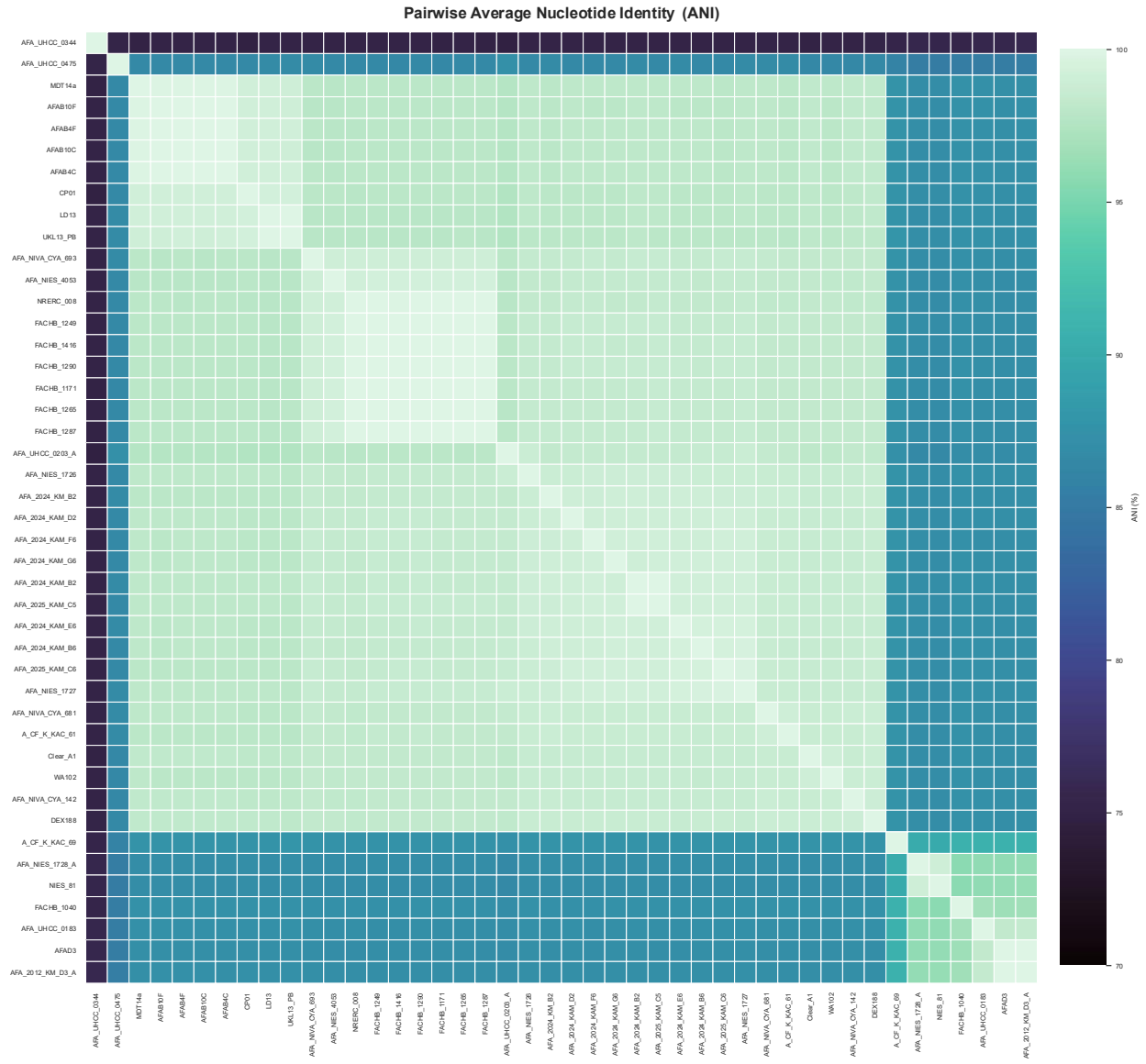


Figure 6: Pairwise Average Nucleotide Identity (ANI) heatmap of 44 Aphanizomenon flos-aquae genomes.

Table 12: Selected ANI metrics.

Comparison / metric	ANI (%)
Pairwise ANI range across dataset (excluding self-comparisons)	74.7–100.0
AFAB10C vs AFAB4C	100.0
AFAB10C vs AFAB10F	100.0

AFAB4C vs AFAB4F	100.0
AFAD3 vs AFA_UHCC_0183	98.4
AFAD3 vs FACHB_1040	96.6
AFAD3 vs AFA_NIES_1728	96.0
AFAD3 vs AFAB10C	87.2
AFAD3 vs A_CF_K_KAC_69	90.7
AFA_UHCC_0344 vs AFA_UHCC_0203	74.7

6.2 Pangenome composition

The anvi'o-based pangenome reconstruction identified 14,296 gene clusters across the full 44-genome dataset (Figure 7). Of these, 2,684 gene clusters were present in 95+% of all analyzed genomes and therefore constituted the core genome. In contrast, 5,983 gene clusters were classified as singletons, meaning that they were detected in only one genome. The remaining 5,629 gene clusters were shared by more than one but not all genomes and therefore comprised the accessory genome.

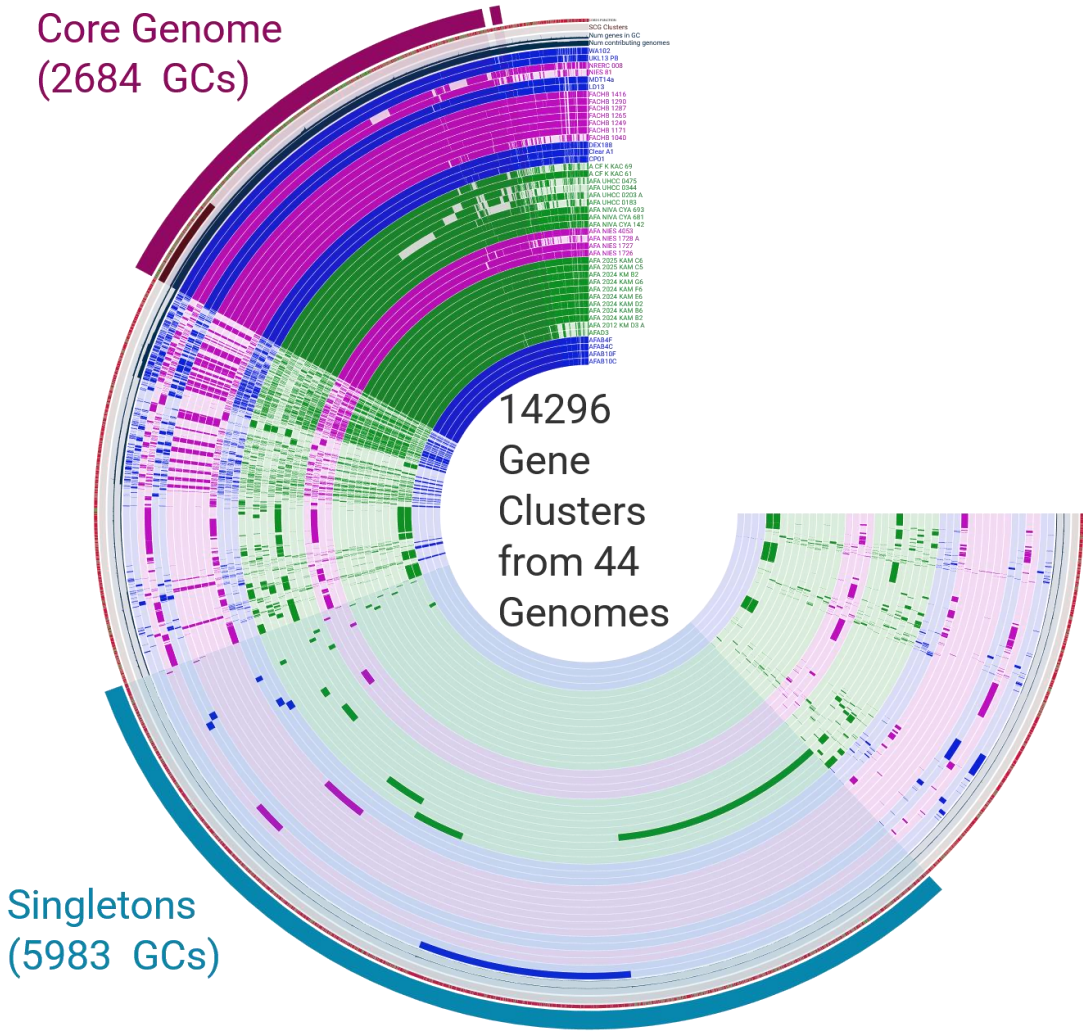


Figure 7: *A. flos-aquae* pangenome

Expressed as proportions of the total pangenome, the core genome accounted for 18.8% of all gene clusters, the accessory genome for 39.4%, and the singleton fraction for 41.9%. Thus, the combined non-core component of the pangenome represented more than four-fifths of all detected gene clusters.

6.3 Rarefaction analysis and Heaps' Law fitting

Rarefaction analysis of the 44-genome Aphanizomenon flos-aquae pangenome showed that the cumulative number of gene clusters continued to increase as additional genomes were added and did not reach a clear plateau within the sampled dataset (Figure 8).

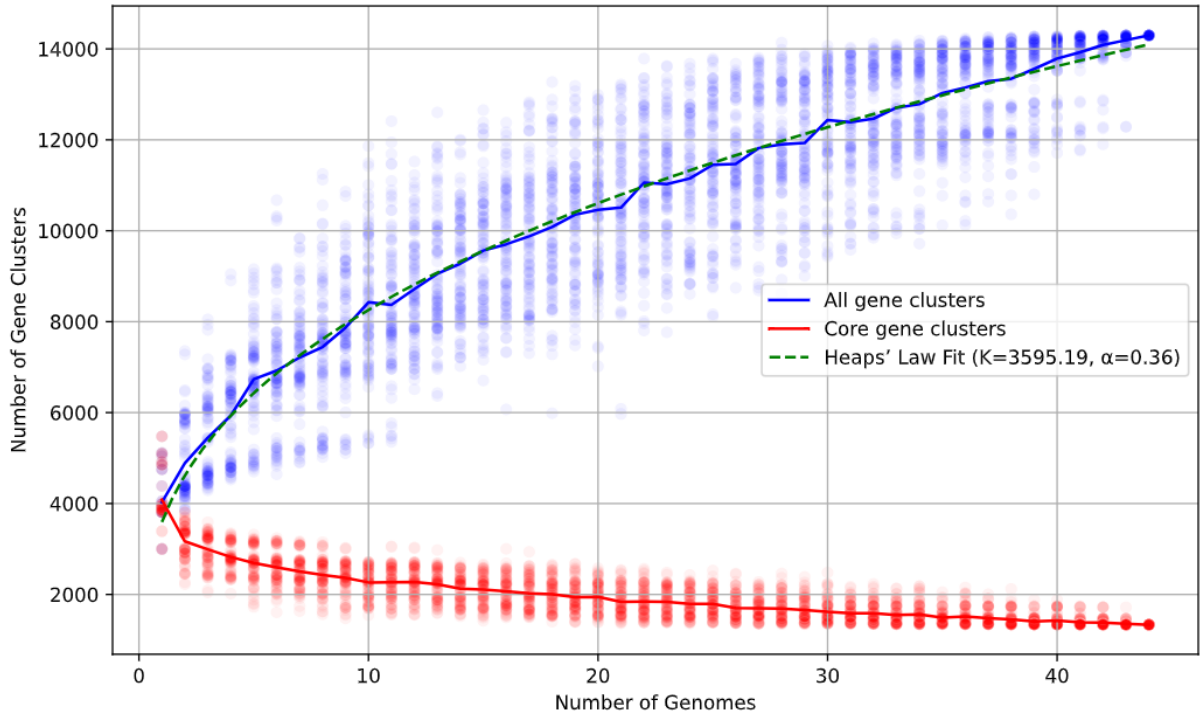


Figure 8: Rarefaction analysis with Heap's law fit.

Fitting Heaps' Law to the rarefaction curve yielded the parameters K = 3595 and α = 0.36.

6.4 Functional characterization of unique *Aphanizomenon flos-aquae* gene clusters

To further examine the variable component of the *Aphanizomenon flos-aquae* pangenome, the three largest strain-specific singleton regions of the pangenome belonging to strains from different geographical groups (excluding AFA_UHCC_0344 due to the large difference in ANI) were functionally annotated and analyzed.

Functional annotations of these unique gene clusters showed that most singleton-associated genes were assigned with no known function. Among the genes that could be annotated, several functional categories were repeatedly represented. These included genes associated with transcription, replication, recombination and repair, cell wall / membrane / envelope biogenesis, signal transduction mechanisms, and defense mechanisms. In addition, annotated singleton genes

were also distributed across several metabolic categories, including functions linked to secondary metabolite biosynthesis and transport (Figure 9).

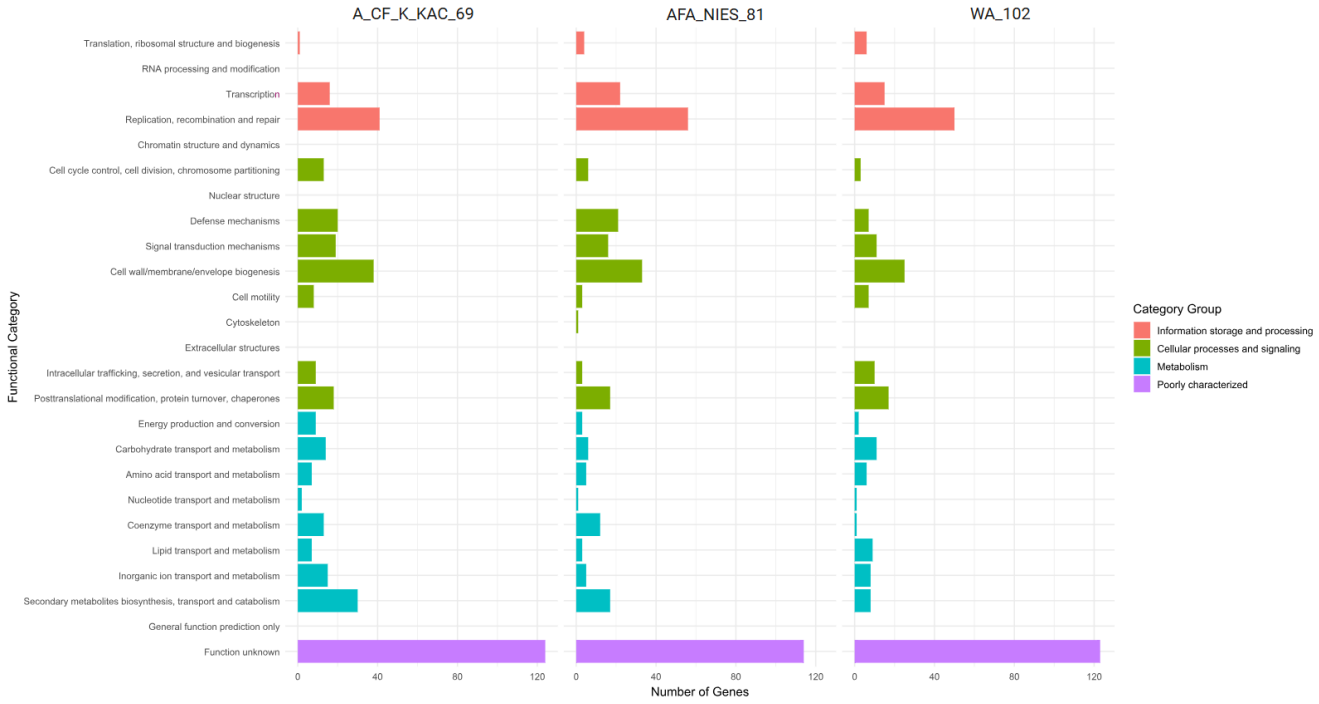


Figure 9: Functional characterization of unique AFA gene clusters.

DISCUSSION

The main aim of this thesis was to characterize metabolic interactions in *Aphanizomenon flos-aquae* microbiomes and to evaluate the role of cyanobacterial genotypes in structuring community composition and predicted metabolic cross-feeding networks. To address this, three objectives were defined:

(1) reconstruct genome-scale metabolic models for *A. flos-aquae* strains and co-occurring heterotrophic bacteria,
(2) integrate individual genome-scale metabolic models into community-level reconstructions, and
(3) assess how genotypic variation among *A. flos-aquae* strains influences microbiome composition and predicted metabolic interactions.

The results addressed all three objectives, although they were not at the same level of resolution. Genome-scale model reconstruction and community integration were completed directly, whereas the role of host genotype was evaluated comparatively through combined taxonomic, pangenomic, ANI, and modelling results.

1. Reconstruction of genome-scale metabolic models

The first objective was achieved through reconstruction of genome-scale metabolic models from MAGs recovered from multiple A. flos-aquae-associated microbiomes. Detailed analysis focused on the AFA_2012_KM_D3 microbiome, which included representative heterotrophic bacteria, some of which, like *Sphingorhabdus_B* are known to play an important role in microbiome metabolism in other systems, such as rhizospheres, particularly when it comes to nitrogen cycling (Xu *et al.*, 2026). All eight draft models supported simulated growth under the selected BG11-based minimal medium, and MEMOTE scores ranged narrowly from 68% to 70%. This indicates that the final model set used for downstream analysis was technically consistent according to a standardized model quality framework (Lieven et al., 2020). The similarity of MEMOTE scores across the eight models is important because it reduces the likelihood that later differences in community-level outputs were driven primarily by strongly unequal model quality.

The medium comparison was also relevant to this objective. Since environmental constraints strongly shape feasible uptake and exchange fluxes in constraint-based models, careful medium

definition is necessary for interpretable simulation results (Diener et al., 2020). In this thesis, the BG11-based minimal medium was therefore used because it reduced nutrient-driven variation introduced by richer alternatives. The final decision to use draft rather than gap-filled models is also directly tied to the first objective. Gap filling can improve formal model connectivity and restore growth, but it may also introduce reactions not clearly supported by genomic evidence (Mendoza *et al.*, 2019). Based on that, the draft models provided the more conservative representation of encoded metabolic potential.

Taken together, these results show that genome-scale model reconstruction was feasible for both the cyanobacterial host and associated heterotrophic bacteria, and that a coherent model set suitable for downstream interaction analysis could be established from the analyzed microbiome, and this approach can successfully be applied to other aquatic host-microbiome systems.

2. Integration of individual genome-scale metabolic models into community-level reconstructions

The cyanobacterial community interaction structure was examined with MICOM, SMETANA, and metage2metabo. These approaches returned different types of outputs, but together they provided complementary views of potential metabolic interactions. As the output of SMETANA was largely similar to MICOM, but consisted of less reactions, it was largely withheld from further analysis.

For AFA_2012_KM_D3, MICOM returned the highest number of predicted exchange reactions (322) involving 57 unique metabolites, whereas SMETANA returned 188 reactions involving 61 unique metabolites. The gapseq-based screening identified a much smaller set of 8 produced metabolites. These differences are consistent with the fact that the methods are based on different modelling frameworks and optimize different quantities.

A complementary perspective was provided by metage2metabo, which does not quantify pairwise exchanges directly but instead evaluates changes in metabolic scope between individual and community contexts. In AFA_2012_KM_D3, all models showed an increase in the number of producible metabolites when analyzed in community context rather than alone. Absolute gains ranged from 88 metabolites in *Hydrogenophaga sp*. to 403 metabolites in *Sediminibacterium sp*., with corresponding relative gains from 10.5% to 141.0%. These values indicate that all AFA_2012_KM_D3 members had broader predicted metabolic scope in community context than

in isolation, although the magnitude of that increase differed across taxa. Taken together, these results show that community integration was successful, but also that interaction outputs are method-dependent. This does not invalidate any one approach; rather, it means that conclusions about predicted interaction structure must be framed in relation to the assumptions and output type of the modelling tool used. In this thesis, MICOM, SMETANA, gapseq, and metage2metabo were therefore informative in different ways rather than interchangeable.

Once the feasibility of community modelling had been established, the next question was whether these predicted interactions were conserved across strain-associated microbiomes or varied from one host-associated community to another.

3. Shared and strain-specific metabolic interaction patterns across microbiomes

Comparison of four reconstructed microbiomes showed that predicted exchange patterns were not uniform across host-associated communities. Instead, MICOM identified both a set of shared metabolites and microbiome-specific sets of uniquely provided and uniquely received compounds.

Across all four microbiomes, shared provided metabolites included Fe^{2+}, H_2O_2, H_2S, and pyruvate, while shared received metabolites included CO_2, CoA, heme, L-arginine, L-cysteine, L-glutamate, L-histidine, L-threonine, NAD, NH_3, and riboflavin. These recurrent metabolites indicate that some predicted exchange features were conserved across multiple strain-associated communities. At the same time, each microbiome retained a distinct unique exchange profile. Among the four, AFA_2012_KM_D3 showed the broadest set of unique provided and unique received metabolites, whereas AFAB4C, AFAB4F, and AFAB10C each had smaller unique metabolite sets.

This shows that community-scale metabolic interaction modelling did not produce a single generic cyanobacterial microbiome interaction network. Instead, even under the same general modelling framework, the predicted metabolite exchange profile varied among host-associated microbiomes. These metabolites remain computationally predicted candidates rather than experimentally demonstrated cross-feeding compounds, but the comparison nevertheless shows that reconstructed microbiomes differ in their inferred interaction structure. To understand why these interaction profiles differed, it is necessary to consider the taxonomic composition of the microbiomes and the genomic variation present in the host strain collection.

The classes of compounds recovered in these analyses - carbon compounds, amino acids, and cofactors - are also consistent with metabolite categories frequently discussed in phytoplankton-bacteria interactions more broadly (Ramanan *et al.*, 2016).

4. Microbiome composition and AFA genomic variation

The taxonomic analysis showed that A. flos-aquae-associated microbiomes shared a recurring higher-rank structure but differed substantially at lower taxonomic levels. Pseudomonadota and Bacteroidota were present in all cultures, and several higher-rank groups recurred across the strain collection. This pattern is consistent with previous observations that phytoplankton-associated microbiomes often contain recurring members of these broad bacterial lineages (Cirri and Pohnert, 2019). However, no strict genus-level core microbiome was detected. Thus, while the microbiomes shared a broad taxonomic framework, they were not compositionally identical.

The geographic analysis showed that region explained a statistically significant but limited fraction of the variation in genus-level microbiome composition. These results indicate that the microbiome dataset contained structure, but not strong discrete clustering by geography. This is important because it means that variation among microbiomes cannot be attributed simply to regional grouping. Host genomic analyses provided a second layer of context. Pairwise ANI values across the 44-genome *A. flos-aquae* dataset ranged from approximately 74.7% to 100.0%, showing that the host strain collection included both highly similar genomes and substantially divergent assemblies. The pangenome contained 14,296 gene clusters, of which 2,684 were core, 5,629 accessory, and 5,983 singleton clusters. The *Prochlorococcus* pangenome reported by Delmont and Eren (2018) comprised 7,385 gene clusters derived from 31 isolate genomes, with a core genome of 766 clusters (10.4% of all clusters) and singletons accounting for 30%. The *A. flos-aquae* pangenome recovered in the present study was notably larger. The higher absolute number of gene clusters in the AFA dataset likely reflects both the larger and more genomically diverse strain collection, indicating that the 44 genomes included substantially more divergent assemblies than a set of closely related isolates within a single species. The proportionally larger singleton fraction in the AFA pangenome further suggests a high degree of strain-specific gene content, consistent with an open pangenome architecture in which accessory and unique genes vastly outnumber the conserved core. Thus, most of the pangenome fell outside the strict core genome, indicating substantial host gene-content variation across the strain collection. This is

consistent with general interpretations of bacterial pangenomes as combinations of a conserved shared backbone and extensive variable content (McInerney, McNally and O'Connell, 2017).

Rarefaction analysis further supported this interpretation of the *A. flos-aquae* pangenome as relatively open. The gene-cluster accumulation curve did not approach a clear plateau across the 44 analyzed genomes, indicating that additional genome sampling would likely continue to reveal novel gene clusters. Consistently, Heaps' Law fitting yielded an α value of 0.36 (K = 3595), which falls within the range commonly interpreted as a relatively open pangenome (Parmigiani, Wittler and Stoye, 2024). In practical terms, this suggests that the genomic diversity captured in the present dataset is substantial but still incomplete, and that the strain collection likely represents only part of the total gene-content diversity associated with *A. flos-aquae*. This interpretation is also consistent with the large accessory and singleton fractions observed here. At the same time, the rarefaction and Heaps' Law estimates should be interpreted cautiously, because they are strongly shaped by the composition of the current genome set. In particular, the broad ANI range within the dataset shows that the analyzed genomes include both very closely related strains and highly divergent assemblies, so the estimated degree of openness reflects the present sampling scheme rather than an absolute species-wide property.

Inspection of the three singleton-enriched sections of the pangenome refined this observation. These regions were dominated by genes of unknown function, but the annotated fraction included genes assigned to transcription, replication, recombination and repair, cell wall / membrane / envelope biogenesis, signal transduction mechanisms, defense mechanisms, and several metabolic categories, including secondary metabolite biosynthesis, transport and catabolism. This shows that the strain-specific component of the host pangenome was not restricted to hypothetical proteins alone, but also contained genes associated with several defined cellular and metabolic functions.

The predominance of unknown-function genes in singleton-rich regions is in line with broader pangenome observations. In a metapangenomic analysis of *Prochlorococcus*, a lower annotation rates for singleton clusters than for core clusters was reported and also emphasized that gene-cluster based genome comparisons can reveal biologically meaningful structure not fully captured by marker-based phylogeny alone (Delmont and Eren, 2018). The present study did not perform an equivalent metapangenomic analysis for *A. flos-aquae*, so the pangenome results here

describe variation in gene-cluster composition across genomes, but not the environmental prevalence of those gene clusters. In this thesis, the pangenome therefore serves primarily as genomic context for strain-associated model variation.

When considered together, the taxonomic, ANI, and pangenome results support the comparative part of the third thesis objective. Host strains differed in genome content, microbiomes differed in taxonomic composition and predicted metabolic interaction profiles also differed across reconstructed communities. However, the present data does not establish direct causal links between specific host gene clusters and specific exchange metabolites or microbiome assemblages. The role of cyanobacterial genotype was therefore evaluated comparatively rather than resolved mechanistically at single-gene level.

This distinction is important for framing the conclusions of the study and leads directly to the main limitations of the work.

5. Considerations and limitations

Several limitations define the scope of the conclusions drawn from this thesis. First, the analysed microbiomes were derived from non-axenic laboratory cultures rather than directly from environmental samples. The associated bacterial communities therefore represent culture-associated microbiomes maintained under laboratory conditions and may not fully reflect the structure of the original field-associated communities.

Second, the metabolic analyses depended on MAG-derived draft models. As a result, all community-level predictions remain sensitive to assembly quality, binning accuracy, annotation completeness, and the specific reconstruction pipeline used. Although model quality was assessed and remained comparatively uniform across the used set, draft models are still simplified representations of encoded metabolic capacity.

Third, community-scale predictions were strongly influenced by modelling assumptions. Medium composition, objective functions, and tool-specific algorithms all shaped the resulting outputs. The differences observed among MICOM, SMETANA, gapseq, and metage2metabo illustrate this clearly. Accordingly, the predicted interaction networks should not be interpreted as direct descriptions of realized metabolism, but as model-based interaction hypotheses derived under defined computational constraints.

A further limitation applies to the host comparative genomics analysis. The pangenome was not linked to metagenomic read recruitment and therefore did not distinguish between gene clusters present in genomes and gene clusters consistently represented in environmental populations, as in metapangenomic studies such as Delmont and Eren (2018). The current pangenome analysis was therefore limited to genomic composition across strains.

Finally, the predicted metabolites identified in this thesis were not validated experimentally. The reported exchanges should therefore be interpreted as candidate metabolites rather than confirmed cross-feeding compounds. Experimental validation would be required to determine whether the predicted transfers occur under culture conditions and whether they contribute measurably to community function.

6. Concluding remarks

In summary, this thesis established a genome-informed framework for modelling metabolic interactions within the cyanobacterial microbiome. Genome-scale metabolic models were reconstructed for *A. flos-aquae* host-associated communities, integrated into community-level analyses, and compared across multiple modelling approaches. These analyses showed that predicted metabolic interaction networks were both community-specific and method-dependent. AFA_2012_KM_D3 provided a case study, with a full host-associated model set, a MICOM-derived exchange network, and measurable community-context metabolic gains across all eight models.

At the same time, the host comparative genomics analyses showed that the *A. flos-aquae* strain collection was genomically heterogeneous, with extensive non-core gene content and singleton-enriched regions dominated by genes of unknown function but also containing annotated functions related to regulation, cellular structure, defense, and metabolism. Together with the taxonomic results, these findings provide context for the observed differences among strain-associated microbiomes and their predicted interaction profiles.

Overall, the study met its core objective by showing that metabolic interaction modelling within *A. flos-aquae* microbiomes is feasible and informative at both individual-community and comparative levels. The resulting predictions define a set of candidate interaction structures and candidate exchanged metabolites that can serve as a basis for future validation and refinement.

CONCLUSIONS

This thesis aimed to characterize metabolic interactions in *Aphanizomenon flos-aquae* microbiomes and to evaluate the role of cyanobacterial genotypes in structuring microbiome composition and predicted metabolic cross-feeding networks.

The first objective was achieved by reconstructing genome-scale metabolic models for *A. flos-aquae* and associated heterotrophic bacteria from metagenome-derived MAGs.

The second objective was achieved by integrating these models into community-level analyses using MICOM, SMETANA, and metage2metabo. These approaches showed that predicted interaction networks differed in size and composition depending on the modelling framework used.

The third objective was addressed comparatively by combining host genomic, microbiome composition, and community modelling analyses. The results showed that host genomes differed in ANI and pangenome composition, microbiomes differed in taxonomic composition and predicted metabolic exchange profiles differed among reconstructed communities, but no exact causative links were established.

Overall, this study shows that metabolic interaction modelling within *A. flos-aquae*-associated microbiomes is feasible and that predicted interaction structure is both community-specific and method-dependent. These results provide a genome-informed basis for future experimental validation of candidate metabolic interactions within cyanobacterial microbiomes.

ACKNOWLEDGEMENTS

Artificial intelligence (Claude Sonnet 4.6) was used in the creation of this work for general grammar, style, and tone correction, as well as for debugging the code used for the bioinformatical analysis.

I would like to express my gratitude to the following individuals and organizations who made this work possible:

Dr. Sigitas Šulčius for the supervision, providing advice and feedback both during the writing of this work, as well as on the interpretation of the results of the research.

Dr. Gediminas Alzbutas for the initial help and advice on the bioinformatical analysis.

The Laboratory of Algology and Microbial Ecology of the Nature Research Centre for providing cyanobacteria cultures, laboratory facilities, as well as carrying out the wet lab work.

The Research Council of Lithuania for providing funding under the "Interactome of *Aphanizomenon flos-aquae* for environmental and biotechnological advancement" **(**INTEC**)** project (P-MIP-25-48), under the Researcher Groups Projects initiative, which allowed to greatly extend the scope of this research.

Vilnius University for organizing the Systems Biology study program, and specifically Dr. Ingrida Domarkienė for exceptional work supervising the program and supporting my educational journey.

On a more personal note, I would be remiss not to mention people special to me for their love and encouragement, in particular:

Olivia Mona Mekionis, whose belief in me has kept my spirits and motivation high during this process, in ways that were more helpful than she might recognize.

Matas Gudauskas for the opportunity to forget the stresses of academic life for a moment, as well as help with translation and proofreading of select sections.

Ieva Laucevičiūtė for the unyielding support over the years, as well as the unexpected help with coming up with some of the ideas that were implemented in this work.

REFERENCES

1. Ackermann, M. (2015) ‘A functional perspective on phenotypic heterogeneity in microorganisms’, *Nature Reviews. Microbiology*, 13(8), pp. 497–508. Available at: https://doi.org/10.1038/nrmicro3491.

2. Afarin, M. and Naeimpoor, F. (2024) ‘Effect of microbial interactions on performance of community metabolic modeling algorithms: flux balance analysis (FBA), community FBA (cFBA) and SteadyCom’, *Bioprocess and Biosystems Engineering*, 47(11), pp. 1833–1848. Available at: https://doi.org/10.1007/s00449-024-03072-7.

3. Anthropic (2026) Claude Sonnet 4.6 [Large language model]. Available at: https://www.anthropic.com.

4. Amin, S.A. *et al.* (2015) ‘Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria’, *Nature*, 522(7554), pp. 98–101. Available at: https://doi.org/10.1038/nature14488.

5. Antoniewicz, M.R. (2015) ‘Methods and advances in metabolic flux analysis: a mini-review’, *Journal of Industrial Microbiology and Biotechnology*, 42(3), pp. 317–325. Available at: https://doi.org/10.1007/s10295-015-1585-x.

6. Aramaki, T. *et al.* (2020) ‘KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold’, *Bioinformatics*, 36(7), pp. 2251–2252. Available at: https://doi.org/10.1093/bioinformatics/btz859.

7. Arévalo, S. *et al.* (2021) ‘Coexistence of Communicating and Noncommunicating Cells in the Filamentous Cyanobacterium Anabaena’, *mSphere*, 6(1), p. 10.1128/msphere.01091-20. Available at: https://doi.org/10.1128/msphere.01091-20.

8. Baart, G.J.E. and Martens, D.E. (2012) ‘Genome-Scale Metabolic Models: Reconstruction and Analysis’, in M. Christodoulides (ed.) *Neisseria meningitidis: Advanced Methods and Protocols*. Totowa, NJ: Humana Press, pp. 107–126. Available at: https://doi.org/10.1007/978-1-61779-346-2_7.

9. Basile, A. *et al.* (2020) ‘Revealing metabolic mechanisms of interaction in the anaerobic digestion microbiome by flux balance analysis’, *Metabolic Engineering*, 62, pp. 138–149. Available at: https://doi.org/10.1016/j.ymben.2020.08.013.

10. Bauer, E. *et al.* (2017) ‘BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities’, *PLOS Computational Biology*, 13(5), p. e1005544. Available at: https://doi.org/10.1371/journal.pcbi.1005544.

11. Benoit, G. *et al.* (2026) ‘High-quality metagenome assembly from nanopore reads with nanoMDBG’, *Nature Communications*, 17(1), p. 3556. Available at: https://doi.org/10.1038/s41467-026-69760-y.

12. Bernstein, D.B. *et al.* (2021) ‘Addressing uncertainty in genome-scale metabolic model reconstruction and analysis’, *Genome Biology*, 22(1), p. 64. Available at: https://doi.org/10.1186/s13059-021-02289-z.

13. Carmichael, W.W., Drapeau, C. and Anderson, D.M. (2000) 'Harvesting of Aphanizomenon flos-aquae Ralfs ex Born. & Flah. var. flos-aquae (Cyanobacteria) from Klamath Lake for human dietary use', *Journal of Applied Phycology*, 12(6), pp. 585–595. Available at: https://doi.org/10.1023/A:1026506713560.

14. Chan, S.H.J., Simons, M.N. and Maranas, C.D. (2017) 'SteadyCom: Predicting microbial abundances while ensuring community stability', *PLOS Computational Biology*, 13(5), p. e1005539. Available at: https://doi.org/10.1371/journal.pcbi.1005539.

15. Chaumeil, P.-A. *et al.* (2022) 'GTDB-Tk v2: memory friendly classification with the genome taxonomy database', *Bioinformatics*, 38(23), pp. 5315–5316. Available at: https://doi.org/10.1093/bioinformatics/btac672.

16. Chen, M.-Y. *et al.* (2021) 'Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation', *The ISME Journal*, 15(1), pp. 211–227. Available at: https://doi.org/10.1038/s41396-020-00775-z.

17. Chen, Y. *et al.* (2021) 'Efficient assembly of nanopore reads via highly accurate and intact error correction', *Nature Communications*, 12(1), p. 60. Available at: https://doi.org/10.1038/s41467-020-20236-7.

18. Cirri, E. and Pohnert, G. (2019) 'Algae–bacteria interactions that balance the planktonic microbiome', *New Phytologist*, 223(1), pp. 100–106. Available at: https://doi.org/10.1111/nph.15765.

19. Coyte, K.Z., Schluter, J. and Foster, K.R. (2015) 'The ecology of the microbiome: Networks, competition, and stability', *Science*, 350(6261), pp. 663–666. Available at: https://doi.org/10.1126/science.aad2602.

20. Croft, M.T. *et al.* (2005) 'Algae acquire vitamin B12 through a symbiotic relationship with bacteria', *Nature*, 438(7064), pp. 90–93. Available at: https://doi.org/10.1038/nature04056.

21. Culp, E.J. and Goodman, A.L. (2023) 'Cross-feeding in the gut microbiome: Ecology and mechanisms', *Cell Host & Microbe*, 31(4), pp. 485–499. Available at: https://doi.org/10.1016/j.chom.2023.03.016.

22. Dal Bello, M. *et al.* (2021) 'Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism', *Nature Ecology & Evolution*, 5(10), pp. 1424–1434. Available at: https://doi.org/10.1038/s41559-021-01535-8.

23. De Coster, W. and Rademakers, R. (2023) 'NanoPack2: population-scale evaluation of long-read sequencing data', *Bioinformatics*, 39(5), p. btad311. Available at: https://doi.org/10.1093/bioinformatics/btad311.

24. Delmont, T.O. and Eren, A.M. (2018) 'Linking pangenomes and metagenomes: the Prochlorococcus metapangenome', *PeerJ*, 6, p. e4320. Available at: https://doi.org/10.7717/peerj.4320.

25. Dewar, A.E. *et al.* (2024) 'Bacterial lifestyle shapes pangenomes', *Proceedings of the National Academy of Sciences*, 121(21), p. e2320170121. Available at: https://doi.org/10.1073/pnas.2320170121.

26. Diener, C., Gibbons, S.M. and Resendis-Antonio, O. (2020) 'MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota', *mSystems*, 5(1), p. 10.1128/msystems.00606-19. Available at: https://doi.org/10.1128/msystems.00606-19.

27. Ding, W., Baumdicker, F. and Neher, R.A. (2018) 'panX: pan-genome analysis and exploration', *Nucleic Acids Research*, 46(1), p. e5. Available at: https://doi.org/10.1093/nar/gkx977.

28. Dukovski, I. *et al.* (2021) 'A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS)', *Nature Protocols*, 16(11), pp. 5030–5082. Available at: https://doi.org/10.1038/s41596-021-00593-3.

29. Eddy, S.R. (1998) 'Profile hidden Markov models.', *Bioinformatics*, 14(9), pp. 755–763. Available at: https://doi.org/10.1093/bioinformatics/14.9.755.

30. Fang, X., Lloyd, C.J. and Palsson, B.O. (2020) 'Reconstructing organisms in silico: genome-scale models and their emerging applications', *Nature Reviews Microbiology*, 18(12), pp. 731–743. Available at: https://doi.org/10.1038/s41579-020-00440-4.

31. Faust, K. and Raes, J. (2012) 'Microbial interactions: from networks to models', *Nature Reviews Microbiology*, 10(8), pp. 538–550. Available at: https://doi.org/10.1038/nrmicro2832.

32. Frioux, C. *et al.* (2020) 'From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes', *Computational and Structural Biotechnology Journal*, 18, pp. 1722–1734. Available at: https://doi.org/10.1016/j.csbj.2020.06.028.

33. Fuster, M. *et al.* (2023) 'Cyanosphere Dynamic During Dolichospermum Bloom: Potential Roles in Cyanobacterial Proliferation', *Microbial Ecology*, 87(1), p. 3. Available at: https://doi.org/10.1007/s00248-023-02317-4.

34. Garcia-Pichel, F. *et al.* (2020) 'What's in a name? The case of cyanobacteria', *Journal of Phycology*, 56(1), pp. 1–5. Available at: https://doi.org/10.1111/jpy.12934.

35. van Gestel, J., Vlamakis, H. and Kolter, R. (2015) 'Division of Labor in Biofilms: the Ecology of Cell Differentiation', *Microbiology Spectrum*, 3(2), p. 10.1128/microbiolspec.mb-0002–2014. Available at: https://doi.org/10.1128/microbiolspec.mb-0002-2014.

36. Gobler, C.J. *et al.* (2016) 'The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms', *Harmful Algae*, 54, pp. 87–97. Available at: https://doi.org/10.1016/j.hal.2016.01.010.

37. Gornicki, P., Scappino, L.A. and Haselkorn, R. (1993) 'Genes for two subunits of acetyl coenzyme A carboxylase of Anabaena sp. strain PCC 7120: biotin carboxylase and biotin carboxyl carrier protein', *Journal of Bacteriology*, 175(16), pp. 5268–5272. Available at: https://doi.org/10.1128/jb.175.16.5268-5272.1993.

38. Grant, M.A.A. *et al.* (2014) 'Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures', *The ISME journal*, 8(7), pp. 1418–1427. Available at: https://doi.org/10.1038/ismej.2014.9.

39. Grossart, H.-P. *et al.* (2020) 'Linking metagenomics to aquatic microbial ecology and biogeochemical cycles', *Limnology and Oceanography*, 65(S1), pp. S2–S20. Available at: https://doi.org/10.1002/lno.11382.

40. Grosskopf, T. and Soyer, O.S. (2014) 'Synthetic microbial communities', *Current Opinion in Microbiology*, 18(100), pp. 72–77. Available at: https://doi.org/10.1016/j.mib.2014.02.002.

41. Gu, C. *et al.* (2019) 'Current status and applications of genome-scale metabolic models', *Genome Biology*, 20, p. 121. Available at: https://doi.org/10.1186/s13059-019-1730-3.

42. Helliwell, K.E. *et al.* (2016) 'Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin B12', *Current biology: CB*, 26(8), pp. 999–1008. Available at: https://doi.org/10.1016/j.cub.2016.02.041.

43. Hessen, D.O. *et al.* (2010) 'Genome streamlining and the elemental costs of growth', *Trends in Ecology & Evolution*, 25(2), pp. 75–80. Available at: https://doi.org/10.1016/j.tree.2009.08.004.

44. Hottes, A.K. *et al.* (2013) 'Bacterial Adaptation through Loss of Function', *PLOS Genetics*, 9(7), p. e1003617. Available at: https://doi.org/10.1371/journal.pgen.1003617.

45. Hugerth, L.W. *et al.* (2015) 'Metagenome-assembled genomes uncover a global brackish microbiome', *Genome Biology*, 16(1), p. 279. Available at: https://doi.org/10.1186/s13059-015-0834-7.

46. Huisman, J. *et al.* (2018) 'Cyanobacterial blooms', *Nature Reviews Microbiology*, 16(8), pp. 471–483. Available at: https://doi.org/10.1038/s41579-018-0040-1.

47. Johnson, D.R. *et al.* (2012) 'Metabolic specialization and the assembly of microbial communities', *The ISME Journal*, 6(11), pp. 1985–1991. Available at: https://doi.org/10.1038/ismej.2012.46.

48. Kang, D.D. *et al.* (2019) 'MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies', *PeerJ*, 7, p. e7359. Available at: https://doi.org/10.7717/peerj.7359.

49. Kaste, J.A.M. and Shachar-Hill, Y. (2024) 'Model validation and selection in metabolic flux analysis and flux balance analysis', *Biotechnology Progress*, 40(1), p. e3413. Available at: https://doi.org/10.1002/btpr.3413.

50. King, Z.A. *et al.* (2016) 'BiGG Models: A platform for integrating, standardizing and sharing genome-scale models', *Nucleic Acids Research*, 44(D1), pp. D515–D522. Available at: https://doi.org/10.1093/nar/gkv1049.

51. Komárek, J. and Komárková, J. (2006) 'Diversity of Aphanizomenon-like cyanobacteria', *Fottea*, 6(1), pp. 1–32.

52. Kononen, K. and Nômmann, S. (1992) 'Spatio-Temporal Dynamics of the Cyanobacterial Blooms in the Gulf of Finland, Baltic Sea', in E.J. Carpenter, D.G. Capone, and J.G. Rueter (eds) *Marine Pelagic Cyanobacteria: Trichodesmium and other Diazotrophs*. Dordrecht: Springer Netherlands, pp. 95–113. Available at: https://doi.org/10.1007/978-94-015-7977-3_7.

53. Kreft, J.-U., Griffin, B.M. and González-Cabaleiro, R. (2020) 'Evolutionary causes and consequences of metabolic division of labour: why anaerobes do and aerobes don't', *Current Opinion in Biotechnology*, 62, pp. 80–87. Available at: https://doi.org/10.1016/j.copbio.2019.08.008.

54. Lakshmanan, M. *et al.* (2014) 'Software applications for flux balance analysis', *Briefings in Bioinformatics*, 15(1), pp. 108–122. Available at: https://doi.org/10.1093/bib/bbs069.

55. Lan, R. *et al.* (2000) 'Intraspecies variation in bacterial genomes: the need for a species genome concept', *Trends in Microbiology*, 8(9), pp. 396–401. Available at: https://doi.org/10.1016/S0966-842X(00)01791-1.

56. Latorre-Pérez, A. *et al.* (2020) 'Assembly methods for nanopore-based metagenomic sequencing: a comparative study', *Scientific Reports*, 10(1), p. 13588. Available at: https://doi.org/10.1038/s41598-020-70491-3.

57. Lee, J.Y. *et al.* (2021) 'Comparative evaluation of Nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis', *Scientific Reports*, 11(1), p. 20740. Available at: https://doi.org/10.1038/s41598-021-00178-w.

58. Louati, I. *et al.* (2015) 'Structural Diversity of Bacterial Communities Associated with Bloom-Forming Freshwater Cyanobacteria Differs According to the Cyanobacterial Genus', *PLOS ONE*, 10(11), p. e0140614. Available at: https://doi.org/10.1371/journal.pone.0140614.

59. Ma, J. *et al.* (2022) 'Cross-feeding between cyanobacterium Synechococcus and Escherichia coli in an artificial autotrophic–heterotrophic coculture system revealed by integrated omics analysis', *Biotechnology for Biofuels and Bioproducts*, 15(1), p. 69. Available at: https://doi.org/10.1186/s13068-022-02163-5.

60. Machado, D. *et al.* (2018) 'Fast automated reconstruction of genome-scale metabolic models for microbial species and communities', *Nucleic Acids Research*, 46(15), pp. 7542–7553. Available at: https://doi.org/10.1093/nar/gky537.

61. Marin, M.G. *et al.* (2025) 'Pitfalls of bacterial pan-genome analysis approaches: a case study of Mycobacterium tuberculosis and two less clonal bacterial species', *Bioinformatics*, 41(5), p. btaf219. Available at: https://doi.org/10.1093/bioinformatics/btaf219.

62. Markle, H.V. and Greenway, D.C. (1996) 'Cobalamin', *Critical Reviews in Clinical Laboratory Sciences*, 33(4), pp. 247–356. Available at: https://doi.org/10.3109/10408369609081009.

63. Mataigne, V. *et al.* (2021) 'Microbial Systems Ecology to Understand Cross-Feeding in Microbiomes', *Frontiers in Microbiology*, 12. Available at: https://doi.org/10.3389/fmicb.2021.780469.

64. Matthews, C.A. *et al.* (2024) 'A gentle introduction to pangenomics', *Briefings in Bioinformatics*, 25(6), p. bbae588. Available at: https://doi.org/10.1093/bib/bbae588.

65. McCloskey, D., Palsson, B.Ø. and Feist, A.M. (2013) 'Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli', *Molecular Systems Biology*, 9, p. 661. Available at: https://doi.org/10.1038/msb.2013.18.

66. McInerney, J.O., McNally, A. and O'Connell, M.J. (2017) 'Why prokaryotes have pangenomes', *Nature Microbiology*, 2(4), p. 17040. Available at: https://doi.org/10.1038/nmicrobiol.2017.40.

67. Mendoza, S.N. *et al.* (2019) 'A systematic assessment of current genome-scale metabolic reconstruction tools', *Genome Biology*, 20(1), p. 158. Available at: https://doi.org/10.1186/s13059-019-1769-1.

68. Moran, N.A., McLaughlin, H.J. and Sorek, R. (2009) 'The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria', *Science*, 323(5912), pp. 379–382. Available at: https://doi.org/10.1126/science.1167140.

69. Morris, J.J., Lenski, R.E. and Zinser, E.R. (2012) 'The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss', *mBio*, 3(2), pp. e00036-12. Available at: https://doi.org/10.1128/mBio.00036-12.

70. Mridha, S. and Kümmerli, R. (2022) 'Enforced specialization fosters mutual cheating and not division of labour in the bacterium Pseudomonas aeruginosa', *Journal of Evolutionary Biology*, 35(5), pp. 719–730. Available at: https://doi.org/10.1111/jeb.14001.

71. Nelson, C. and Garcia-Pichel, F. (2021) 'Beneficial Cyanosphere Heterotrophs Accelerate Establishment of Cyanobacterial Biocrust', *Applied and Environmental Microbiology*, 87(20), p. e0123621. Available at: https://doi.org/10.1128/AEM.01236-21.

72. Nelson, C., Giraldo-Silva, A. and Garcia-Pichel, F. (2021) 'A symbiotic nutrient exchange within the cyanosphere microbiome of the biocrust cyanobacterium, Microcoleus vaginatus', *The ISME journal*, 15(1), pp. 282–292. Available at: https://doi.org/10.1038/s41396-020-00781-1.

73. Nilsson, A. and Nielsen, J. (2017) 'Genome scale metabolic modeling of cancer', *Metabolic Engineering*, 43, pp. 103–112. Available at: https://doi.org/10.1016/j.ymben.2016.10.022.

74. O'Brien, E.J., Monk, J.M. and Palsson, B.O. (2015) 'Using Genome-scale Models to Predict Biological Capabilities', *Cell*, 161(5), pp. 971–987. Available at: https://doi.org/10.1016/j.cell.2015.05.019.

75. Opdam, S. *et al.* (2017) 'A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models', *Cell Systems*, 4(3), pp. 318-329.e6. Available at: https://doi.org/10.1016/j.cels.2017.01.010.

76. Pacheco, A.R., Moel, M. and Segrè, D. (2019) 'Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems', *Nature Communications*, 10(1), p. 103. Available at: https://doi.org/10.1038/s41467-018-07946-9.

77. Pan, S., Zhao, X.-M. and Coelho, L.P. (2023) 'SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing', *Bioinformatics*, 39(Supplement_1), pp. i21–i29. Available at: https://doi.org/10.1093/bioinformatics/btad209.

78. Parks, D.H. *et al.* (2018) 'A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life', *Nature Biotechnology*, 36(10), pp. 996–1004. Available at: https://doi.org/10.1038/nbt.4229.

79. Parks, D.H. *et al.* (2022) 'GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy', *Nucleic Acids Research*, 50(D1), pp. D785–D794. Available at: https://doi.org/10.1093/nar/gkab776.

80. Parmigiani, L., Wittler, R. and Stoye, J. (2024) 'Revisiting pangenome openness with *k*-mers', *Peer Community Journal*, 4. Available at: https://doi.org/10.24072/pcjournal.415.

81. Pearson, W.R. (2014) 'BLAST and FASTA Similarity Searching for Multiple Sequence Alignment', in D.J. Russell (ed.) *Multiple Sequence Alignment Methods*. Totowa, NJ: Humana Press, pp. 75–101. Available at: https://doi.org/10.1007/978-1-62703-646-7_5.

82. Pell, J. *et al.* (2012) 'Scaling metagenome sequence assembly with probabilistic de Bruijn graphs', *Proceedings of the National Academy of Sciences*, 109(33), pp. 13272–13277. Available at: https://doi.org/10.1073/pnas.1121464109.

83. Pérez-Carrascal, O.M. *et al.* (2020) 'Single-colony sequencing reveals phylosymbiosis, co-phylogeny, and horizontal gene transfer between the cyanobacterium Microcystis and its microbiome'. bioRxiv, p. 2020.11.17.386995. Available at: https://doi.org/10.1101/2020.11.17.386995.

84. Planavsky, N.J. *et al.* (2014) 'Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event', *Nature Geoscience*, 7(4), pp. 283–286. Available at: https://doi.org/10.1038/ngeo2122.

85. Raines, C.A. (2003) 'The Calvin cycle revisited', *Photosynthesis Research*, 75(1), pp. 1–10. Available at: https://doi.org/10.1023/A:1022421515027.

86. Rajaniemi, P. *et al.* (2005a) 'Phylogenetic and morphological evaluation of the genera Anabaena, Aphanizomenon, Trichormus and Nostoc (Nostocales, Cyanobacteria)', *International Journal of Systematic and Evolutionary Microbiology*, 55(1), pp. 11–26. Available at: https://doi.org/10.1099/ijs.0.63276-0.

87. Ramanan, R. *et al.* (2016) 'Algae–bacteria interactions: Evolution, ecology and emerging applications', *Biotechnology Advances*, 34(1), pp. 14–29. Available at: https://doi.org/10.1016/j.biotechadv.2015.12.003.

88. Rolff, C., Almesjö, L. and Elmgren, R. (2007) 'Nitrogen fixation and abundance of the diazotrophic cyanobacterium Aphanizomenon sp. in the Baltic Proper', *Marine Ecology Progress Series*, 332, pp. 107–118. Available at: https://doi.org/10.3354/meps332107.

89. Roothans, N., van Loosdrecht, M.C.M. and Laureni, M. (2025) 'Metabolic labour division trade-offs in denitrifying microbiomes', *The ISME Journal*, 19(1), p. wraf020. Available at: https://doi.org/10.1093/ismejo/wraf020.

90. Rouli, L. *et al.* (2015) 'The bacterial pangenome as a new tool for analysing pathogenic bacteria', *New Microbes and New Infections*, 7, pp. 72–85. Available at: https://doi.org/10.1016/j.nmni.2015.06.005.

91. Sakaki, K. *et al.* (2020) 'A suicide enzyme catalyzes multiple reactions for biotin biosynthesis in cyanobacteria', *Nature Chemical Biology*, 16(4), pp. 415–422. Available at: https://doi.org/10.1038/s41589-019-0461-9.

92. Sánchez-Baracaldo, P. *et al.* (2022) 'Cyanobacteria and biogeochemical cycles through Earth history', *Trends in Microbiology*, 30(2), pp. 143–157. Available at: https://doi.org/10.1016/j.tim.2021.05.008.

93. Santo, É. do E. *et al.* (2022) 'Obtaining Bioproducts from the Studies of Signals and Interactions between Microalgae and Bacteria', *Microorganisms*, 10(10), p. 2029. Available at: https://doi.org/10.3390/microorganisms10102029.

94. Sato, N. (2021) 'Are Cyanobacteria an Ancestor of Chloroplasts or Just One of the Gene Donors for Plants and Algae?', *Genes*, 12(6). Available at: https://doi.org/10.3390/genes12060823.

95. Schirrmeister, B.E., Gugger, M. and Donoghue, P.C.J. (2015) 'Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils', *Palaeontology*, 58(5), pp. 769–785. Available at: https://doi.org/10.1111/pala.12178.

96. Scoglio, G.D., Jackson, H.O. and Purton, S. (2024a) 'The commercial potential of Aphanizomenon flos-aquae, a nitrogen-fixing edible cyanobacterium', *Journal of Applied Phycology*, 36(4), pp. 1593–1617. Available at: https://doi.org/10.1007/s10811-024-03214-0.

97. Scoglio, G.D., Jackson, H.O. and Purton, S. (2024b) 'The commercial potential of Aphanizomenon flos-aquae, a nitrogen-fixing edible cyanobacterium', *Journal of Applied Phycology*, 36(4), pp. 1593–1617. Available at: https://doi.org/10.1007/s10811-024-03214-0.

98. Sereika, M. *et al.* (2022) 'Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing', *Nature Methods*, 19(7), pp. 823–826. Available at: https://doi.org/10.1038/s41592-022-01539-7.

99. Setubal, J.C. (2021) 'Metagenome-assembled genomes: concepts, analogies, and challenges', *Biophysical Reviews*, 13(6), pp. 905–909. Available at: https://doi.org/10.1007/s12551-021-00865-y.

100. Sieber, C.M.K. *et al.* (2018) 'Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy', *Nature Microbiology*, 3(7), pp. 836–843. Available at: https://doi.org/10.1038/s41564-018-0171-1.

101. Sirithanakorn, C. and Cronan, J.E. (2021) 'Biotin, a universal and essential cofactor: synthesis, ligation and regulation', *FEMS Microbiology Reviews*, 45(4). Available at: https://doi.org/FEMS%20Microbiology%20Reviews.

102. Smith, N.W. *et al.* (2019) 'The Classification and Evolution of Bacterial Cross-Feeding', *Frontiers in Ecology and Evolution*, 7. Available at: https://doi.org/10.3389/fevo.2019.00153.

103. Strunecký, O., Ivanova, A.P. and Mareš, J. (2023) 'An updated classification of cyanobacterial orders and families based on phylogenomic and polyphasic analysis', *Journal of Phycology*, 59(1), pp. 12–51. Available at: https://doi.org/10.1111/jpy.13304.

104. Succurro, A. and Ebenhöh, O. (2018) 'Review and perspective on mathematical modeling of microbial ecosystems', *Biochemical Society Transactions*, 46(2), pp. 403–412. Available at: https://doi.org/10.1042/BST20170265.

105. Traxler, M.F. *et al.* (2013) 'Interspecies Interactions Stimulate Diversification of the Streptomyces coelicolor Secreted Metabolome', *mBio* [Preprint]. Available at: https://doi.org/10.1128/mbio.00459-13.

106. Vaser, R. *et al.* (2017) 'Fast and accurate de novo genome assembly from long uncorrected reads', *Genome Research*, 27(5), pp. 737–746. Available at: https://doi.org/10.1101/gr.214270.116.

107. Verma, D., Satyanarayana, T. and Dias, P.J. (2024) 'Editorial: Microbial comparative genomics and pangenomics: new tools, approaches and insights into gene and genome evolution', *Frontiers in Genetics*, 15. Available at: https://doi.org/10.3389/fgene.2024.1490645.

108. Vosloo, S. *et al.* (2021) 'Evaluating de Novo Assembly and Binning Strategies for Time Series Drinking Water Metagenomes', *Microbiology Spectrum*, 9(3), pp. e01434-21. Available at: https://doi.org/10.1128/Spectrum.01434-21.

109. Wang, D. *et al.* (2023) 'Cross-Feedings, Competition, and Positive and Negative Synergies in a Four-Species Synthetic Community for Anaerobic Degradation of Cellulose to Methane', *mBio*, 14(2), p. e0318922. Available at: https://doi.org/10.1128/mbio.03189-22.

110. Wang, Z. *et al.* (2024) 'Effective binning of metagenomic contigs using contrastive multi-view representation learning', *Nature Communications*, 15(1), p. 585. Available at: https://doi.org/10.1038/s41467-023-44290-z.

111. Wick, R.R., Judd, L.M. and Holt, K.E. (2023) 'Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing', *PLOS Computational Biology*, 19(3), p. e1010905. Available at: https://doi.org/10.1371/journal.pcbi.1010905.

112. Wienhausen, G. *et al.* (2022) 'The overlooked role of a biotin precursor for marine bacteria - desthiobiotin as an escape route for biotin auxotrophy', *The ISME Journal*, 16(11), pp. 2599–2609. Available at: https://doi.org/10.1038/s41396-022-01304-w.

113. Xu, X. *et al.* (2026) 'Metagenomic insights into rhizosphere microbiome dynamics of *Oenanthe javanica* in ecological floating beds under different hydrodynamic regimes', *Journal of Contaminant Hydrology*, 277, p. 104795. Available at: https://doi.org/10.1016/j.jconhyd.2025.104795.

114. Yue, Y. *et al.* (2020) 'Evaluating metagenomics tools for genome binning with real metagenomic datasets and CAMI datasets', *BMC Bioinformatics*, 21(1), p. 334. Available at: https://doi.org/10.1186/s12859-020-03667-3.

115. Zengler, K. and Palsson, B.O. (2012) 'A road map for the development of community systems (CoSy) biology', *Nature Reviews. Microbiology*, 10(5), pp. 366–372. Available at: https://doi.org/10.1038/nrmicro2763.

116. Zhang, C. and Hua, Q. (2016) 'Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine', *Frontiers in Physiology*, 6, p. 413. Available at: https://doi.org/10.3389/fphys.2015.00413.

117. Zimmermann, J., Kaleta, C. and Waschina, S. (2021) ‘gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models’, *Genome Biology*, 22(1), p. 81. Available at: https://doi.org/10.1186/s13059-021-02295-1.

118. Zomorrodi, A.R. and Maranas, C.D. (2012) ‘OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities’, *PLOS Computational Biology*, 8(2), p. e1002363. Available at: https://doi.org/10.1371/journal.pcbi.1002363.

SUMMARY

Institution: Vilnius University

Study program: Systems Biology

Name and surname: Petr Sharygin

Title of thesis: Modelling Metabolic Interactions Within the Cyanobacteria Microbiome

This thesis investigated metabolic interactions within *Aphanizomenon flos-aquae*-associated microbiomes and evaluated whether cyanobacterial genotype is linked to microbiome composition and predicted metabolic cross-feeding. Cyanobacterial microbiomes are increasingly recognized as important for host ecology, but their interaction structure remains less understood than their taxonomic composition.

To address this, non-axenic *A. flos-aquae* cultures were analyzed using genome-resolved metagenomics, comparative genomics, and community-scale metabolic modelling. Metagenome-assembled genomes were reconstructed for cyanobacterial hosts and associated heterotrophic bacteria, taxonomic composition was compared across strain-associated microbiomes, and draft genome-scale metabolic models were built for selected microbiome members. Community-level analyses were then performed using MICOM, SMETANA, and metage2metabo. In parallel, host genomic diversity was assessed using Average Nucleotide Identity and pangenome analysis.

The results showed that the microbiomes shared a recurring higher-rank taxonomic structure, dominated by Pseudomonadota and Bacteroidota, but varied substantially at lower taxonomic levels. For the AFA_2012_KM_D3 microbiome, as well as several other ones, a complete set of draft metabolic models was reconstructed for the host and seven associated heterotrophic bacteria. Community modelling identified both shared and microbiome-specific candidate exchange metabolites, while the number and type of predicted interactions differed among modelling frameworks. Comparative genomics showed substantial host genomic variation, including a large accessory and singleton fraction in the *A. flos-aquae* pangenome.

Overall, the study showed that metabolic interaction modelling within *A. flos-aquae*-associated microbiomes is feasible and that predicted interaction networks are both community-specific and

method-dependent. These results provide a genome-informed basis for future experimental validation of candidate metabolic interactions in cyanobacterial microbiomes.

SUMMARY IN LITHUANIAN

Institucija: Vilniaus universitetas

Studijų programa: Sistemų biologija

Vardas ir pavardė: Petr Sharygin

Darbo pavadinimas: Metabolinių Sąveikų Modeliavimas Melsvabakterių Mikrobiome

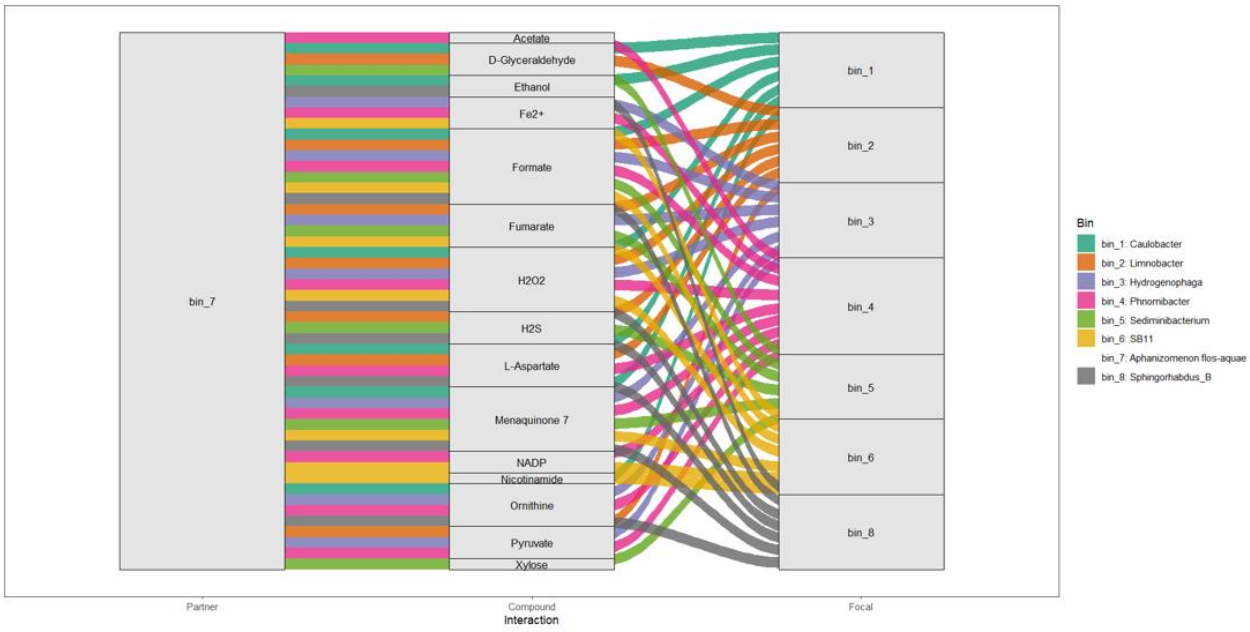
Šiame darbe tirtos metabolinės sąveikos su *Aphanizomenon flos-aquae* susijusiuose mikrobiomuose ir vertinta, ar melsvabakterių genotipas yra susijęs su mikrobiomo sudėtimi bei prognozuojamu metabolitų tarpusavio apsikeitimu. Melsvabakterių mikrobiomai vis dažniau pripažįstami svarbiais šeimininko ekologijai, tačiau jų sąveikų struktūra išlieka mažiau ištirta nei jų taksonominė sudėtis.

Šiai problemai spręsti neakseninės *A. flos-aquae* kultūros buvo tiriamos taikant genomų rekonstrukcija grįstą metagenomiką, tai pat lyginamąją genomiką ir bendruomenės masto metabolinį modeliavimą. Tyrimo metu buvo rekonstruoti iš metagenomų surinkti melsvabakterių šeimininkų ir susijusių heterotrofinių bakterijų genomai, su skirtingomis padermėmis palyginta susijusių mikrobiomų taksonominė sudėtis, o atrinktiems mikrobiomo nariams sudaryti preliminarūs genomo masto metaboliniai modeliai. Taip pat, naudojant MICOM, SMETANA ir metage2metabo metodus, buvo atlikta bendruomenės lygmens analizė. Šeimininko genominė įvairovė buvo lygiagrečiai vertinama taikant vidutinio nukleotidų tapatumo (ANI) ir pangenomo analizės metodus.

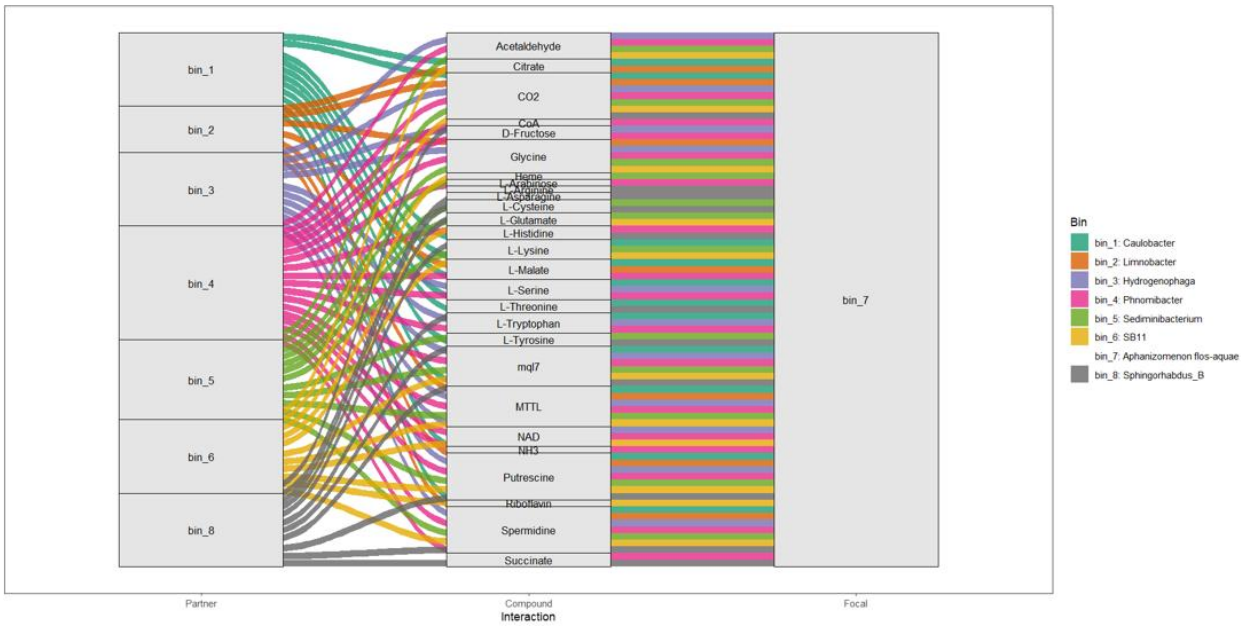
Tyrimų rezultatai parodė, kad mikrobiomams buvo būdinga pasikartojanti aukštesnių taksonominių rangų struktūra, kurioje dominavo Pseudomonadota ir Bacteroidota tipo bakterijos, tačiau žemesniuose taksonominiuose lygmenyse aiškiai dominuojančių bakterijų nenustatyta. AFA_2012_KM_D3 mikrobiomui buvo rekonstruotas pilnas preliminarių metabolinių modelių rinkinys šeimininkui ir septynioms susijusioms heterotrofinėms bakterijoms. Bendruomenės modeliavimu identifikuoti tiek bendri, tiek konkrečiam mikrobiomui būdingi potencialūs apsikeitimo metabolitai, o prognozuotų sąveikų skaičius ir tipas skyrėsi priklausomai nuo taikytos modeliavimo sistemos. Lyginamoji genomika parodė didelę šeimininko genominę įvairovę, įskaitant didelę papildomųjų ir vienetinių genų dalį *A. flos-aquae* pangenome.

Apibendrinant, tyrimas parodė, kad metabolinių sąveikų modeliavimas jau minėtuose mikrobiomuose yra įmanomas, o prognozuojami sąveikų tinklai yra tiek specifiški konkrečioms bendruomenėms, tiek priklausomi nuo taikomo metodo. Šių tyrimų rezultatai sudaro genomine informacija pagrįstą pagrindą būsimam galimų metabolinių sąveikų melsvabakterių mikrobiomuose eksperimentiniam patvirtinimui.

APPENDIX 1: FULL METABOLIC INTERACTIONS NETWORK OF AFA_2012_KM_D3



Supplementary figure 1: Metabolites provided by AFA.



Supplementary figure 2: Metabolites received by AFA.

APPENDIX 2: MICROBIOME COMPOSITION OF STRAINS AFA_2012_KM_D3, AFAB4C, AFAB4F, AND AFAB10C.

Supplementary table 1: Microbiome composition of strain AFA_2012_KM_D3

Phylum	Class	Order	Family	Genus	Species
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*Caulobacter*	
Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	*Limnobacter*	*sp002954425*
Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	*Hydrogenophaga*	
Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	*Phnomibacter*	
Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	*Sediminibacterium*	
Bacteroidota	Bacteroidia	NS11-12g	UBA8524	*SB11*	*sp008933805*
Cyanobacteria	Cyanobacteriia	Cyanobacteriales	Aphanizomenonaceae	*Aphanizomenon*	*flos-aquae*
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingorhabdus_B*	

Supplementary table 2: Microbiome composition of strain AFAB4F

Phylum	Class	Order	Family	Genus	Species
Pseudomonadota	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingorhabdus_B*	*sp023910635*
Cyanobacteriota	Cyanobacteriia	Cyanobacteriales	Nostocaceae	*Dolichospermum*	*flosaquae*
Armatimonadota	Fimbriimonadia	Fimbriimonadales	Fimbriimonadaceae	*Fimbriimonas*	
Pseudomonadota	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	*PBB1*	
Bacteroidota	Bacteroidia	NS11-12g	UKL13-3	*UKL13-3*	*sp001602415*
Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	*JAAFHG01*	

Pseudomonadota	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*CAISGS01*	
Pseudomonadota	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	*Bosea*	*sp023910605*
Planctomycetota	Planctomycetia	Pirellulales	Pirellulaceae	*Pirellula_B*	
Pseudomonadota	Gammaproteobacteria	Nevskiales	Nevskiaceae	*Nevskia*	*sp026398815*
Pseudomonadota	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Erythrobacter*	
Pseudomonadota	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	*Undibacterium*	*sp009937955*
Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	*Flavobacterium*	
Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	*Flavipsychrobacter*	*sp027484425*
Cyanobacteriota	Cyanobacteriia	PCC-6307	Cyanobiaceae	*Cyanobium*	

Supplementary table 3: Microbiome composition of strain AFAB4C

Phylum	Class	Order	Family	Genus	Species
Pseudomonadota	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*CAISGS01*	
Pseudomonadota	Gammaproteobacteria	Burkholderiales	Usitatibacteraceae	*UKL13-2*	*sp001602455*
Bacteroidota	Bacteroidia	NS11-12g	UKL13-3	*UKL13-3*	*sp001602415*
Cyanobacteriota	Cyanobacteriia	Cyanobacteriales	Nostocaceae	*Dolichospermum*	*flosaquae*
Pseudomonadota	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingorhabdus_B*	*sp023910635*

Supplementary table 4: Microbiome composition of strain AFAB10C

Phylum	Class	Order	Family	Genus	Species
Bacteroidota	Bacteroidia	NS11-12g	UKL13-3	*UKL13-3*	*sp001602415*

Pseudomonadota	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	*Sphingorhabdus_B*	*sp023910635*
Cyanobacteriota	Cyanobacteriia	Cyanobacteriales	Nostocaceae	*Dolichospermum*	*flosaquae*
Pseudomonadota	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	*CAISGS01*	
Pseudomonadota	Gammaproteobacteria	Burkholderiales	Usitatibacteraceae	*UKL13-2*	*sp001602455*
Planctomycetota	Phycisphaerae	Phycisphaerales	UBA1924	*UBA2396*	