Proteomes in the flow: proteomic insights into engineered water environments

Tugui, G.

DOI
10.4233/uuid:c4f2dcf4-ccb5-4bc6-b6cc-06d6f4efa87f

Publication date
2025

Document Version
Final published version

Citation (APA)
Tugui, G. (2025). *Proteomes in the flow: proteomic insights into engineered water environments*. [Dissertation (TU Delft), Delft University of Technology]. Ridderprint bv. https://doi.org/10.4233/uuid:c4f2dcf4-ccb5-4bc6-b6cc-06d6f4efa87f

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Proteomes in the flow: proteomic insights into engineered water environments

Claudia G. Tugui

Proteomes in the flow: proteomic insights into engineered water environments

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on Wednesday **24th September 2025** at **10:00**

by

Claudia Gabriela TUGUI

Master of Science in Microbiology, Radboud University, The Netherlands
Born in Ploieşti, Romania

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof.dr.ir. M.C.M van Loosdrecht	Delft University of Technology, promotor
Dr. M. Pabst	Delft University of Technology, promotor

Independent members:

Prof. dr. H. Schmitt	Delft University of Technology, NL
Prof. dr. G.J. Medema	Delft University of Technology, NL
Prof. dr. T. Curtis	Newcastle University, UK
Dr. J. A. A. Demmers	Erasmus University, NL
Dr. T. Muth	Robert Koch Institute, GER
Prof.dr.ir. P.A.S. Daran-Lapujade	Delft University of Technology, NL, reserve member

The research presented in this thesis was performed at the Environmental Biotechnology Section, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands. This work was supported by NWO Spinoza prize awarded to Mark van Loosdrecht and by Evides Waterbedrijf N.V.

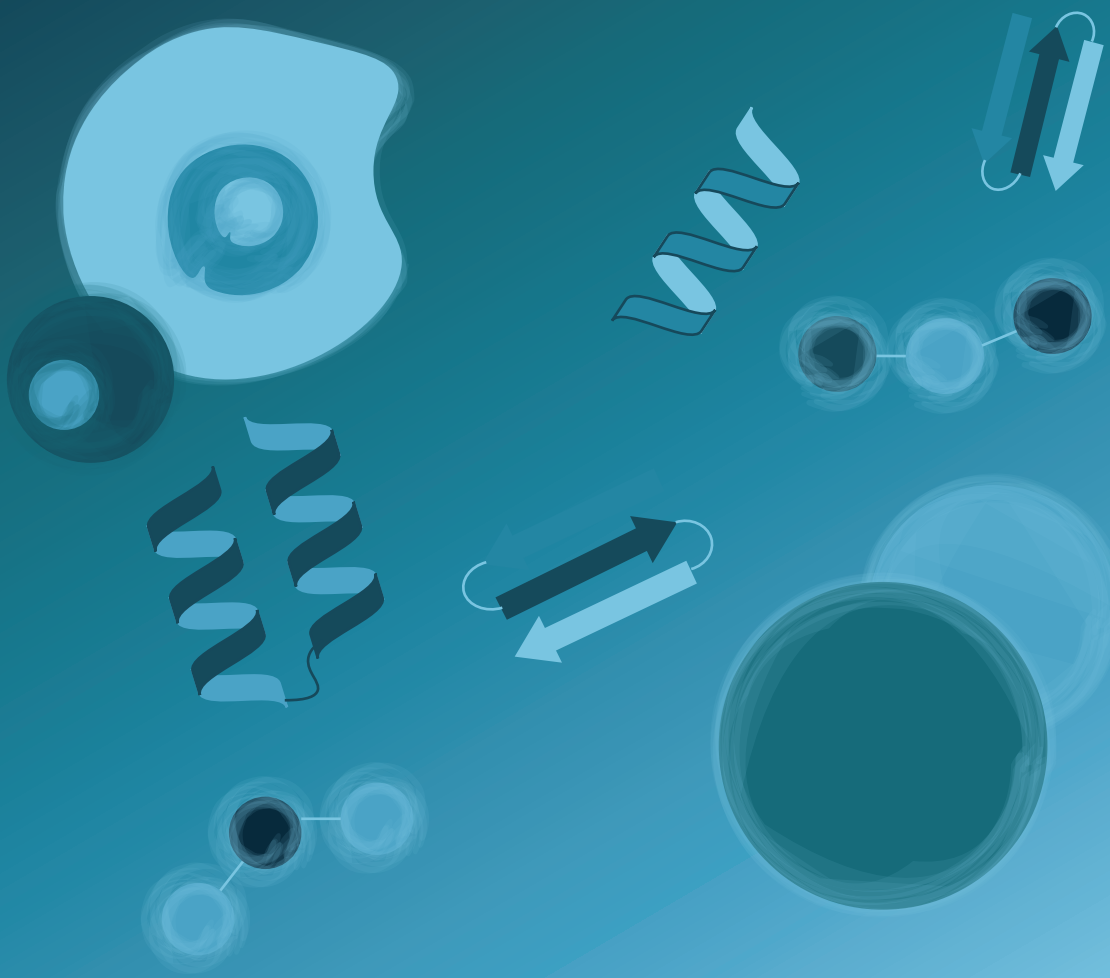
Cover: Claudia Tugui
Layout: Claudia Tugui
Printed by: Ridderprint |www.ridderprint.nl

An electronic version of this dissertation is available at https://repository.tudelft.nl/

Table of Contents

Summary

Proteomes in the flow: proteomic insights into engineered water environments



Summary

Microbes have evolved to thrive in diverse and extreme environments. Understanding these microbes in engineered ecosystems, such as those used for wastewater treatment and drinking water production, is crucial for elucidating their roles in sanitation, nutrient cycling, and overall system stability. Traditional methods, such as microscopy and in situ staining, provide limited insight into microbial diversity and function, particularly for low-abundance species. The rise of culture-independent techniques like 16S rRNA and whole metagenome sequencing has revolutionized microbial ecology, enabling deeper analysis of taxonomic profiles, but does not answer questions beyond the metabolic potential.

Mass spectrometry is a powerful technology which enables the identification and characterization of proteins at a large scale from small amounts of cell material, referred to as proteomics. This has gained wide interest in the scientific community and industries, as proteins mediate fundamental processes in cells, such as enzymatic catalysis, molecular transport, signaling, cell division, and defense mechanisms.

Most recent advances in mass spectrometry allowed to transition the field of proteomics to investigate microbes, and complete microbial communities. Advantageously, microbial proteomics provides insights into the active metabolic pathways in microbes and microbial communities, thereby complementing the information obtained from DNA-based approaches. While microbial proteomics has already been widely employed in various fields of research, including medical applications and biotechnology e.g. for understanding cell factories, much less has been done on environmental microbes, including those found in engineered ecosystems like wastewater and drinking water production systems.

This thesis aims to advance the application of microbial proteomics for studying microbial communities in engineered ecosystems, specifically within wastewater and drinking water environments. By analyzing the proteins present in these environments, this approach provides critical insights into the expressed metabolic functions of individual microbes and the protein biomass composition of microbial ecosystems. Additionally, investigating the microbial secretome provides new insights into metabolic versatility of microorganisms in nutrient-poor environments.

In **Chapter 2** the potential of *Aeromonas* to degrade and use the carbohydrate polymer chitin is investigated in order to understand how this organism can survive in nutrient poor environments such as in drinking water distribution systems. Members of the *Aeromonas* genus are commonly found in natural aquatic ecosystems including in non-chlorinated drinking water distribution systems. Earlier studies suggested that *Aeromonas* may utilize chitin from invertebrates like *Asellus aquaticus* as an energy source, however their underlaying metabolic routes are only poorly understood. Since the secret behind degradation of biological polymers such as chitin relies in the production and export of

specific enzymes called glycoside hydrolases, this chapter investigates both, the cellular proteome and the secretome of two *Aeromonas* species, *A. bestiarum* and *A. rivuli.* Not only were both species able to degrade the polymer, they also efficiently utilized this polymer as sole carbon and nitrogen source. Proteomic analysis revealed a significant reorganization of both the intracellular and secreted proteome when *Aeromonas* was grown on chitin compared to glucose. This included the expression of dedicated glycoside hydrolases and transporters. The ability of *Aeromonas* to efficiently degrade and grow on chitin has prompted the question whether and how this organism can also grow on other biopolymers.

The exploration of *Aeromonas'* capabilities to degrade and utilize other ecologically relevant biopolymers is investigated in **Chapter 3**. *Aeromonas* bestiarum was cultivated on fucoidan, starch, xylan, cellulose, dextrin, collagen, pullulan, pectin and an EPS extract. *A. bestiarum* was able to grow on pullulan, dextrin, starch, collagen and the EPS extract. Proteomic analysis of the secretome revealed many dedicated enzymes related to the degradation of these biopolymers. However, the secretome analysis revealed also that approximately one quarter of the secreted proteins are of unknown function, highlighting current challenges in characterizing proteomes from poorly investigates microbes. Despite this, the study demonstrated *Aeromonas'* metabolic versatility and its potential to survive in oligotrophic environments. An organism capable of degrading multiple biological polymers is called a generalist, positioning it in a crucial place within the food chain, which sheds light on its survival in oligotrophic environments.

However, alongside *Aeromonas*, drinking water distribution systems usually harbor a surprisingly rich microbial ecosystem. The study of such complex environments and their microbial inhabitants requires a spectrum of different methods. Therefore, in **Chapter 4** the complete microbial ecosystem present in loose deposits collected from seven different locations across a drinking water distribution system, are investigated. Metagenomics and metaproteomics identified an extremely complex and diverse array of organisms, from small animals and protists to bacteria and viruses, which share a limited environmental niche. The majority of genera also showed the potential to degrade more than one biopolymer. Metaproteomics furthermore provided insights into the protein biomass composition of these communities. This study is the first to combine whole metagenome sequencing and microbial community proteomics (a.k.a. metaproteomics), providing insight into the complex drinking water ecosystem, and revealing potential survival mechanisms under highly oligotrophic conditions.

Microbial community proteomics proved to be extremely useful in obtaining a more complete picture of microbes present in the drinking water sediments. However, an orthogonal view on complex ecosystems can also be beneficial for exploring other microbial environments. In **Chapter 5** microbial community proteomics was investigated to provide an alternative view on wastewater. Wastewater-based surveillance has become routine for

tracking pathogens, antibiotic resistance genes, and population-level exposure to pharmaceuticals and chemicals. However, despite the presence of a wide range of proteins in wastewater, large-scale monitoring of protein biomarkers is currently not performed due to analytical challenges. This study developed new sample preparation and data analysis approaches, to enable fully untargeted analysis of proteins from across all domains of life in wastewater. A broad spectrum of microbes from different origins was detected, along with important human proteins that could potentially serve as biomarkers to indicate population health. This study brings microbial community proteomics closer to routine applications in wastewater surveillance.

Samenvatting

Micro-organismen zijn geëvolueerd om te gedijen in diverse en extreme omgevingen. Het begrijpen van deze microben in technische ecosystemen, zoals gebruikt worden voor afvalwaterzuivering en drinkwaterproductie, is cruciaal om hun rol in reiniging, nutriëntenkringloop en algehele systeemstabiliteit te begrijpen. Traditionele methoden, zoals microscopie en in situ-kleuring, bieden beperkte inzichten in microbiële diversiteit en functie, vooral voor soorten met een lage abundantie. De opkomst van cultuur-onafhankelijke technieken zoals 16S rRNA- en volledige metagenoomsequencing heeft de microbiële ecologie gerevolutioneerd en maakt diepere analyses van taxonomische profielen mogelijk, maar beantwoordt geen vragen over metabole activiteit.

Massaspectrometrie is een krachtige technologie die het mogelijk maakt om eiwitten op grote schaal te identificeren en karakteriseren uit kleine hoeveelheden celmateriaal, ook wel proteomics genoemd. Dit heeft brede belangstelling gewekt in de wetenschappelijke gemeenschap en industrieën, omdat eiwitten fundamentele processen in cellen aansturen, zoals enzymatische katalyse, moleculair transport, signaaloverdracht, celdeling en verdedigingsmechanismen.

De meest recente vooruitgangen in massaspectrometrie hebben het veld van proteomics in staat gesteld om microben en volledige microbiële gemeenschappen te bestuderen. Microbiële proteomics biedt hierbij waardevolle inzichten in de actieve metabole routes binnen microben en microbiële gemeenschappen, en vormt zo een waardevolle aanvulling op DNA-gebaseerde benaderingen. Hoewel microbiële proteomics al breed is toegepast in diverse onderzoeksgebieden, waaronder medische toepassingen en biotechnologie (bijvoorbeeld voor het begrijpen van cel-fabrieken), is er aanzienlijk minder onderzoek gedaan naar natuurlijke microben, waaronder die in procestechnische ecosystemen zoals afvalwater- en drinkwaterzuiveringssystemen.

Dit proefschrift beoogt de toepassing van microbiële proteomics te bevorderen voor het bestuderen van microbiële gemeenschappen in procestechnische ecosystemen, specifiek

binnen afvalwater- en drinkwateromgevingen. Door de aanwezige eiwitten in deze microbiële gemeenschappen te analyseren, kunnen cruciale inzichten in de tot expressie gebrachte metabole functies van individuele microben en in de eiwitbiomassasamenstelling van microbiële ecosystemen worden verkregen. Daarnaast verschaft het onderzoek naar het microbiële secretoom nieuwe inzichten in de metabole veelzijdigheid van micro-organismen in nutriënt-arme omgevingen.

In Hoofdstuk 2 wordt het potentieel van *Aeromonas* onderzocht om het koolhydraatpolymeer chitine af te breken en te gebruiken; dit om te begrijpen hoe dit organisme kan overleven in nutriënt-arme omgevingen zoals in drinkwaterdistributiesystemen. Leden van het *Aeromonas*-geslacht worden vaak aangetroffen in natuurlijke aquatische ecosystemen, waaronder niet-gechloreerde drinkwaterdistributiesystemen. Eerdere studies suggereerden dat *Aeromonas* chitine van ongewervelde dieren zoals Asellus aquaticus als energiebron kan gebruiken, maar de onderliggende metabole routes zijn slechts beperkt begrepen. Aangezien het geheim van de afbraak van biologische polymeren zoals chitine ligt in de productie en export van specifieke enzymen, genaamd glycosidehydrolasen, wordt in dit hoofdstuk zowel het cellulaire proteoom als het secretoom van twee *Aeromonas*-soorten onderzocht: *A. bestiarum* en *A. rivuli*. Beide soorten bleken het polymeer niet alleen te kunnen afbreken, maar gebruikten het ook efficiënt als enige koolstof- en stikstofbron. Proteome analyse toonde een significante reorganisatie van zowel het intracellulaire als het uitgescheiden proteoom wanneer *Aeromonas* groeide op chitine in plaats van glucose. Dit omvatte de expressie van specifieke glycosidehydrolasen en transporters. Het vermogen van *Aeromonas* om chitine efficiënt af te breken en erop te groeien leidde tot de vraag of en hoe dit organisme ook op andere biopolymeren kan groeien.

De verkenning van de capaciteiten van *Aeromonas* om andere ecologisch relevante biopolymeren af te breken en te gebruiken wordt besproken in Hoofdstuk 3. *Aeromonas* bestiarum werd gekweekt op fucoidan, zetmeel, xylaan, cellulose, dextrine, collageen, pullulan, pectine en een EPS-extract. A. bestiarum kon groeien op pullulan, dextrine, zetmeel, collageen en het EPS-extract. Proteoom analyse van het secretoom toonde vele specifieke enzymen aan die betrokken zijn bij de afbraak van deze biopolymeren. De analyse van het secretoom onthulde echter ook dat ongeveer een kwart van de uitgescheiden eiwitten een onbekende functie heeft, wat de huidige uitdagingen benadrukt bij het karakteriseren van proteomen van weinig onderzochte microben. Desondanks toonde de studie de metabole veelzijdigheid van *Aeromonas* aan en zijn potentieel om te overleven in oligotrofe omgevingen. Een organisme dat meerdere biologische polymeren kan afbreken wordt een generalist genoemd, en dit plaatst het op een cruciale plek binnen de voedselketen en verklaart zijn overleving in nutriënt-arme milieus.

Naast *Aeromonas* herbergen drinkwaterdistributiesystemen een verrassend rijk microbieel ecosysteem. De studie van zulke complexe omgevingen en hun microbiële bewoners vereist

een spectrum aan verschillende methoden. Daarom worden in Hoofdstuk 4 de volledige microbiële ecosystemen onderzocht die aanwezig zijn in losse afzettingen, verzameld op zeven verschillende locaties binnen een drinkwaterdistributiesysteem. Metagenomica en metaproteomica identificeerden een uiterst complexe en diverse verzameling organismen, van kleine dieren en protisten tot bacteriën en virussen, die een beperkte ecologische niche delen. De meeste geslachten toonden bovendien het potentieel om meer dan één biopolymeer af te breken. Metaproteoom analyse gaf bovendien inzicht in de eiwitbiomassasamenstelling van deze gemeenschappen. Deze studie is de eerste die volledige metagenoomsequencing en microbiële gemeenschapproteoom analyse (ook wel metaproteomica genoemd) combineert, wat inzicht biedt in het complexe drinkwater-ecosysteem en mogelijke overlevingsmechanismen onder sterk oligotrofe omstandigheden onthult.
Microbiële gemeenschapproteomics bleek uiterst nuttig om een vollediger beeld te krijgen van de microben die aanwezig zijn in de drinkwatersedimenten. Een orthogonaal perspectief op complexe ecosystemen kan echter ook voordelig zijn bij het verkennen van andere microbiële omgevingen. In Hoofdstuk 5 werd microbiële gemeenschapproteoom analyse onderzocht om een alternatief perspectief te bieden op afvalwater. Surveillance op basis van afvalwater is routine geworden voor het volgen van pathogenen, genen voor antibioticaresistentie, en blootstelling van de bevolking aan geneesmiddelen en chemicaliën. Ondanks de aanwezigheid van een breed scala aan eiwitten in afvalwater, wordt grootschalige monitoring van eiwitbiomarkers momenteel niet uitgevoerd vanwege analytische uitdagingen. Deze studie ontwikkelde nieuwe benaderingen voor monstervoorbereiding en data-analyse, om een volledig ongerichte analyse van eiwitten uit alle domeinen van het leven in afvalwater mogelijk te maken. Een breed spectrum aan microben van verschillende oorsprong werd gedetecteerd, samen met belangrijke humane eiwitten die potentieel kunnen dienen als biomarkers om de gezondheid van de bevolking te evalueren. Deze studie brengt microbiële gemeenschapproteoom analyse dichter bij routinetoepassingen in afvalwatergebaseerde epidemiologie.

Rezumat

Microbii au evoluat pentru a prospera în medii diverse și extreme. Înțelegerea acestor microbi în ecosisteme artificiale, cum ar fi cele utilizate pentru tratarea apelor uzate și producerea apei potabile, este esențială pentru a clarifica rolurile lor în salubritate, ciclul nutrienților și stabilitatea generală a sistemului. Metodele tradiționale, precum microscopia și colorarea in situ, oferă informații limitate despre diversitatea și funcția microbiană, în special pentru speciile cu abundență redusă. Apariția tehnicilor independente de cultură, cum ar fi secvențierea 16S rRNA și a întregului metagenom, a revoluționat ecologia

microbiană, permițând analize mai profunde ale profilurilor taxonomice, dar fără a oferi răspunsuri dincolo de potențialul metabolic.

Spectrometria de masă este o tehnologie avansată ce permite identificarea și caracterizarea proteinelor la scară largă, pornind de la cantități mici de material celular – proces cunoscut sub numele de proteomică. Aceasta a atras un larg interes în comunitatea științifică și în industrie, deoarece proteinele mediază procese fundamentale celulare, precum cataliza enzimatică, transportul molecular, semnalizarea, diviziunea celulară și mecanismele de apărare.

Progresele recente în spectrometria de masă au facilitat trecerea domeniului proteomicii către studiul microbilor și al comunităților microbiene complete. Proteomica microbiană oferă perspective avantajoase asupra căilor metabolice active din microbii individuali și din comunitățile microbiene, completând astfel informațiile obținute prin metode bazate pe ADN. Deși proteomica microbiană a fost deja utilizată pe scară largă în diverse domenii de cercetare, inclusiv în aplicații medicale și biotehnologie (de exemplu, pentru înțelegerea „fabricilor celulare"), s-a făcut mult mai puțină cercetare în ceea ce privește microbii prezenţi in mediul ambiental, inclusiv cei prezenți în ecosisteme artificiale precum sistemele de tratare a apelor uzate și de producere a apei potabile.

Această teză își propune să avanseze aplicarea proteomicii pentru studiul comunităților microbiene din ecosisteme artificiale, în special în medii de ape uzate și apă potabilă. Prin analiza proteinelor prezente în aceste medii, această abordare oferă perspective esențiale asupra funcțiilor metabolice exprimate ale microbilor individuali și asupra compoziției biomasei proteice a ecosistemelor microbiene. În plus, investigarea secretomului microbian oferă noi perspective asupra versatilității metabolice a microorganismelor în medii sărace în nutrienți.

În Capitolul 2 este investigat potențialul speciei *Aeromonas* de a degrada și utiliza polimerul chitina, pentru a înțelege cum poate acest organism să supraviețuiască în medii sărace în nutrienți, cum ar fi sistemele de distribuție a apei potabile. Membrii genului *Aeromonas* sunt frecvent întâlniți în ecosisteme acvatice naturale, inclusiv în sistemele de apă potabilă în care nu s-a folosit clor. Studii anterioare au sugerat că *Aeromonas* ar putea utiliza chitina provenită de la nevertebrate, precum *Asellus aquaticus*, ca sursă de energie, însă căile lor metabolice subiacente sunt puțin înțelese. Deoarece secretul degradării polimerilor biologici precum chitina constă în producerea și exportul unor enzime specifice numite hidrolaze, acest capitol investighează atât proteomul celular, cât și secretomul a două specii de *Aeromonas*: A*. bestiarum* și *A. rivuli*. Ambele specii nu doar că au degradat chitina, ci au și utilizat eficient acest polimer ca singură sursă de carbon și azot. Analiza proteomică a arătat o reorganizare semnificativă a proteomului intracelular și secretat atunci când *Aeromonas* a fost crescută pe chitină comparativ cu glucoza. Aceasta a inclus expresia unor glicozidaze și transportori dedicați. Capacitatea speciei *Aeromonas* de a degrada și crește

eficient pe chitină a ridicat întrebarea dacă și cum ar putea acest organism să crească utilizând și alți biopolimeri ca sursă de hrană.

Explorarea capacităților speciei *Aeromonas* de a degrada și utiliza alți biopolimeri cu relevanţă ecologică este discutată în Capitolul 3. *Aeromonas* bestiarum a fost cultivată pe fucoidan, amidon, xilan, celuloză, dextrină, colagen, pullulan, pectină și un extract EPS. *A. bestiarum* a fost capabilă să crească pe pullulan, dextrină, amidon, colagen și extractul EPS. Analiza proteomică a secretomului a indicat prezenţa a numeroase enzime dedicate degradării acestor biopolimeri. Totuși, analiza a arătat și că aproximativ un sfert din proteinele secretate au funcție necunoscută, subliniind provocările actuale în caracterizarea proteomului microbilor puţin studiați. Cu toate acestea, studiul a demonstrat versatilitatea metabolică a speciei *Aeromonas* și potențialul său de supraviețuire în medii oligotrofe. Un organism capabil să degradeze mai mulți polimeri biologici este numit „generalist", ceea ce îl plasează într-o poziție crucială în lanțul trofic și oferă explicații pentru supraviețuirea sa în medii oligotrofe.

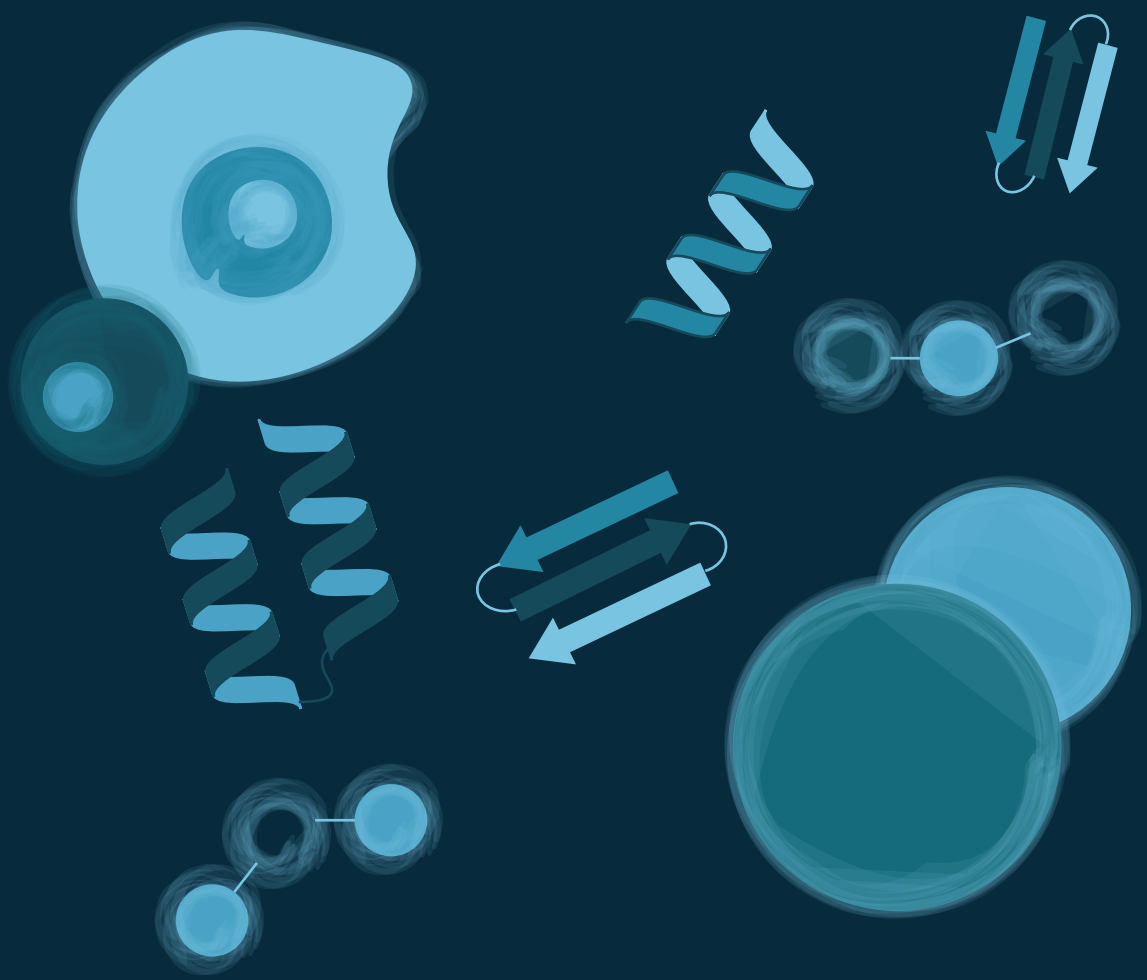
Totuși, pe lângă *Aeromonas*, sistemele de distribuție a apei potabile găzduiesc de obicei un ecosistem microbian surprinzător de bogat. Studierea acestor medii complexe și a locuitorilor lor microbieni necesită o gamă variată de metode. Prin urmare, în Capitolul 4 este investigat ecosistemul microbian prezent în depozitele sedimentare colectate din șapte locații diferite ale unui sistem de distribuție a apei potabile. Metagenomica și metaproteomica au identificat o gamă extrem de complexă și diversificată de organisme, de la animale mici și protiste la bacterii și viruși, care împart o nișă ecologică limitată. Majoritatea genurilor au arătat de asemenea potențialul de a degrada mai mult de un biopolimer. Metaproteomica a oferit și informații despre compoziția biomasei proteice a acestor comunități. Acesta este primul studiu care combină secvențierea completă a metagenomului cu proteomica comunităților microbiene (a.k.a. metaproteomică), oferind perspective asupra ecosistemului complex al apei potabile și dezvăluind posibile mecanisme de supraviețuire în condiții extrem de oligotrofe.

Proteomica comunităților microbiene s-a dovedit extrem de utilă în obținerea unei imagini mai complete a microbilor prezenți în depozitele sedimentare prezente în apa potabilă. Totuși, o viziune ortogonală asupra ecosistemelor complexe poate fi de asemenea benefică pentru explorarea altor medii microbiene. În Capitolul 5, proteomica comunităților microbiene a fost investigată pentru a oferi o viziune alternativă asupra apelor reziduale. Monitorizarea apei reziduale a devenit o practică de rutină pentru urmărirea agenților patogeni, a genelor de rezistență la antibiotice și a expunerii populației la produse farmaceutice și substanțe chimice. Cu toate acestea, în ciuda prezenței unei game largi de proteine în apele uzate, monitorizarea la scară largă a biomarkerilor proteici nu este în prezent realizată din cauza provocărilor analitice. Acest studiu a dezvoltat noi metode de pregătire a probelor și de analiză a datelor, pentru a permite analiza complet imparţială a proteinelor din toate domeniile vieții prezente în apele reziduale. A fost detectat un spectru

larg de microbi de diverse origini, împreună cu proteine umane importante care ar putea servi drept biomarkeri pentru a indica starea de sănătate a populației. Acest studiu aduce proteomica comunităților microbiene mai aproape de aplicațiile de rutină în supravegherea apelor reziduale.

Chapter 1

Introduction to proteomes in the flow: proteomic insights into engineered water environments



General introduction

From acid mine drainage, to wastewater, bottom of the ocean and human microbiome, microbes have evolved to thrive in different and sometimes extreme ecosystems. However, the vast majority of all these bacterial species cannot be easily studied since they cannot be cultured in laboratory settings. This presents a challenge in understanding the processes that take place in natural but also engineered ecosystems, such as those used for wastewater treatment, for bioremediation of water and soil, or for producing drinking water, which is of utmost importance in times of growing world population.

The understanding of the bacterial communities that are present in anthropogenic water cycles, like the treatment of wastewater or the production of the drinking water, is gaining more interest across the globe. Microbes are involved in water treatment through their employment in biological sand filters which are effective in protecting the consumers against pathogens. Since the resident bacteria from the biofilm growing on sand filters tend to be more robust and adapted to the environment than the potential pathogens contained in the water [1].

The bacterial communities that are employed in wastewater treatment are highly diverse. Bacterial communities are involved in recycling nitrogen, phosphorus and consuming carbon compounds lowering their concentration in the effluent. Thus, the risk of eutrophication of water source where the effluent is discharged is reduced [2].

There is a strong need to understand these complex environments and to develop methods that provide detailed information on which microbes are present and what functions they perform within these ecosystems. Routinely used techniques like microscopy and in situ staining can offer fast information about the morphology and even taxonomy of certain microbes but it does not provide any information about their metabolic activity. Also, microbes that are in low abundance and that may not be well characterized, are usually overlooked in this type of analysis. Therefore, there is a need for methods that unbiasedly capture a wide range of microbes while also providing information about their metabolic functions as much as possible. In recent decades, the advancement of culture-independent techniques has gained significant momentum.

The earliest attempt at taxonomic profiling was realized by the sequencing of 16S rRNA genes, a highly conserved gene among bacterial and archaeal species. Later, the advent of next generation sequencing techniques allowed to target whole metagenomes at low cost within a short time. This allowed the sequencing of microbes directly from environmental samples. This approach helped in unraveling the composition of microbial communities and their metabolic potential.

Additionally, tremendous developments in mass spectrometric instrumentation over the past decade have allowed the transition from conventional cellular proteomics to the field

of microbial ecology. Microbial community proteomics, often referred to as metaproteomics, aims to measure all expressed proteins form a microbial community. This provides direct information about the actively expressed metabolic pathways present at a certain time. These 'omics methods are relatively new, therefore, these methods still need further advancements and validation in order to efficiently apply them to complex environments like wastewater and drinking water. Environments with complex matrixes like wastewater or sediments but also low particle environments such as drinking water, represent a big challenge for meta 'omics techniques.
The further advancement of 'omics methods, especially metagenomic and metaproteomics, and the integration of these methods, promise to provide important insights into the microbial processes present in different water environments.

DNA based approaches

The genome refers to an organism's complete set of genetic material, whereas the metagenome encompasses all the genomes present within a complex ecosystem, originating from at least two different organisms [3].
The foundation of genetics dates back to the 19th century, when Gregor Mendel discovered the principles of heredity through pea plant experiments, laying the groundwork for modern genomics. In the 20th century, advancements in technology led to the discovery of DNA's structure by Watson and Crick, followed by the development of genetic analysis methods such as Sanger sequencing and polymerase chain reaction (PCR). Genomics gained significant momentum in the 1990s with the launch of the Human Genome Project, which, using the available technology, took 13 years to complete, culminating in 2003 [4].
The term metagenomics was introduced by Handelsman in 1998 to describe a biosynthetic gene cluster [5]. The early 21st century saw the rise of next-generation sequencing [6] technologies like Illumina™, PacBio™, and Oxford Nanopore™, which drastically improved sequencing speed, accuracy, and cost-effectiveness. Today, genome sequencing has become a routine analysis, and thanks to Nanopore sequencing, we can now recover complete genomes even from highly complex environments.
Currently, two main sequencing approaches dominate: short-read sequencing (e.g., Illumina™) and long-read sequencing (e.g., Oxford Nanopore™ and PacBio™), each with distinct advantages and limitations. NGS revolutionized DNA sequencing by increasing throughput while reducing sequencing time. Illumina remains the gold standard for short-read sequencing, widely used in metagenomics, while long-read platforms have enabled high-contiguity genome assemblies, improved strain differentiation, and more comprehensive microbial profiling.

Whole metagenomic sequencing present significant challenges due to computationally demanding pipelines, complexities in sample preparation, and challenges in validation and normalization across different samples, often resulting in high variation even within the same sample batches. Furthermore, the accuracy of taxonomic and functional classification remains highly dependent on the specific environment being studied and the availability of representative microbial sequences in reference databases. It has been also noted that metagenomic analysis is often biased toward bacteria, making it more challenging to accurately identify Archaea and viruses [7]. However, depending on the sequencing technology, data processing pipelines, and reference databases used, discrimination between closely related species and strains may be a challenge. Finally, metagenomics will benefit from the recent advancements in machine learning and artificial intelligence. The continuous development of computer power and as a result of more powerful algorithms will improves taxonomic and functional classification, and it may lead to a better integration with other omics data [8, 9].

Mass spectrometry based microbial proteomics

While J.J. Thomson developed the first mass spectrometer at the end of the 19th century, it was not this invention but rather his discovery of the electron—enabled by the first mass spectrometer—that earned him the Nobel Prize in Physics in 1906 [10, 11]. Modern versions of mass spectrometers were developed by Arthur Jeffrey Dempster in 1918 and F.W. Aston in 1919. During the Second World War, A. Nier developed a mass spectrometer that played a crucial role in enriching uranium-235, proving its responsibility for nuclear fission while contributing to the Manhattan Project [12]. Nearly two decades later, K. Biemann utilized mass spectrometry to analyze and determine the structure of natural products [13]. The invention of the well-known tandem mass spectrometry was coined by D. Hunt, an invention that opened the way to the study of structural information of the ions [14]. Only the later invention of soft ionization techniques like electrospray ionization and laser desorption ionization allowed the study of biological molecules. The soft ionization allows the analysis of big biomolecules without further fragmentation [11].

The increasing importance of mass spectrometry in biological sciences called for more powerful analyzers. This led first to the development of time-of-flight mass analyzers, and later to the Orbitrap mass analyzer in the early 2000's by Makarov [15, 16].The Orbitrap is an ion trap mass analyzer that consists of an outer-barrel like electrode and a spindle like inner electrode [15].

The ions are first collected in the C-trap, from where they are injected into the Orbitrap (**Figure 1A**). In the Orbitrap mass analyzer the ions oscillate around a central, spindle-shaped electrode. Ions are trapped in an electrostatic field and orbit the central electrode while

simultaneously moving back and forth along its axis. The frequency of these axial oscillations is detected and converted into mass-to-charge ratios with high resolution and accuracy [15, 17]. The two main methods of analyzing complex protein mixtures with mass spectrometry are data-dependent acquisition (DDA) and data-independent acquisition (DIA). DDA selects the most intense precursor ion in MS1 for further fragmentation. This provides high quality spectra for individual peptides. In DIA analysis, all the ions within a predefined m/z window are further fragmented [18]. This allows DIA analysis to have increased reproducibility and throughput compared to DDA which can lack enough resolution for detecting low abundant peptides [19, 20].

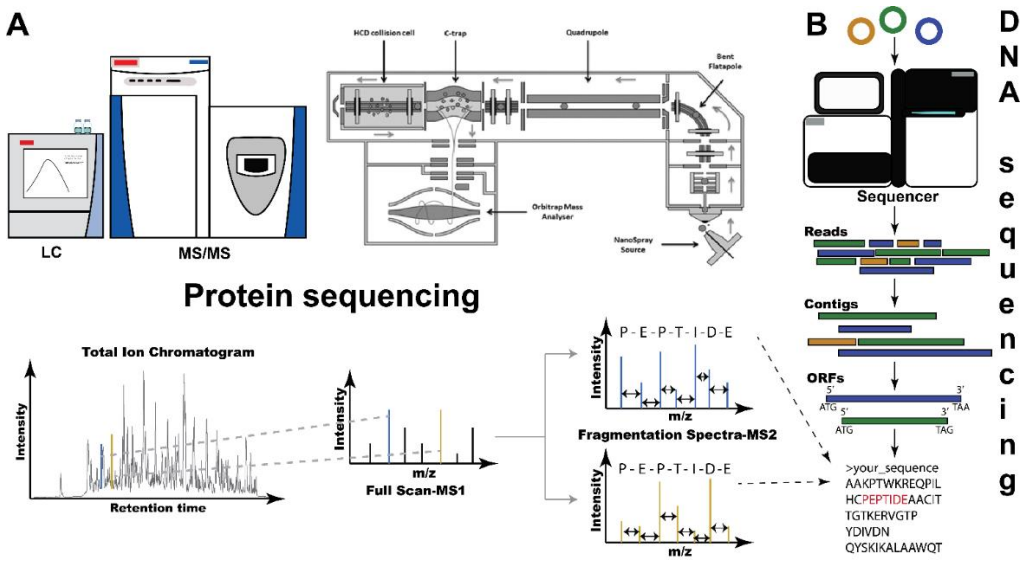


Figure 1: Example high-resolution tandem mass spectrometer and concept of peptide sequencing.
A. Protein sequencing: Liquid-chromatography (LC) coupled to tandem mass-spectrometry (MS/MS)- The upper left picture shows the outline of the quadrupole-Orbitrap mass spectrometer, which was employed in the current thesis. The arrows indicate the path taken by the ions inside the mass spectrometer, and the dots represent the ions (figure adapted from Michalski *et al*., 2011). The chromatogram bottom left) shows the total ion intensity chromatogram after chromatographic separation and mass spectrometric detection of a proteolytic digest of proteins obtained from a whole cell lysate. The full scan mass spectrum (MS1) shows mass peaks derived from peptides that elute at a given time point during the chromatographic separation, for which the mass-to-charge (m/z) ratio and the ion intensity are measured. Further the MS2 (fragmentation) spectra are obtained by isolating a certain m/z value by the quadrupole mass filter and subsequent fragmentation by HCD. The fragments are then measured in the Orbitrap mass analyser. The blue and yellow lines show the fragmentation spectra from a theoretical peptide. The sequencing spectrum finally represents the amino acid sequence of the peptide, as shown for the theoretical sequence "PEPTIDE" (right graphs).
B. **DNA Sequencing:** In whole metagenome sequencing experiments the reads obtained from the

sequencing experiment are assembled into contigs. The ORFs (open reading frames) are identified from the contigs and are further used to translate them into protein sequences. The generated protein sequences from the metagenomic experiment can serve as a database for retrieving and annotating proteins from proteomic experiments.

Two main strategies for the annotation of mass spectrometric fragmentation spectra (MS/MS) with amino acid sequences exist. The first one is database dependent, and it involves the in-silico digestion of the proteins in the database and the generation of MS/MS spectra for each peptide in the proteins. These theoretical generated spectra are further compared to the experimental MS2 spectra with the help of an algorithm. This approach is called database searching, which ultimately provides peptide spectrum matches (PSMs). After statistical validation, the significant PSMs are retained and further annotated to protein sequence templates. The second approach to annotate fragmentation spectra with amino acid sequences is by *de novo* sequencing. *De novo* sequencing determines the peptide sequence directly from the observed fragment ions without relying on a reference sequence database. The amino acid sequence is determined from the N and C-terminus by looking for distinct mass differences between fragment ions, which indicate the presence of certain amino acids [21]. However, the validation of the correctness of the identified peptide sequence is more challenging, and usually only based on positional confidence scores. In the last decade database-independent protein identification using *de novo* sequencing has also gained attention in metaproteomic studies. Advancements in machine learning models proved extremely useful to enhance the accuracy of de novo sequencing in complex samples [22].

Most recent technical advancements in mass spectrometry include Thermo Fisher's Orbitrap Astral mass spectrometer, which may provide twice the proteome coverage and quadruple the throughput compared to previous Orbitrap mass analyser-based mass spectrometers. The Astral mass analyser traps and fragments ions at rates up to 200 Hz, then aligns and accelerates them through an over 30-meter-long asymmetric electrostatic track, achieving a spectral resolution of more than 80K (m/z 524) with high dynamic range, single-ion sensitivity and excellent linearity [23].

Other developments include the Bruker's timsTOF Pro mass spectrometer, which integrates Trapped Ion Mobility Spectrometry (TIMS) with Parallel Accumulation Serial Fragmentation (PASEF), achieving high-speed, high-sensitivity proteomics without compromising mass resolution [24]. These innovations represent significant improvements over earlier instruments, providing deeper insights into complex microbial samples and facilitating the discovery of low-abundance proteins and microbes.

The advancement in the last decade of NanoPore sequencing technology in genome sequencing fuelled the idea of applying this technology in sequencing intact or partially cleaved proteins thus offering an alternative to top-down proteomics method. Although this

method is still in its infancy, several studies provided proof of concept for sequencing proteins using nanopore technologies [25, 26].

Cellular proteomics benefitted from all the advancements both in technical and biochemical fields. The number of proteins that can be identified from whole cell lysates employing short one-dimensional chromatographic separations and shotgun proteomics has increased considerably over the last decades. The sensitivity of mass spectrometric setups, and advancements in microfluidic sample preparation increased to such degree that proteomic studies of single (mammalian) cells has become possible [27].

Analysis of biological samples have improved greatly overtime with the advancements in mass-spectrometry, especially the field of cellular proteomics. This allowed identification and quantification of proteins that are present in cells or tissues with high accuracy and certainty. Two main experimental approaches are that allow the analysis of proteins through mass-spectrometry: top-down and bottom-up. The top-down approach involves the analysis of intact proteins through mass-spectrometry. However, due to the different hydrophilicities of different proteins and low throughput, this method has currently limited applications. Bottom-up proteomics, however, allows for the identification of proteins and even peptide modifications without compromising accurate identification or the need of a special set-up. Bottom-up shotgun proteomics is the main used approach in mass spectrometry-based proteomics where proteins are first digested into smaller peptides before analysis. First proteins are extracted from cells or tissues and then enzymatically digested into peptides. The most common enzyme used for proteolytic digestion is trypsin due to its stability and specificity [28]. Trypsin cleaves after arginine (R) and lysine (K) residues, at the C-terminal side of these amino acids, unless followed by proline (P). This results in easy to analyse and interpret the proteomic data with minimal loss of information. The generated peptides are first chromatographically separated, then ionized and guided into the mass spectrometer, where their mass is measured before they are fragmented to obtain sequence information.

However, metaproteomics, aims to identify all proteins of all organisms present in a microbial community. While advanced metaproteomic setups can identify over 10,000 proteins in short runs, subsequent data processing is challenged by the large data volumes and the increased complexity of protein inference. Furthermore, efficiently extracting the proteins from all different organisms present in the community, as well as within a cell is usually not achievable. For example, some microbes have more rigid cell walls than others. Additionally, extracting membrane-embedded proteins is typically more challenging due to their increased hydrophobicity and often poor solubility [29].

When studying expressed proteins, proteomics is commonly preferred over transcriptomics unless the target may be mRNA, in a metatranscriptomic analysis [30] or miRNA and its regulatory function [31]. The discrepancy between proteomics and transcriptomics usually arises because mRNA levels often correlate poorly with the actual protein abundances [32, 33].

This may be due to differences in half-life of the mRNA and proteins as well as from regulatory mechanisms that affect translation and protein stability [34, 35].

Nevertheless, obtaining quantitative information in metaproteomics is challenging. In complex communities, a microbe's protein biomass does not necessarily correlate with its cellular abundance (number of cells), as larger cells tend to have more protein mass than smaller ones [36]. Finally, a frequently overlooked process is the post-translational modification of proteins, which can significantly influence their structure and activity. Although metaproteomics enables the detection of metabolic pathway expression and may even allow for the quantification of enzymes within these pathways, protein abundance does not always directly correlate with activity. For example, it is well known that post-translational modifications, such as phosphorylation, can switch enzyme activity on or off [37]. However, such modifications are typically not captured in conventional metaproteomics experiments.

When performing metaproteomic analysis of microbial communities, a metagenome-based reference sequence database is essential to achieve a high proteome coverage (**Figures 1A** and **B**). Typically, the same sample is subjected to whole metagenome sequencing, where genes are identified after assembly and/or binning and then converted into a protein reference sequence database. This database is used for database searching, linking them to peptide sequences, proteins, and finally taxonomies. Alternatively, protein sequences can be retrieved from public databases such as UniProtKB or NCBI based on the taxa identified in the metagenomic analysis (**Figure 2**) [38]. However, these databases are often large and may still be incomplete, as only a small fraction of all microbial species have been sequenced to date. Due to their size, these databases can increase the likelihood of both false-positive and false-negative annotations. Additionally, two-step database searches are often used in combination with very large databases to enable more efficient data handling [39].

Alternative strategies for database construction use de novo sequencing, where taxonomic classification of the de novo sequences helps refine and focus large, generic reference sequence databases [22]. These database-independent approaches using de novo peptide identification proved to provide highly comparable taxonomic profiles compared to conventional database searching methods [40, 41].

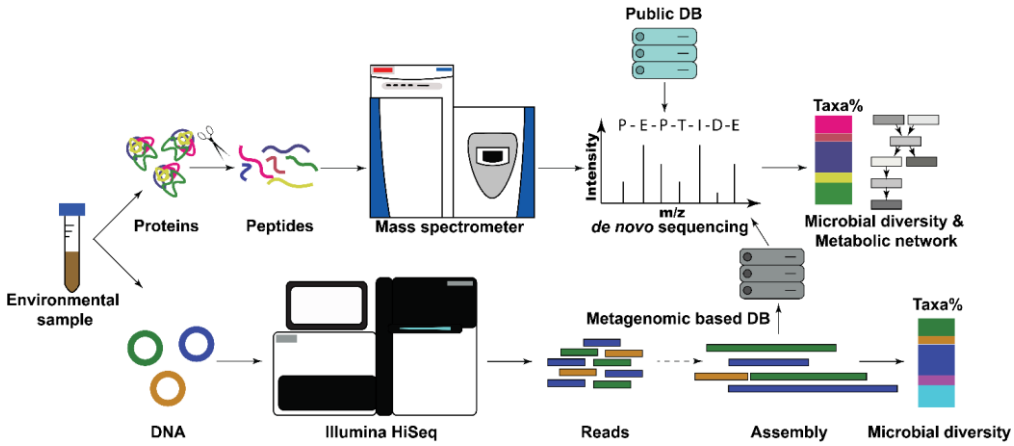


Figure 2: Outline of a multi-metaomic workflow combining metagenomics and metaproteomics for characterizing microbial communities from complex environments. The genomic sequences derived from the metagenomic study are used as a database for the annotation of the proteins from the metaproteomic experiment. Alternatively, a large, generic reference sequence database can be used to annotate the spectra, which can be further focused using *de novo* sequencing.

Often, multiple techniques are required to fully capture the microbial processes. Metagenomics provides an accurate representation of microbial diversity, and quantitative data typically correlate with the number of cells per organism [36]. In addition, whole metagenome sequencing, including the generation of metagenome-assembled genomes (MAGs), enables strain-level resolution and unambiguous identification of the metabolic processes of individual microorganisms [42-44]. On the other hand, metaproteomics directly measures the proteins without prior amplification which serves as a direct proxy for the biomass contribution of individual organisms in complex communities [36, 45]. However, the reconstruction of metabolic fluxes requires complementary metabolomic studies. Such metabolic studies can be combined with stable isotope probing experiments, which allow tracking the flow of carbon or nitrogen through metabolic pathways and provide insights into the utilization of nutrient sources within microbial communities [46].
This allows discrimination between pathways present in different organisms in the environment or enrichments emphasizing the metabolic contribution of individual species [47]. Also, labelled substrates directly show which microorganisms consume which substrate, and how its byproducts may influence the overall community. This allows for the construction and investigation of so-called “food webs” within communities [46, 48-50].

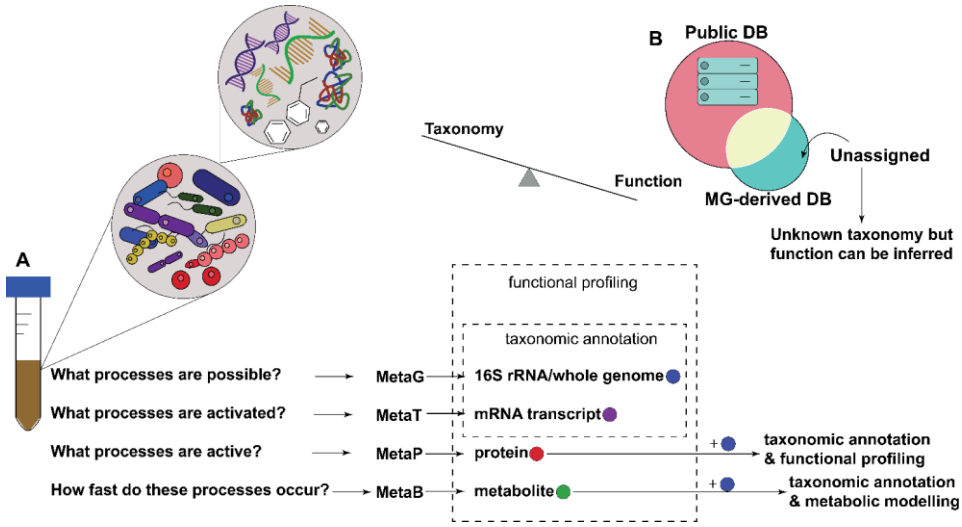


Figure 3. A. The analysis of an environmental sample harboring a complex microbial community using different omics technologies. While metagenomics (MetaG) reveals the identity of microbes in a community, metaproteomics (MetaP) and metatranscriptomics (MetaT) identify the active biochemical processes that occur; metabolomics (MetaB) on the other hand, provides valuable information about the speed at which these biochemical processes occur. Various biochemical methods can be used to extract taxonomic and functional insights from bacterial communities. Metagenomics leverages 16S rRNA gene sequencing and whole-genome sequencing to identify microbial taxa and predict their potential functions in the environment. Metatranscriptomics focuses on cellular mRNA to determine which bacterial processes are actively expressed and which organisms are involved. Metaproteomics analyzes the proteins present in environmental samples, providing insight into active biochemical pathways and their ecological relevance. Meanwhile, metabolomics assesses metabolism, offering valuable information about dynamics and rates of biochemical processes. Both metaproteomics and metabolomics can be combined with metagenomics in order to improve the taxonomic annotation and functional profiling in case of metaproteomics, or the taxonomic annotation and metabolic modelling in case of metabolomics. **B.** The taxonomic classification is often challenging due to the incompleteness of public databases. A large fraction of retrieved sequences often cannot be assigned to a specific species or strain. Nevertheless, in some cases, the gene or protein function can still be identified, which also provides more informative insights than a taxonomic name. In the last decade multiple bioinformatic tools have been developed that allow function inference of a protein based on sequence, peptide residues or protein folding

The majority of organisms cannot be cultured in laboratory environments. Therefore, culture-independent methods are essential for studying environmental communities [51]. Different omics methods address different questions and combining them is often necessary to fully understand what these organisms are doing in their natural environments (**Figure 3A**). The combination of metagenomics and metaproteomics provides both taxonomic identity (and diversity) and insights into metabolic responses to dynamic environments [52-53]

(**Figure 2** and **Figure 3A**). On the other hand, the addition of metabolomics provides insights into metabolic fluxes in complex environments [54].

Identification of new microbes, especially in complex environments like soil, or water, can be achieved either through selective culturing of microbes on different substrates until only one microbial colony survives or through metagenomic analysis and the recovery of full metagenome assembled genomes (MAGs). Until now, advancements in metagenomic allowed the construction of MAGs from different environments [55-57]. Still, the high complexity of microbes present in different environments exceeds the current technical capabilities of genome-based analysis [58]. As a result, a large fraction of sequences remains without taxonomic or functional annotations. However, these unassigned sequences may hold important information about community function. For example, many bacteria in soil are not sequenced, however enzymes involved in central carbon metabolism or transport systems that are well studied and preserved in nature, are identified and their function is annotated despite the lack of deep taxonomic information, like genus or species [59-61]. (**Figure 3B**). In recent years, the focus has been shifted from identifying bacteria and trying to isolate them to trying to predict the function of genes that are recovered from whole metagenome studies. While already multiple software tools are available, the advancement of functional prediction tools remains important. Widely employed tools are: Sequence-based approaches such as HMM (Hidden Markov Model) [62], InterPro [63], BLAST [64], Pfam [65], dbCAN [66], PROSITE [67], PRINTS [68], structure-based approaches including AlphaFold [69], AlphaFold-Multimer [70], SWISS-Model [71], I-TASSER [72], machine learning and deep learning approaches such as DeepGOPlus [73], ProtTrans [74], DeepFRI [75], ProtVec [76], ProteinBERT [77], network-based approaches, such as STRING Database [78], GeneMANIA [79] and evolutionary and phylogenetic approaches including OrthoFinder [80] and EggNOG-mapper [81].

The water cycle

The earth's surface is covered in 70% water from each, only 1% can be used by humans for consumption and it is the only chemical that can naturally occur in all three forms of state: liquid, solid and vapor. For life, water is the perfect reaction medium due to its characteristics. It can store large amounts of heat, is electrical and pH wise neutral. Water in a liquid state allowed the occurrence of life on earth in all its forms. In order for the water stock to be replenished, nature created the water cycle.

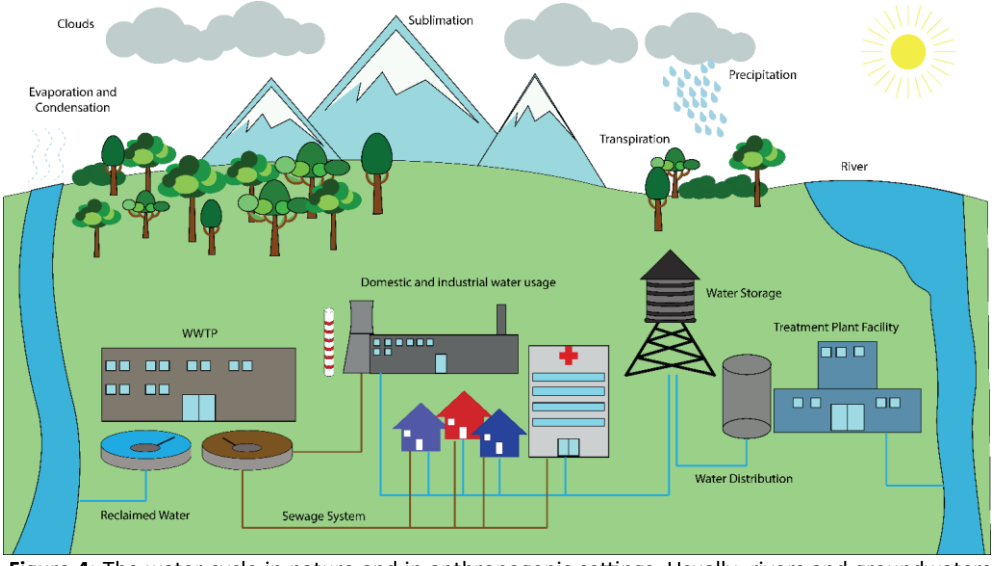


Figure 4: The water cycle in nature and in anthropogenic settings. Usually, rivers and groundwaters represent the main water sources for potable water; the treated water is further distributed to the cities, farms or for industrial purposes. The wastewater is again collected through the sewage system and transported to a wastewater treatment plant; the water that is reclaimed after treatment is either fed to a river or ground stream or used for industrial and agricultural purposes. Naturally the freshwater is replenished through precipitations in the form of rain and snow; the clouds appear as a result of snow sublimation, plant transpiration, water evaporation followed by condensation.

The anthropological water cycle involves the use of water resources for domestic consumers as well as for use in agriculture and industries. In order to ensure the replenishment of the used water with as little loss as possible and avoid contamination of other water sources, wastewater treatment plants were implemented. Thus, the reclaimed water obtained from the water treatment can be used for agricultural or industrial fields or feed back into the natural environment. From rivers and groundwater this water can be taken up again by the treatment plants and turned into potable water (**Figure 4**). Despite cleaning efforts and purification methods employed in wastewater treatments plants, the resulting cleaning water may not meet the required standards to be used further for further treatment and turned into potable water. Few countries like Singapore, Israel and Australia and Namibia have implemented a full water reclaim system, reusing their reclaimed water in agriculture and for non-potable consumers [82].

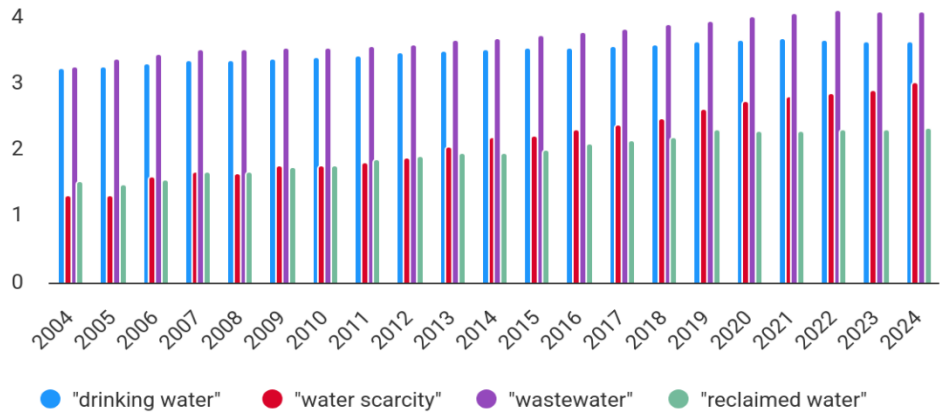


Figure 5: Occurrences of key words “drinking water”, “water scarcity”, “wastewater” and “reclaimed water” in the NCBI PubMed database as of 05.03.2025. The numbers are represented as a log function of occurrences per year according to the PubMed database. Y-axis values are expressed on a $\log_{10}$ scale, and the x-axis shows the year of publication.

In recent years, water scarcity began to be recognized as an emerging global problem. All over the globe, heatwaves and drought change the overall water demand among the population and industries with agriculture being one of the most important ones. In case of extreme weather like drought, water demand increases, and new sources of water need to be found and exploited to fulfill the demand [83]. However, due to lack of precipitation, the water sources cannot be replenished which in the end leads to water scarcity. In the past 20 years, there has been an increase in the scientific field for improving water technologies

in order to be able to reuse as much of the treated water that we produce in the context of emerging water scarcity (**Figure 5**).

Drinking water

The history of water sanitation started over 3500 years ago in ancient Greece with the employment of boiling and charcoal filtering. Later, in ancient Egypt coagulants called alum were employed to filter out negatively charged components from the water followed by the boiling of water [84]. In 1804 in Scotland, the first water treatment plant that used sand filtration was built to ensure potable water for residents [85].

In the mid of 19th century, John Snow discovered the main source of the cholera outbreak that was occurring in London at that time. Until then it was believed that this pathogen was airborne, but he was the one who discovered that the source of the pathogen was actually the drinking water (DW) that was supplied by a water pump in the city [86]. The importance of water sanitation thus gained attention.

In the current worldwide context of water resources, the quality of drinking water has become a prerequisite standard in the developed world. The organoleptic qualities of drinking water, like taste, odor and smell, are highly monitored to ensure safe and high-quality water for consumption. However, all these quality aspects can be altered as a consequence of bacterial growth in the drinking water systems. Thus, in the last decade, studies have been conducted in order to unravel the reasons behind regrowth of microorganisms in such oligotrophic environments. It has been stated that a reason for the regrowth of bacteria and other microorganisms (**Figure 6**) in the drinking water distribution system (DWDS) may be the existence of dissolved inorganic and organic matter. These compounds enter the distribution system and serve as food source for bacteria residing in the water and biofilms. Further, the bacteria and the forming biofilm may serve as food source for protozoa and small invertebrates that graze on them [87]. Nonetheless, it has not been revealed yet how microorganisms interact with each other or differently adjusting depending on the amount of present substrate.

aBased on different environmental dynamics and bacterial population, the drinking water microenvironment can be split into four different phases: bulk water (water present in the pipe), pipe wall biofilm (attached to the surface of the pipe material), suspended solids (particles that flow through the pipe) and loose deposits (accumulation of particles at the bottom of the pipe)(**Figure 6**) [88]. All these microenvironments share more or less the same bacterial community, as it is supposed that the upstream system is seeding the downstream system [89].

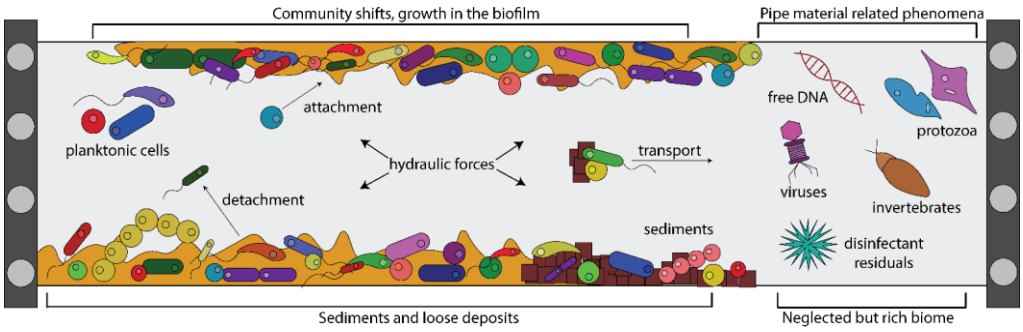


Figure 6: The microbial diversity in drinking water distribution systems and the phenomena attributed to bacterial regrowth (after Proctor and Hammes[89]); The drinking water system do not contain only bacteria but also viruses, protozoa and even small invertebrates; the microenvironments in which bacteria may reside include the biofilm, sediments and even the bulk water; The hydraulic forces present in the distribution system favors the biofilm detachment and dissemination of microorganisms across the system; residual disinfectant in the drinking water system can greatly impact the bacterial community

One method to predict the microbial regrowth potential in drinking water is by measuring *Aeromonas sp* in the drinking water systems. In the Netherlands, this bacterium is used as an indicator strain, but some representatives of the genus can also be potential pathogens like *A. hydrophyla*, *A. veronii*, *A. sobria*. Currently it is not known how *Aeromonas sp*. can survive in such highly oligotrophic environments, but it is believed that the presence of complex carbon molecules, like biopolymers can serve as a source of food for these bacteria [90].

Routinely, drinking water is measured using HPCs (heterotrophic plate counts) and sometimes 16S rRNA amplicon sequencing. The major drawback with these methods is that they are highly selective for specific microbes, and it does not reflect the actual diversity of drinking water. HPC mainly focuses on the identification of active cells capable of regrowth that can indicate the biological stability of the drinking water. It requires between 3 to 7 days to deliver results, which makes it unsuitable to monitor in real- time the variation of the bacterial community, it also gives higher error rates when dealing with low number of culturing bacteria [91]. This may be due to the different growing conditions or medium compositions used in different labs. Lastly, with HPC, the bacterial diversity that can be targeted is about 1% due to the high number of VBNC (viable but non culturable) bacteria or the rich nutrient medium that can select specific bacterial strains [92].

16S rRNA on other hand cannot discriminate between active and dormant bacteria or the free-floating DNA. Also the different GC content in bacteria, alignment of primers and different numbers of the target genes lead to PCR biases [93-95]. This indiscriminatory approach cannot encompass the full diversity of the drinking water community in all its niches.

Flow cytometry is a culture-independent technique widely used in assessing the quality of drinking water by monitoring the particle load, but it can also be applied for the study of wastewater. Even RT-FCM was used for real time monitoring of the microbial dynamics in tap water [96]. Compared to biochemical essays like ATP measurement or HPC, flow cytometry has a higher reproducibility [97]. However, flow cytometry falls short when it needs to discriminate between bacterial flocks, clumps or in the case of the drinking water, in case of particle associated bacteria [98].

Table 1: Different methods are employed in the surveillance of drinking water biostability, varying in their degree of automation and reproducibility, duration of use, and whether they are culture-dependent or independent. HPC = Heterotrophic Plate Counts.

Method	Automation	Reproducibility	Used since approx.	Approach
HPC	x	x	>70 years	Culture-dependent
Flow cytometry	v	v	>30 years	Culture-independent
Metatranscriptomics	v	v	>20 years	Culture-independent
Metagenomics	v	v	>20 years	Culture-independent
Metaproteomics	v	v	>20 years	Culture-independent
16S rRNA amplicon sequencing	v	v	>60 years	Culture-independent

Over the last decades, the development of methods capable of capturing a broad range of microbial diversity has improved the monitoring and study of complex environments, such as drinking water distribution systems (**Table 1**). In the field of drinking water surveillance, the methods employed for measuring microbial activity and thus appreciating the number of viable bacteria is usually comprised of ATP assays, AOC (assimilable organic carbon) assay and enzymatic tests. The ATP measurement assay is a well-known, routinely used method to check the viability of microorganisms in drinking water either by measuring the total ATP or by measuring only the free-floating molecule. However, this approach cannot measure the total amount of bacteria nor the diversity of them, not to mention that it only works in the case of active bacteria. Also, waterborne pathogens are overlooked since they cannot be identified through this approach.

The AOC measurement is another routinely used method to assess drinking water biostability and it refers to the amount of carbon that can be used as food source for bacteria. The method involves inoculation of drinking water, that assumably contains AOC, with 2 bacterial strains *Pseudomonas fluorescens* strain P-17 and *Spirillum* strain NOX [97]. The amount of AOC is measured by the growth of the two microorganisms. Different interactions between microorganisms can influence the survival of bacteria in DW. The use of 2 bacterial strains to measure the AOC is not accurate since other bacteria can use a wider range of carbon molecules increasing their survivability in the DWDS.

Enzymatic activity tests quantify the presence of specific enzymes by measuring the byproduct of enzymatic degradation. They are widely used to detect the presence of coliforms *E. coli* in source water by measuring the activity of β-D-galactosidase and β-D-glucuronidase [99]. Other enzymatic assays target chitinases, xylosidase, cellobiohydrolase, α- and β-glucosidase [100]. Although enzymatic assays are fast, still they lack specificity since they cannot distinguish between pathogens from commensals. Also, this method is prone to erroneous measurements due to inhibition factors that may be present in the environment that can interact with enzymatic activity. FISH (Fluorescence in situ hybridization) is a visualization technique that can target specific microorganisms based on their 16S rRNA.

The techniques mentioned above, whether targeted or untargeted, fail to capture the full extent of microbial diversity, in environments such as drinking water. Therefore, high-throughput and highly accurate methods are essential for studying microbial dynamics in drinking water systems. Genomic techniques that target a broad spectrum of microbes in drinking water have been utilized in previous studies. Moreover, advancements in DNA extraction and analysis from complex ecosystems like drinking water have enabled the identification of numerous bacterial genera and classes associated with this environment [98, 101-104].

The fraction of drinking water analyzed is also a crucial factor, as the most metabolically active community is found in the loose deposit fraction, followed by biofilm, as indicated by ATP measurements [101]. Additionally, the bacterial community in the suspended particle fraction closely resembles that of loose deposits.

In the quest for understanding the biological dynamics of the drinking water system and further improve sanitation methods, several studies have focused on the identification of non-bacterial residents of drinking water like amoeba [105, 106], fungi [107] and even small invertebrates [108]. The presence of amoebas in this environment is of high relevance since it is known to graze on the biofilms present and can harbor different bacteria and even pathogens providing shelter against sanitation methods [105, 109].

Apart from genomic methods, metatranscriptomics has been employed in drinking water research to identify active microbial populations [110]. Multiple analytical methods have been employed in measuring and monitoring the drinking water system in a quest to identify the

main actors and factors that contribute to a decrease in its biological stability. So far, this presumably oligotrophic system proved to be more diverse than expected.

Aeromonas sp. beyond the drinking water system

In the Netherlands, *Aeromonas sp.* is used as a strain indicator for assessing the biological stability of the drinking water. Since its discovery in 1962, different representatives of the genus have been discovered. *Aeromonas* species are mostly found in freshwater, reservoirs but also in brackish water, sewage and even wastewater. *Aeromonas* are mesophilic gram-negative bacteria, that can tolerate a wide range of temperatures from 0˚C to 45˚C. At the moment of 2020, 36 members have been described in the *Aeromonas* genus from which 19 are considered emerging potential pathogens for human health [111]. Members of this genus are potential pathogens and may even pose an economic threat to the fish industry due to the pathogenicity of *A. salmonicida* [112]. However, *Aeromonas* infections in humans have been merely attributed to rainy seasons, floods or bathing in natural lakes end even with leech therapy., as it has been discovered that *A. veronii* resides in the mucus of leeches [113]. In the drinking water system, *Aeromonas sp.* has been recovered from biofilms which offer them protection against disinfection methods [114-117]. The mesophilic *Aeromonas*, in particular *A. hydrophila*, are able to utilize a wide range of low-molecular-weight compounds (amino acids, carbohydrates and carboxylic acids), peptides and long-chain fatty acids [118]. Members of the genus have been identified residing in the drinking water system, more precisely they have been found as part of the biofilm community attached to the pipe walls [119, 120].

Studies showed that *Aeromonas* can survive and grow in drinking water supplies; it can resist water treatment strategies such as rapid/slow sand filtration and use activated granular carbon and even hyperchlorination [121-125]. Low numbers of mesophilic *Aeromonas* have been found in final waters in 20 Dutch plants [126]. In the drinking water, the capacity of *Aeromonas* to degrade other chitin has been tested, although no growth has been observed [6]. *Aeromonas* species are known to possess chitinases [127-129] but the question is whether it can degrade chitin since there is a difference between a chitinolytic and a chitinotrophic organism.

Other biopolymers in water environments

Carbohydrate polymers are one of the most prevalent carbon sources for microbes on Earth. There is no accumulation of these polymers in soil or marine ecosystems because they are completely degraded and recycled. However, the diverse chemical and structural properties of carbohydrate polymers require microorganisms to produce a specialized set of enzymes in order to degrade and utilize them as a nutrient source. It has been well documented in the last decade on the presence of bacteria that can degrade polymers, including cellulose, the most abundant polymer in nature [130, 131] and chitin, the second most abundant, highly present in aquatic environments [132-134].

Their roles vary from serving as storage polymers to acting as structural components in plant cell walls and the shells of invertebrates dwelling on the ocean floor. The transformation of different carbohydrate polymers in aquatic and terrestrial ecosystems is crucial for the global carbon cycle. For example, in aquatic ecosystems, macroalgae sequester approximately 173 Tg of carbon dioxide per year [135].

Bacterial regrowth in the drinking water poses high risks especially for its potential to harbor potential pathogens. However, how bacteria can grow in such a nutrient poor environment is still debated. Still the drinking water microbial community contains, apart from bacteria, fungi and even small invertebrates. However, despite the low nutrient availability in such an environment and extensive treatment, some bacteria can regrow and maintain their stability across the drinking water distribution system. This may indicate the presence of some nutrient sources that are stable, degradable and potentially carbon rich.

The presence of other organisms like fungi and even invertebrates may indicate the presence of this carbon rich source. The bacterial degradation of such recalcitrant polymers indicated a highly complex and diverse enzymatic process that a bacterium must possess to feed on the polymer. As previously mentioned, fungi and invertebrates that occur in DWS contain such polymers especially chitin and chitosan. Chitin is the structural component of the exoskeleton of invertebrates and of the cell wall of fungi [136]. The presence of slow degradable polymers in the DWS may be the reason why bacterial regrowth occurs in the first place [137, 90]. Since there are no chemical methods to detect the direct presence of polymers in drinking water, it is extremely hard to assess its concentration. Once a carbohydrate polymer is degraded, its byproducts are rapidly and easily taken up by various bacteria, including the ones incapable of polymer degradation [138].

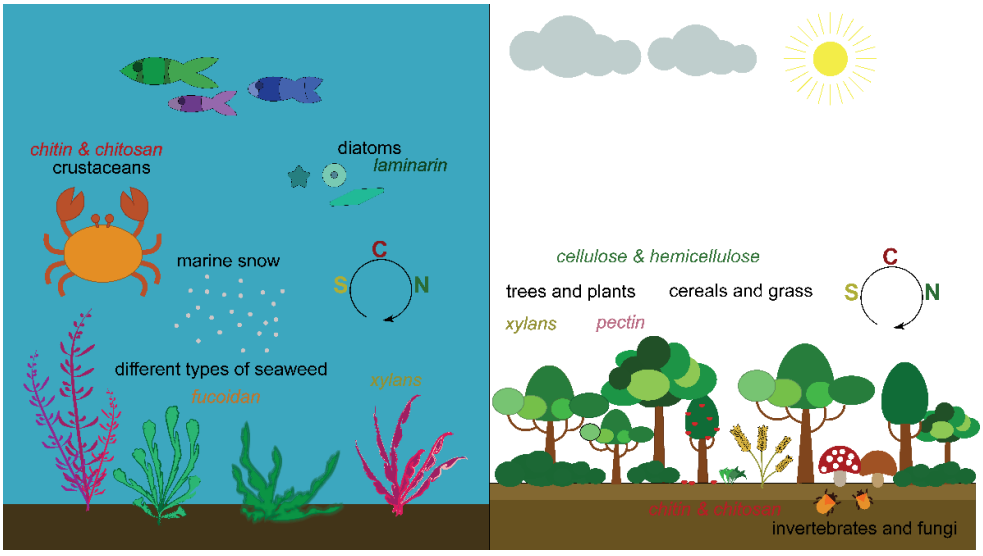


Figure 7: Polymer diversity across marine and soil environments. Some polymers are specific to the marine environment like laminarin and fucoidan, while pectin and cellulose are predominant in soil environments.

In marine environments, various polymers such as chitin, laminarin [139, 140], fucoidan [141], and xylan [142] are available to microbes (**Figure 7**). These polymers are structural components of algae, including diatoms, and invertebrates like crustaceans [134]. With the decay of animals and plants in aquatic ecosystems, these polymers become accessible through a phenomenon known as marine snow. Marine snow particles vary in size and composition, including detritus, dead or decaying organisms, fecal matter, microorganisms, and other debris. Marine snow plays a pivotal role in ocean ecosystems by distributing organic matter and nutrients from the surface waters to the deeper ocean layers. This process, known as the biological pump, transfers carbon and energy from the surface to the deep ocean, influencing global carbon cycling. It also serves as a crucial food source for animals, bacteria, and archaea dwelling in the deep ocean and on the ocean floor [143].

The presence of specific biopolymers in ecological niches leads to the specialization of microorganisms, enabling them to thrive in these environments. Conversely, some microbes demonstrate high adaptability to various carbon sources, making them versatile across a broader spectrum of biopolymers. However, their degradation often requires multiple enzymes and binding proteins. The hydrolysis products must then be taken up through generalized or specialized transporters and subsequently channeled into the central carbon metabolism.

Wastewater

In The Netherlands, approximately 3.94 *10^6 m^3 of wastewater is produced daily by households and industries that requires proper biological treatment before being reused or released into the environment [144]. Globally, 359.4×109 m3 yr−1 of wastewater is produced from which approximately 50% undergoes treatment [145].

Technological efforts have been made in order to capitalize the resources that can be extracted from wastewater. Since 98-99% wt of the composition wastewater is water, the idea of recovering water for domestic use has been exploited. However, there is enormous potential for recovering organic compounds like VFAs (volatile fatty acids), methane, alginate, EPS, cellulose, feedstock but also inorganic compounds like nitrogen and phosphorus that can be used in agriculture [146, 147]. The reclaimed water that is obtained from the wastewater treatment, however, cannot satisfy the biological and chemical safety requirements to become potable but it can be used for other domestic uses like cleaning and flushing although in some instances the treatment is advanced enough to allow domestic consumption [148]. Since the wastewater effluent has been extensively studied for recovery of valuable compounds, the influent has been scarcely studied and mostly in the context of antibiotic resistance and pharmaceutical byproducts of drugs. The influent can deliver however, important information regarding the population health thanks to the development of screening techniques.

The advent of technologies capable of screening multiple species lead to increased interest in investigating more complex environments that were hardly accessible before due to the lack of resolution and accuracy. The most extensive use of mass spectrometry technology in wastewater was to monitor drug consumption across population by measuring drug metabolites that are eliminated through urine. This approach helped the authorities to know exactly the extent of drug consumption without disclosure of private information over consumers or false declarations coming from citizens [149].

The first time when viral genetic material was amplified from sewage was the was during a polio outbreak in Israel in 2013 [150]. Also, a link between the amount of viral genetic material and the extent of the disease across the population was observed.

The SARS-CoV pandemic outbreak from 2019 brought into perspective the possibility to monitor population's health indirectly with the scope of preventing outbreaks and install early quarantine. Surveillance of wastewater during the pandemic proved to be extremely useful in detecting virus variants even before a potential outbreak [151, 152]. Usually, genetic techniques are employed for monitoring specific pathogens in wastewater. However, the matrix of this type of sample interferes and hampers DNA amplification. Moreover, DNA based techniques do not target the free-floating proteins of human origin that can serve as biomarkers. Mass spectrometry-based proteomics can target these biomarkers, which can

be even expanded to target certain bacteria. Until now, only a few articles investigated metaproteomics to studying wastewater [153-156]. Urine represent less than 1% of the total wastewater [157] but it is a good source for investigating huma biomarkers. Urine is a product of homeostasis, it reflects changes in chemical compositions across multiple organs [158]. Among the most prevalent human biomarkers are immunoglobulins, uromodulin, interleukins [159]. Depending on the disease, some biomarkers have to be present exceeding a certain threshold. When monitoring the wastewater, inferring the amount of a certain biomarker and trying to interpret it as a sign of a disease becomes cumbersome, since thousands of individuals contribute to the biomarker pool in a wastewater. Therefore, the focus should be on monitoring those biomarkers whose direct presence showcase a certain disease. In the case of predicting viral or bacterial outbreaks, the presence of a certain pathogen in the wastewater can be a first sign of a potential outbreak. However, extensive monitoring over a certain period of time may indeed help in predicting the potential epidemic outbreak by measuring the increase of the bacterial amount over time.

Thesis goal and outline of chapters

Natural and engineered water environments are complex microbial ecosystems. Given their crucial role in managing one of Earth's most valuable resources, both wastewater and drinking water systems have become the focus of extensive research. Advanced meta-omics approaches, particularly microbial proteomics, are especially promising for exploring these environments. Key questions arise which can be addressed with advanced microbial omics methods, such how microbes like *Aeromonas* survive in nutrient-poor settings like drinking water distribution systems - specifically, whether they can grow on various biopolymers as their sole carbon and nitrogen sources. At the same time, there is growing interest in developing routine metaproteomics workflows to investigate other environments, such as wastewater, with the potential to expand the repertoire of public health markers.
Therefore, the primary objectives of this thesis are:
i) to investigate *Aeromonas* ability to utilize the carbohydrate polymer chitin, as found in small invertebrates in drinking water distribution systems,
ii) to investigate *Aeromonas* ability to utilize different biopolymers as found in various water environments
iii) to further advance metaproteomics in studying complex water ecosystems through improvements in sample preparation and data analysis.
Chapter 2 examines the ability of *Aeromonas* to degrade and utilize chitin as its sole carbon and nitrogen source. Proteomic analysis of both the secretome and intracellular proteins revealed a large reorganization of the proteome and the spectrum of secreted enzymes, when *Aeromonas* was grown on chitin compared to glucose.

Chapter 3 explores *Aeromonas'* metabolic versatility by investigating growth on various biopolymers. The bacterium successfully grew on multiple polymers including pullulan, starch, dextrin, EPS extract, and collagen. A spectrum of secreted enzymes dedicated to the degradation of these polymers was identified. This study further underscores its broad metabolic adaptability, reinforcing its classification as a generalist.

Chapter 4 explores the broader context of the drinking water microbiome present in sediments of drinking water distribution systems, using a combination of metagenomics and metaproteomics. The analysis revealed a highly diverse and complex bacterial ecosystem. In addition to bacteria, proteins from invertebrates, fungi, protists, and amoebas were identified, underscoring the largely unexplored nature of this ecosystem. Furthermore, the range of detected glycosyl hydrolase genes offered new insights into the survival strategies of microorganisms in these oligotrophic environments.

Chapter 5 pushes the boundaries of metaproteomics to the study of raw wastewater. The complex background matrix of wastewater presents significant challenges in protein identification and quantification. However, a newly developed sample preparation method suitable for multiplexing, combined with a de novo sequencing-based data processing pipeline, enabled fully untargeted detection of a broad spectrum of microbes across all domains of life, along with over 200 human proteins. This includes potential pathogens and human biomarkers relevant to disease surveillance and public health monitoring.

References

1. Webster TM, Fierer N. Microbial Dynamics of Biosand Filters and Contributions of the Microbial Food Web to Effective Treatment of Wastewater-Impacted Water Sources. 2019;85(17):e01142-19.
2. Michael T. Madigan KSB, Daniel H. Buckley, W. Matthew Sattley and David A. Stahl. Brock Biology of Microorganisms. 15th edition ed. 330 Hudson Street, NY NY 10030: Pearson; 2018 28 Mar 2018. 1064 p.
3. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiology and molecular biology reviews : MMBR. 2004;68(4):669-85.
4. Gibbs RA. The Human Genome Project changed everything. Nature reviews Genetics. 2020;21(10):575-6.
5. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chemistry & biology. 1998;5(10):R245-9.
6. van Bel N, van der Wielen P, Wullings B, van Rijn J, van der Mark E, Ketelaars H, et al. Aeromonas species from non-chlorinated distribution systems and their competitive planktonic growth in drinking water. Appl Environ Microbiol. 2021;87(5).

7. Meyer F, Fritz A, Deng Z-L, Koslicki D, Lesker TR, Gurevich A, et al. Critical Assessment of Metagenome Interpretation: the second round of challenges. Nature Methods. 2022;19(4):429-40.
8. Harris ZN, Dhungel E, Mosior M, Ahn T-H. Massive metagenomic data analysis using abundance-based machine learning. Biology Direct. 2019;14(1):12.
9. Mathieu A, Leclercq M, Sanabria M, Perin O, Droit A. Machine Learning and Deep Learning Applications in Metagenomic Taxonomy and Functional Annotation. 2022;13.
10. Yates JR, III. The Revolution and Evolution of Shotgun Proteomics for Large-Scale Proteome Analysis. Journal of the American Chemical Society. 2013;135(5):1629-40.
11. Smoluch M, Silberring J. A Brief History of Mass Spectrometry. Mass Spectrometry2019. p. 5-8.
12. Nier AO. A Mass Spectrometer for Routine Isotope Abundance Measurements. 1940;11:212-6.
13. Biemann K. Four decades of structure determination by mass spectrometry: from alkaloids to heparin. Journal of the American Society for Mass Spectrometry. 2002;13(11):1254-72.
14. Hunt DF, Buko AM, Ballard JM, Shabanowitz J, Giordani AB. Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. Biomedical mass spectrometry. 1981;8(9):397-408.
15. Makarov A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Analytical Chemistry. 2000;72(6):1156-62.
16. Eliuk S, Makarov A. Evolution of Orbitrap Mass Spectrometry Instrumentation. Annual review of analytical chemistry (Palo Alto, Calif). 2015;8:61-80.
17. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new mass spectrometer. Journal of mass spectrometry : JMS. 2005;40(4):430-43.
18. Doerr A. DIA mass spectrometry. Nature Methods. 2015;12(1):35-.
19. Zhao J, Yang Y, Xu H, Zheng J, Shen C, Chen T, et al. Data-independent acquisition boosts quantitative metaproteomics for deep characterization of gut microbiota. npj Biofilms and Microbiomes. 2023;9(1):4.
20. Pietilä S, Suomi T, Elo LL. Introducing untargeted data-independent acquisition for metaproteomics of complex microbial samples. ISME Communications. 2022;2(1):51.
21. Marcotte EM. How do shotgun proteomics algorithms identify proteins? Nature Biotechnology. 2007;25(7):755-7.
22. Van Den Bossche T, Beslic D, van Puyenbroeck S, Suomi T, Holstein T, Martens L, et al. Metaproteomics Beyond Databases: Addressing the Challenges and Potentials of De Novo Sequencing.n/a(n/a):e202400321.
23. Rethink what is possible Orbitrap Astral mass spectrometer [Available from: https://pragolab.cz/documents/br-001728-ms-orbitrap-astral-mass-spectrometer-br001728-en.pdf.
24. Koch HM, Kaspar-Schoenefeld S, Goedecke N, Raether O, Drechsler N, Krause M, et al., editors. PASEFTM on a timsTOF Pro defines new performance standards for

shotgun proteomics with dramatic improvements in MS/MS data acquisition rates and sensitivity2018.
25. Yu L, Kang X, Li F, Mehrafrooz B, Makhamreh A, Fallahi A, et al. Unidirectional single-file transport of full-length proteins through a nanopore. Nature Biotechnology. 2023;41(8):1130-9.
26. Ouldali H, Sarthak K, Ensslen T, Piguet F, Manivet P, Pelta J, et al. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nature Biotechnology. 2020;38(2):176-81.
27. Huffman RG, Leduc A, Wichmann C, Di Gioia M, Borriello F, Specht H, et al. Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. Nature Methods. 2023;20(5):714-22.
28. Steen H, Mann M. The abc's (and xyz's) of peptide sequencing. Nature Reviews Molecular Cell Biology. 2004;5(9):699-711.
29. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Current opinion in structural biology. 2008;18(5):581-6.
30. Hou S, Pfreundt U, Miller D, Berman-Frank I, Hess WR. mdRNA-Seq analysis of marine microbial communities from the northern Red Sea. Scientific Reports. 2016;6(1):35470.
31. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.
32. Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS letters. 2009;583(24):3966-73.
33. Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R, et al. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Molecular systems biology. 2011;7:511.
34. Bashiardes S, Zilberman-Schapira G, Elinav E. Use of Metatranscriptomics in Microbiome Research. Bioinformatics and biology insights. 2016;10:19-25.
35. Huch S, Nersisyan L, Ropat M, Barrett D, Wu M, Wang J, et al. Atlas of mRNA translation and decay for bacteria. Nature Microbiology. 2023;8(6):1123-36.
36. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass contributions in microbial communities via metaproteomics. Nat Commun. 2017;8(1):1558.
37. Cappelletti V, Hauser T, Piazza I, Pepelnjak M, Malinovska L, Fuhrer T, et al. Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. Cell. 2021;184(2):545-59.e22.
38. Blakeley-Ruiz JA, Kleiner M. Considerations for constructing a protein sequence database for metaproteomics. Computational and Structural Biotechnology Journal. 2022;20:937-52.
39. Jagtap P, Goslinga J, Kooren JA, McGowan T, Wroblewski MS, Seymour SL, et al. A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies. Proteomics. 2013;13(8):1352-7.

40. Kleikamp HBC, Pronk M, Tugui C, Guedes da Silva L, Abbas B, Lin YM, et al. Database-independent de novo metaproteomics of complex microbial communities. Cell systems. 2021;12(5):375-83.e5.
41. Kleikamp HBC, van der Zwaan R, van Valderen R, van Ede JM, Pronk M, Schaasberg P, et al. NovoLign: metaproteomics by sequence alignment. ISME Communications. 2024;4(1).
42. Jansson JK, Hofmockel KS. The soil microbiome—from metagenomics to metaphenomics. Current Opinion in Microbiology. 2018;43:162-8.
43. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochemical and biophysical research communications. 2016;469(4):967-77.
44. Rubio-Rincón FJ, Weissbrodt DG, Lopez-Vazquez CM, Welles L, Abbas B, Albertsen M, et al. "Candidatus Accumulibacter delftensis": A clade IC novel polyphosphate-accumulating organism without denitrifying activity on nitrate. Water Research. 2019;161:136-51.
45. Kleikamp HBC, Grouzdev D, Schaasberg P, van Valderen R, van der Zwaan R, Wijgaart RV, et al. Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. Water Res. 2023;246:120700.
46. von Bergen M, Jehmlich N, Taubert M, Vogt C, Bastida F, Herbst F-A, et al. Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. The ISME Journal. 2013;7(10):1877-85.
47. Lawson CE, Nuijten GHL, de Graaf RM, Jacobson TB, Pabst M, Stevenson DM, et al. Autotrophic and mixotrophic metabolism of an anammox bacterium revealed by in vivo 13C and 2H metabolic network mapping. The ISME Journal. 2021;15(3):673-87.
48. Kleiner M. Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. mSystems. 2019;4(3).
49. Taubert M, Vogt C, Wubet T, Kleinsteuber S, Tarkka MT, Harms H, et al. Protein-SIP enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial consortium. 2012;6(12):2291-301.
50. Pan C, Fischer CR, Hyatt D, Bowen BP, Hettich RL, Banfield JF. Quantitative tracking of isotope flows in proteomes of microbial communities. Molecular & cellular proteomics : MCP. 2011;10(4):M110.006049.
51. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. 2012;109(19):E1173-E82.
52. Corbera-Rubio F, Laureni M, Koudijs N, Müller S, van Alen T, Schoonenberg F, et al. Meta-omics profiling of full-scale groundwater rapid sand filters explains stratification of iron, ammonium and manganese removals. Water Research. 2023;233:119805.
53. Wang T, Li L, Figeys D, Liu Y-Y. Pairing metagenomics and metaproteomics to characterize ecological niches and metabolic essentiality of gut microbiomes. ISME Communications. 2024;4(1).

54. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nature Medicine. 2019;25(6):968-76.
55. Singh NK, Wood JM, Patane J, Moura LMS, Lombardino J, Setubal JC, et al. Characterization of metagenome-assembled genomes from the International Space Station. Microbiome. 2023;11(1):125.
56. Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, et al. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nature Communications. 2018;9(1):870.
57. Barnum TP, Figueroa IA, Carlström CI, Lucas LN, Engelbrektson AL, Coates JD. Genome-resolved metagenomics identifies genetic mobility, metabolic interactions, and unexpected diversity in perchlorate-reducing communities. The ISME Journal. 2018;12(6):1568-81.
58. Anthony WE, Allison SD, Broderick CM, Chavez Rodriguez L, Clum A, Cross H, et al. From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology. Environmental Microbiome. 2024;19(1):56.
59. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. 2012;109(52):21390-5.
60. Sorokin DY, Merkel AY, Messina E, Tugui C, Pabst M, Golyshin PN, et al. Anaerobic carboxydotrophy in sulfur-respiring haloarchaea from hypersaline lakes. The ISME Journal. 2022;16(6):1534-46.
61. Waschulin V, Borsetto C, James R, Newsham KK, Donadio S, Corre C, et al. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. The ISME Journal. 2022;16(1):101-11.
62. Eddy SR. Accelerated Profile HMM Searches. PLoS computational biology. 2011;7(10):e1002195.
63. Blum M, Andreeva A, Florentino Laise C, Chuguransky Sara R, Grego T, Hobbs E, et al. InterPro: the protein sequence classification resource in 2025. Nucleic acids research. 2024;53(D1):D444-D56.
64. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403-10.
65. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic acids research. 2020;49(D1):D412-D9.
66. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic acids research. 2012;40(W1):W445-W51.
67. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, et al. PROSITE: a documented database using patterns and profiles as motif descriptors. Briefings in bioinformatics. 2002;3(3):265-74.
68. Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ. PRINTS--a database of protein motif fingerprints. Nucleic acids research. 1994;22(17):3590-6.

69. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.
70. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. 2022:2021.10.04.463034.
71. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic acids research. 2003;31(13):3381-5.
72. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic acids research. 2015;43(W1):W174-81.
73. Kulmanov M, Hoehndorf R. DeepGOPlus: improved protein function prediction from sequence. Bioinformatics. 2019;36(2):422-9.
74. Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, et al. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. IEEE transactions on pattern analysis and machine intelligence. 2022;44(10):7112-27.
75. Gligorijević V, Renfrew PD, Kosciolek T, Leman JK, Berenberg D, Vatanen T, et al. Structure-based protein function prediction using graph convolutional networks. Nature Communications. 2021;12(1):3168.
76. Asgari E, Mofrad MR. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. PLoS One. 2015;10(11):e0141287.
77. Brandes N, Ofer D, Peleg Y, Rappoport N, Linial M. ProteinBERT: a universal deep-learning model of protein sequence and function. Bioinformatics. 2022;38(8):2102-10.
78. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic acids research. 2023;51(D1):D638-d46.
79. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research. 2010;38(Web Server issue):W214-20.
80. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019;20(1):238.
81. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution. 2021;38(12):5825-9.
82. Santos ASP, Pachawo V, Melo MC, Vieira JMP. Progress on legal and practical aspects on water reuse with emphasis on drinking water – an overview. Water Supply. 2021;22(3):3000-14.
83. Cárdenas Belleza GA, Bierkens MFP, van Vliet MTH. Sectoral water use responses to droughts and heatwaves: analyses from local to global scales for 1990–2019. Environmental Research Letters. 2023;18(10):104008.
84. Baker MN. The Quest for Pure Water: the History of Water Purification from the Earliest Records to the Twentieth Century. 2nd Edition ed. Denver: Literary Licensing, LLC (October 13, 2012); 1981.

85. Buchan J. Crowded with Genius: The Scottish Enlightenment: Edinburgh's Moment of the Mind. First ed: Harper Perennial; 2004.
86. Tulchinsky TH. John Snow, Cholera, the Broad Street Pump; Waterborne Diseases Then and Now: Case Studies in Public Health. 2018:77-99. doi: 10.1016/B978-0-12-804571-8.00017-2. Epub 2018 Mar 30.
87. van Lieverloo JHM, Hoogenboezem W, Veenendaal G, van der Kooij D. Variability of invertebrate abundance in drinking water distribution systems in the Netherlands in relation to biostability and sediment volumes. Water Research. 2012;46(16):4918-32.
88. Liu G, Bakker GL, Li S, Vreeburg JH, Verberk JQ, Medema GJ, et al. Pyrosequencing reveals bacterial communities in unchlorinated drinking water distribution system: an integral study of bulk water, suspended solids, loose deposits, and pipe wall biofilm. Environ Sci Technol. 2014;48(10):5467-76.
89. Proctor CR, Hammes F. Drinking water microbiology—from measurement to management. Current Opinion in Biotechnology. 2015;33:87-94.
90. Hijnen WAM, Brouwer-Hanzens A, Schurer R, Wagenvoort AJ, van Lieverloo JHM, van der Wielen PWJJ. Influence of biopolymers, iron, biofouling and Asellus aquaticus on Aeromonas regrowth in three non-chlorinated drinking water distribution systems. Journal of Water Process Engineering. 2024;61:105293.
91. Jongenburger I, Reij MW, Boer EPJ, Gorris LGM, Zwietering MH. Factors influencing the accuracy of the plating method used to enumerate low numbers of viable micro-organisms in food. International Journal of Food Microbiology. 2010;143(1):32-40.
92. Safford HR, Bischel HN. Flow cytometry applications in water treatment, distribution, and reuse: A review. Water Res. 2019;151:110-33.
93. Duhaime MB, Deng L, Poulos BT, Sullivan MB. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. Environmental microbiology. 2012;14(9):2526-37.
94. Klappenbach JA, Saxman PR, Cole JR, Schmidt TM. rrndb: the Ribosomal RNA Operon Copy Number Database. Nucleic acids research. 2001;29(1):181-4.
95. Wu JH, Hong PY, Liu WT. Quantitative effects of position and type of single mismatch on single base primer extension. Journal of microbiological methods. 2009;77(3):267-75.
96. Props R, Rubbens P, Besmer M, Buysschaert B, Sigrist J, Weilenmann H, et al. Detection of microbial disturbances in a drinking water microbial community through continuous acquisition and advanced analysis of flow cytometry data. Water Research. 2018;145:73-82.
97. Pick FC, Fish KE, Biggs CA, Moses JP, Moore G, Boxall JB. Application of enhanced assimilable organic carbon method across operational drinking water systems. PLOS ONE. 2019;14(12):e0225477.
98. Liu G, Ling FQ, Magic-Knezev A, Liu WT, Verberk JQJC, Van Dijk JC. Quantification and identification of particle-associated bacteria in unchlorinated drinking water

from three treatment plants by cultivation-independent methods. Water Research. 2013;47(10):3523-33.
99. Fiksdal L, Tryland I. Application of rapid enzyme assay techniques for monitoring of microbial water quality. Current Opinion in Biotechnology. 2008;19(3):289-94.
100. Lautenschlager K, Hwang C, Ling F, Liu W-T, Boon N, Köster O, et al. Abundance and composition of indigenous bacterial communities in a multi-step biofiltration-based drinking water treatment plant. Water Research. 2014;62:40-52.
101. Liu G, Bakker GL, Li S, Vreeburg JHG, Verberk JQJC, Medema GJ, et al. Pyrosequencing Reveals Bacterial Communities in Unchlorinated Drinking Water Distribution System: An Integral Study of Bulk Water, Suspended Solids, Loose Deposits, and Pipe Wall Biofilm. Environmental science & technology. 2014;48(10):5467-76.
102. Liu G, Van der Mark EJ, Verberk JQ, Van Dijk JC. Flow cytometry total cell counts: a field study assessing microbiological water quality and growth in unchlorinated drinking water distribution systems. Biomed Res Int. 2013;2013:595872.
103. Vavourakis CD, Heijnen L, Peters M, Marang L, Ketelaars HAM, Hijnen WAM. Spatial and Temporal Dynamics in Attached and Suspended Bacterial Communities in Three Drinking Water Distribution Systems with Variable Biological Stability. Environ Sci Technol. 2020;54(22):14535-46.
104. Brumfield KD, Hasan NA, Leddy MB, Cotruvo JA, Rashed SM, Colwell RR, et al. A comparative analysis of drinking water employing metagenomics. PLoS One. 2020;15(4):e0231210.
105. Delafont V, Brouke A, Bouchon D, Moulin L, Héchard Y. Microbiome of free-living amoebae isolated from drinking water. Water Res. 2013;47(19):6958-65.
106. Corsaro D, Pages GS, Catalan V, Loret JF, Greub G. Biodiversity of amoebae and amoeba-associated bacteria in water treatment plants. Int J Hyg Environ Health. 2010;213(3):158-66.
107. Kanzler D, Buzina W, Paulitsch A, Haas D, Platzer S, Marth E, et al. Occurrence and hygienic relevance of fungi in drinking water. Mycoses. 2008;51(2):165-9.
108. Ketelaars HAM, Wagenvoort AJ, Peters M, Wunderer J, Hijnen WAM. Taxonomic diversity and biomass of the invertebrate fauna of nine drinking water treatment plants and their non-chlorinated distribution systems. Water Res. 2023;242:120269.
109. Bichai F, Payment P, Barbeau B. Protection of waterborne pathogens by higher organisms in drinking water: a review. Can J Microbiol. 2008;54(7):509-24.
110. Shen L, Zhang Z, Wang R, Wu S, Wang Y, Fu S. Metatranscriptomic data mining together with microfluidic card uncovered the potential pathogens and seasonal RNA viral ecology in a drinking water source. Journal of Applied Microbiology. 2023;135(1).
111. Fernández-Bravo A, Figueras MJ. An Update on the Genus Aeromonas: Taxonomy, Epidemiology, and Pathogenicity. Microorganisms. 2020;8(1).
112. Schubert RJBsmodb, 8th edn. Williams, Wilkins B. Genus II. Aeromonas Kluyver and van Niel 1936, 398. 1974.

113. Ott BM, Cruciger M, Dacks AM, Rio RVM. Hitchhiking of host biology by beneficial symbionts enhances transmission. Scientific Reports. 2014;4(1):5825.
114. September S, Els F, Venter S, Brözel VJJow, health. Prevalence of bacterial pathogens in biofilms of drinking water distribution systems. 2007;5(2):219-27.
115. Farkas A, Drăgan-Bularda M, Ciatarâş D, Bocoş B, Ţigan ŞJJoW, Health. Opportunistic pathogens and faecal indicators in drinking water associated biofilms in Cluj, Romania. 2012;10(3):471-83.
116. Chauret C, Volk C, Creason R, Jarosh J, Robinson J, Warnes CJCjom. Detection of Aeromonas hydrophila in a drinking-water distribution system: a field and pilot study. 2001;47(8):782-6.
117. Jahid IK, Ha S-DJFP, Disease. Inactivation kinetics of various chemical disinfectants on Aeromonas hydrophila planktonic cells and biofilms. 2014;11(5):346-53.
118. van der Kooij D, Hijnen WA. Nutritional versatility and growth kinetics of an Aeromonas hydrophila strain isolated from drinking water. Applied and environmental microbiology. 1988;54(11):2842-51.
119. Egorov AI, Best JM, Frebis CP, Karapondo MS. Occurrence of Aeromonas spp. in a random sample of drinking water distribution systems in the USA. Journal of water and health. 2011;9(4):785-98.
120. Bomo AM, Storey MV, Ashbolt NJ. Detection, integration and persistence of aeromonads in water distribution pipe biofilms. Journal of water and health. 2004;2(2):83-96.
121. Burke V, Robinson J, Gracey M, Peterson D, Partridge K. Isolation of Aeromonas hydrophila from a metropolitan water supply: seasonal correlation with clinical isolates. Applied and environmental microbiology. 1984;48(2):361-6.
122. Havelaar AH, During M, Versteegh JF. Ampicillin-dextrin agar medium for the enumeration of Aeromonas species in water by membrane filtration. The Journal of applied bacteriology. 1987;62(3):279-87.
123. Huys G, Kämpfer P, Altwegg M, Kersters I, Lamb A, Coopman R, et al. Aeromonas popoffii sp. nov., a mesophilic bacterium isolated from drinking water production plants and reservoirs. International journal of systematic bacteriology. 1997;47(4):1165-71.
124. Stelzer W, Jacob J, Feuerpfeil I, Schulze E. [The occurrence of aeromonads in a drinking water supply system]. Zentralblatt fur Mikrobiologie. 1992;147(3-4):231-5.
125. Janda JM, Abbott SL. Evolving concepts regarding the genus Aeromonas: an expanding Panorama of species, disease presentations, and unanswered questions. Clin Infect Dis. 1998;27(2):332-44.
126. Havelaar A, Versteegh J, During MJZfHuUIJoH, Medicine E. The presence of Aeromonas in drinking water supplies in The Netherlands. 1990;190(3):236-56.
127. Ueda M, Fujiwara A, Kawaguchi T, Arai M. Purification and some properties of six chitinases from Aeromonas sp. no. 10S-24. Bioscience, biotechnology, and biochemistry. 1995;59(11):2162-4.
128. Pentekhina I, Hattori T, Tran DM, Shima M, Watanabe T, Sugimoto H, et al. Chitinase system of Aeromonas salmonicida, and characterization of enzymes

involved in chitin degradation. Bioscience, biotechnology, and biochemistry. 2020;84(9):1936-47.
129. Stumpf AK, Vortmann M, Dirks-Hofmeister ME, Moerschbacher BM, Philipp B. Identification of a novel chitinase from Aeromonas hydrophila AH-1N for the degradation of chitin within fungal mycelium. FEMS Microbiology Letters. 2018;366(1).
130. Delmer DP. CELLULOSE BIOSYNTHESIS: Exciting Times for A Difficult Field of Study. Annual review of plant physiology and plant molecular biology. 1999;50:245-76.
131. Somerville C. Cellulose synthesis in higher plants. Annual review of cell and developmental biology. 2006;22:53-78.
132. McCarthy M, Pratum T, Hedges J, Benner R. Chemical composition of dissolved organic nitrogen in the ocean. Nature. 1997;390(6656):150-4.
133. Aluwihare LI, Repeta DJ, Pantoja S, Johnson CG. Two chemically distinct pools of organic nitrogen accumulate in the ocean. Science (New York, NY). 2005;308(5724):1007-10.
134. Peters WJCB, Biochemistry PPBC. Occurrence of chitin in Mollusca. 1972;41(3):541-50.
135. Krause-Jensen D, Lavery P, Serrano O, Marbà N, Masque P, Duarte CM. Sequestration of macroalgal carbon: the elephant in the Blue Carbon room. 2018;14(6):20180236.
136. Nunes CS, Philipps-Wiemann P. Chapter 18 - Chitinases. In: Nunes CS, Kumar V, editors. Enzymes in Human and Animal Nutrition: Academic Press; 2018. p. 361-78.
137. Hijnen WAM, Schurer R, Bahlman JA, Ketelaars HAM, Italiaander R, van der Wal A, et al. Slowly biodegradable organic compounds impact the biostability of non-chlorinated drinking water produced from surface water. Water Res. 2018;129:240-51.
138. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. The ISME Journal. 2019;13(5):1119-32.
139. Biersmith A, Benner RJMC. Carbohydrates in phytoplankton and freshly produced dissolved organic matter. 1998;63(1-2):131-44.
140. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, et al. Laminarin is a major molecule in the marine carbon cycle. 2020;117(12):6599-607.
141. Deniaud-Bouët E, Kervarec N, Michel G, Tonon T, Kloareg B, Hervé C. Chemical and enzymatic fractionation of cell walls from Fucales: insights into the structure of the extracellular matrix of brown algae. Annals of Botany. 2014;114(6):1203-16.
142. Ebringerová A, Heinze TJMrc. Xylan and xylan derivatives–biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. 2000;21(9):542-56.
143. Decho AW, Gutierrez T. Microbial Extracellular Polymeric Substances (EPSs) in Ocean Systems. 2017;8.
144. Agency EE. [Available from: https://water.europa.eu/freshwater/countries/uwwt/netherlands.

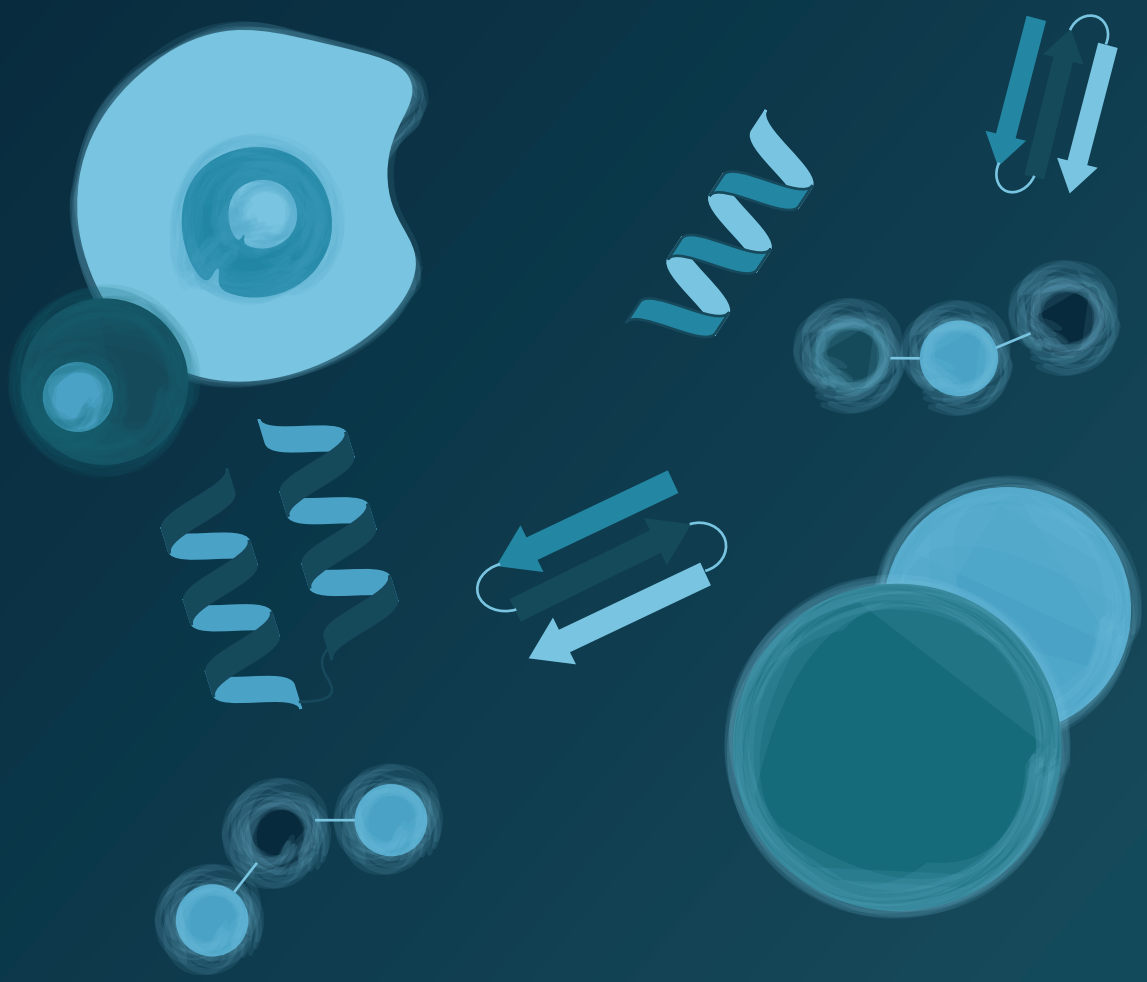
145. Jones ER, van Vliet MTH, Qadir M, Bierkens MFP. Country-level and gridded estimates of wastewater production, collection, treatment and reuse. Earth Syst Sci Data. 2021;13(2):237-54.
146. Ostermeyer P, Capson-Tojo G, Hülsen T, Carvalho G, Oehmen A, Rabaey K, et al. Resource recovery from municipal wastewater: what and how much is there? 2022. p. 1-19.
147. Hao X, Li J, Liu R, van Loosdrecht MCM. Resource Recovery from Wastewater: What, Why, and Where? Environmental science & technology. 2024;58(32):14065-7.
148. Deng S, Yan X, Zhu Q, Liao C. The utilization of reclaimed water: Possible risks arising from waterborne contaminants. Environmental Pollution. 2019;254:113020.
149. Huizer M, ter Laak TL, de Voogt P, van Wezel AP. Wastewater-based epidemiology for illicit drugs: A critical review on global data. Water Research. 2021;207:117789.
150. Manor Y, Shulman LM, Kaliner E, Hindiyeh M, Ram D, Sofer D, et al. Intensified environmental surveillance supporting the response to wild poliovirus type 1 silent circulation in Israel, 2013. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 2014;19(7):20708.
151. Spurbeck RR, Minard-Smith A, Catlin L. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. The Science of the total environment. 2021;789:147829.
152. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. Environmental Science & Technology Letters. 2020;7(7):511-6.
153. Carrascal M, Abian J, Ginebreda A, Barceló D. Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. Science of The Total Environment. 2020;747:141145.
154. Carrascal M, Abian J, Ginebreda A, Barceló D. Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. The Science of the total environment. 2020;747:141145.
155. Carrascal M, Sánchez-Jiménez E, Fang J, Pérez-López C, Ginebreda A, Barceló D, et al. Sewage Protein Information Mining: Discovery of Large Biomolecules as Biomarkers of Population and Industrial Activities. Environmental Science & Technology. 2023;57(30):10929-39.
156. Sánchez-Jiménez E, Abian J, Ginebreda A, Barceló D, Carrascal M. Shotgun proteomics to characterize wastewater proteins. MethodsX. 2023;11:102403.
157. Randall DG, Naidoo V. Urine: The liquid gold of wastewater. Journal of Environmental Chemical Engineering. 2018;6(2):2627-35.
158. Owens GL, Barr CE, White H, Njoku K, Crosbie EJ. Urinary biomarkers for the detection of ovarian cancer: a systematic review. Carcinogenesis. 2022;43(4):311-20.

159. Stephenson S, Eid W, Wong CH, Mercier E, D'Aoust PM, Kabir MP, et al. Urban wastewater contains a functional human antibody repertoire of mucosal origin. Water Research. 2024;267:122532.

Chapter 2

Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin

Claudia G. Tugui, Dimitry Y. Sorokin, Wim Hijnen, Julia Wunderer, Kaatje Bout, Mark C.M. van Loosdrecht and Martin Pabst



Essentially as published in:
Tugui, C. G., Sorokin, D. Y., Hijnen, W., Wunderer, J., Bout, K., van Loosdrecht, M. C., & Pabst, M. (2025). Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin. RSC Chemical Biology. Electronic supplementary information (ESI) is available via: https://doi.org/ 10.1039/d4cb00200h

Abstract

Members of the *Aeromonas* genus are commonly found in natural aquatic ecosystems. However, they are also frequently present in non-chlorinated drinking water distribution systems. High densities of these bacteria indicate favorable conditions for microbial regrowth, which is considered undesirable. Studies have indicated that the presence of *Aeromonas* is associated with loose deposits and the presence of invertebrates, specifically Asellus aquaticus. Therefore, a potential source of nutrients in these nutrient poor environments is chitin, the structural shell component in these invertebrates. In this study, we demonstrate the ability of two *Aeromonas* strains, commonly encountered in drinking water distribution systems, to effectively degrade and utilize chitin as a sole carbon and nitrogen source. We conducted a quantitative proteomics study on the cell biomass and secretome from pure strain cultures when switching the nutrient source from glucose to chitin, uncovering a diverse array of hydrolytic enzymes and metabolic pathways specifically dedicated to the utilization of chitin. Additionally, a genomic analysis of different *Aeromonas* species suggests a general ability of this genus to degrade and utilize a variety of carbohydrate biopolymers. This study indicates the relation between the utilization of chitin by *Aeromonas* and their association with invertebrates such as A. aquaticus in loose deposits in drinking water distribution systems.

Key words: *Aeromonas bestiarum*, *Aeromonas rivuli*, chitin, secretome, CAZy enzymes, drinking water

Introduction

The genus *Aeromonas* comprises a group of Gammaproteobacteria that are widely distributed in aquatic environments [1]. Some members of this genus have the potential to cause diseases in humans and other animals [2]. *Aeromonas* are also commonly found in engineered ecosystems, such as drinking water distribution systems (DWDS). However, in these environments, the bacterium is generally considered non-pathogenic [3]. Nevertheless, elevated levels of *Aeromonas* in non-chlorinated drinking water distribution systems (DWDS) are considered an indicator of favorable growth conditions for microbial growth, which may lead to compromised characteristics, including changes in taste, odour, colour, the presence of large invertebrates, as well as potential occurrence of opportunistic pathogens. Additionally, in the context of water resource management and global warming, maintaining the quality of drinking water is a persistent challenge. Therefore, studying the regrowth of microorganisms, such as *Aeromonas*, in DWDS is a crucial research objective.

The nutritional versatility of *Aeromonas* in the oligotrophic drinking water distribution systems has been investigated earlier, which showed a high affinity for amino acids, long-chain fatty acids [3, 4] and biopolymers such as starch and chitin [3, 5]. However, experiments have shown that competitive planktonic growth in drinking water is not very likely [3]. *Aeromonas* is commonly present in loose deposits and in bulk water of groundwater [6, 7] and surface water of drinking water distribution systems. This niche is also shared by invertebrates, including *A. aquaticus* especially found in drinking water distribution systems worldwide, which may be a nutrient source for microbes [8, 9]. As a result, it has been speculated that *Aeromonas* may take advantage of the carbohydrate polymer chitin, which is a key component of the exoskeletons of these invertebrates.

Chitin is generally considered the second most abundant biopolymer in nature and is the most prevalent in aquatic environments such as the oceans. As a result, this polymer plays a crucial role in the global carbon and nitrogen cycles [10]. The ability to utilize this carbohydrate polymer as nutrient source is therefore widely found among microorganisms [11-13]. However, its role in environments such as the drinking water distribution systems is only poorly understood to date.

The utilization of this carbohydrate polymer is known to require a cascade of enzymatic reactions to make it accessible for uptake and growth. The chitin backbone is made of β-(1-4)-linked N-acetylglucosamine (GlcNAc) units. The hydrolytic degradation of chitin usually starts outside the cell via different (endo- and exo-) chitinases and associated hydrolytic enzymes. Additional lytic polysaccharide monooxygenases (LPMOs) can cleave crystalline chitin via an oxidation reaction. This generates regions with decreased crystallinity which subsequently become more accessible for other chitinases [14]. Endochitinases cleave the chitin chain at random points to produce oligomers such as chitotetriose, chitotriose and

chitobiose. Exochitinases and N-acetylhexoseamindases produce the smaller chitobiose and GlcNAc monomers [10]. Chitin oligomers may also be deacetylated, which can be converted by chitosanases and hexosaminidases into glucosamine forms, and which may be even further cleaved by (non-specific) cellulases [10, 15, 16]. The uptake of the GlcNAc oligomers and monomers can be facilitated through dedicated porins, ABC and PTS transporters [17]. Further catabolism of the GlcNAc in the cytoplasm commonly starts with phosphorylation by a GlcNAc kinase [10]. Alternatively, PTS sugar transporters perform transport and phosphorylation simultaneously [10, 18]. The resulting GlcNAc-P can then be deacetylated and deaminated to produce fructose-6P, a metabolite which can directly enter glycolysis [10]. Additionally, the conversion of GlcNAc into GlcNH2 releases acetate, which further results in the production of acetyl-CoA, and the deamination releases ammonia, which can be incorporated into the proteinogenic amino acid glutamine [10]. Glutamine furthermore plays an important role as a nitrogen carrier and storage within bacterial cells [19]. Finally, the oxidative degradation of chitin by LPMOs produces oxidized GlcNAc mono- and oligomers. These oxidation products, have been reported to be directly deacetylated and deaminated to produce 2-keto-3-deoxygluconate, acetate and ammonia without prior phosphorylation [13]. This is supposed to be a common utilization route for crystalline chitin in marine Gammaproteobacteria [13]. Several bacteria, including some *Aeromonas* strains, have already been investigated for the expression and activity of a range of chitinases, such as *A. hydrophila* [18], *Aeromonas sp*. [20] and *Aeromonas caviae* [21]. Furthermore, the suspended growth of some *Aeromonas* strains on low concentrations of chitin and nitrate was demonstrated recently [3].

However, the underlaying metabolic routes and the cellular and secretome response to chitin as sole carbon and nitrogen source, has not been investigated to date. Therefore, we demonstrate the efficient growth on chitin as sole carbon and nitrogen source for two selected *Aeromonas* species, *Aeromonas* bestiarum and *Aeromonas rivuli*, which are frequently found in non-chlorinated drinking water distribution systems [3]. We performed an extensive quantitative proteomics study on their chitin degradation, uptake and catabolic network. The identified upregulated secreted enzymes, porins and transporters provide strong evidence for a dedicated chitin utilization network in *Aeromonas* members. Finally, a broader study on biopolymer degradation routes suggests that the genus *Aeromonas* possesses the ability to break down a range of different biopolymers. In summary, the study brings fundamental insights into the metabolic ability of Aeromonas in degrading and utilizing the carbohydrate polymer chitin as sole carbon and nitrogen source. This provides a better understanding of how these microbes can survive in nutrient-poor environments such as drinking water distribution systems.

Materials and Methods

Growth of *Aeromonas* on glucose and chitin. Two *Aeromonas* strains (*A. bestiarum* DSM 13956 and *A. rivuli* DSM 22539) were purchased from DSMZ (Leibniz, Germany, https://www.dsmz.de/dsmz). The cells were reactivated, and grown on glucose and peptone rich TSB medium, as recommended by DSMZ. From these cultures, glycerol (30%) stocks were obtained which were stored at -80°C until further use. The two *Aeromonas* strains, *A. rivuli* and *A. bestiarum* were inoculated on M9 minimum salt medium (Na2HPO4 6.78 g/L, KH2PO4 3 g/L, NaCl 0.5 g/L, NH4Cl 0.21 g/L) and acidic trace element solution (22) combined with 20% MgSO4 to (2 mL to 1 L of medium), and 10% CaCl2 (1 mL/L of medium), pH=6.7. Glucose was added to a concentration of 13.6 mM, and amorphous chitin, prepared through 'HCl decrystallization' from powdered defatted/deproteinated crab shells (Sigma Aldrich/Merck, US)[23], was added to a final concentration of 0.5 g/L. The cultures were grown aerobically at 33°C on a rotary shaker (Edmund Bühler GmbH, Germany) at 100 rpm. The growth was monitored using an Ultrospec 10 Cell Density Meter (Biochrom Ltd, UK). For both strains, we conducted biological triplicates of both growth conditions (glucose and chitin), resulting in a total of 12 shake-flask experiments. Microscopy. Light microscopy was performed using a Zeiss Axio Imager M2 microscope equipped with an Axiocam 305 color camera (Carl Zeiss, Germany). The microscope setting possesses a 63x and 100x oil immersion objective lens and phase contrast capabilities. The proprietary Zeiss software for image capture and analysis was Zen 3.3. Cell harvesting and supernatant concentration. From every culture, every day 1 mL of cell broth was harvested and centrifuged (14000 rpm, 10 minutes) to separate the cell pellet from the supernatant. The resulting supernatant (approx. 1 mL) and cell pellet was stored separately at -20°C, until further processed. For the analysis of chitin hydrolysis products, 50 mL of the supernatant was further filtered through sterile syringe filters (0.2 µm Sartorius) and then concentrated using a speedvac concentrator to a final volume of 500 µL. The time point with the highest OD was selected for proteome analysis. Cell lysis, protein extraction, and proteolytic digestion. Analysis of the cell biomass proteome. The cell pellets from the biological triplicates of each strain (*A. bestiarum* and *A. rivuli*) and contrasted growth conditions (glucose and chitin) were resuspended in 175 µL 50 mM TEAB buffer (with 1% NaDOC) and 175 µL B-PER buffer (Thermo Scientific, Germany) by vortexing. Then acid washed glass beads (105–212µm, Sigma Aldrich), were added and the mixtures were vortexed thoroughly, sonicated for 15 minutes and then frozen at -80°C for 15 minutes. Thereafter, the samples were thawed in a Thermocycler at 40°C and under shaking at 1000 rpm for 15 minutes. Afterwards, the samples were spun down at 14000 rpm. The supernatant was collected, and Trichloroacetic acid (TCA) was added (1 volume TCA to 4 volumes supernatant). The mixture was vortexed and incubated at 4°C for 20 minutes, then spun down at 14000 rpm for 15 minutes at 4°C.

The obtained protein pellets were once washed with ice cold acetone and then dissolved in 6 M urea (in 100 mM ammonium bicarbonate, ABC). Further, the disulfide bridges were reduced by the addition of 10 mM Dithiothreitol (DTT) and incubation for one hour at 37°C under shaking at 300 rpm. Thereafter, 20 mM iodoacetamide (IAA) was added. The mixture was kept in the dark for 30 minutes. 200 mM ABC buffer was then added to the samples to obtain a solution with <1 M Urea. Finally, proteolytic digestion was performed by adding Trypsin (0.1 µg/µL in 1 mM HCl, Sequencing Grade Modified Trypsin, Promega) at a ratio of 50:1 (w:w, Protein:Trypsin) to the sample. The proteolytic digestion was performed overnight at 37°C, under gentle shaking at 300 rpm. Peptides were desalted using an OASIS HLB solid phase extraction well plate (Waters, UK) according to the instructions of the manufacturer, speed vac dried and stored at -20°C until further processed. Analysis of the secreted proteome. 600 µL of the collected supernatants from the biological triplicates of each strain (*A. bestiarum* and *A. rivuli*) and both growth conditions (glucose and chitin) were processed with the same protocol as described for the cell pellets (see above) albeit starting directly with the TCA protein precipitation step. Quantitative shotgun proteomics and analysis of the secreted proteome. Approx. 500 ng of proteolytic digest were analyzed in duplicate injections using an EASY nano-LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 25% solvent B over 180 min, then from 25% to 55% B over 60 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 240 minutes. Solvent A was H2O containing 0.1% formic acid, and solvent B consisted of 80% ACN in H2O and 0.1% formic acid. The mass spectrometer was operated in data-dependent acquisition mode. Full MS scans were acquired from m/z 380–1250 at a mass resolution of 70 K with a maximum injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 most intense precursor ions were selected for fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at a resolution of 17.5 K with an AGC target of 2E5 and IT of 75 ms, 2.0 m/z isolation width and normalized collision energy (NCE) of 28. Processing of mass spectrometric raw data. Database searching of the shotgun proteomics raw data was performed using proteome reference databases from A. bestiarum and A. rivuli, obtained from UniprotKB (UP000224937 formerly annotated to *Aeromonas sp. CA23*, now redundant to UP001220108 from *A. bestiarum*) and NCBI (NCBI taxonomy ID: 648794) including cRAP protein sequences (https://thegpm.org/crap/), using PEAKS Studio X (Bioinformatics Solutions Inc., Canada). The database searching allowed 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, carbamidomethylation as fixed and methionine oxidation and N/Q deamidation as variable modifications. Peptide spectrum matches were filtered for 1% false discovery rates (FDR) and identifications with ≥2 unique peptides were considered as significant. The resulting protein-level FDRs were for all

samples below 1%. To confirm the accuracy of database searching when using the reference proteomes of the *Aeromonas* species, we conducted additional searches on selected replicates (one from A. bestiarum and one from A. rivuli) using a database augmented with the Drosophila melanogaster reference proteome (UP000000803), included either in its native form or with its sequences shuffled (**SI DOC Table 1** and **2**, **SI DOC Figure 6**). Quantitative analysis of the changes between chitin and glucose-grown conditions, and the cell pellet and secreted proteome abundances was performed using the PEAKSQ module (Bioinformatics Solutions Inc., Canada). Normalization was based on the total ion current (TIC), and only proteins with at least 2 unique peptides and identified in at least 2 out of 3 biological replicates were considered (prefiltering criterion). Peptide spectrum matches were filtered with a 1% false discovery rate (FDR). ANOVA was used to determine the statistical significance of the changes between the conditions, expressed as -10xlog10(p), where p corresponds to the significance testing p-value. The adjusted p-values (Benjamini-Hochberg correction) were calculated using the PEAKS Q module. The complete quantitative proteomics results including all statistical parameters are provided in SI_EXCEL_DOC_1. Annotation of structural components, functions and estimation of protein abundance. Results from PEAKSQ were further used for the analysis of expressed functions and metabolic pathways. Data processing and visualization was performed using Python 3.11.3. Furthermore, SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) [24] was used for the prediction of signal peptides in order to confirm secreted proteins (prediction > 0.9 was required to accept Sec or TAT annotation). The transporters were identified by using a regex search using the terms: “PTS”, “porin”, “permease”, “transporter”, “transport”. Different ABC transporters and their subunits were grouped into general “ABC transporter” categories in the figures. The reference proteomes were furthermore annotated with Kegg Orthology (KO) numbers using BlastKoala (https://www.kegg.jp /blastkoala/), and the KEGG pathway map was obtained from Kegg-mapper (https://www.kegg.jp /kegg/mapper/) using the annotated proteins. Sequence alignment to confirm the presence of key enzymes was conducted using the BLASTp tool (https://blast.ncbi.nlm.nih.gov/) with default parameters. The reference sequences used to identify key metabolic enzymes involved in chitin metabolism are provided in SI_EXCEL_Table_2. Additionally, manual analysis of sequence domains for putative enzymes and transporters was performed using the InterPro database (https://www.ebi.ac.uk/interpro). emPAI indices were calculated according to the formula: emPAI=10^(#observed/#observable)-1, where Nobserved is the number of peptides measured in the experiment and Nobservable is the number of theoretical peptides that a protein can produce [25]. The considered mass range for theoretical peptides was 800–3000 Da. For the multisubunit enzymes, the ratio glucose/chitin was determined by first averaging the area of the subunits and then the ratio was determined by diving the area of the respective enzymes. For homologue enzymes, only the variant with the highest

sequence coverage is shown in the figures. The complete list of identified enzymes for each metabolic pathway is provided in the SI-doc (**SI Table 3**). The analysis for potential other carbohydrate polymer degrading genes of different *Aeromonas* strains was performed using the HMM 3.3.2 tool (http://hmmer.org/) and the Carbohydrate Active Enzymes (CAZy) database (http://www.cazy.org/) as described above. For this, the following proteomes were obtained from UniProt: *Aeromonas hydrophyla* (TaxID 644, UP000000756), *Aeromonas media* (TaxID 651, UP000502657), *Aeromonas caviae* (TaxID 648, UP000280168), *Aeromonas encheleia* (TaxID 73010, UP000275277), *Aeromonas molluscorum* (TaxID 271417, UP000013526), *Aeromonas salmonicida* (TaxID 645, UP000077360), *Aeromonas schubertii* (TaxID 652, UP000058114), *Aeromonas taiwanensis* (TaxID 633417, UP000297311), *Aeromonas veronii* (TaxID 654, UP000237142). The relevant CAZy database families for the different carbohydrate polymers are shown in the Supplementary Information Table 1. Chitin degradation assay. 30 µL unfiltered supernatant of A. bestiarum and A. rivuli cultures were incubated with 1 mg of chitin suspended in 220 µL 1% PBS (phosphate buffered saline, Sigma Aldrich/Merck, US) in LC-MS grade water. Additional control samples were prepared containing only supernatant in 1% PBS, or chitin in 1% PBS. The samples were incubated at 33°C, 300 rpm for 18.5 hours. The samples were further spun down at 10K rpm using a bench top centrifuge for 2.5 minutes. GlcNAc oligomer standards were prepared from chitin following hydrolysis using HCl. For this, 250 µL of 5 M HCl was mixed with 1 mg of chitin. The mixture was incubated at 37°C, under shaking at 300 rpm for 18 hours. The sample was centrifuged at 14000 rpm for 10 minutes. The supernatant was then purified before MS analysis using 25 mg HyperSep™ Hypercarb™ solid phase extraction cartridges (Thermo Fisher Scientific, Germany). The Porous graphitic carbon PGC material was washed with 500 µL 50% acetonitrile (0.1% formic acid), and then equilibrated with 2 x 500 µL H2O. samples were then loaded on the PGC cartridge and washed with 1 x 500 µL H2O. Sugars were eluted with 1 x 300 µL 50% acetonitrile (0.1% formic acid), collected in an Eppendorf tube and speed-vac dried. Samples were stored at -20°C until further analyzed. MS analysis of chitin hydrolysis products. Solid Phase Extraction SPE purified samples from the release assays were analyzed using an ACQUITY UPLC system (Waters, UK) equipped with a Hypercarb™ separation column (100 x 1 mm, Thermo Scientific, Germany) which was connected to a QE Focus™ hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). Solvent A was 100% water (0.1% formic acid) and solvent B was 100% acetonitrile (0.1% formic acid). A gradient was maintained at 100 µL/min flow rate from 2.5% B to 40% B over 8 minutes, and constant 40% B over 5 minutes, before equilibrating back to 2.5% B. The mass spectrometer was operated in ES+ (3.25 kV), where full MS scans were acquired from 250–1500 m/z at 70K resolution and an AGC target of 1e6. The m/z values for the GlcNAc mono and oligomers are GlcNAc= C8H16NO6+, 222.09721; GlcNAc-GlcNAc= C16H29N2O11+, 425.17659; GlcNAc-GlcNAc-GlcNAc= C24H43N3O16+, 629.26378; and for the oxidized forms GlcNAc1A= C8H16NO7+,

238.09213; GlcNAc-GlcNAc1A= C16H29N2O12+, 441.1715, and for the native deacetylated forms GlcNAc-GlcNH2 = C14H27N2O10+, 383.16602. Additional parallel reaction monitoring (PRM) was performed for the native GlcNAc mono and oligomers: m/z 222 ([M+H]+, HexNAc), m/z 425 ([M+H]+, (HexNAc)2), m/z 629 ([M+H]+, (HexNAc)3), m/z 833 ([M+H]+, (HexNAc)4), and m/z 1033 ([M+H]+, (HexNAc)5) using an isolation window of 2 m/z, an AGC target of 1e5, 100 ms max IT, 2 micro scans and 35K resolution. MS raw data were analyzed using XCalibur 4.1, where the area for the MS1 precursor ion or the most abundant fragment ion for each compound was integrated.

Results

A quantitative proteomics study on *Aeromonas* grown on glucose and chitin.

Two *Aeromonas* strains, *A. bestiarum* and *A. rivuli* were cultured in the presence of either glucose or amorphous chitin and subjected to a quantitative proteomics study. The cell culture supernatants were furthermore subjected to a chitin degradation study in order to identify the size distribution of the chitin degradation products (**Figure 1**). Culturing experiments were performed in biological triplicates for both strains, resulting in a total of 12 shake flask experiments. Both *Aeromonas* strains showed immediate growth on chitin (**SI Figure 1**) Nevertheless, microscopy images showed for both strains homogeneous cultures with rod-shaped cells approximately 1–2 µm in length (**SI Figure 2**).

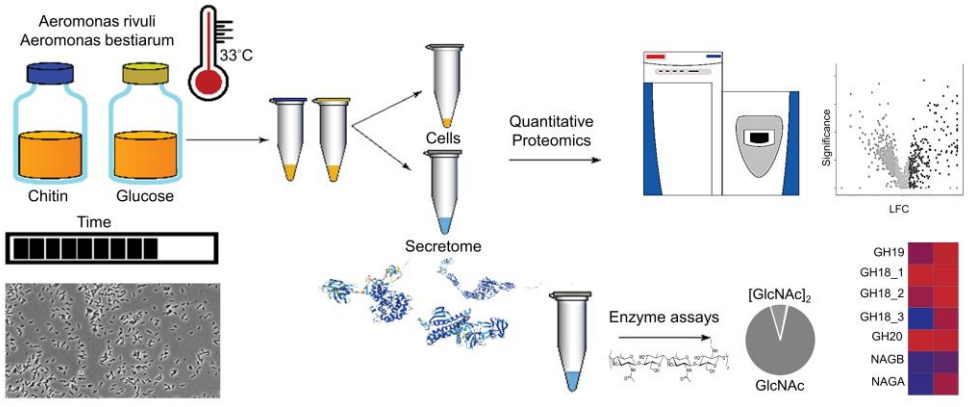


Figure 1: The graph outlines the employed workflow used to study the chitin degradation and utilization routes in *A. bestiarum* and *A. rivuli*. Both *Aeromonas* strains were cultured at 33°C with either glucose or chitin in biological triplicates. The growth of the bacteria was followed using OD660 and light microscopy over time. Quantitative proteomics was then employed to identify the enzymes secreted by the bacteria and to reveal their uptake and metabolic routes when switching the growth substrate from glucose to chitin. Annotation of hydrolytic enzymes was furthermore performed using

the CAZy reference database, and enriched and depleted functions and pathways were identified using the STRING tool. Finally, chitin degradation assays were performed using the supernatants of both strains to determine the size distribution of the chitin hydrolysis products.

Samples were selected from the plateau of the growth curves (between days 2–3) for subsequent proteomics and chitin degradation experiments. After centrifugation of the cell suspension, the cell pellet was separated from the supernatant and processed separately. All 24 samples (2 strains, 3 biological replicates for growth on glucose and chitin, and for each condition biomass and supernatant samples) were subjected to shotgun proteomic experiments. The reproducibility of the biological experiments was further confirmed by principal component analysis and hierarchical clustering of the obtained proteomics profiles (**SI Figures 3 and 4**). Additionally, annotation of hydrolytic enzymes was achieved using the CAZy reference sequence database and InterPro protein signature databases. Finally, the fresh supernatants, separated from the cell pellets, were incubated with chitin and then analyzed to identify the size distribution of the chitin degradation products.

Cellular proteome and secretome response when switching to growth on chitin.

When analyzing the proteomes of both *Aeromonas strains*, we found several enzymes, transporters and metabolic pathways expressed that indicate involvement in the utilization of chitin, including a spectrum of significant abundance changes in the secretome, when comparing *Aeromonas* grown on glucose and on chitin (**Figure 2**). In the cell biomass several abundant enzymes were detected during growth on chitin, hinting towards their involvement in the degradation, uptake or catabolization of this biopolymer. For *A. rivuli*, 3 different glycoside hydrolases (2 x GH20 and a GH13), one carbohydrate binding protein and 4 different outer membrane porins were detected. For example, the porin OprD (WP_042041362) increased 46 times in abundance in cells grown on chitin. Additionally, 3 PTS transporters (including one GlcNAc-specific transporter) and a range of different substrate-binding proteins related to different ABC transporters were found highly expressed **(SI EXCEL DOC 1**, **Figure 1**). The same was observed for *A. bestiarum* where 4 different glycoside hydrolases (2 x GH18 and 2 x GH20), 2 carbohydrate binding proteins and an outer membrane porin (OprD) were highly expressed. Like in *A. rivuli*, OprD (A0A291U719) was one of the most abundant proteins and it increased over 60 times in abundance in response to chitin. Furthermore, we detected one up-regulated GlcNAc PTS system and several substrate-binding proteins related to different ABC transporters (**SI EXCEL DOC 1**, **Figure 2**).

Quantitative comparison of the secretomes from cultures grown on glucose and chitin showed a range of upregulated extracellular hydrolytic enzymes and chitin binding proteins. For example, in *A. rivuli*, 5 chitin-specialized glycoside hydrolases (2 x GH18, GH19, GH20 and GH84), a putative carbohydrate esterase-deacetylase (CE12) and 2 sugar binding

proteins (one GlcNAc binding protein, GbpA), a range of outer membrane porins and several ABC transporters were also abundantly present in the secretome of cells which were grown on chitin.

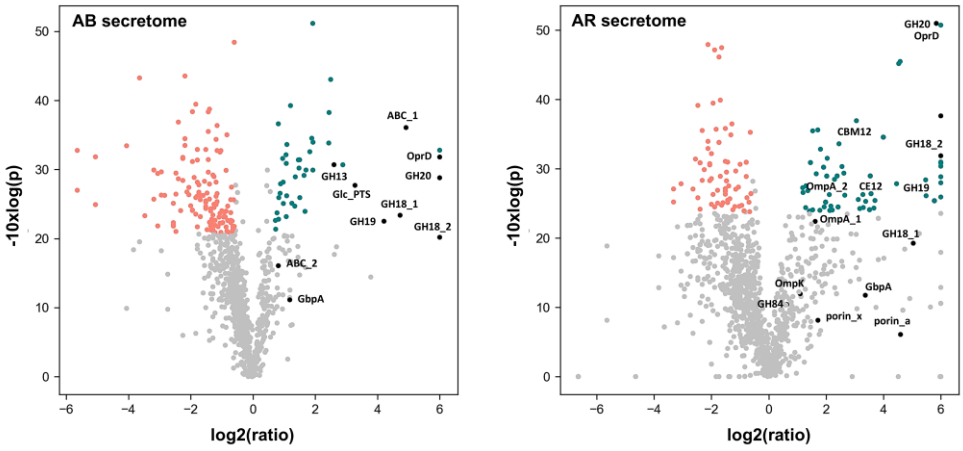


Figure 2: The volcano plots show the downregulated (red) and overexpressed (green) proteins in the secretomes of *A. bestiarum* ('Ab' left plot) and *A. rivuli* ('Ar' right plot) when comparing growth on glucose with cells grown on chitin. The x-axis shows the log2(fold-change) of the ratio chitin/glucose, and the y-axis showsthe -10log(p) value of each fold-change, whereas p represents the statistical significance of the fold change. Only proteins meeting the criteria of a fold change > 1.5 and an adjusted p-value < 0.05 are highlighted (labelled proteins are indicated with a black dot). *A. bestiarum* (Ab) secretome: GH18_1 A0A291TVA9, GH18_2 A0A291U6Z3, GH19 A0A291U6T4, GH20 A0A291U507, GH13 A0A291U830, GbpA A0A291TWD8, OprD A0A291U719, ABC_1 A0A291U0Y6, ABC_2 A0A291U6T2, Glc_PTS A0A291U5N4. *A. rivuli* (Ar) secretome: GH18_1 WP_042040344, GH18_2 WP_224432263, GH19 WP_084218236, CBM12 WP_232301920, GH20 WP_224432421, GH84 WP_224432541, GbpA WP_232302089, CE12 WP_042041941, porin_x WP_042044143, OmpA_1 WP_042040660, OmpK WP_224431060, OprD WP_042041362, OmpA_2 WP_042040662, porin_a WP_042043439. The fold changes and significance values for all proteins (cells and secretome) are reported in the SI EXCEL DOC 1. Similarly, the supernatant from *A. bestiarum* revealed 5 different glycoside hydrolases (2 x GH18, 2 x GH20 and one GH13) a GlcNAc binding protein (GbpA), one outer membrane porin, several other proteins related to PTS systems and different solute binding proteins that significantly increased abundance (**SI EXCEL DOC 1, Figure 1**).

Chitin degradation and utilization routes in *A. rivuli* and *A. bestiarum*.

The secretome and cellular proteome compared between growth on glucose and chitin as well as the oligo- and monomer ratios obtained from the chitin degradation assays allowed to elaborate on the putative chitin degradation routes in both *Aeromonas* strains. Analysis of chitin degradation products revealed domination of GlcNAc dimer and monomers for both strains (**SI Figure 5**) with only trace amounts of oxidation products (i.e. GlcNAc1A) and deacetylated forms. Consequently, the hydrolysis product profile already indicated that

both *Aeromonas* strains predominantly utilize GlcNAc dimer- and monomers (**Figure 3C**). Furthermore, we identified a similar set of 2 x GH18s (WP_042040344.1, WP_224432263.1, and A0A291U6Z3, A0A291TVA9), one GH19 (WP_084218236.1 and A0A291U6T4) and one GH20 (WP_224432421.1 and A0A291U507) in both *Aeromonas* strains (**SI EXCEL DOC 1** and **Figure 3B**). However, while both GH18s from *A. rivuli* and *A. bestiarum* showed a very high sequence identity (>80%), the GH19s had only a sequence identity of less than 50%, and the GH20s less than 25% (**SI EXCEL DOC 2**). Family 18 and 19 glycoside hydrolases are endo-acting enzymes that break down chitin at internal sites, forming chitobiose, chitotriose, and chitotetraoses. Family GH20 includes N-acetyl- glucosaminidases which act on non-reducing ends to either release dimers (chitobiose) or to further break down multimer products into GlcNAc [26]. The GH18 and GH19 families do not share sequence similarity. GH18 chitinases cleave the chitin into β-anomer products, whereas GH19 hydrolyze chitin to α-anomer by using the inverting mechanism [27, 28]. Interestingly, both strains express 2 different GH18 chitinases, one with an approx. MW of 90 kDa and a second with a MW of approx. 105 kDa. Analysis of the amino acid sequence against the InterPro databases showed that the smaller GH18 chitinase (*A. rivuli*, WP_224432263.1, 90 kDa) includes next to the glycoside hydrolases family 18 domain (PF00704) also a Chitinase A N-terminal domain (PF08329), a PKD/REJ-like domain (PF02010) and two carbohydrate-binding modules family 5/12 (IPR003610, also annotated with PF02839). The latter (PF02839) is known to specifically bind to insoluble chitin [29]. The amino acid sequence of the larger GH18 (A. rivuli WP_042040344, 105 kDa) contained next to the glycoside hydrolases family 18 domain (PF00704) a Chitinase C domain (PF06483) and a bacterial Ig domain (PF17957), and two carbohydrate-binding family 5/12 modules (IPR003610). Nevertheless earlier studies on the different chitinases (ChiA, ChiB, ChiC, ChiD, ChiE, ChiF, ChiG, and ChiH) demonstrated that these show different hydrolytic activities against the different forms of chitin [30, 31]. Intriguingly, in *A. rivuli* the GH18 Chitinase C was higher expressed, while in *A. bestiarum* the GH18 Chitinase A was more abundant (**Figure 3B**, GH18_1 Chitinase A domain; GH18_2 Chitinase C domain). Additionally, two others, but lower-abundance glycoside hydrolases (GH13 and GH84) were observed in either *A. rivuli*, or *A.* bestiarum growing on chitin. While GH13 family enzymes are specialized on alpha-glucans, the detected GH84 (WP_224432541) contains a beta-N-acetylglucosaminidase catalytic domain. Yet, the presence of GH18, 19 and 20 hydrolases seems to be sufficient to explain effective growth of the studied species on insoluble chitin. Furthermore, their abundance profile differences between both strains explains the observed ratio differences between the GlcNAc mono- and dimers in the chitin degradation experiment. Also, one carbohydrate binding protein that was detected in the supernatant of both strains (*A. bestiarum* A0A291TWD8, *A. rivuli* WP_224431348.1) was classified as CAZy family AA10 (formerly CBM33), which is a copper-dependent lytic polysaccharide monooxygenase (LPMO) (**SI EXCEL DOC 2** and **3**). These enzymes catalyze the cleavage of 1,4-glycosidic bonds found in different types of plant cell

wall polysaccharides and chitin. LPMOs function on the crystalline regions of polysaccharides, thereby making crystalline chitin better available to other hydrolytic enzymes. The LMPOs found in *A. rivuli* and *A. bestiarum* showed a sequence identity of >70% (**SI EXCEL DOC 2**). However, no significant changes in the expression levels could be detected, in contrast to several of the chitinases (**SI EXCEL DOC1**). One possible explanation for this is that the growth experiments utilized amorphous chitin, eliminating the need for accelerating degradation of crystalline chitin. In agreement with this, only trace quantities of GlcNAc1A were detected in the chitin hydrolysis assays (**SI Figure 5**). Interestingly, N-acetyl-hexosaminidases (GH20) have been reported to be located mainly in the periplasm. However, in particular in the secretome of *A. bestiarum* we detected larger quantities of a GH20 (A0A291U507, **SI EXCEL DOC 1**), which according to SignalP prediction (SignalP-6.0) however contains a Sec/SPI signal, which could allow to cross the outer membrane in Gram-negative bacteria. Previous studies have shown that chitin monomers and dimers uptake can be enhanced through outer membrane porins (**Figure 4**). Specifically, Kitaoku *et. al*., reported on a chitoporin specific to chitin found in *Vibrio spp* [32]*.* Homologues can also be found in other proteomes, such as *Escherichia. coli* and *A. veronii*. To further investigate this, we performed BLAST search on the genomes of *A. bestiarum* and *A. rivuli* using the reported chitoporins from *A. veronii* and *E. coli* (**SI EXCEL DOC 2**). Thereby, we found a high sequence identity with the OprD family outer membrane porins in both *Aeromonas* strains (*A. rivuli* WP_042041362 and *A. bestiarum* A0A291U719), which were also highly upregulated during growth on chitin. Therefore, it can be hypothesized that the OprDs identified in both strains are chitoporins which further support the uptake of GlcNAc monomers and dimers into the periplasm. The GlcNAc dimers can then be further cleaved by periplasmatic GH20s into monomers. In fact, we found abundant GH20s in the cellular proteomics experiments of both strains (*A. rivuli* WP_042039723 and *A. bestiarum* A0A291U507, **SI EXCEL DOC 1**). Different mechanism have been reported which transport monomers and dimers through the inner cell membrane (**Figure 4**) [13, 18, 32, 33].

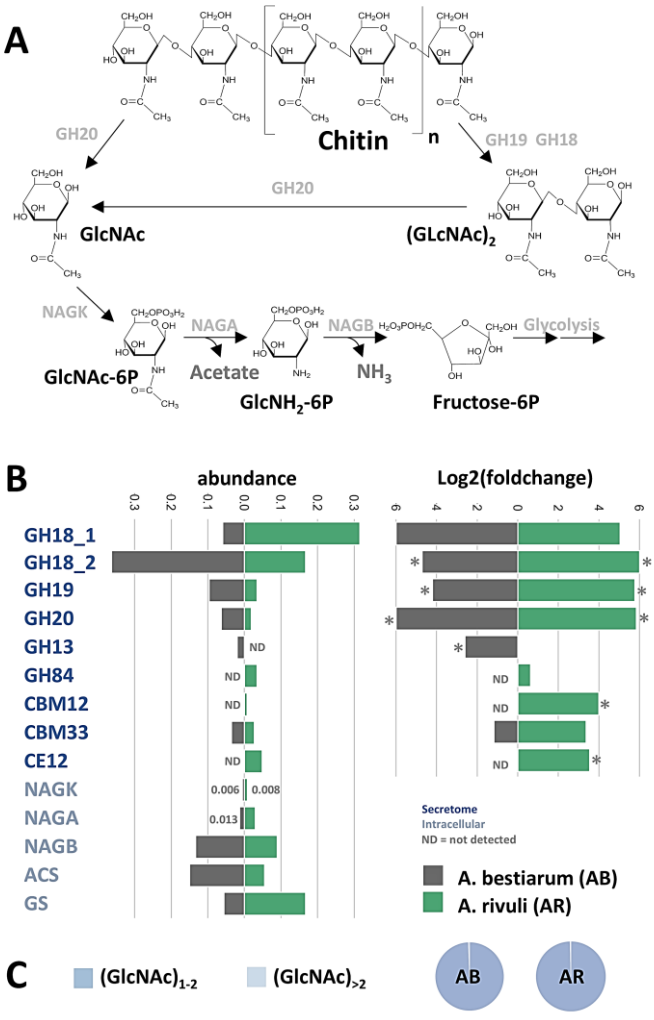


Figure 3: A) The diagram outlines the primary routes of chitin degradation, based on observed enzymes in the secretome and cellular enzymes. First, chitin is cleaved into GlcNAc oligomers by several endochitinases (GH18_1 GH18_2 and GH19), followed by terminal cleavage through exo-chitinase (GH20). After uptake, GlcNAc oligomers are further broken down into monomers and GlcNAc is converted to GlcNAc-6P by NAGK within the cell, and then to GlcNH2-6P by NAGA, before finally transforming into Fructose-6-phosphate (F6P) by NAGB. F6P can then enter glycolysis. The released acetate can also be used to produce acetyl-CoA and ammonia can be incorporated into glutamine. The different potential uptake and transport routes for both *Aeromonas* species are illustrated in **Figure 4**. B) The bar graphs represent the observed abundance (emPAI index, total of all shown scaled to 1, when grown on chitin) and the Log2(foldchange) between growth on chitin and glucose, for the identified glycoside hydrolases and catabolic enzymes in both *Aeromonas* strains. Significant changes are marked with an asterisk. Protein accessions, emPAI indices, FC values and statistical parameters can be found in the **SI EXCEL Table 1** [9]. C) Chitin degradation assays using the secreted enzymes from both cell cultures revealed that primarily GlcNAc monomers and dimers are produced for both strains.

In fact, several substrate-binding domain-containing proteins related to different ABC transporters (which potentially transport the GlcNAc monomers and dimers) were found to be more abundant in the secretome when grown on chitin (**SI EXCEL DOC 1**). GlcNAc

monomers are then directly phosphorylated, while the dimers are cleaved and phosphorylated by a GlcNAc specific kinase (*A. rivuli* WP_042041492, *A. bestiarum* A0A291U705), which then releases GlcNAc and GlcNAcP (**Figure 4**). Moreover, we also identified GlcNAc-specific phosphotransferase system (PTS) which was upregulated during growth on chitin (*A. rivuli* WP_042039724 and *A. bestiarum* A0A291U5N4, which was annotated as 'glucose specific' albeit having 96% sequence identify with WP_042039724), this transporter simultaneously phosphorylates and transports GlcNAc monomers into the cytoplasm. Li *et al.*, (2007) reported on a putative $(GlcNAc)_2$ catabolic operon in *V. cholerae* [15]. This organism is from the same Gammaproteobacteria class as the here studied *Aeromonas* strains. This operon includes next to the chitin sensor *chiS* also the chitin binding protein '$(GlcNAc)_2$ periplasmic binding protein' and a related ABC-type $(GlcNAc)_2$ transporter [34]. The periplasmic binding protein binds chitin oligomers (e.g. GlcNAc dimers) upon which it dissociates from *chiS*. This triggers expression of the chitinolytic genes. BLAST search confirmed that homologues genes (with high sequence similarity) are also present in both investigated *Aeromonas* strains (**SI EXCEL DOC 2**). Furthermore, the putative $(GlcNAc)_2$ periplasmic binding protein (*A. rivuli* WP_042041499.1 and *A. bestiarum* A0A291U0Y6) of the ABC-type $(GlcNAc)_2$ transporter was strongly upregulated and abundant during growth on chitin (**SI EXCEL DOC 1**).

Finally, a putative diacetyl-chitobiose specific PTS was detected in the genome of *A. rivuli*. This system facilitates the phosphorylation of GlcNAc dimers alongside their translocation across the inner membrane. However, a homologous gene could not be identified in the genome of *A. bestiarum*. Such PTS transporters have previously been described for other Gammaproteobacteria such as *Serratia marcescens* or *E. coli* [35, 36]. However, the known subunits (EIIA WP_224432960.1, EIIB WP_224432412.1 and EIIC WP_042041548.1) were barely detected and upregulation in response to chitin could therefore not be confirmed. In the cytoplasm, GlcNAc dimers are then supposedly cleaved by another GH20. Indeed, the cellular proteomics experiments revealed a strongly upregulated GH20 (*A. rivuli* WP_224432421.1, *A. bestiarum* A0A291U0V8) which does not contain a signal peptide (and which therefore likely resides in the cytoplasm, **SI EXCEL DOC 1**). The GlcNAc monomers are then further converted into glucosamine-6-phosphate by GlcNAc kinase (NAGK, for *A. rivuli* either WP_224430983.1 or WP_042040690.1, *A. bestiarum* A0A291TVT4), then into glucosamine-6-phosphate by N-acetylglucosamine-6-phosphate deacetylase (NAGA, *A. rivuli* WP_042039726.1, *A. bestiarum* A0A291U5J9) and finally into fructose-6-phospate (F6P) by Glucosamine-6-phosphate deaminase (NAGB, *A. rivuli* WP_042039725.1, *A. bestiarum* A0A291U4P9) which can enter glycolysis, and the resulting pyruvate (in form of acetyl-CoA) the TCA cycle (**Figure 4**).

Noteworthy, during the intracellular conversation of GlcNAc to glucosamine, acetate is also released which can be further converted to acetyl-CoA by the acetyl-CoA synthetase. In fact, the acetyl-CoA synthetase/acetate-CoA ligase was found being expressed in *A. bestiarum*

and *A. rivuli* when grown on chitin (*A. rivuli* WP_224432291.1, *A. bestiarum* A0A291U0K7). Another pathway that can utilize acetate is the glyoxylate cycle, which was found being expressed in both *Aeromonas* strains. Chitin furthermore can be an important nitrogen source in environments where nitrate and ammonia concentrations are low, such as in drinking water systems.

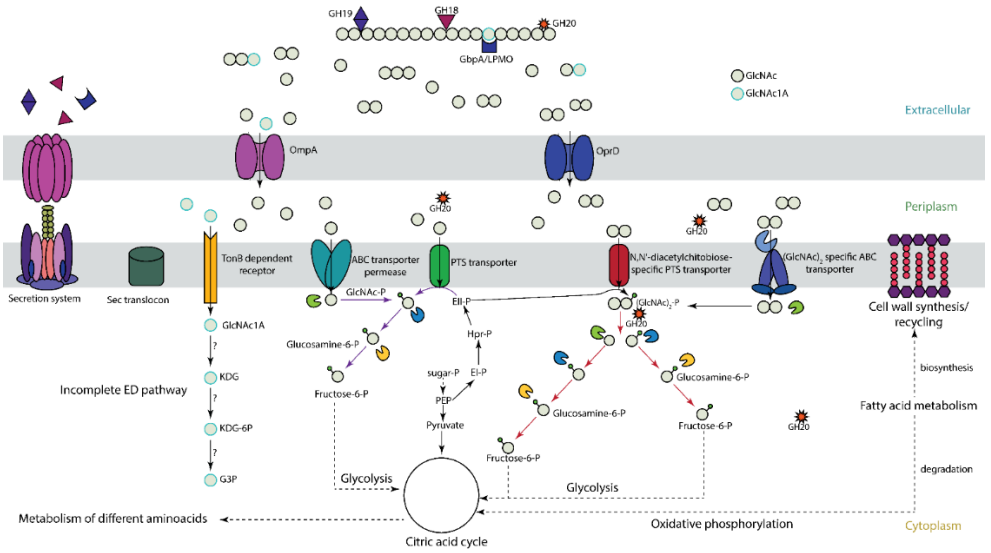


Figure 4: Putative chitin degradation and uptake routes present in *A. bestiarum* and *A. rivuli* based on identified hydrolytic enzymes, transporters, and metabolic enzymes. Chitin is degraded outside the cell by endo- and exo-chitinases (GH18 and GH19) and beta-hexosaminidases (GH20) (see also Figure 3). The hydrolysis products, GlcNAc mono- and oligomers are then transported into the periplasm through dedicated outer membrane porins such OprD, which was highly upregulated when grown on chitin. The chitin degradation assays furthermore demonstrated that mainly GlcNAc mono- and dimers are produced (SI Figure 5). In the periplasm, GlcNAc dimers are further cleaved by N-acetyl hexosaminidases (GH20). Hydrolysis products may then be further transported into the cytoplasm through a GlcNAc-specific phospho-transferase system (PTS), bind to substrate-binding proteins related to different ABC transporters, or are transported through a dedicated $(GlcNAc)_2$ ABC-type transporter. Interestingly, *A. rivuli* possesses an additional $(GlcNAc)_2$ specific PTS transporter. Finally, oxidative degradation products (GlcNAc1A) produced through LPMOs are transported inside the cell where it can be directly deacetylated and deaminated to produce KDG without activation by phosphorylation. However, for *Aeromonas* this route could not be confirmed in the present study. Finally, GlcNAc dimers get cleaved by dedicated GH20s and GlcNAc gets further phosphorylated, where GlcNAc-6P is converted into $GlcNH_2$-6P and finally into fructose-6-phosphate (F6P), by glucosamine-6-phosphate deaminase, allowing it to enter glycolysis.

Glucosamine-6-phosphate deaminase (NAGB) releases ammonia which can be assimilated through incorporation into glutamine, a proteogenic amino acid and an important nitrogen

carrier [37]. This reaction is facilitated by glutamine synthase [3], which was found being expressed in both strains, but significant changes in response to growth on chitin were not detected (*A. rivuli* WP_042042627.1, *A. bestiarum* A0A291TY36). Finally, we wanted to explore whether both *Aeromonas* strains have the potential to utilize oxidative degradation products (GlcNAc1A) as recently proposed by Jiang *et al.* (2022)[13]. The oxidative degradation of chitin through LPMOs accelerates the bioconversion of crystalline chitin. Jiang *et al*., demonstrated that GlcNAc1A can be directly deacetylated and deaminated to produce KDG without activation by phosphorylation. KDG is also the precursor of a key intermediate of the Entner-Doudoroff pathway (KDG-6P), which pathway however was found to be incomplete in the genome of both *Aeromonas* strains (**SI EXCEL DOC 4**). Nevertheless, the utilization of oxidation products through this route is believed to be common in marine Gammaproteobacteria. Albeit *A. bestiarum* and *A. rivuli* have been isolated from fresh waters, other members of the genus *Aeromonas* are widely distributed in estuarine and marine environments. Therefore, we searched for genes involved in degradation and catabolism of GlcNAc1A (LPMOs, OngA/B/C, KdgK and KdgA) as described by Sheng *et al*., [13] in the genomes of *A. bestiarum* and *A. rivuli*. Interestingly, while homologues for LPMOs (*A. rivuli* WP_224431348.1, *A. bestiarum* A0A291TWD8), and KdgK (*A. rivuli* WP_042040847.1, *A. bestiarum* A0A291U805) and KdgA (*A. rivuli* A. WP_224431468.1, *A. bestiarum* A0A291TWZ9) were found, only more distant or no significant alignments were obtained for the key genes OngC or OngB (**SI EXCEL DOC 2**). While both putative LPMOs were found to be expressed when grown on (amorphous) chitin, the putative KdgK and KdgA homologues were not detected in the proteomics experiments. Nevertheless, at present, it cannot be excluded that other deacetylases and deaminases expressed by *Aeromonas* facilitate the conversion of GlcNAc1A.

The potential to degrade a broader spectrum of carbohydrate biopolymers.

The conducted growth and proteomics experiments clearly demonstrate that both *Aeromonas* strains have the ability to efficiently degrade and grow on chitin as the sole carbon and nitrogen source. However, in highly oligotrophic environments such as drinking water distribution systems, the survival of *Aeromonas* may be further enhanced by the ability to degrade, uptake, and utilize of a wider range of different biopolymers.

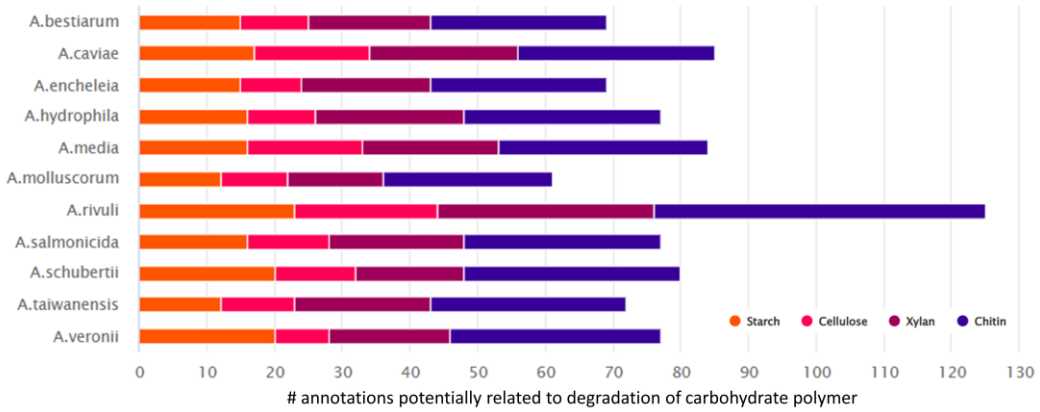


Figure 5: The bar graph displays the frequency of genes associated with CAZy families involved in hydrolyzing starch, cellulose, xylan, and chitin. The analysis concentrated on *Aeromonas* strains, which are commonly found in drinking water distribution systems. A large spectrum of glycoside hydrolases (GHses) related to the degradation of starch, cellulose, xylan and chitin were identified in the genomes of the selected *Aeromnas* strains (**SI DOC TABLE 1**).

In order to investigate this, we analyzed the genomes of several additional *Aeromonas* species commonly found in drinking water distribution systems, for dedicated CAZy genes. This revealed a range of genes which are potentially involved in the degradation of different carbohydrate polymers such as starch, cellulose or xylan (**Figure 5**, **SI Figure 7**). However, it is important to note that the *in-silico* identification of glycoside hydrolases (and related binding proteins) is only a potential, and their actual ability to hydrolyze and utilize specific polysaccharides requires further cultivation experiments.

Discussion

Aeromonads are frequently found in the sediments and loose deposits of drinking water distribution systems [3, 6, 7, 38-41]. These deposits contain a variety of organic and inorganic suspended solids, and harbor microscopic fungi, as well as small and larger invertebrates, such as *A. aquaticus* [42-45]. These organisms are known to produce biopolymers, including chitin, which can serve as a source of carbon and nitrogen for microbes. The ability to use chitin as a nutrient source is commonly found among bacteria [11-13]. However, its specific role within environments such as drinking water distribution systems is not yet fully understood.

The growth experiments conducted here demonstrate that two frequently found *Aeromonas* strains in drinking water distribution systems, *A. bestiarum* and *A. rivuli*,

efficiently degrade and utilize chitin as their sole carbon and nitrogen source. This finding demonstrates that *Aeromonas* is not only chitinolytic, but also chitinotrophic. Previous studies have not shown significant growth for *A. rivuli* (and several other strains) in sterilized drinking water and low concentrations of chitin [3]. In our study, *A. rivuli* also showed a slower growth compared to *A. bestiarum*, despite the use of a more soluble amorphous chitin and potentially higher substrate concentrations. Our quantitative proteomics analysis of the secretome and cell biomass further revealed that chitin oligomers induced the (over)expression of dedicated glycoside hydrolases, sugar binding proteins and transporters (see **Figure 1**). Additional metabolic routes are also required because GlcNAc has to be converted into fructose 1,6-bisphosphate before it can enter glycolysis, and additional acetate and ammonia are released which also provides energy, organic carbon and a nitrogen source [26]. Consequently, we identified a dedicated network of chitinases (2x GH18s and one GH19), β-N-acetyl hexosaminidases (GH20) and binding proteins. One GH18 contains a Chitinase A and the second a Chitinase C domain and a carbohydrate-binding module (PF02839) which can bind to insoluble chitin, which was particularly abundant in *A. bestiarum*. *Aeromonas* species also expressed a GH19, which are commonly only found in plants, some actinobacteria and some chitin degrading bacteria [46, 47]. The few known bacterial GH19s are supposedly acquired from plants through horizontal gene transfer [48]. The GH18s in *A. rivuli* and *A. bestiarum* show a high sequence identify, but the GH19 and GH20s appeared to be only distantly related. Nevertheless, the supernatants from both strains efficiently degraded chitin into monomers and dimers. GlcNAc dimers and oligomers may provide a competitive advantage since the GlcNAc monomers can be easily utilized by other microorganisms present in the community. Nevertheless, there are also some microbes such as *E.coli* which – albeit not being able to hydrolyze chitin – possess transporters and enzymes dedicated to utilization of GlcNAc dimers (chitobiose) [49].

Furthermore, both strains strongly overexpressed a putative chitin specific outer membrane porin (OprD) in response to chitin, which apparently supports the transport of GlcNAc oligomers and monomers through the outer membrane. Membrane transporters are commonly challenging to study because these proteins often show a very hydrophobic nature, are difficult to solubilize, digest and consequently detect. Nevertheless, our study detected several substrate-binding proteins related to different ABC and PTS transporters. Based on these transporters, the uptake mechanisms are diverse and supposedly slightly differ between both strains. For example, *A. rivuli* possesses an additional chitobiose specific PTS sugar transporter, which is not present in the genome of *A. bestiarum*. Interestingly, several transporters and porins have also been identified in the secretome. While we cannot exclude the presence of outer membrane vesicles, a possible reason for the detection of these porins and transporters could be a small number of lysed cells releasing proteins into the supernatant. The recently reported oxidative pathway opens another route to also utilize the oxidation products (GlcNAc1A) generated by LPMOs.

However, albeit homologues to LPMOs and other key metabolic enzymes of this pathway are present in both strains, the amorphous chitin used in this study apparently did not lead to relevant amounts of GlcNAc1A. Nevertheless, initial growth experiments on shrimp chitin flakes, which show a high degree of crystallization, did not result in growth of *A. bestiarum* either according to microscopic investigations. Therefore, the activity of the identified putative LPMOs – which support the degradation of crystalline chitin in other marine microbes [13] – could not be confirmed in this study. Furthermore, both strains secreted also several lipases, peptidases and proteases, which commonly help to make nutrient sources accessible [50, 51]. This is also in agreement with the previous finding that feeding Gram-negative bacteria complex sugars can increases the production of lipases [52], which may aid in biofilm formation and play a role in virulence [53, 54]. A high growth affinity of *Aeromonas* for long-chain fatty acids in drinking water has been also demonstrated earlier [4].

Nevertheless, albeit *Aeromonas* can efficiently degrade and grow on chitin, likely also several other microbes contribute to the degradation of chitin and other biopolymers in such environments [55, 56].

For example, earlier studies also showed that *Aeromonas* is only a minor part of the microbial community present in the drinking water distribution system [6, 57], and other microbes could make crystalline chitin more accessible within the food web. The metabolic end products may furthermore support other heterotrophic bacteria in the drinking water environment. Such interactions have been reported for other niches, such as the soil environments, recently [58-60].

In summary, this study demonstrates how *Aeromonas* can grow on the carbohydrate polymer chitin available in the biomass of invertebrates such as *A. aquaticus,* often found in the loose deposits in drinking water distribution system. Additionally, the quantitative proteomics data reveal a dedicated chitin degradation and uptake network, providing a valuable resource for further investigation of the identified hydrolytic enzymes, transporters, and catabolic enzymes. A deeper understanding of the metabolic routes in these microbes supports the development of better water sanitation strategies.

Conflict of interest

All authors declare that they have no conflicts of interest.

Data availability

Shotgun proteomic data, reference sequence databases and database searching files for this article are available via the ProteomeXchange consortium database with the dataset identifier PXD047459.

Acknowledgements

The authors would like to thank Dita Heikens for her support in the proteomics laboratory, and all colleagues in the Department of Biotechnology, as well as Professor Bert van der Wal, for valuable discussions. They would also like to acknowledge Evides (The water company N.V.) and the NWO Spinoza prize awarded to Mark van Loosdrecht for funding and support.

References

1. Holt J, Krieg N, Sneath P, Staley J, Williams SJBsModB. Genus Aeromonas. 1994:190-1.
2. Janda JM, Abbott SL. Evolving concepts regarding the genus Aeromonas: an expanding Panorama of species, disease presentations, and unanswered questions. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 1998;27(2):332-44.
3. van Bel N, van der Wielen P, Wullings B, van Rijn J, van der Mark E, Ketelaars H, et al. Aeromonas species from non-chlorinated distribution systems and their competitive planktonic growth in drinking water. Appl Environ Microbiol. 2020;87(5).
4. Van der Kooij D, Hijnen W. Nutritional versatility and growth kinetics of an Aeromonas hydrophila strain isolated from drinking water. Applied and Environmental Microbiology. 1988;54(11):2842-51.
5. Van der Kooij D, Visser A, Hijnen W. Growth of Aeromonas hydrophila at low concentrations of substrates added to tap water. Applied and Environmental Microbiology. 1980;39(6):1198-204.
6. van der Wielen PW, Lut MC. Distribution of microbial activity and specific microorganisms across sediment size fractions and pipe wall biofilm in a drinking water distribution system. Water Science and Technology: Water Supply. 2016;16(4):896-904.
7. Liu G, Tao Y, Zhang Y, Lut M, Knibbe W-J, van der Wielen P, et al. Hotspots for selected metal elements and microbes accumulation and the corresponding water quality deterioration potential in an unchlorinated drinking water distribution system. Water research. 2017;124:435-45.
8. Christensen SC, Nissen E, Arvin E, Albrechtsen HJ. Distribution of Asellus aquaticus and microinvertebrates in a non-chlorinated drinking water supply system--effects of pipe material and sedimentation. Water Res. 2011;45(10):3215-24.

9. van Bel N, van Lieverloo JHM, Verschoor AM, Pap-Veldhuizen L, Hijnen WAM, Peeters ETHM, et al. Survival and Growth of Asellus aquaticus on Different Food Sources from Drinking Water Distribution Systems. 2024;2(3):192-211.
10. Beier S, Bertilsson S. Bacterial chitin degradation-mechanisms and ecophysiological strategies. Front Microbiol. 2013;4:149.
11. Reguera G, Leschine SB. Chitin degradation by cellulolytic anaerobes and facultative aerobes from soils and sediments. FEMS Microbiology Letters. 2001;204(2):367-74.
12. Yang C, Rodionov DA, Li X, Laikova ON, Gelfand MS, Zagnitko OP, et al. Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of Shewanella oneidensis. J Biol Chem. 2006;281(40):29872-85.
13. Jiang W-X, Li P-Y, Chen X-L, Zhang Y-S, Wang J-P, Wang Y-J, et al. A pathway for chitin oxidation in marine bacteria. Nature Communications. 2022;13(1):5899.
14. Courtade G, Aachmann FL. Chitin-active lytic polysaccharide monooxygenases. Targeting Chitin-containing Organisms. 2019:115-29.
15. Li X, Wang L-X, Wang X, Roseman S. The chitin catabolic cascade in the marine bacterium Vibrio cholerae: Characterization of a unique chitin oligosaccharide deacetylase. Glycobiology. 2007;17(12):1377-87.
16. Xia W, Liu P, Liu J. Advance in chitosan hydrolysis by non-specific cellulases. Bioresource technology. 2008;99(15):6751-62.
17. Bassler B, Yu C, Lee Y, Roseman S. Chitin utilization by marine bacteria. Degradation and catabolism of chitin oligosaccharides by Vibrio furnissii. Journal of Biological Chemistry. 1991;266(36):24276-86.
18. Lan X, Zhang X, Hu J, Shimosaka M. Cloning, expression, and characterization of a chitinase from the chitinolytic bacterium Aeromonas hydrophila strain SUWA-9. Biosci Biotechnol Biochem. 2006;70(10):2437-42.
19. Cruzat V, Macedo Rogero M, Noel Keane K, Curi R, Newsholme P. Glutamine: metabolism and immune function, supplementation and clinical translation. Nutrients. 2018;10(11):1564.
20. Ueda M, Fujiwara A, Kawaguchi T, Arai M. Purification and some properties of six chitinases from Aeromonas sp. no. 10S-24. Biosci Biotechnol Biochem. 1995;59(11):2162-4.
21. Sitrit Y, Vorgias CE, Chet I, Oppenheim AB. Cloning and primary structure of the chiA gene from Aeromonas caviae. J Bacteriol. 1995;177(14):4187-9.
22. Pfennig N, Lippert KDJAfM. Über das vitamin B 12-bedürfnis phototropher Schwefelbakterien. 1966;55:245-56.
23. Sorokin DY, Toshchakov SV, Kolganova TV, Kublanov IV. Halo (natrono) archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates. Frontiers in microbiology. 2015;6:942.

24. Nielsen H, Tsirigos KD, Brunak S, von Heijne G. A Brief History of Protein Sorting Prediction. Protein J. 2019;38(3):200-16.
25. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein*S. Molecular & Cellular Proteomics. 2005;4(9):1265-72.
26. Adrangi S, Faramarzi MA. From bacteria to human: a journey into the world of chitinases. Biotechnology advances. 2013;31(8):1786-95.
27. Brameld KA, Goddard III WA. The role of enzyme distortion in the single displacement mechanism of family 19 chitinases. Proceedings of the National Academy of Sciences. 1998;95(8):4276-81.
28. Lv C, Gu T, Ma R, Yao W, Huang Y, Gu J, et al. Biochemical characterization of a GH19 chitinase from Streptomyces alfalfae and its applications in crystalline chitin conversion and biocontrol. International Journal of Biological Macromolecules. 2021;167:193-201.
29. Ikegami T, Okada T, Hashimoto M, Seino S, Watanabe T, Shirakawa M. Solution structure of the chitin-binding domain of Bacillus circulans WL-12 chitinase A1. Journal of Biological Chemistry. 2000;275(18):13654-61.
30. Kawase T, Yokokawa S, Saito A, Fujii T, Nikaidou N, Miyashita K, et al. Comparison of enzymatic and antifungal properties between family 18 and 19 chitinases from S. coelicolor A3 (2). Bioscience, biotechnology, and biochemistry. 2006;70(4):988-98.
31. Nguyen-Thi N, Doucet N. Combining chitinase C and N-acetylhexosaminidase from Streptomyces coelicolor A3 (2) provides an efficient way to synthesize N-acetylglucosamine from crystalline chitin. Journal of Biotechnology. 2016;220:25-32.
32. Kitaoku Y, Fukamizo T, Kumsaoad S, Ubonbal P, Robinson RC, Suginta W. A structural model for (GlcNAc) 2 translocation via a periplasmic chitooligosaccharide-binding protein from marine Vibrio bacteria. Journal of Biological Chemistry. 2021;297(3).
33. Suginta W, Chumjan W, Mahendran KR, Janning P, Schulte A, Winterhalter M. Molecular uptake of chitooligosaccharides through chitoporin from the marine bacterium Vibrio harveyi. PLoS One. 2013;8(1):e55126.
34. Li X, Roseman S. The chitinolytic cascade in Vibrios is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. Proceedings of the National Academy of Sciences. 2004;101(2):627-31.
35. Uchiyama T, Kaneko R, Yamaguchi J, Inoue A, Yanagida T, Nikaidou N, et al. Uptake of N, N'-diacetylchitobiose [(GlcNAc) 2] via the phosphotransferase system is

essential for chitinase production by Serratia marcescens 2170. Journal of bacteriology. 2003;185(6):1776-82.

36. Keyhani NO, Wang L-X, Lee Y, Roseman S. The Chitin Disaccharide, N, N'-Diacetylchitobiose, Is Catabolized byEscherichia coli and Is Transported/Phosphorylated by the Phosphoenolpyruvate: Glycose Phosphotransferase System. Journal of Biological Chemistry. 2000;275(42):33084-90.
37. Kleiner D. Bacterial ammonium transport. FEMS Microbiology Letters. 1985;32(2):87-100.
38. Krovacek K, Faris A, Baloda SB, Lindberg T, Peterz M, Mnsson I. Isolation and virulence profiles of Aeromonas spp. from different municipal drinking water supplies in Sweden. Food Microbiology. 1992;9(3):215-22.
39. Lechevallier MW, Evans TM, Seidler RJ, Daily OP, Merrell BR, Rollins DM, et al. Aeromonas sobria in chlorinated drinking water supplies. Microb Ecol. 1982;8(4):325-33.
40. Pablos M, Rodríguez-Calleja JM, Santos JA, Otero A, García-López M-L. Occurrence of motile Aeromonas in municipal drinking water and distribution of genes encoding virulence factors. International Journal of Food Microbiology. 2009;135(2):158-64.
41. Vavourakis CD, Heijnen L, Peters MCFM, Marang L, Ketelaars HAM, Hijnen WAM. Spatial and Temporal Dynamics in Attached and Suspended Bacterial Communities in Three Drinking Water Distribution Systems with Variable Biological Stability. Environmental Science & Technology. 2020;54(22):14535-46.
42. Novak Babič M, Gunde-Cimerman N, Vargha M, Tischner Z, Magyar D, Veríssimo C, et al. Fungal Contaminants in Drinking Water Regulation? A Tale of Ecology, Exposure, Purification and Clinical Relevance. Int J Environ Res Public Health. 2017;14(6).
43. Ketelaars HAM, Wagenvoort AJ, Peters M, Wunderer J, Hijnen WAM. Taxonomic diversity and biomass of the invertebrate fauna of nine drinking water treatment plants and their non-chlorinated distribution systems. Water Res. 2023;242:120269.
44. Ren X, Li J, Zhou Z, Zhang Y, Wang Z, Zhang D, et al. Impact of invertebrates on water quality safety and their sheltering effect on bacteria in water supply systems. Environmental Pollution. 2023;330:121750.
45. Prest EI, Martijn BJ, Rietveld M, Lin Y, Schaap PG. (Micro) Biological Sediment Formation in a Non-Chlorinated Drinking Water Distribution System. Water. 2023;15(2):214.

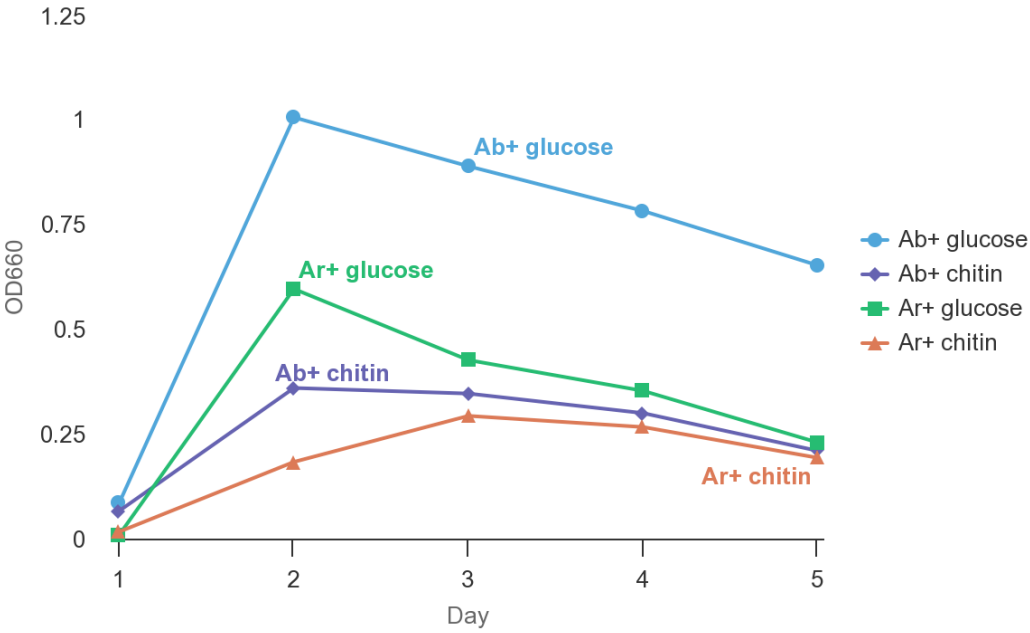
46. Zhang A, Mo X, Zhou N, Wang Y, Wei G, Hao Z, et al. Identification of Chitinolytic Enzymes in Chitinolyticbacter meiyuanensis and Mechanism of Efficiently Hydrolyzing Chitin to N-Acetyl Glucosamine. 2020;11.
47. Lau N-S, Furusawa G. Polysaccharide degradation in Cellvibrionaceae: Genomic insights of the novel chitin-degrading marine bacterium, strain KSP-S5-2, and its chitinolytic activity. Science of The Total Environment. 2024;912:169134.
48. Udaya Prakash N, Jayanthi M, Sabarinathan R, Kangueane P, Mathew L, Sekar K. Evolution, homology conservation, and identification of unique sequence signatures in GH19 family chitinases. Journal of molecular evolution. 2010;70:466-78.
49. Walter A, Friz S, Mayer C. Chitin, Chitin Oligosaccharide, and Chitin Disaccharide Metabolism of Escherichia coli Revisited: Reassignment of the Roles of ChiA, ChbR, ChbF, and ChbG. Microbial Physiology. 2021;31(2):178-94.
50. Sorokin DY, Rakitin AL, Gumerov VM, Beletsky AV, Sinninghe Damsté JS, Mardanov AV, et al. Phenotypic and genomic properties of Chitinispirillum alkaliphilum gen. nov., sp. nov., a haloalkaliphilic anaerobic chitinolytic bacterium representing a novel class in the phylum Fibrobacteres. Frontiers in Microbiology. 2016;7:407.
51. Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Damsté JS, Muyzer G, et al. Genome analysis of C hitinivibrio alkaliphilus gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum T ermite G roup 3. Environmental microbiology. 2014;16(6):1549-65.
52. Jaeger KE, Ransac S, Dijkstra BW, Colson C, van Heuvel M, Misset O. Bacterial lipases. FEMS Microbiol Rev. 1994;15(1):29-63.
53. Qin S, Xiao W, Zhou C, Pu Q, Deng X, Lan L, et al. Pseudomonas aeruginosa: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics. Signal Transduction and Targeted Therapy. 2022;7(1):199.
54. Nguyen M-T, Luqman A, Bitschar K, Hertlein T, Dick J, Ohlsen K, et al. Staphylococcal (phospho)lipases promote biofilm formation and host cell invasion. International Journal of Medical Microbiology. 2018;308(6):653-63.
55. Siracusa V. Microbial degradation of synthetic biopolymers waste. Polymers. 2019;11(6):1066.
56. Hoppe H-G, Kim S-J, Gocke K. Microbial decomposition in aquatic environments: combined process of extracellular enzyme activity and substrate uptake. Applied and Environmental microbiology. 1988;54(3):784-90.
57. van der Wielen PW, Bakker G, Atsma A, Lut M, Roeselers G, de Graaf B. A survey of indicator parameters to monitor regrowth in unchlorinated drinking water. Environmental Science: Water Research & Technology. 2016;2(4):683-92.

58. Wieczorek AS, Schmidt O, Chatzinotas A, Von Bergen M, Gorissen A, Kolb S. Ecological functions of agricultural soil bacteria and microeukaryotes in chitin degradation: a case study. Frontiers in microbiology. 2019;10:1293.
59. Karwautz C, Zhou Y, Kerros M-E, Weinbauer MG, Griebler C. Bottom-up control of the groundwater microbial food-web in an alpine aquifer. Frontiers in Ecology and Evolution. 2022;10:854228.
60. Taubert M, Stähly J, Kolb S, Küsel K. Divergent microbial communities in groundwater and overlying soils exhibit functional redundancy for plant-polysaccharide degradation. PLoS One. 2019;14(3):e0212937.

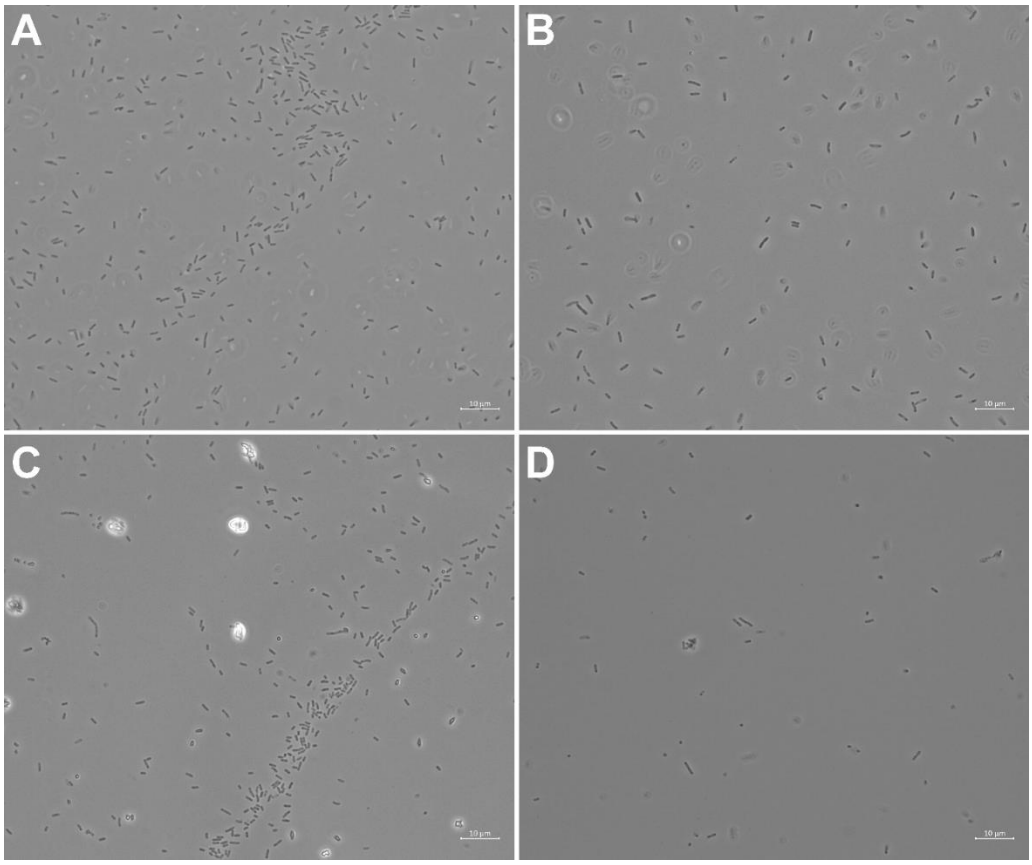
Supplementary information material to:

Exploring the metabolic potential of *Aeromonas* to utilise the carbohydrate polymer chitin

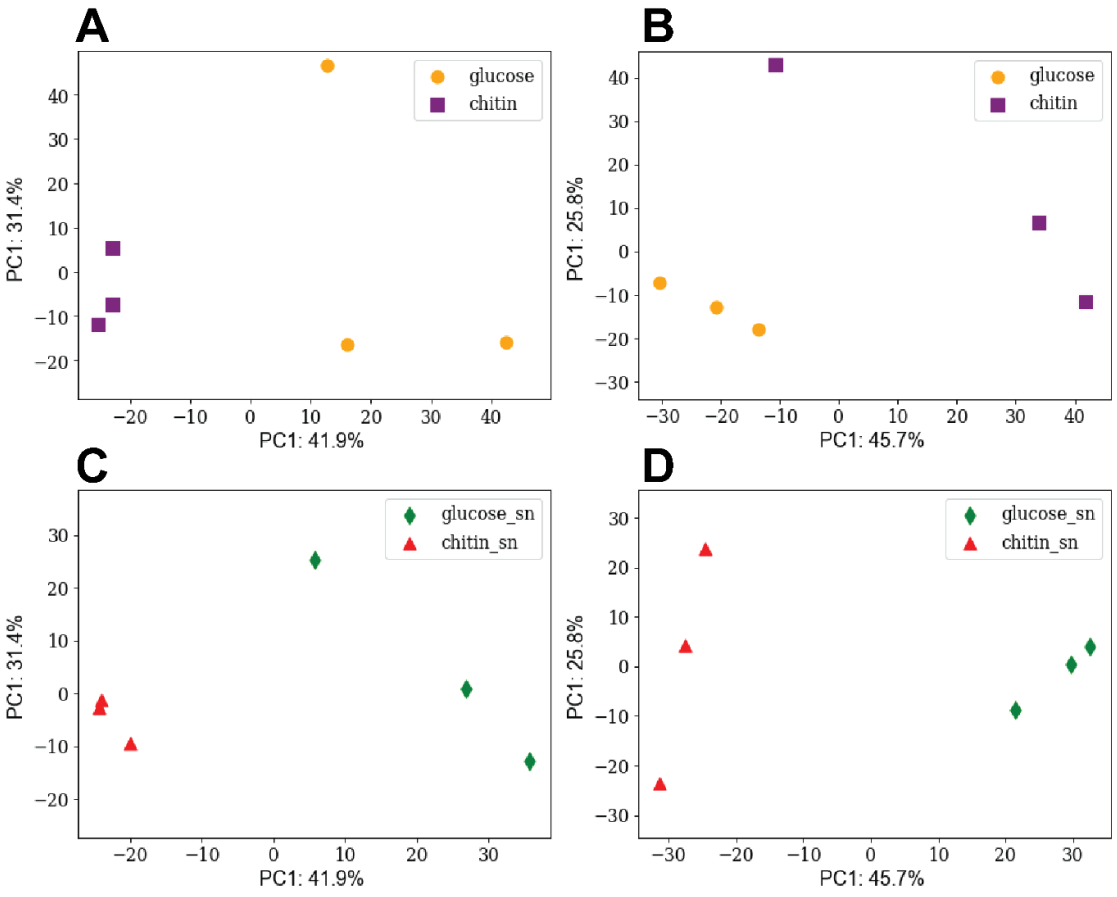
TABLE OF CONTENTS



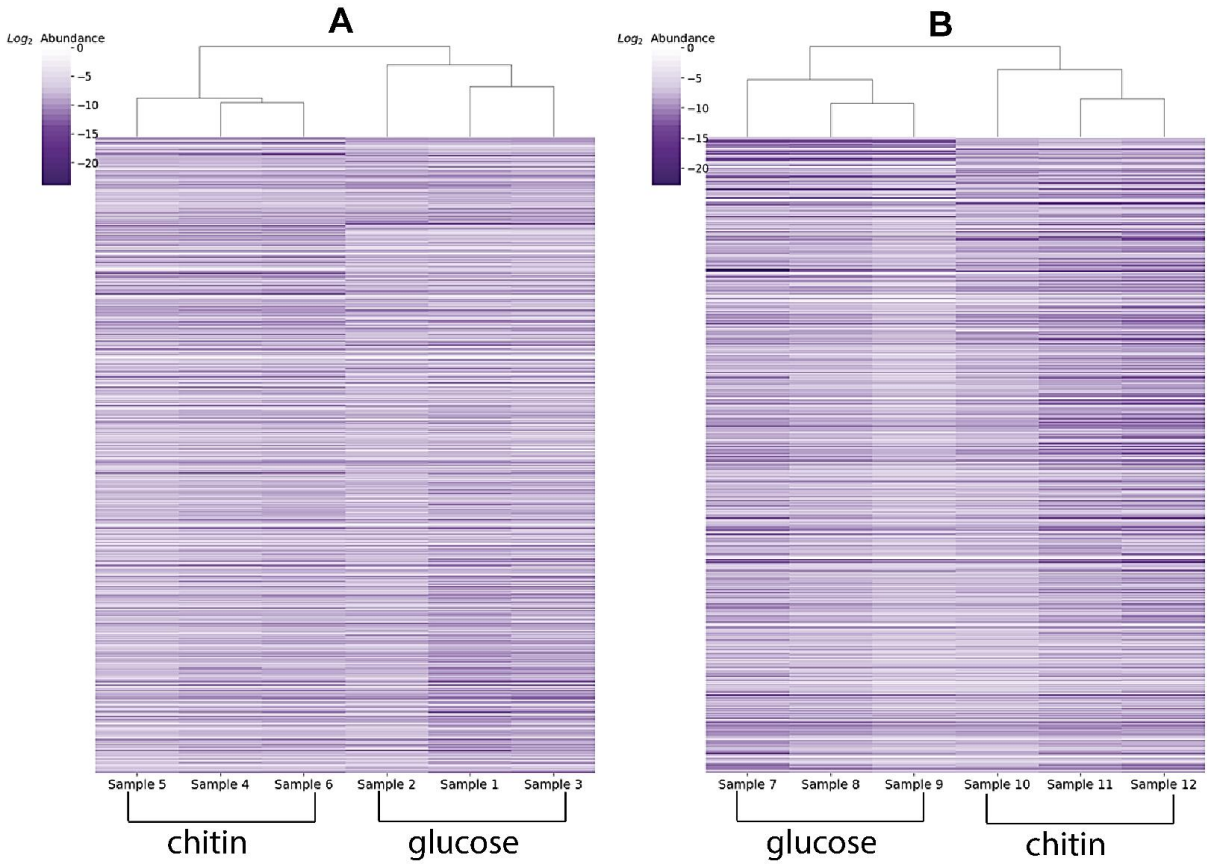
SI Figure 1. OD660 measurements for *A. bestiarum* (Ab) and *A. rivuli* (Ar) cultures grown on glucose or chitin over 5 days.



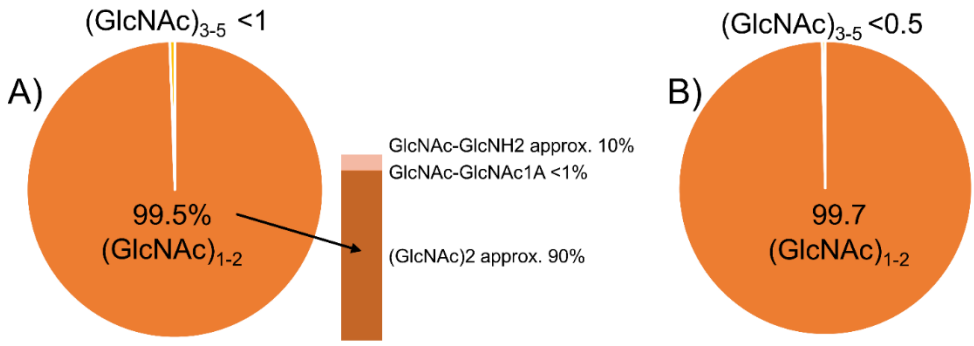
SI Figure 2. Light microscopy images (at 100x magnification) of cultures from (A) *A. bestiarum* grown on chitin, (B) *A. bestiarum* grown on glucose, (C) *A. rivuli* grown on glucose, and (D) *A. rivuli* grown on chitin.



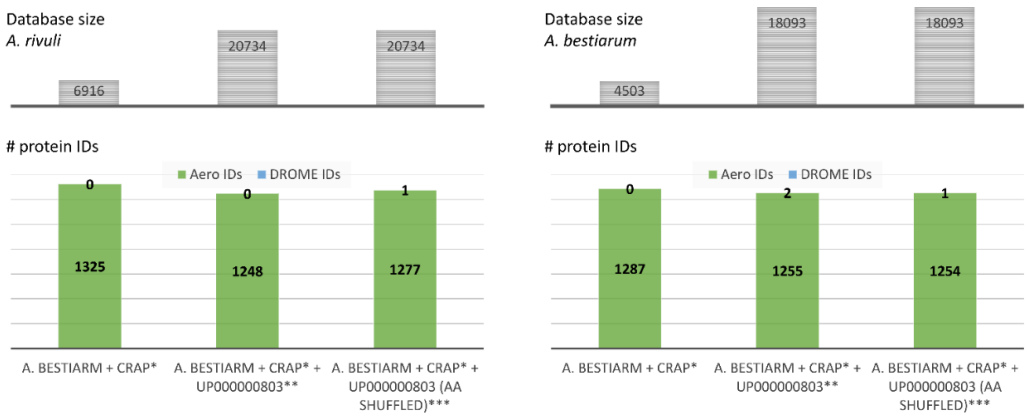
SI Figure 3: The graphs display the principal component analysis (PCA) of the profiles acquired from the triplicate growth experiments of (A) *A. bestiarum* biomass, (B) *A. rivuli* biomass, (C) *A. bestiarum* secretome, and (D) *A. rivuli* secretome. Supernatant is abbreviated as "sn".



SI Figure 4: The heatmaps display the hierarchical clustering of the cellular proteome profiles acquired from the triplicate growth experiments of (A) *A. bestiarum*, and (B) *A. rivuli*.



SI Figure 5: The pie charts display the distribution between different GlcNAc hydrolysis products (monomers = $(GlcNAc)_1$, dimers = $(GlcNAc)_2$, trimers = $(GlcNAc)_3$, and tetramers = $(GlcNAc)_4$) obtained by exposing chitin to the cell culture supernatants of A) *A. bestiarum* and B) *A. rivuli*. The abundances are based on the different fragment ion intensities of the respective hydrolysis products. Interestingly, the ratios $(GlcNAc)_2$ to $(GlcNAc)_1$ is inverted between both *Aeromonas* species. For both strains small quantities of oxidized forms could be detected. However, all were <1% in abundance compared to the main hydrolysis product. For *A. bestiarum* partially deacetylated forms could also be detected (approx. 10%). The m/z values for the native hydrolysis products are: GlcNAc = 222.09721, C8H16NO6+; GlcNAc-GlcNAc = C16H29N2O11+, 425.17659; and for the oxidized forms are: GlcNAc1A = 238.09213, C8H16NO7+; GlcNAc-GlcNAc1A = C16H29N2O12+, 441.1715, and for the native deacetylated forms are: GlcNAc-GlcNH2 = 383.16602, C14H27N2O10+.



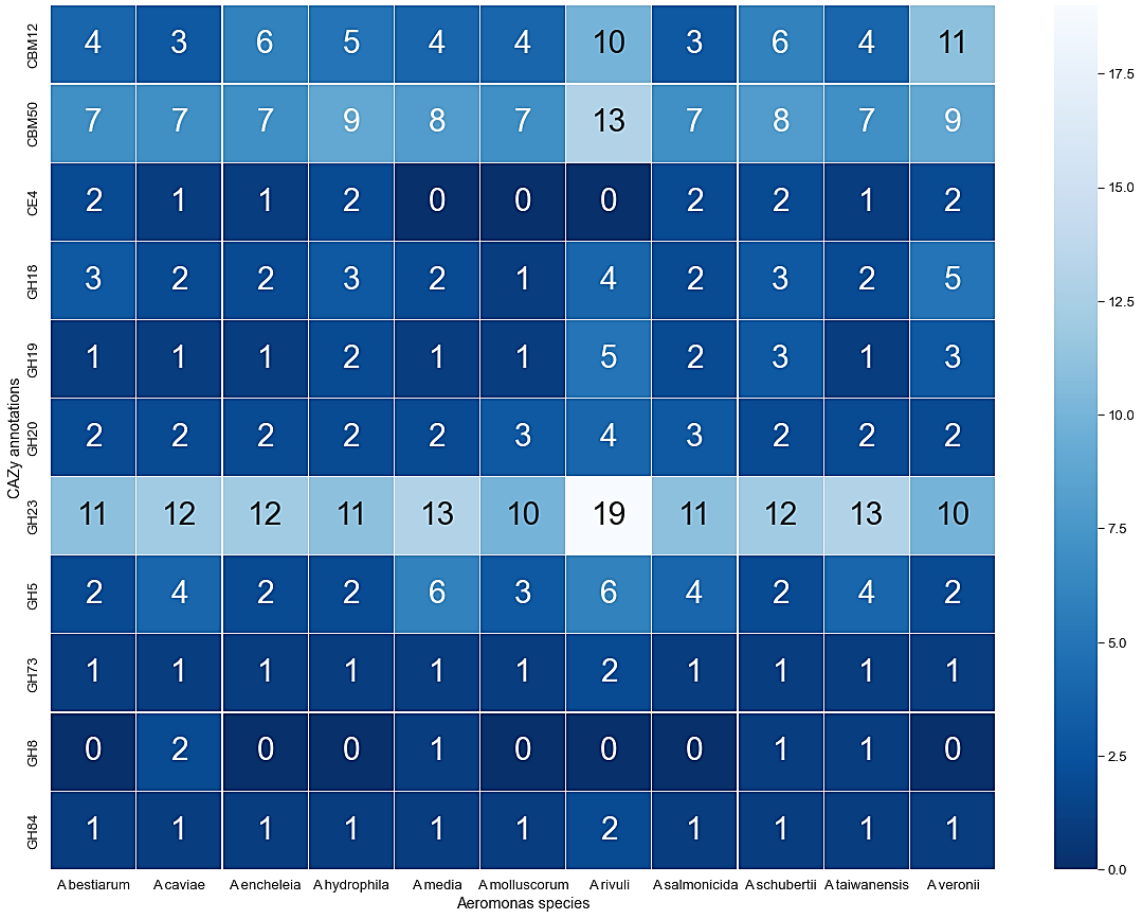
SI Figure 6: Evaluation of database searching accuracy. After increasing the database size from 4,503 to 18,093 entries for *A. bestiarum* and from 6,916 to 20,734 entries for *A. rivuli* (following the incorporation of the reference proteome UP000000803 from Drosophila melanogaster (DROME), either directly or after randomizing the amino acid sequence), the number of identified proteins decreased only slightly, by an average of 2.52% for *A. bestiarum* and 4.75% for *A. rivuli*. Additionally, the number of incorrect matches to the decoy DROME proteome remained below 1% for both strains, averaging 0.12% for A. bestiarum and 0.04% for A. rivuli.

SI Table 1: Evaluation of database searching accuracy for *A. bestiarum* (grown on glucose).

A. bestiarum **(pellet, grown on Glc)**	**$DB size**	**PSMs (total)**	**PSMs (DROME#)**	**#Protein IDs (total)**	**#Protein IDs (DROME)**	**% less Protein IDs**	**PSM FDR (PEAKS)**	**Protein FDR (PEAKS)**	**% DROME IDs**
A. bestiarm **+ cRAP***	4503	30139	0	1287	0	0.00%	1.0	0.4	X
A. bestiarm **+ cRAP* + UP000000803****	18093	28117	5	1255	2	2.49%	1.0	0.0	0.16%
A. bestiarm **+ cRAP* + UP000000803 (AA shuffled)*****	18093	27918	2	1254	1	2.56%	1.0	0.0	0.08%

SI Table 2: Evaluation of database searching accuracy for *A. rivuli* (grown on glucose).

A. rivuli (pellet, grown on Glucose)	**$DB size**	**PSMs (total)**	**PSMs (DROME#)**	**#Protein IDs (total)**	**#Protein IDs (DROME)**	**% less Protein IDs**	**PSM FDR (PEAKS)**	**Protein FDR (PEAKS)**	**% DROME IDs**
A. rivuli **+ cRAP***	6916	13638	0	1325	0	0	1	0.1	X
A. rivuli **+ cRAP* + UP000000803****	20734	12888	0	1248	0	5.81%	1.0	0.0	0.00%
A. rivuli **+ cRAP* + UP000000803 (AA shuffled)*****	20734	13327	2	1277	1	3.62%	1.0	0.0	0.08%



SI Figure 7: The heatmap shows the number of identified glycoside hydrolases (GH), related carbohydrate-binding modules (CBM) and chitin esterases (CE) in the genomes of different *Aeromonas* species. These are potentially involved in the breakdown of chitin and chitosan. The analyzed species are frequently found in drinking water distribution systems.

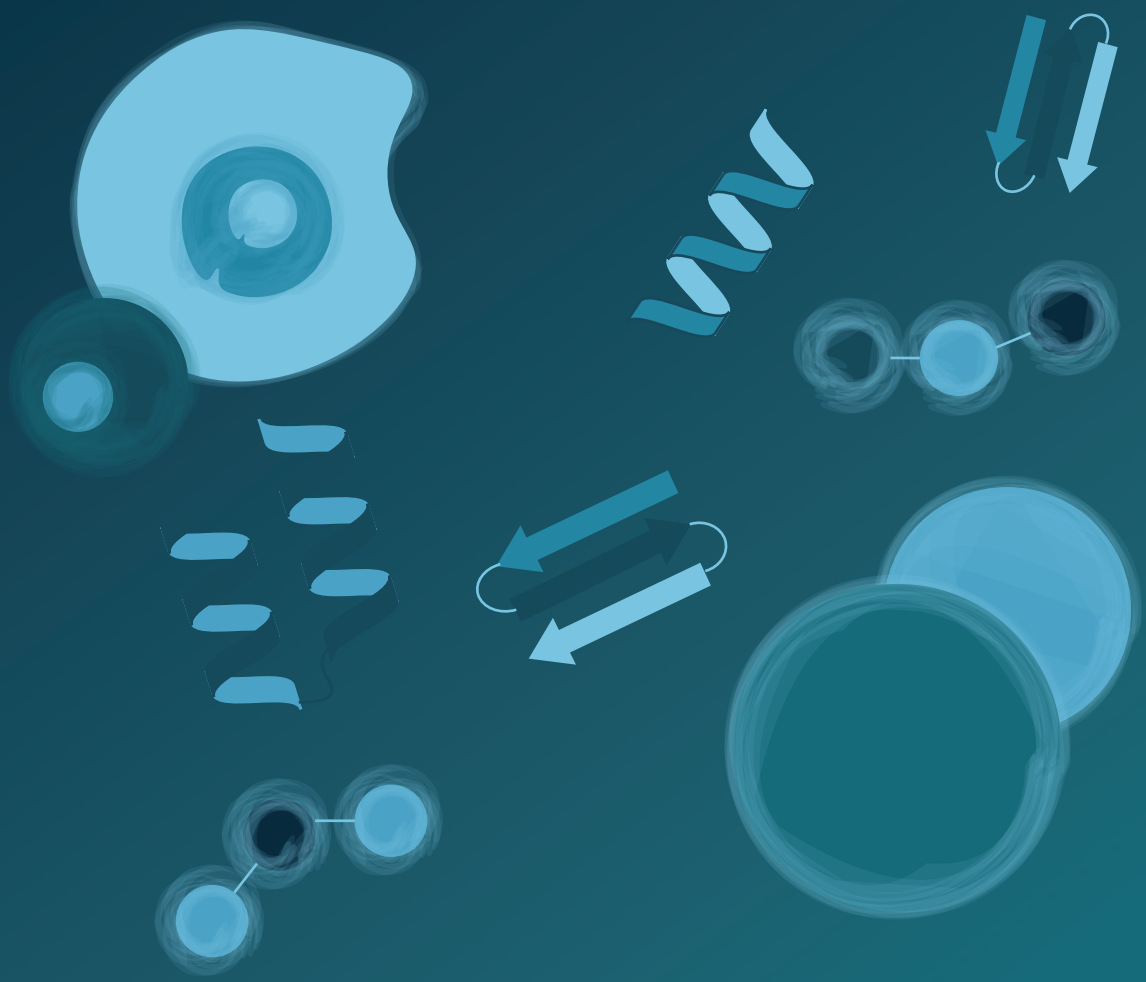
SI Table 3: The table lists CAZy database families used to search for the potential to degrade various carbohydrate biopolymers in *Aeromonas* genomes.

Biopolymer	CAZy targets
Chitin	CBM12, CBM14, CBM18, CBM50, CBM55, GH18, GH19, GH23, GH48, CBM32, GH18, GH20, GH73, GH84, GH85, GH89, GH111, GH116, GH163, GH5, GH7, GH8, GH46, GH75, GH80, CE4
Xylan	AA10, AA14, CBM2, CBM4, CBM6, CBM9, CBM13, CBM15, CBM22, CBM31, CBM35, CBM36, CBM42, CBM54, CBM59, CBM60, CBM72, CBM91, CE1, CE2, CE3, CE4, CE5, CE6, CE7, CE12, GH3, GH5, GH8, GH10, GH11, GH18, GH26, GH30, GH43, GH51, GH67, GH98, GH115, GH141, GT8, GT43, GT47, GT61
Cellulose	AA9, AA10, AA15, AA16, CBM1, CBM2, CBM3, CBM4, CBM6, CBM8, CBM9, CBM10, CBM16, CBM17, CBM28, CBM30, CBM37, CBM44, CBM46, CBM49, CBM59, CBM63, CBM64, CBM72, GH5, GH8, GT2
Starch	AA13, CBM20, CBM21, CBM25, CBM26, CBM34, CBM45, CBM53, CBM69, CBM74, CBM82, CBM83, GT5, GT35, GH13, GH14, GH57, GH126, GH15, GH57, GH97, GH119
Chitosan	GH3, GH5, GH7, GH8, GH18, GH46, GH75, GH80

Chapter 3

Versatile hydrolytic potential in Aeromonas bestiarum

Claudia G. Tugui, Dimitry Y. Sorokin, Mark C.M. van Loosdrecht and Martin Pabst



Data availability: SI Figures and Tables can be found at the end of the chapter. Additional SI data including shotgun proteomic data, database search files and tables for this article are available via the DataverseNL project: https://-doi.org/10.34894/ZSLMYQ

Abstract

Native polysaccharides are one of the most prevalent primary carbon sources for microbes on Earth. There is no accumulation of these polymers in soils or marine ecosystems because they are completely degraded and recycled. The polysaccharide-degrading microorganisms produce a wide spectrum of specialized hydrolytic enzymes allowing them to degrade and utilize these polymers as a carbon and energy source. *Aeromonas* species are widespread in aquatic environments, including highly oligotrophic niches such as drinking water distribution systems. Several *Aeromonas* strains have been shown to efficiently degrade and utilize the carbohydrate polymer chitin recently. *Aeromonas* has been linked to loose deposits in drinking water distribution systems, where small invertebrates with chitin-containing shells are found. However, a recent genomic study suggests that *Aeromonas* species may be capable of degrading a wider range of biopolymers, which could expand their nutrient sources and help them thrive in nutrient-poor environments. Here we investigate the ability of *Aeromonas bestiarum* to grow on various biopolymers, including fucoidan, starch, xylan, cellulose, dextrin, pullulan, pectin, collagen, and a complex biopolymer extract. We perform a proteomics study on the spectrum of secreted proteins, that potentially facilitates the degradation and enable the subsequent uptake of these biopolymers. Our study revealed characteristic enzyme profiles for the individual biopolymers, along with a spectrum of proteins that are consistently expressed, regardless of the substrate. For approximately 20% of all secreted proteins we could not obtain a functional classification, emphasizing the gap in our understanding of protein functions in microbes.

Key words: *Aeromonas bestiarum*, polymers, secretome, CAZy enzymes

Introduction

Polysaccharides are widespread in nature, accounting for a significant portion of biological carbon [1]. Their roles vary from serving as storage polymers to acting as structural components in biofilms, plant cell walls and the shells of invertebrates dwelling both in terrestrial and aquatic habitats. The transformation of different carbohydrate polymers is crucial for the global carbon cycle. For example, in aquatic ecosystems, macroalgae sequester approximately 173 Tg of carbon dioxide per year [2]. This carbon is successfully recycled by the marine microbiome and fauna in such a way that it remains within the aquatic environment. Despite their resistance to degradation, these sugar polymers also serve as excellent substrates for various microorganisms and animals across different ecosystems.

In marine environments, various polymers such as chitin, laminarin, alginate, fucoidan, and xylan are available to microbes. These polysaccharides are structural components of plants, algae, including diatoms, and invertebrates like crustaceans and insects. With the decay of animals and plants in aquatic ecosystems, these polymers become accessible through a phenomenon known as marine snow. Marine snow particles vary in size and composition, including detritus, dead or decaying organisms, faecal matter, microorganisms, and other debris. Marine snow plays a pivotal role in ocean ecosystems by distributing organic matter and nutrients from the surface waters to the deeper ocean layers. This process, known as the biological pump, transfers carbon and energy from the surface to the deep ocean, influencing global carbon cycling. It also serves as a crucial food source for animals, bacteria, and archaea dwelling in the deep ocean and on the ocean floor [3].

The presence of specific biopolymers in ecological niches leads to the specialization of microorganisms, enabling them to thrive in these environments. Conversely, some microbes demonstrate high adaptability to various carbon sources, making them versatile across a broader spectrum of biopolymers. However, their polymer metabolism often requires multiple hydrolytic enzymes and binding modules. After hydrolysis, the products must be taken up through generalized or specialized transporters and subsequently channelled into the central carbon metabolism to generate energy and building blocks. The secretion of proteins involved in these processes can be achieved through the conserved general secretion (Sec) system or the twin-arginine transport system [4-6]. The Sec system translocates proteins in an unfolded state, while the Tat system exports proteins that are already folded. Both systems recognize a specific amino acid pattern known as a signal peptide present in the proteins destined for secretion [7]. Bacterial proteins can also be secreted through a non-classical pathway, typically utilized by proteins lacking a signal peptide. It is not fully understood how these proteins are recognized by the non-classical secretion pathway to facilitate their secretion [8]. It has been shown in some cases that such

proteins may possess sequence patterns or specific peptide subunits that assist in their recognition [9-11]. The entire collection of these secreted proteins is referred to as the secretome. Understanding the secretome and the available transport mechanisms in bacteria under certain conditions reveals how microbes respond to nutrient availability in their environment.

The genus *Aeromonas* comprises Gram-negative Gammaproteobacteria widely distributed in aquatic ecosystems, including both natural environments such as rivers and lakes, and engineered, oligotrophic environments such as drinking water distribution systems [12]. Some representatives of the genus, such as *A. hydrophila*, *A. sobria*, and *A. veronii*, are potential pathogens that may cause diseases in humans and fish (e.g., *A. salmonicida*) [13]. However, *Aeromonas* species found in the drinking water distribution networks in the Netherlands were found to be non-pathogenic. Until now, the ability of *Aeromonas* to survive in nutrient-poor environments like drinking water distribution systems has not been fully explored. Only recently it was demonstrated that *Aeromonas* can efficiently grow on the biopolymer chitin, a key component of the exoskeletons of certain invertebrates (in preparation). *Aeromonas* is commonly found in loose deposits within drinking water distribution systems, a niche also occupied by invertebrates, including *Asellus aquaticus*. This underscores the possibility that *Aeromonas* may exploit chitin or other polymeric components of these invertebrates [14-17]. It has been indicated that *Aeromonas* might be capable of degrading a broader spectrum of biopolymers beyond just chitin [18,19]. This could potentially expand the range of available nutrient sources, enhancing the ability to thrive in nutrient-poor environments. Elevated levels of *Aeromonas* in drinking water systems are considered an indicator of favourable microbial growth conditions and therefore signify compromised water characteristics, including changes in taste, odor, color, the presence of large invertebrates, and the regrowth of opportunistic pathogens [20,21, 16]. Understanding the metabolic strategies of these microbes to survive in oligotrophic environments is critically important. This is especially relevant in the context of water resource management and global warming, where studying the regrowth of microorganisms such as *Aeromonas* becomes a crucial research objective. In the current study, we demonstrate that *A. bestiarum* is capable of degrading and utilizing many different carbohydrate polymers present in both aquatic and soil environments, focusing on the enzymes produced to allow this growth. The ability of utilizing a wide variety of substrates may indicate high versatility of the microorganisms in fluctuating environments.

Materials and Methods

***A. bestiarum* growth experiments.** *A. bestiarum* DSM 13956 was purchased from DSMZ (Leibniz, Germany). The cells were obtained in lyophilized form in ampoules. The strain was furthermore reactivated, and grown glucose and peptone rich TSB medium as recommended by DSMZ. From these cultures, glycerol (30%) stocks were obtained and stored at -80C. *Aeromonas bestiarum* was cultured in the presence of different polymers. All the polymers used in this experiment were in powder form and they were added to each culture bottle prior to autoclavation. The tested polymers were pullulan (500 mg to 58 mL medium), collagen (250 mg to 58 mL medium), cellulose (500 mg to 58 mL medium), xylan (500 mg to 58 mL medium), starch (500 mg to 58 mL medium), pectin (500 mg to 58 mL medium), dextrin (500 mg to 58 mL medium), fucoidan (250 mg to 58 mL medium) and aerobic granular sludge exopolymeric substances extract (120 mg to 58 mL medium, which is a complex mixture of proteins, carbohydrates and lipids, and other metabolites). Pullulan (P4516), starch (S9765), collagen (C9879), pectin (93854) and fucoidan (F8315) were purchased from Merck, xylan from corncob was purchased from Roth Chemikalen (8659.1), and the complex biopolymer extract ("EPS) from aerobic granular sludge was obtained as described by Felz et al., 2016 [22]. *A. bestiarum* was cultured on M9 minimum salt medium (Na_2HPO4 6.78g/L, KH_2PO_4 3g/L, NaCl 0.5g /L, NH_4Cl 1g/L) containing Mg-trace element and $CaCl_2$ solution [23] at a final pH = 6.69. For the inoculation of the cultures containing polymers, a starting glucose culture (molar concentration of 0.03M) was made and once this one reached OD600 = 1, 2 mL were inoculated in the other cultures and grown on 33˚ C with moderate shaking. Also, another glucose culture was made from the initial culture in order to supervise the culture purity.

Cell harvesting and supernatant concentration. This was performed according to Tugui *et al*., 2024 [18]: From every culture, every day 1 mL of cell broth was harvested and centrifuged (14K rpm, 10 minutes) to separate the cell pellet from the supernatant. The resulting supernatant (approx. 1 mL) and cell pellet was stored separately at -20°C, until further processed.

Cell lysis, protein extraction, and proteolytic digestion. This was performed according to Tugui *et al*., 2024 [18]: 600 µL of the collected supernatants from the culturing experiment of *A. bestiarum* were taken and centrifuged. TCA was added to the supernatant (1 volume TCA to 4 volumes supernatant), and the mixture was vortexed and incubated at 4°C for 20 minutes, then spun down at 14000 rpm for 15 minutes at 4°C. The obtained protein pellets were once washed with ice cold acetone and then dissolved in 6 M urea (in 100 mM ammonium bicarbonate, ABC). Further, the disulfide bridges were reduced by the addition of 10 mM DTT and incubation for one hour at 37°C under shaking at 300 rpm. Thereafter, 20 mM IAA was added. The mixture was kept in the dark for 30 minutes. 200 mM ABC buffer

was then added to the samples to obtain a solution with <1 M Urea. Finally, proteolytic digestion was performed by adding Trypsin (0.1 µg/µL in 1 mM HCl, Sequencing Grade Modified Trypsin, Promega) at a ratio of 50:1 (w:w, Protein:Trypsin) to the sample. The proteolytic digestion was performed overnight at 37°C, under gentle shaking at 300 rpm. Peptides were desalted using an OASIS HLB solid phase extraction well plate (Waters, UK) according to the instructions of the manufacturer, speed vac dried and stored at -20°C until further processed**.**

Shotgun proteomics of secreted proteome. This was performed according to Tugui *et al*., 2024 [18]: Approx. 500 ng of proteolytic digest were analyzed in duplicate injections using an EASY nano-LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 25% solvent B over 180 min, then from 25% to 55% B over 60 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 240 minutes. Solvent A was H_2O containing 0.1% formic acid, and solvent B consisted of 80% ACN in H_2O and 0.1% formic acid. The mass spectrometer was operated in data-dependent acquisition mode. Full MS scans were acquired from m/z 380–1250 at a mass resolution of 70 K with a maximum injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 most intense precursor ions were selected for fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at a resolution of 17.5 K with an AGC target of 2E5 and IT of 75 ms, 2.0 m/z isolation width and normalized collision energy [24] of 28.

Processing of mass spectrometric raw data. This was performed according to Tugui *et al*., 2024 [18]: Database searching of the shotgun proteomics raw data was performed using proteome reference databases from *A. bestiarum*, obtained from UniprotKB (UP000224937) using PEAKS Studio X (Bioinformatics Solutions Inc., Canada). The database searching allowed 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, carbamidomethylation as fixed and methionine oxidation and N/Q deamidation as variable modifications. Peptide spectrum matches were filtered for 1% false discovery rates (FDR) and identifications with ≥2 unique peptides were considered as significant. Quantitative analysis of the changes between chitin and glucose-grown conditions, and the cell pellet and secreted proteome abundances was performed using the PEAKSQ module (Bioinformatics Solutions Inc., Canada). Normalization was based on the total ion current (TIC), and only proteins with at least 2 unique peptides were considered. Peptide spectrum matches were filtered with a 1% false discovery rate (FDR). ANOVA was used to determine the statistical significance of the changes between the conditions, expressed as -10xlog10(p), where p corresponds to the significance testing p-value.

Proteomics data analysis and visualization. This was performed according to Tugui *et al*., 2024 [18]: Briefly, results from PEAKSQ were further used for the analysis of enriched functions and metabolic pathways. Data processing and visualization was performed using

Python 3.11.3 and the charting library Highcharts (https://www.highcharts.com/demo). Furthermore, SignalP 6.0 (https://services.healthtech.dtu.dk/services/ SignalP-6.0/) [25] was used for the prediction of signal peptides in order to confirm secreted proteins. Only proteins that showed a significance value > 0.9 at any of the Sec or TAT categories were kept for further analysis. Further the non-classical secreted protein were identified with the help of the SecretomeP 2.0 software (https://services.healthtech.dtu.dk/services/SecretomeP-2.0/)(8) with the selection for Gram-negative bacteria. All the proteins with a SecP score higher than 0.5 were chosen as part of further analysis. The HMM 3.3.2 tool (http://hmmer.org/) was used with the Pfam (http://pfam-legacy.xfam.org)[26] and the Carbohydrate Active Enzymes (CAZy) (http://www.cazy.org/) [24] databases to identify the glycolytic enzymes and their associated modules. For this, default parameters were used to perform an HMM scan. The results were filtered for e-values <10E-5 and for independent e-values of <10E-5. The "uncharacterized proteins" were annotated with the help of SWISS-MODEL (https://swissmodel.expasy.org/)[27] and InterPro (https://www.ebi.ac.uk/interpro/)[28]. For the SWISS-MODEL the results with the best identity and similarity scores were chosen.

Microscopy. Light microscopy was performed according to Tugui *et al*., 2024 [18], using a Zeiss Axio Imager M2 microscope equipped with an Axiocam 305 color camera (Carl Zeiss, Germany). The microscope setting possesses a 63x and 100x oil immersion objective lens and phase contrast capabilities. The proprietary Zeiss software for image capture and analysis was Zen 3.3. Microscopy images were taken after 1, 2 and 4 (for complex biopolymer extract ("EPS"), starch, dextrin, pullulan) and 5 (for pectin, xylan, cellulose, collagen) days post incubation.

Results

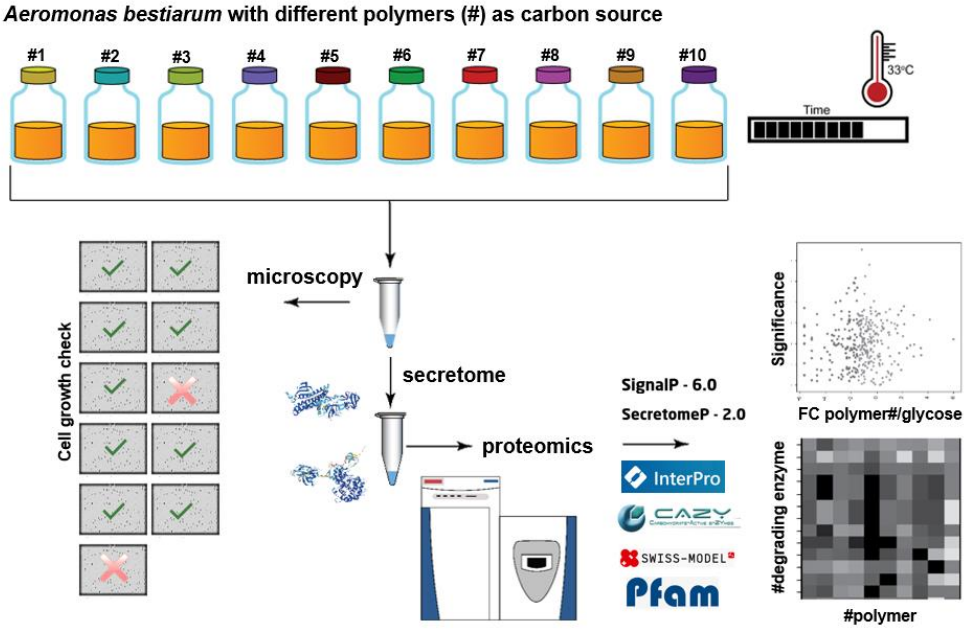


Figure 1. Workflow for screening the capacity of *Aeromonas bestiarum* for the degradation and utilization of various biopolymers (which are listed in Table 1). The growth of the bacteria was monitored using OD600 and light microscopy over time. Subsequently, quantitative proteomics was employed to identify the enzymes secreted. The functional classification of the identified proteins was conducted using InterPro, CAZy, and Swiss-Model. Confirmation of secretion signals in the proteins was performed using SignalP and SecretomeP, available at https://services.healthtech.dtu.dk/services/.

Aeromonas bestiarum was cultured in the presence of 10 different biopolymers, utilized as the sole carbon source (**Figure 1**). These included the carbohydrate polymers: i) cellulose, ii) xylan, iii) starch, iv) pectin, v) dextrin, vi) pullulan, vii) fucoidan, along with three control nutrient sources: i) glucose, ii) collagen, and iii) an extract of biopolymers from aerobic granular sludge. The glucose condition, which required no additional hydrolytic enzymes for growth, served as the baseline. In contrast, collagen, a protein and not a carbohydrate polymer, was included as another contrasting condition. Collagen, the main structural protein in the extracellular matrix of various connective tissues, is the most abundant protein in mammals. It forms a triple helix of elongated fibrils and is glycosylated at various positions along the protein backbone, making it a challenging biopolymer to degrade and utilize. Additionally, we included an extract of complex biopolymers from aerobic granular sludge, which is a consortium of different bacteria forming solid granules during wastewater treatment and contains a diverse mixture of carbohydrate polymers, proteins, lipids, and

other small molecules [29, 30]. For *Aeromonas bestiarum*, it has been previously confirmed that it can degrade and efficiently utilize the biopolymer chitin for growth [18].

The growth of *Aeromonas* on the different biopolymers was initially assessed using light microscopy and by measuring the OD600. Clear growth was observed for starch, dextrin, alginate, pullulan, collagen, and the complex biopolymer extract ("EPS") from aerobic granular sludge. However, no growth was observed for cellulose, xylan, pectin, and fucoidan after a certain period. It is noteworthy that in the case of the complex biopolymer extract ("EPS"), growth stagnated around an OD600 of 0.45 for several days (**SI Figure 1**). In the case of the cellulose and xylan samples, the OD measurement proved to be inaccurate due to the powdered structure of the polymer that can give false positive results. No growth was, however, observed under the microscope. The cell broth was centrifuged, and the supernatant was analyzed using shotgun proteomics to identify the secreted enzymes. The functional classification of the identified proteins was conducted using InterPro with the Pfam and CAZy databases, and Swiss-Model. The presence of secretion signals in the identified proteins was confirmed using SignalP and SecretomeP. Across all conditions, the identified extracellular proteins accounted for 5.79% of the complete *A. bestiarum* genome (22 proteins **Figure 2**). Approximately 21% of the extracellular proteins could not be functionally classified.

Polymer	Chemical structure	Common ecological niches	Growth observed	#ID proteins in secretome (>=2 unique peptides; FDR=1%)	#CAZy	#Peptidases	#Lipases	#Proteins unknown function
Cellulose	$(C_6H_{10}O_5)_n$	terrestrial	No	67	4	3	1	7
Pectin	$(C_6H_8O_6)n+CH_3$	terrestrial	No	36	2	4	1	4
Fucoidan	$(C_6H_{12}O_5)n$ +S	aquatic	No	56	4	5	3	5
Pullulan	$(C_6H_{12}O_5)n$	terrestrial	Yes	148	14	7	2	23
Starch	$(C_6H_{10}O_5)n$	terrestrial	Yes	119	13	5	2	18
Dextrin	$(C_6H_{10}O_5)n$	terrestrial	Yes	173	14	9	2	28
Xylan	$C_5H_{10}O_5$	aquatic and terrestrial	No	14	2	1	1	1

Table 1. Overview of *A. bestiarum* growth experiments on different sugar polymers, their chemical structure, ecological prevalence and the number of different enzymes that were secreted in each condition. For an enzyme to be considered as secreted it must have an enzyme concentration ratio polymer/glucose over 1.

Secretion of glycoside hydrolases and polymer-specific carbohydrate binding proteins.

In the current study, *A. bestiarum* secreted less than half of the proteins that are identified as exoproteins on the genome. In case of the glycoside hydrolases, only a quarter was discovered from the genomic potential, half of the carbohydrate esterases and half of polysaccharide lyases (**Figure 2**). The overall secreted CAZy enzymes, peptidases and lipases across all the growth conditions are presented in Table 1. For all polymers where growth was observed, *A. bestiarum* secreted nine different glycoside hydrolases (A0A291TZV4, A0A291U830, A0A291U6W6, A0A291U507, A0A291U6T4, A0A291U6Z3, A0A291TWR0, A0A291TVA9, A0A291U5G7), five different carbohydrate esterases (CE1, CE3, CE9, CE12, CE16; A0A291U367, A0A291U063, A0A291TWQ3, A0A291U6X0, one protein A0A291U063 has double function of CE3 and CE12), two different glycosyl transferases (two GT41, and one GT51; A0A291U4X9, A0A291U4U2, A0A291TZL0), a polysaccharide lyase (PL22; A0A291U056), and one auxiliary activity identified as a lytic polysaccharide monooxygenase (LPMO, AA10; provide accession number) (**SI Figure 2**). Among the glycoside hydrolases, the most commonly identified function is attributed to GH13, with four different enzymes from *A. bestiarum* linked to this function: A0A291TZV4, A0A291U830, A0A291U6W6, and A0A291TWR0. This is followed by two expressed GH18 enzymes, with the accession numbers A0A291U6Z3 and A0A291TVA9. Recently, Tugui *et al*.[18] demonstrated that *A. bestiarum* can degrade and efficiently utilize chitin as a sole

carbon and nitrogen source. The potential for degrading other biopolymers was indicated as well, due to the presence of dedicated genes in the genome and specific enzymes, such as pullulanases and collagenases, in the secretome when grown under different conditions (17).

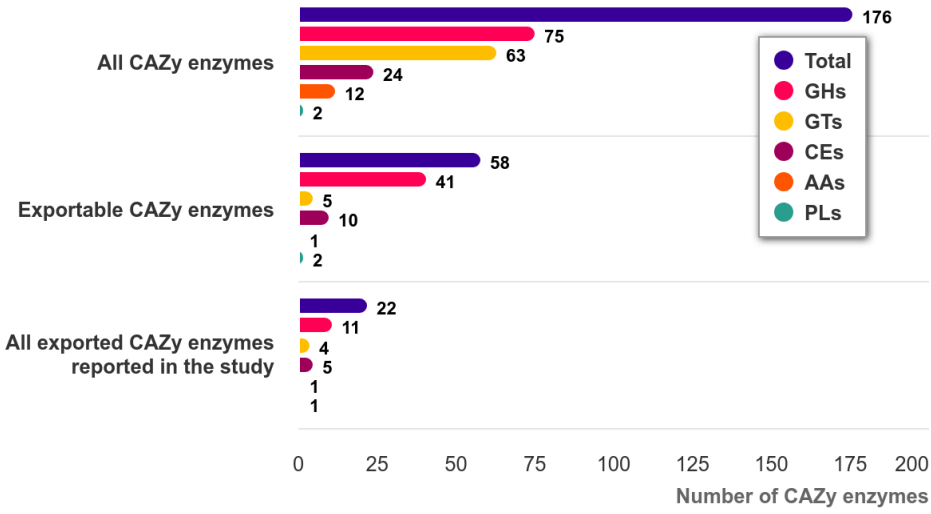


Figure 2. The total number of carbohydrate-active enzymes annotated in *A. bestiarum* genome, the CAZy that can be exported outside the cell and the ones secreted across all 10 growth conditions: fucoidan, complex biopolymer extract, pullulan, starch, xylan, dextrin, cellulose, pectin, collagen, and chitin. The annotation was performed using the HMM V3.4 scan method [31] and the CAZy database (Carbohydrate Active Enzymes database, http://www.cazy.org/) [24] using the default parameters. The enzymes are grouped into different classes: glycoside hydrolases (GH), carbohydrate esterases (CEs), glycosyltransferases (GTs), polysaccharide lyases (PLs), and auxiliary activities (AAs). The auxiliary CBMs were not taken into consideration

A. bestiarum was able to grow on both pullulan and collagen as also seen in the microscopy images (**SI DOC Figure 3**). When grown on pullulan, dedicated enzymes, including two GH13s — an alpha-amylase and another alpha-glycosidase (A0A291U6W6, A0A291TWR0) — were greatly overexpressed. The specific pullulanase responsible for degrading the pullulan backbone (A0A291TW15) was almost three times overexpressed compared to growth on glucose. This protein also contains the carbon-binding motif CBM69. This enzyme was also significantly expressed during growth on starch and dextrin (**SI Figure 8**). The growth on pullulan is therefore supported by several abundant enzymes related to pullulan degradation Henrissat [32-34].

The secretome profile when growing on starch and dextrin showed a strong similarity regarding the overexpressed carbohydrate active enzymes. In both cases, two alpha-

amylase GH13s (A0A291U6W6 and A0A291TWR0), a GT41 (A0A291TZF6), and a PL22 (A0A291U056) were overexpressed compared to glucose grown cells (**Figure 3**). In the supernatant of cells grown on collagen, several enzymes annotated by CAZy were expressed, like two alpha-amylases GH13 (A0A291U830 and A0A291U6W6), which were around two to three times overexpressed compared to cells grown on glucose (**Figure 3**). Additionally, two GH18s were observed in the secretome, which were also overexpressed in the collagen-grown cells (A0A291U6Z3, A0A291TVA9). Collagen is both N- and O-glycosylated on asparagine, hydroxyproline, and hydroxylysine, respectively [35]. Enzymes for cleaving sugars from glycan residues, such as α-galactosidases (GH36, A0A291TVU8, A0A291U2L4) are present in the genome of *A. bestiarum*. However, according to annotations by SignalP and SecretomeP, none of them have an export signal, indicating they act intracellularly on smaller glycopeptide fragments. Alternatively, enzymes like GH103/GH23 (A0A291U5G7) which were found to be overexpressed during growth on collagen, could be involved in releasing carbohydrate residues from the amino acid backbone extracellularly. According to CAZy, GH103 is a peptidoglycan lyase, and GH23 has double catalytic activity, capable of degrading both chitin and peptidoglycans. Interestingly, no overexpression was observed for the putative collagenase (A0A291TWL8) when grown on collagen, compared to the glucose sample **(Figure 3)**. However, this enzyme was present in all conditions and expression levels likely sufficient to degrade the supplied collagen and support growth. For the complex extracellular polymeric substances extract, *Aeromonas* showed expression of alpha-amylase (GH13, A0A291U6W6), and a β-N-acetylhexosaminidase (GH20 and GH138 annotations, A0A291U507), an enzyme with double glycolytic function **(Figure 3)**.

In the case of fucoidan, pectin, cellulose, and xylan no growth was observed. Growth on fucoidan did not occur, likely due to a bottleneck in the degradation of fucoidan and other sulphated compounds, the need to produce and secrete sulfatases that cleave the sulphated groups from the polymer backbone, allowing the compounds to be hydrolysed and transported inside the cell [36-37]. No sulfatases or related PLs were discovered in *A. bestiarum* genome. Pectin, another complex carbohydrate, was not degraded by *A. bestiarum*. The genome of *A. bestiarum* encodes 2 PL22s known for their involvement in pectin metabolism in phytopathogenic plants and intestinal bacteria [38]. In this study, the secreted PL22 (**Figure 4**) showed no overexpression in the pectin-amended medium. Even with partial degradation of the pectin, the uptake of negatively charged polymers may require specialized anionic porins [4].

Cellulose and xylan were two other polymers that *A. bestiarum* could not degrade. The degradation of cellulose is usually reserved for specialized microbes that encode multiple beta-1,4 glucanases of the GH5 and GH9 families in their genome [39,40]. In this study, *A. bestiarum* was grown on crystalline cellulose. In another experiment, the bacterium was incubated with amorphous cellulose, and growth could be observed under the microscope;

however, no secretion of classical cellulases was detected (data not shown). Similarly, powdered xylan did not yield any growth. Annotation using the dbCAN [41] pipeline indicated that their function is more correctly classified as GH13.

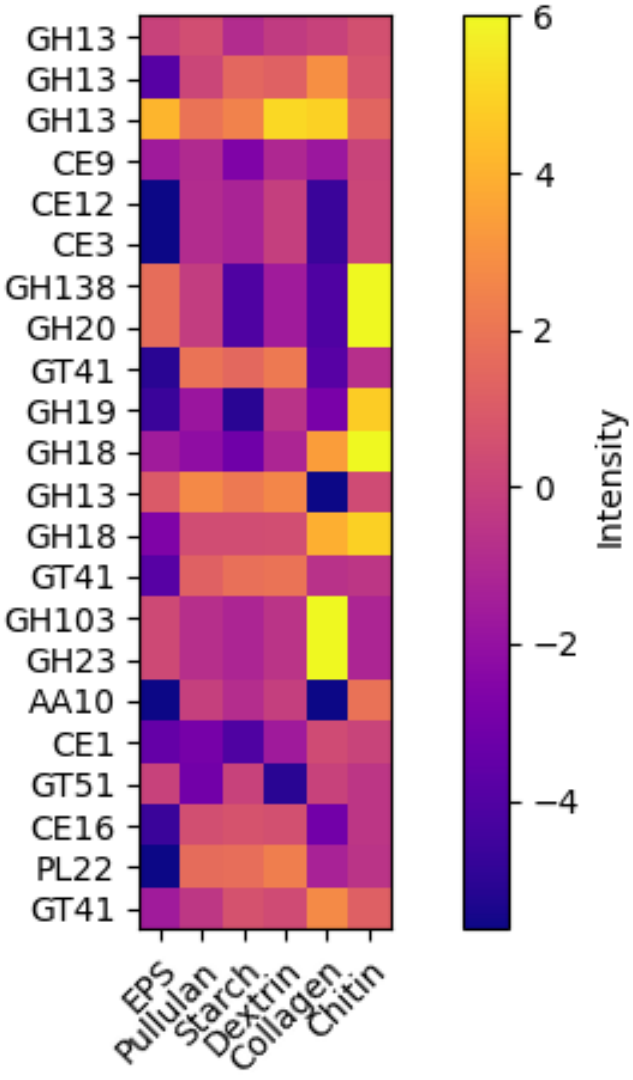


Figure 3. Abundance of different carbohydrate-active enzymes (grouped according to CAZy families) found in the secretome of *A. bestiarum* when grown on various polymers: fucoidan, Complex biopolymer extract, pullulan, starch, xylan, cellulose, dextrin, pectin, collagen, and chitin. The intensity is represented by the log2(fold change) compared to growth on glucose (note: CBMs are not included in this graph).

Secretion of peptidases, proteases, lipases and other proteins.

Across all growth conditions, *A. bestiarum* secreted a total of 21 peptidases, proteases, and lipases with about half being overexpressed in one or more growth conditions. Among the lipases, a lipase (A0A291TY84) was overexpressed in media amended with starch and chitin. A murein L,D-transpeptidase (A0A291U6G2), identified by Pfam as a patatin-like phospholipase, was overexpressed in the complex biopolymer extract, starch, and collagen-grown cells. The native patatin found in potatoes not only exhibits antioxidant activity but also serves as an acyl hydrolase, as well as β-1,3-glucanase activity, which facilitates the degradation of carbohydrate polymers. The other overexpressed proteins were mainly

peptidases. Growth on several sugar polymers, such as pullulan, starch, and dextrin, showed elevated amounts of peptidases, for example, peptidase M35 (A0A291U7T1) was overexpressed almost threefold in media amended with pullulan, starch, and dextrin. Similarly, peptidase S8 (A0A291U1M7) was overexpressed in pullulan and starch growth conditions (**Figure 4**). The collagen and complex biopolymer extract growth conditions showed a slightly different overexpression profile. Growth on collagen exhibited high levels of peptidase S8 (A0A291U1M7), a Zn-dependent exopeptidase M28 (A0A291TZZ0), a neutral metalloproteinase, another peptidase M8 (A0A291TVA6), a serine protease, and a patatin-like phospholipase (A0A291TZ48) (**Figure 4**). Other proteins were also overexpressed, but their functional annotation was too generic to clearly relate them to the degradation of the biopolymers. For example, A0A291U0D8, identified as a lipoprotein, was overexpressed in pullulan, starch, dextrin, and to a lesser degree in chitin. Another putative lipoprotein, A0A291TY64, was up to four times overexpressed under the same growth conditions (**SI EXCEL Tables 6, 7**).

Interestingly, the presence of moonlighting proteins was observed in the secretome of *A. bestiarum* under various conditions (**SI Figure 4**). For example, proteins such as glyceraldehyde-3-phosphate dehydrogenase (A0A291TZV8), phosphoenolpyruvate carboxykinase (ATP) (A0A291TW09), the acetyltransferase component of the pyruvate dehydrogenase complex (A0A291TZ51), superoxide dismutase (A0A291U5A9), and putative glucose-6-phosphate 1-epimerase (A0A291TZU4) were observed in the secretome, but none of them was overexpressed in any of the growth conditions (compared to glucose). Interestingly, a cadherin-like protein with an IPT/TIG domain and an outer membrane autotransporter barrel domain (A0A291U7Z2) was highly abundant when grown on complex biopolymer extract and collagen (**SI Figure 4**). This enzyme family is believed to help the microorganisms adhere to insoluble polysaccharides and aid in their degradation [42].

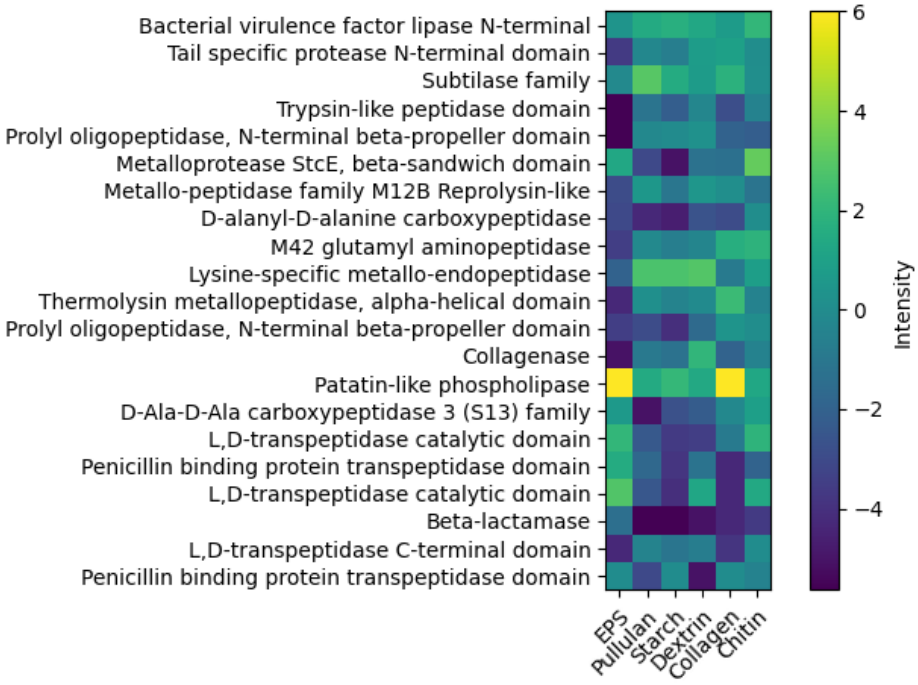


Figure 4. Abundance of secreted lipases, peptidases, and proteases detected in the secretome *of A. bestiarum* when grown on various polymers: fucoidan, complex biopolymer extract ("EPS"), pullulan, starch, xylan, cellulose, dextrin, pectin, collagen, and chitin. The intensity is represented by the log2(fold-change) compared to growth on glucose (note: CBMs are not shown).

Additional functional classification of uncharacterised proteins. Approximately 40 secreted proteins (20% of the entire secretome) could not be functionally classified.

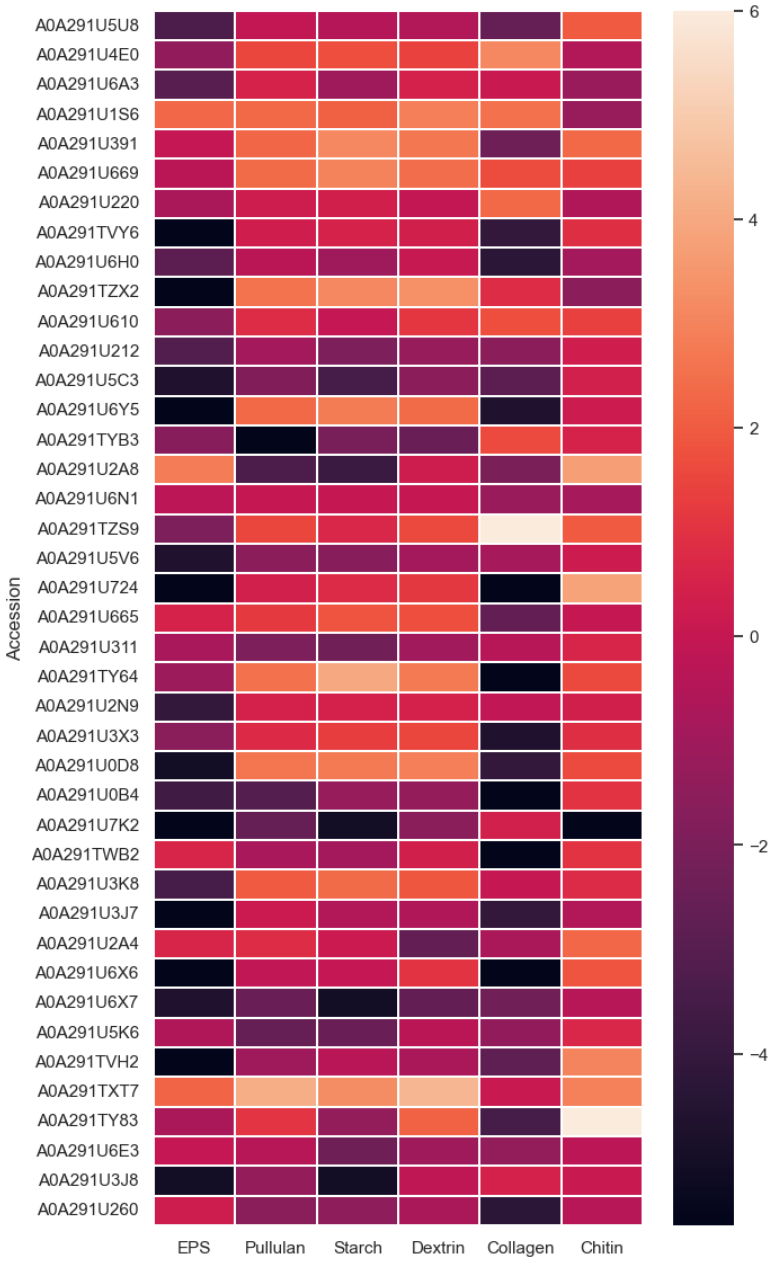


Figure 5. The heatmap displays proteins without functional annotations, identified in the *A. bestiarum* secretome across growth on different polymers: fucoidan, complex biopolymer extract ("EPS"), pullulan, starch, xylan, cellulose, dextrin, pectin, collagen, and chitin. The abundance in each condition is expressed as log2(fold-change) compared to growth on glucose.

These proteins were all labelled as 'uncharacterized protein' and also failed to provide any clues through HMMER searches using the CAZy and Pfam databases. Several of these

uncharacterized proteins were overexpressed in one or more growth conditions **(Figure 5)**. Consequently, additional searches using SWISS-MODEL and InterPro were performed, which provided hints about the possible functions of some proteins **(Table 2 and SI DOC Table 7)**. Although no functional attributes could be derived for most of the proteins, several others offered interesting insights. For instance, a glycosidase (A0A291U4E0) was overexpressed in growth on pullulan, starch, and collagen compared to the glucose-amended medium. A TonB system biopolymer transport component (A0A291TZS9) was overexpressed in cells grown on collagen, dextrin, pullulan, and chitin. A periplasmic heavy metal sensor (A0A291TY83) was overexpressed in xylan, cellulose, dextrin, and chitin conditions. Surprisingly, A0A291U6A3—a putative collagenase—was expressed in all conditions **(Table 2, SI Table 7)**.

Table 2. Selected proteins in the secretome of *A. bestiarum* that received additional functional annotations using SWISS-MODEL and InterPro.

Accession	SWISS-MODEL annotation	InterPro domain
A0A291U4E0	-	six-hairpin glycosidase
A0A291TZS9	TonB system biopolymer transport component	-
A0A291U6A3	GlyGly-CTERM sorting domain-containing protein	metallo peptidase (collagenase)
A0A291U220	MtrB/PioB family decaheme-associated outer membrane protein	porin
A0A291U3J7	-	Lipid A acylation PagP
A0A291U5K6	-	signal peptide
A0A291U6H0	molecular chaperone	-
A0A291TY83	periplasmic heavy metal sensor	-
A0A291U260	autotransporter domain-containing protein	-
A0A291U6X7	molybdopterin-binding oxidoreductase	-

Comparison of secretion profiles across growth conditions.

The secretome profiles obtained from the different growth conditions included carbohydrate-active enzymes, peptidases, proteases, and lipases, which varied between the conditions. Therefore, we performed a principal component analysis (PCA) to compare the similarity of the secreted proteome profiles between the different polymers (**Figure 6**). This analysis showed that most conditions clearly separate from each other in terms of the expressed secretome profile. However, some conditions clustered together, such as growth on chitin, collagen, and starch. Surprisingly, dextrin and starch are clearly separated from each other, even though the CAZy enzymes displayed similar profiles between them.

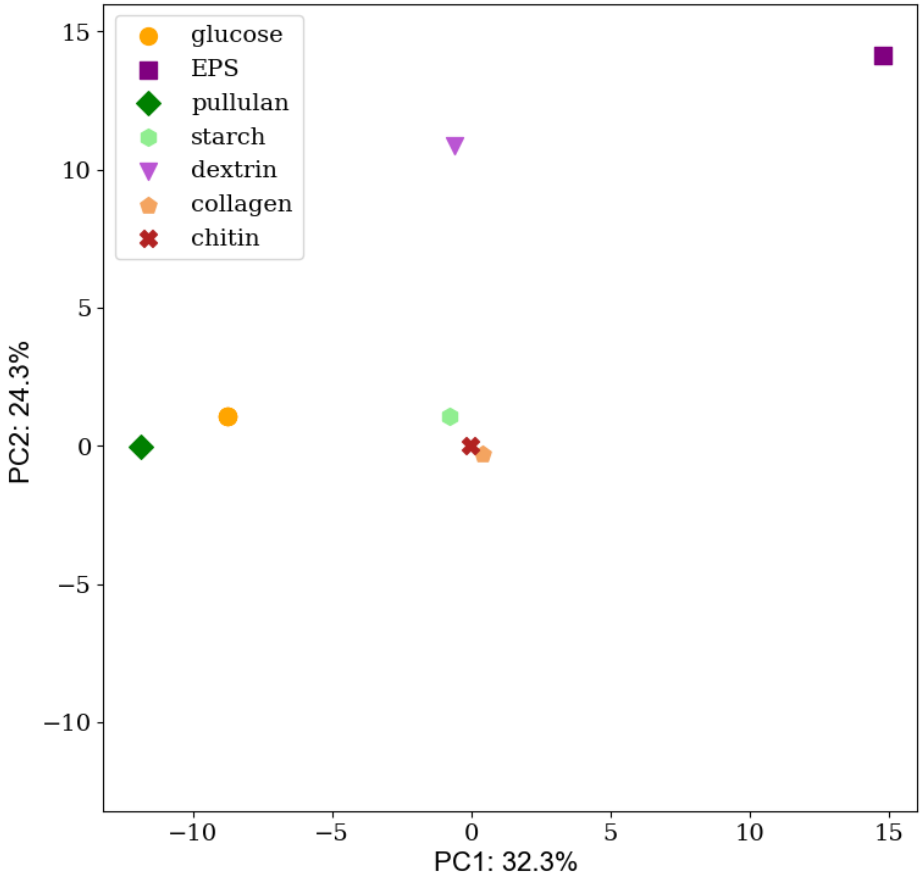


Figure 6. Principal component analysis (PCA) comparing the secreted proteome profiles across different polymer growth conditions (complex biopolymer extract ("EPS"), pullulan, starch, dextrin, collagen and chitin) for *A. bestiarum*.

Discussion and Conclusion

In the current study, we demonstrate that *A. bestiarum* can degrade and grow on a range of different carbohydrate polymers, collagen, and complex biopolymer extracts. This metabolic flexibility supports its survival in highly oligotrophic and dynamic environments, such as the drinking water distribution system. The growth on multiple polymers makes *A. bestiarum* a generalist in the microbial world. These polymers can be found in both aquatic and terrestrial ecosystems as they are also component of fungi (chitin, pullulan), invertebrates both aquatic and terrestrial (chitin, collagen), vertebrates (collagen) and terrestrial plants (starch, dextrin). The source of these polymers overlaps with the classical

niches of *Aeromonas* which are fresh water, sediments and estuaries but also drinking water systems [43-45]. *Aeromonas* has been a well-known resident of the sediments in drinking water pipelines, being able to feed on low molecular compounds like amino acids [14]. However, the presence of other organisms like fungi and even invertebrates may serve as a carbon source to microorganisms like *Aeromonas*.

In addition to GHs necessary for polymer degradation and uptake, *A. bestiarum* secreted multiple proteins and lipoproteins known to play roles in pathogenicity. Hemolysin, a toxin secreted by the *Aeromonas* genus, disrupts the cell walls of red blood cells and is a major reason why species such as *A. hydrophila* and *A. veronii* are considered potential pathogens [46,47]. In this study, *A. bestiarum* produced hemolysin (A0A291U597) under all conditions, but it was overexpressed only when growing on chitin and dextrin. Another toxin, aerolysin (A0A291TWK7), was overexpressed in chitin-amended medium. Additionally, the presence of so-called moonlighting proteins, which have dual functions in the cytoplasm and in the secretome, is another indicator of pathogenicity. Such proteins are usually secreted through a non-classical secretion pathway that does not involve the Sec/Tat translocation system. Among the moonlighting proteins secreted, some are known to be involved in pathogenicity. They have been intensively studied in known pathogenic bacteria and have been shown to confer fitness and adaptability of the pathogen in relation to its host [48,49].

Another enzyme that can be considered a pathogenic marker is collagenase. The unchanged expression levels of collagenase across all conditions compared to growth on collagen may be explained by the fact that the levels were already sufficiently high to support growth. Another excreted protein - lipid A acylation PagP (A0A291U3J7) has been shown to help pathogenic bacteria evade host immune systems, thereby increasing pathogenicity [50].

The degradation of polymers like the complex biopolymer extract led to the production of several enzymes involved in lipid and protein degradation. The enzyme that was most overexpressed for the complex biopolymer extract compared to glucose was a serine protease (A0A291TZ48) annotated as a patatin-like phospholipase, according to both Pfam and InterPro. Furthermore, several peptidases and proteinases, such as the L, D-transpeptidase (A0A291U6G2) and a YkuD domain-containing L, D-transpeptidase (A0A291TWB4) were observed.

Approximately a fifth of all the recovered secreted proteins were of unknown function. Some of those were even overexpressed in polymer condition with no growth observed indicating an important role either in cellular signalling, partial degradation or transportation (**SI Figure 5).** Still, only a fourth of the potentially exportable CAZy enzymes encoded in the genome were recovered in the current study (**Figure 3**). Thus, the questions are in which other conditions are the other CAZy enzymes overexpressed and what other type of substrates can *A. bestiarum* degrade.

Many proteins in the secretome remain to be functionally classified and fully characterized, especially in the past decades where secretome analysis has primarily focused on the

interaction between hosts and pathogens, particularly in the medical field. However, the secretome is more than just a means for a microorganism to infect a host; it is how it communicates with the environment and other microorganisms. The secretome is key to the survivability of bacteria in diverse environments, including highly oligotrophic niches like drinking water, the deep sea, or deep-subsurface soil. The ability to express and secrete a wide range of enzymes targeting multiple types of polymers may indicate *Aeromonas*' capability to adapt to a wide range of habitats, both terrestrial and aquatic, showcasing high metabolic versatility. In cases where *Aeromonas* was unable to degrade certain polymers, it is possible that this bacterium can partially degrade these polymers but requires the initial degradation to be led by a more specialized microbe that possesses all the necessary enzymatic machinery to convert the polymer into smaller subunits. Most of the bacteria possessing the ability to degrade polymers usually produce multiple copies of hydrolases with the same CAZy function that work synergistically assuring and improving degradation [36].

In the current study *Aeromonas bestiarum* showcased its capacity to degrade multiple polymers classifying it as a generalist in the microbial world. Although it was not able to degrade other polymers like cellulose, pectin or xylan commonly abundant in nature, its genome shows that it can produce some enzymes that were previously shown to be involved in the degradation of these polymers. Other species of the genus, however, has been proved to degrade cellulose [19,51] showcasing metabolic versatility of the genus. In a complex community where different niches of microorganisms are present, *Aeromonas* may not be able to be the main degrader of those polymers. Instead, it may be capable to play an intermediate role by degrading smaller oligomers of the polymers. If specialists are the first and the most important in degrading recalcitrant polymers like chitin, fucoidan, xylan or cellulose, classifying them as first degraders, *Aeromonas* can play the role of a second degrader, by degrading smaller oligomers, with lower crystallisation degree and higher water solubility. These generalists may be more resistant and adaptable to different environmental niches even oligotrophic ones like the drinking water. This dynamic between bacteria may reveal the relationships among different microbes: those that initiate degradation, those that continue the degradation, those that take up polymer oligomers, and those that scavenge the byproducts of other microbes [52,53]. Further research is needed to discover the strategies of each polymer-degrading bacterium, whether the enzymes are fully exported or attached to the external cell wall [54]. This understanding could elucidate the presence of diverse microbial communities in oligotrophic environments like drinking water.

In this study, we examined the growth of *Aeromonas* on various biopolymers, focusing on the secreted enzymes that facilitate degradation and subsequently enable their update. While we identified a range of biopolymer-specific enzymes, further research on the bacterial secretome is needed to provide functional insights into the many secreted

proteins whose functions remain unknown. Additionally, the important role of generalist organisms as secondary degraders in complex communities should not be overlooked, as they may drive overall community diversity and adapt to dynamic environments.

Acknowledgements

The authors would like to thank Dita Heikens for her support with the proteomics work and Lemin Chen for providing the complex biopolymer extract ("EPS"). They also acknowledge the NWO Spinoza Prize awarded to Mark van Loosdrecht for funding and support. ChatGPT was used to assist with language editing and proofreading.

Conflict of interest

All authors declare that they have no conflicts of interest.

References

1. Kabir SF, Rahman A, Yeasmin F, Sultana S, Masud RA, Kanak NA, et al. Chapter One - Occurrence, distribution, and structure of natural polysaccharides. In: Naeem M, Aftab T, Khan MMA, editors. Radiation-Processed Polysaccharides: Academic Press; 2022. p. 1-27.
2. Krause-Jensen D, Duarte CM. Substantial role of macroalgae in marine carbon sequestration. Nature Geoscience. 2016;9(10):737-42.
3. Decho AW, Gutierrez T. Microbial Extracellular Polymeric Substances (EPSs) in Ocean Systems. Front Microbiol. 2017;8:922.
4. Blot N, Berrier C, Hugouvieux-Cotte-Pattat N, Ghazi A, Condemine G. The Oligogalacturonate-specific Porin KdgM of Erwinia chrysanthemi Belongs to a New Porin Family*. Journal of Biological Chemistry. 2002;277(10):7936-44.
5. van Dijl J, Hecker M. Bacillus subtilis: from soil bacterium to super-secreting cell factory. Microbial Cell Factories. 2013;12(1):3.
6. Kang Z, Yang S, Du G, Chen J. Molecular engineering of secretory machinery components for high-level secretion of proteins in Bacillus species. J Ind Microbiol Biotechnol. 2014;41(11):1599-607.
7. Natale P, Brüser T, Driessen AJM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2008;1778(9):1735-56.
8. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiology. 2005;5(1):58.

9. Pallen MJ. The ESAT-6/WXG100 superfamily -- and a new Gram-positive secretion system? Trends Microbiol. 2002;10(5):209-12.
10. Bottai D, Gröschel MI, Brosch R. Type VII Secretion Systems in Gram-Positive Bacteria. Curr Top Microbiol Immunol. 2017;404:235-65.
11. Zhao L, Chen J, Sun J, Zhang D. Multimer recognition and secretion by the non-classical secretion pathway in Bacillus subtilis. Scientific Reports. 2017;7(1):44023.
12. Percival SL, Williams DW. Chapter Three - Aeromonas. In: Percival SL, Yates MV, Williams DW, Chalmers RM, Gray NF, editors. Microbiology of Waterborne Diseases (Second Edition). London: Academic Press; 2014. p. 49-64.
13. Janda JM, Abbott SL. The genus Aeromonas: taxonomy, pathogenicity, and infection. Clin Microbiol Rev. 2010;23(1):35-73.
14. van der Kooij D, Hijnen WA. Nutritional versatility and growth kinetics of an Aeromonas hydrophila strain isolated from drinking water. Appl Environ Microbiol. 1988;54(11):2842-51.
15. van der Kooij D, Visser A, Hijnen WA. Growth of Aeromonas hydrophila at Low Concentrations of Substrates Added to Tap Water. Appl Environ Microbiol. 1980;39(6):1198-204.
16. van Bel N, van der Wielen P, Wullings B, van Rijn J, van der Mark E, Ketelaars H, et al. Aeromonas species from non-chlorinated distribution systems and their competitive planktonic growth in drinking water. Appl Environ Microbiol. 2021;87(5).
17. Claudia GT, Dimitry YS, Wim H, Julia W, Kaatje B, Mark CMvL, et al. Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin. bioRxiv. 2024:2024.02.07.579344.
18. Tugui CG, Sorokin DY, Hijnen W, Wunderer J, Bout K, van Loosdrecht MCM, et al. Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin. RSC Chemical Biology. 2025;6(2):227-39.
19. Jiang Y, Xie C, Yang G, Gong X, Chen X, Xu L, et al. Cellulase-producing bacteria of Aeromonas are dominant and indigenous in the gut of Ctenopharyngodon idellus (Valenciennes). 2011;42(4):499-505.
20. van der Wielen PWJJ, Bakker G, Atsma A, Lut M, Roeselers G, de Graaf B. A survey of indicator parameters to monitor regrowth in unchlorinated drinking water. Environmental Science: Water Research & Technology. 2016;2(4):683-92.
21. Prest EI, Hammes F, van Loosdrecht MCM, Vrouwenvelder JS. Biological Stability of Drinking Water: Controlling Factors, Methods, and Challenges. 2016;7.
22. Felz S, Al-Zuhairy S, Aarstad OA, van Loosdrecht MC, Lin YM. Extraction of Structural Extracellular Polymeric Substances from Aerobic Granular Sludge. J Vis Exp. 2016(115).

23. Pfennig N, Lippert KD. Über das Vitamin B12-Bedürfnis phototropher Schwefelbakterien. Archiv für Mikrobiologie. 1966;55(3):245-56.
24. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research. 2021;50(D1):D571-D7.
25. Nielsen H, Tsirigos KD, Brunak S, von Heijne G. A Brief History of Protein Sorting Prediction. Protein J. 2019;38(3):200-16.
26. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-d9.
27. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-w303.
28. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar Gustavo A, et al. InterPro in 2022. Nucleic Acids Research. 2022;51(D1):D418-D27.
29. Flemming H-C, Wingender J. The biofilm matrix. Nature Reviews Microbiology. 2010;8(9):623-33.
30. Flemming H-C, Neu Thomas R, Wozniak Daniel J. The EPS Matrix: The "House of Biofilm Cells". Journal of Bacteriology. 2007;189(22):7945-7.
31. Eddy SR. Accelerated Profile HMM Searches. PLoS computational biology. 2011;7(10):e1002195.
32. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280 (Pt 2)(Pt 2):309-16.
33. Janecek S, Svensson B, Henrissat B. Domain evolution in the alpha-amylase family. J Mol Evol. 1997;45(3):322-31.
34. Domań-Pytka M, Bardowski J. Pullulan degrading enzymes of bacterial origin. Crit Rev Microbiol. 2004;30(2):107-21.
35. van Huizen NA, Ijzermans JNM, Burgers PC, Luider TM. Collagen analysis with mass spectrometry. 2020;39(4):309-35.
36. Sichert A, Corzett CH, Schechter MS, Unfried F, Markert S, Becher D, et al. Verrucomicrobia use hundreds of enzymes to digest the algal polysaccharide fucoidan. Nature Microbiology. 2020;5(8):1026-39.
37. Orellana LH, Francis TB, Ferraro M, Hehemann J-H, Fuchs BM, Amann RI. Verrucomicrobiota are specialist consumers of sulfated methyl pentoses during diatom blooms. The ISME Journal. 2022;16(3):630-41.
38. Abbott DW, Gilbert HJ, Boraston AB. The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate beta-elimination. J Biol Chem. 2010;285(50):39029-38.

39. Yamane K, Suzuki H. Cellulases of Pseudomonas fluorescens var. cellulosa. Methods in Enzymology. 160: Academic Press; 1988. p. 200-10.
40. Nakamura K, Kitamura K. Cellulases of Cellulomonas uda. Methods in Enzymology. 160: Academic Press; 1988. p. 211-6.
41. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40(Web Server issue):W445-51.
42. Fraiberg M, Borovok I, Weiner RM, Lamed R, Bayer EA. Bacterial Cadherin Domains as Carbohydrate Binding Modules: Determination of Affinity Constants to Insoluble Complex Polysaccharides. In: Himmel ME, editor. Biomass Conversion: Methods and Protocols. Totowa, NJ: Humana Press; 2012. p. 109-18.
43. Matyar F, Kaya A, Dinçer S. Distribution and antibacterial drug resistance of Aeromonas spp. from fresh and brackish waters in Southern Turkey. Annals of Microbiology. 2007;57(3):443-7.
44. Evangelista-Barreto NS, de Carvalho FC, Vieira RH, Dos Reis CM, Macrae A, Rodrigues Ddos P. Characterization of Aeromonas species isolated from an estuarine environment. Braz J Microbiol. 2010;41(2):452-60.
45. Kivanc M, Yilmaz M, Demir F. The occurrence of Aeromonas in drinking water, tap water and the porsuk river. Braz J Microbiol. 2011;42(1):126-31.
46. Farmer J, Arduino M, Hickman-Brenner F. The genera Aeromonas and Plesiomonas. 1992.
47. Lujan-Hernandez J, Schultz KS, Rothkopf DM. Rapidly Progressive Soft Tissue Infection of the Upper Extremity With Aeromonas veronii Biovar sobria. J Hand Surg Am. 2020;45(11):1091.e1-.e4.
48. Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. Infect Immun. 2011;79(9):3476-91.
49. Egea L, Aguilera L, Giménez R, Sorolla MA, Aguilar J, Badía J, et al. Role of secreted glyceraldehyde-3-phosphate dehydrogenase in the infection mechanism of enterohemorrhagic and enteropathogenic Escherichia coli: interaction of the extracellular enzyme with human plasminogen and fibrinogen. Int J Biochem Cell Biol. 2007;39(6):1190-203.
50. Bishop RE. The lipid A palmitoyltransferase PagP: molecular mechanisms and role in bacterial pathogenesis. 2005;57(4):900-12.
51. Islam F, Roy N. Screening, purification and characterization of cellulase from cellulase producing bacteria in molasses. BMC Research Notes. 2018;11(1):445.
52. Pontrelli S, Szabo R, Pollak S, Schwartzman J, Ledezma-Tejeida D, Cordero OX, et al. Metabolic cross-feeding structures the assembly of polysaccharide degrading communities. 2022;8(8):eabk3076.

53. Goldfarb KC, Karaoz U, Hanson CA, Santee CA, Bradford MA, Treseder KK, et al. Differential Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying Chemical Recalcitrance. 2011;2.
54. Reintjes G, Arnosti C, Fuchs B, Amann R. Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. The ISME Journal. 2019;13(5):1119-32.

Supplementary information material to:

Versatile hydrolytic potential in *Aeromonas bestiarum*

TABLE OF CONTENTS

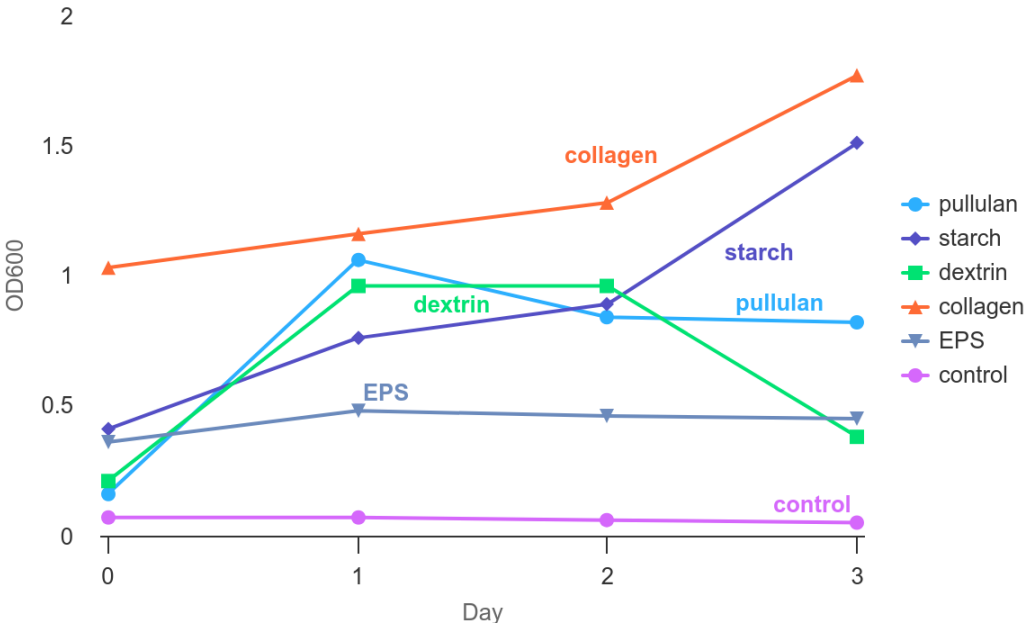


Figure 1. OD600 measurement of the *A. bestiarum* cultures grown on 9 different polymers: pullulan, starch, dextrin, xylan, cellulose, pectin, EPS

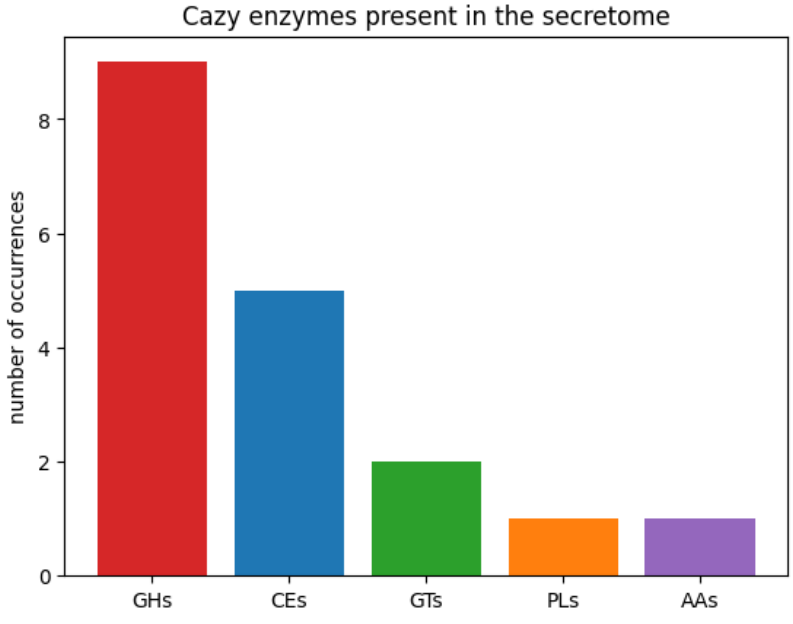


Figure 2. The number of different CAZy enzymes that were secreted in all 10 growth conditions: fucoidan, complex biopolymer extract, pullulan, starch, xylan, dextrin, cellulose, pectin collagen and chitin

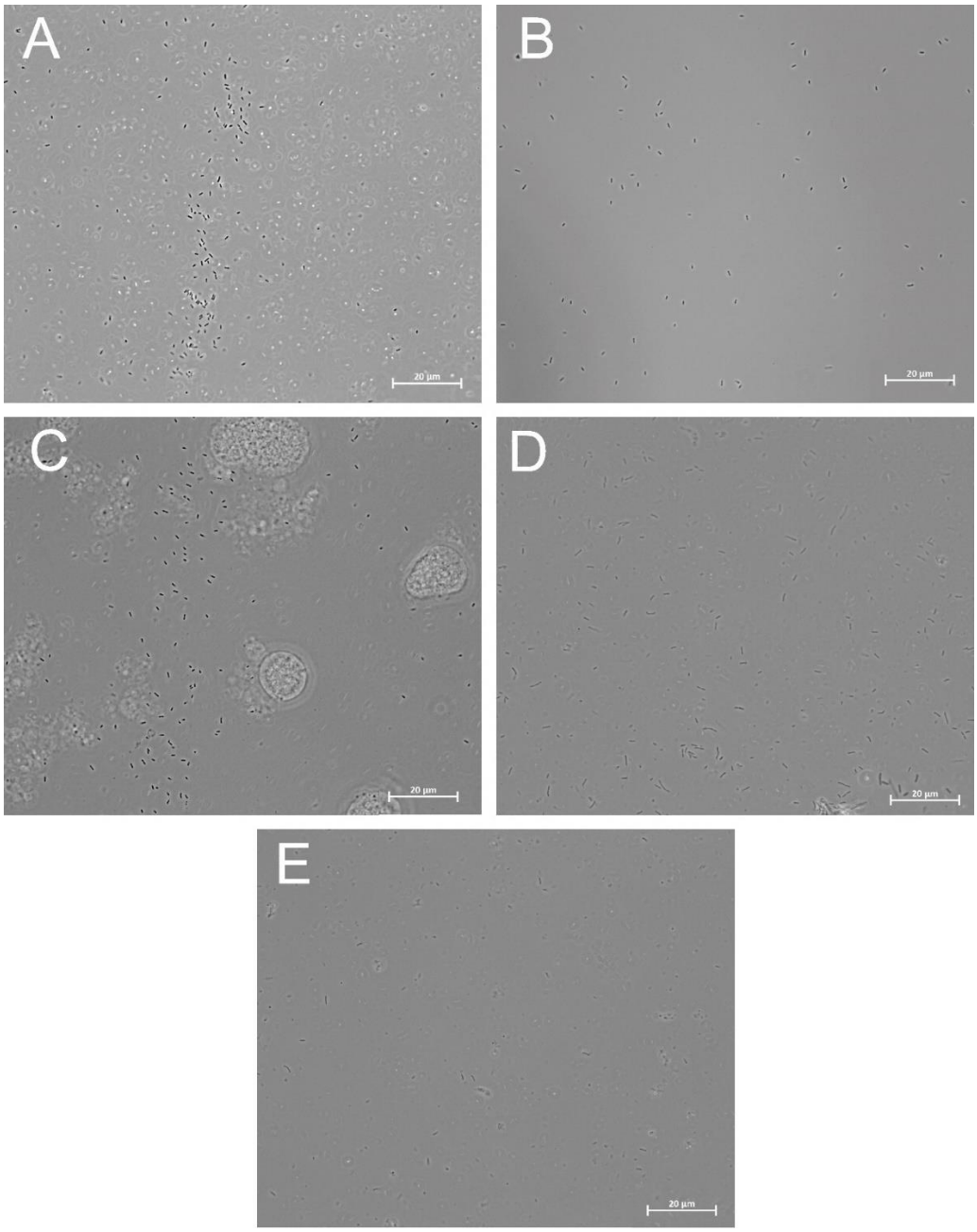


Figure 3. Microscopic pictures of *A. bestiarum* grown on pullulan (A); dextrin (B), starch(C), collagen (D) and complex biopolymer extract ("EPS") (E). The pictures were taken at a 63x magnification

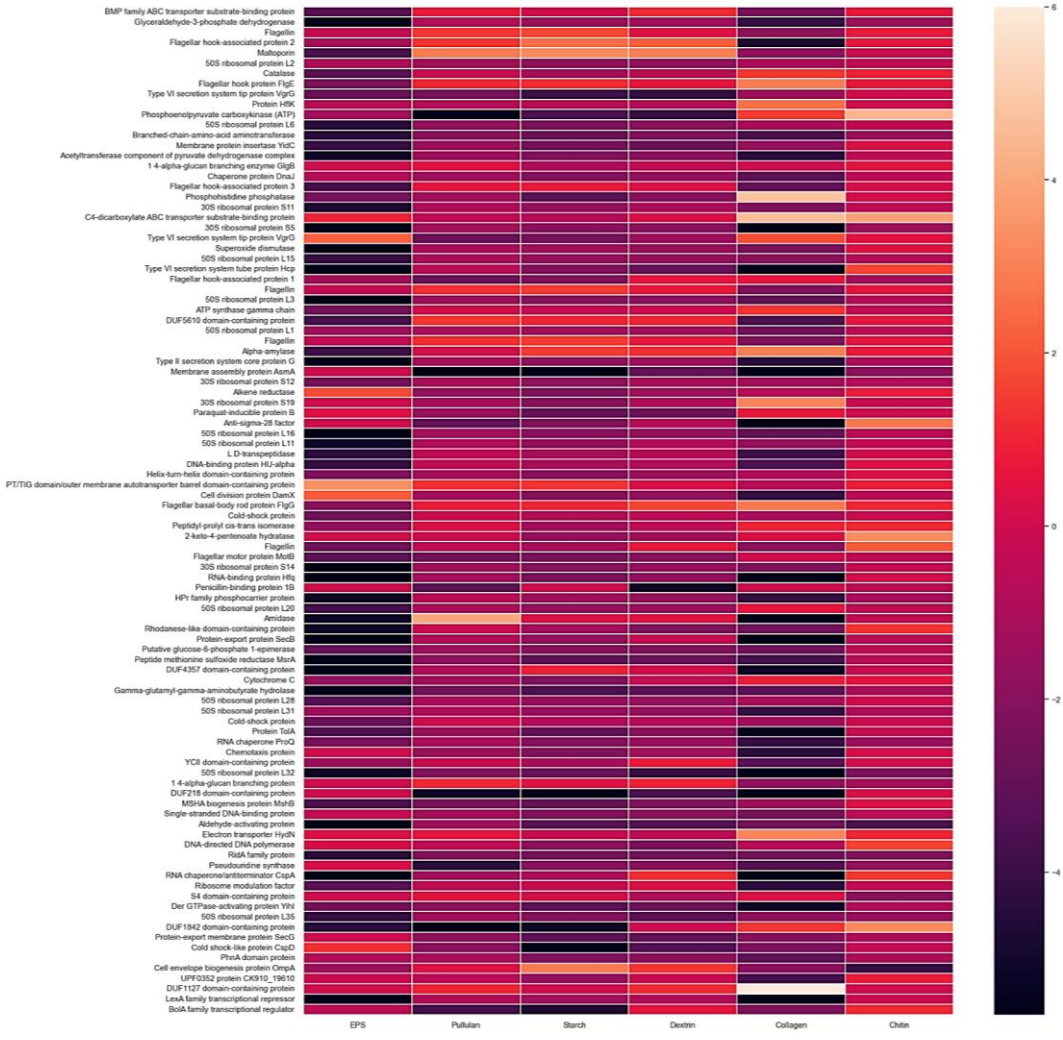


Figure 4. Relative abundance of all the protein secreted in a non-classical pathway, that were identified in all the growth conditions

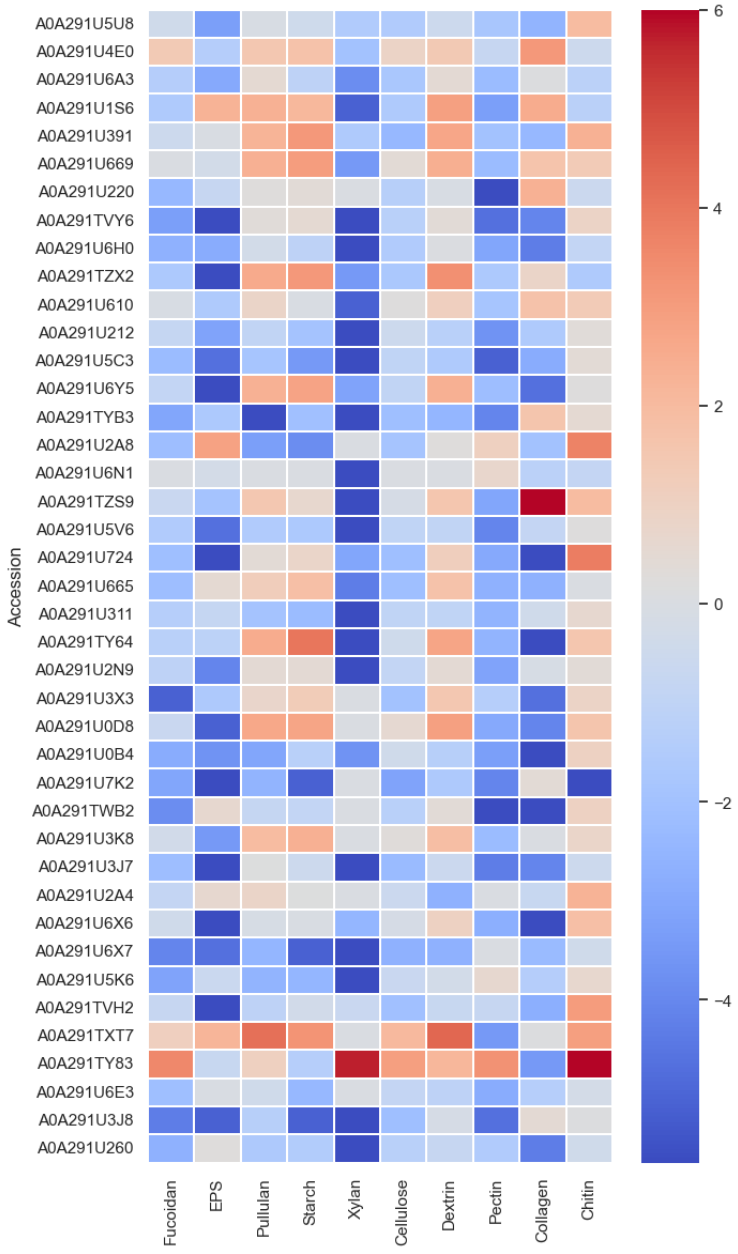


Figure 5. Heatmap represents the relative abundance of the proteins labelled as "uncharacterized" across all 10 polymer growth conditions in the genome *of A bestiarum*

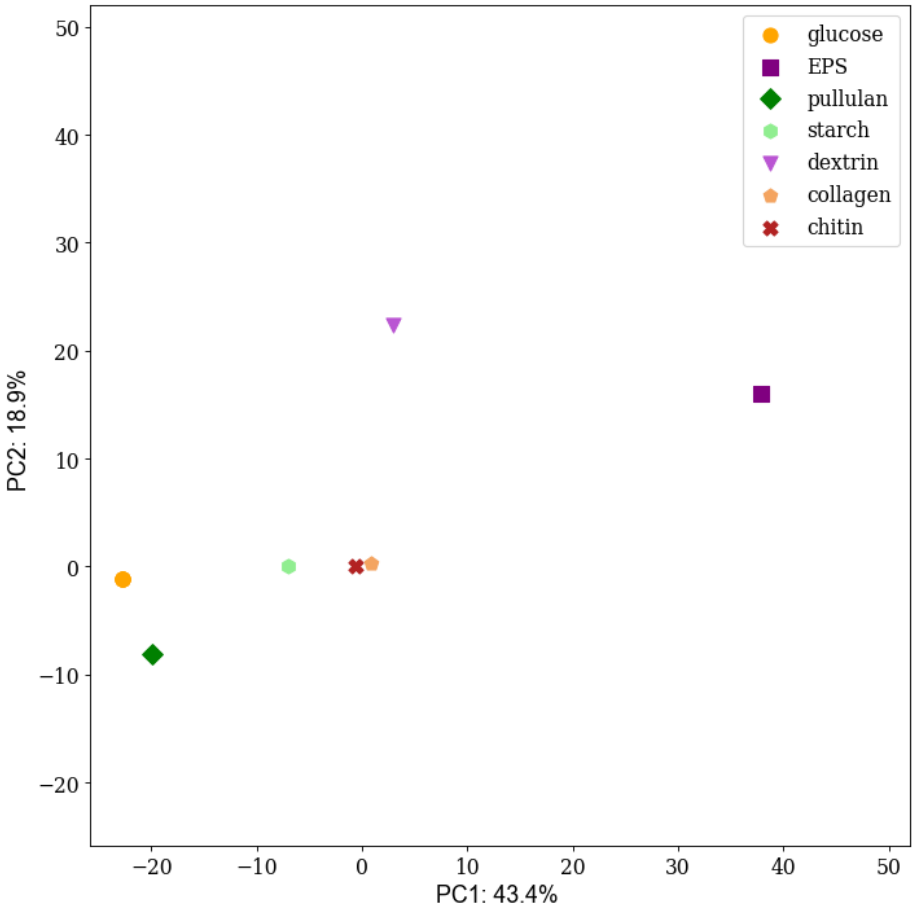


Figure 6. PCA analysis of the supernatant of all 11 conditions. Each condition is plotted relative to glucose. All the proteins that were identified in the supernatant fraction was part of the analysis.

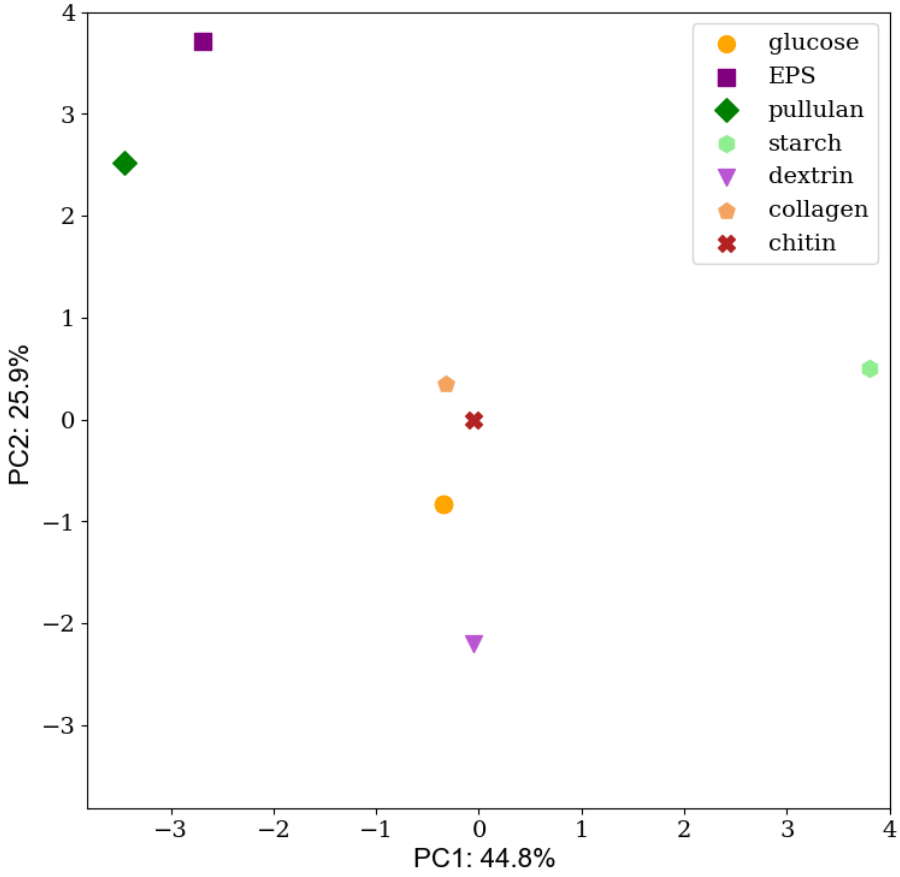


Figure 7. PCA analysis of all the CAZy enzymes annotated enzymes in all 11 conditions. Each condition is plotted relative to glucose. All the proteins that were identified in the supernatant fraction was part of the analysis.

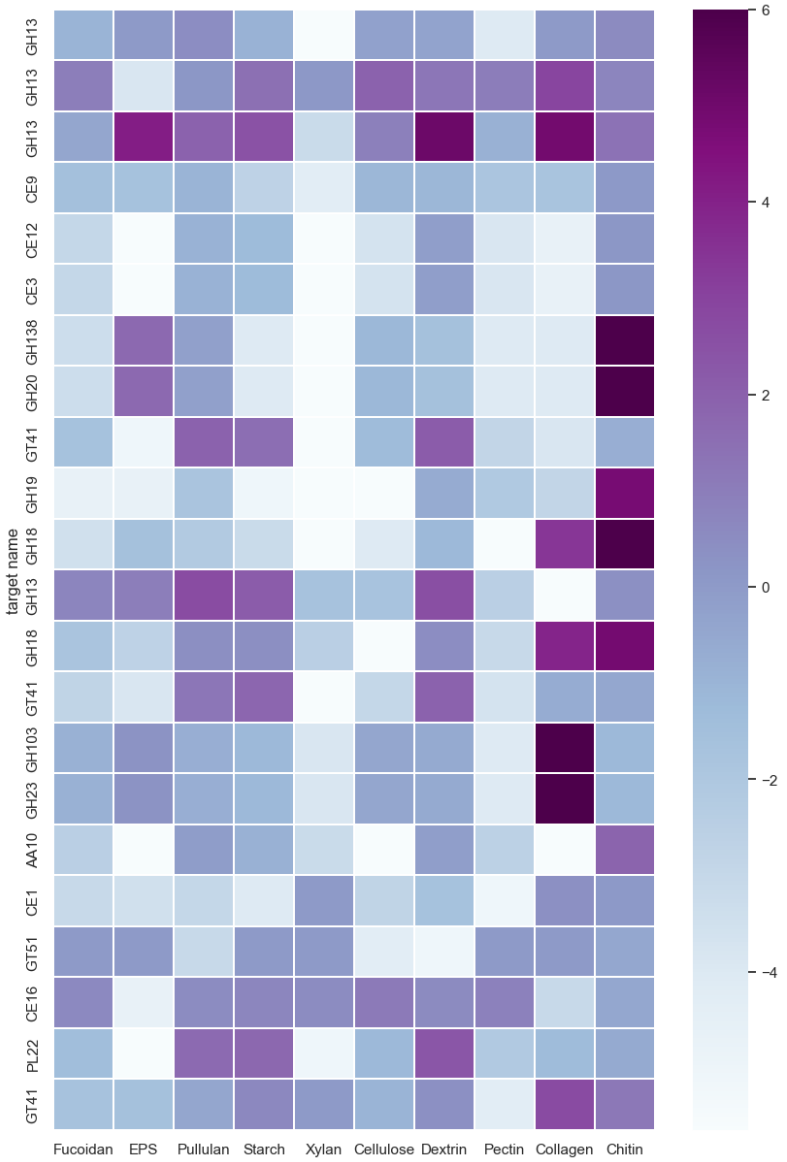
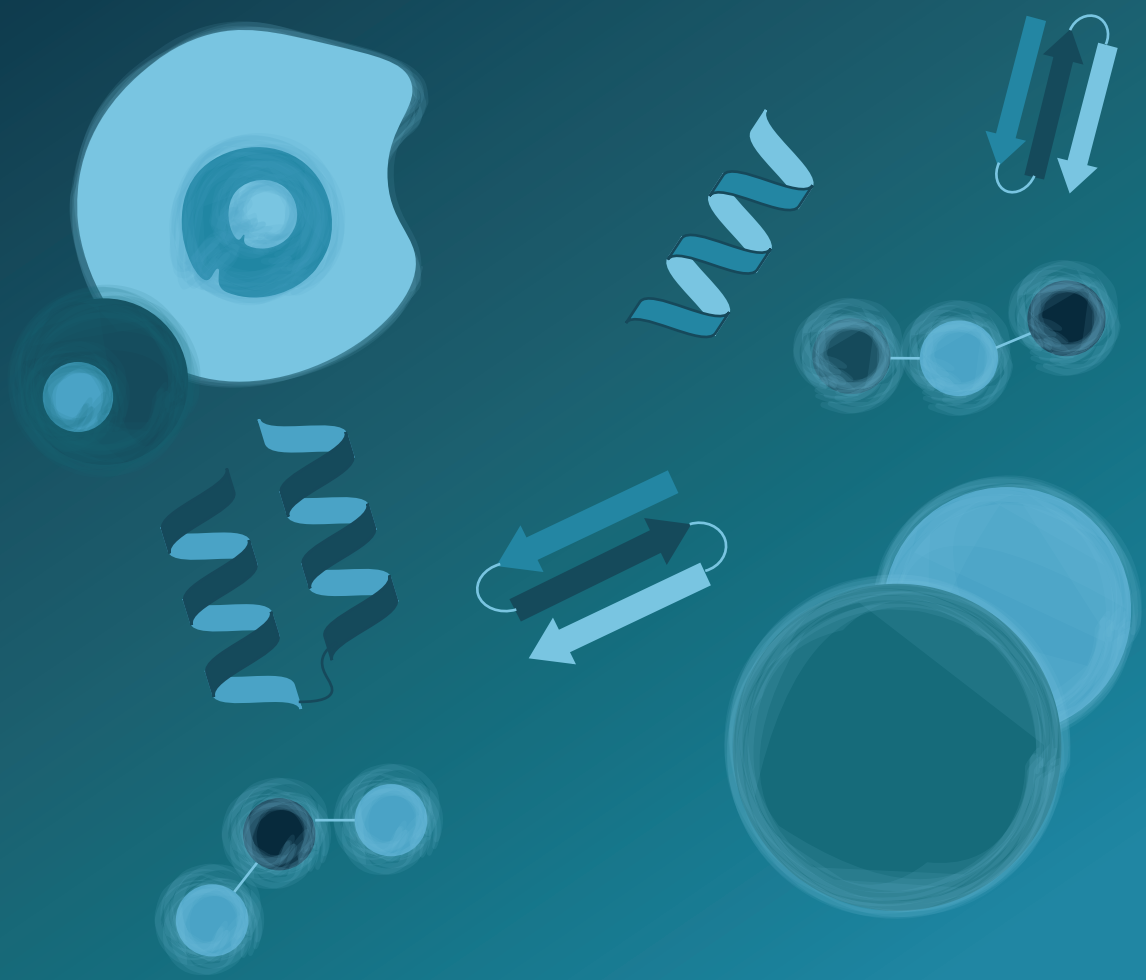


Figure 8. Relative abundance of all the identified CAZy proteins across all the growth conditions.

Chapter 4

Studying the microbiome of drinking water loose deposits using metagenomics and metaproteomics

Claudia G. Tugui, Wim Hijnen, Mark C.M. van Loosdrecht and Martin Pabst



Data availability: SI Figures and Tables can be found at the end of the chapter. Additional SI data including shotgun proteomic data, metagenomic sequence databases and tables, and database search files for this article are available via the DataverseNL project: https://doi.org/10.34894/KNEBSG

Abstract

Drinking water distribution systems harbor a surprisingly rich microbiome, regardless of the water treatment applied. How bacteria manage to thrive in this oligotrophic environment remains a topic of debate. Indicator organisms are used worldwide to assess drinking water quality and detect bacterial regrowth. One possible explanation for bacterial regrowth is the presence of certain biopolymers. In the Netherlands, *Aeromonas*—a microbe known for its ability to grow on the biopolymer chitin—is used as an indicator strain to signal bacterial regrowth potential. However, microbial communities in these environments are highly diverse, with functional capacities relevant to biopolymer degradation.

In this study, sediment samples from seven different locations across three different drinking water distribution systems supplied by drinking water produced from the same surface water from large reservoirs were analyzed using metagenomics and metaproteomics. A highly diverse microbial community was observed at all locations, with a core microbiome shared across all loose deposits samples. Metagenomic analysis indicated the presence of over 100 bacterial genera, including many well-known species commonly found in water and soil. Additionally, taxa from plants, animals, archaea, and protozoa were detected. The functional profile, particularly glycoside hydrolases, suggested that many bacteria have the potential to degrade and utilize multiple biopolymers. Metaproteomic analysis further revealed that a significant portion of the identified proteins in certain locations were non-bacterial, underscoring the role of protists, animals, and fungi in shaping the microbial community within the drinking water distribution system.

Cluster analysis of microbial diversity and polymer degradation potential showed distinct groupings of locations based on previously reported *Aeromonas* regrowth (with one exception). In addition to unraveling the microbial structure and protein biomass composition within these locations, the findings also support the hypothesis that biopolymers in the supplied drinking water serve as a source for the growth and maintenance of biofilms in loose deposits.

Key words: metagenomics, metaproteomics, drinking water, loose deposits, *Aeromonas*, CAZy enzymes

Introduction

Access to safe, high-quality drinking water is now considered a basic standard in developed countries. Drinking water quality parameters such as microbial and chemical contaminants, turbidity, taste, odor and smell, are monitored to ensure high-quality and safe water for consumption. Nevertheless, supply of biologically unstable drinking water can severely impact on the microbial and esthetical quality of the water in the distribution system which is called regrowth. Thus, in the last decade, much research has been devoted to unravelling the reasons behind regrowth of microorganisms in these oligotrophic environments especially when the water is distributed with a residual chlorine [1]. A reason for the regrowth of bacteria and other microorganisms in the drinking water distribution systems is the presence of easily or slowly biodegradable organic matter. Trace levels of these compounds in the drinking water may serve as food sources for bacteria residing in sediments and biofilms in the pipes [2, 3]. In turn, these biofilms may serve as food source for protozoa and small invertebrates which graze on them [4]. Furthermore, in surface water supplies slowly biodegradable biopolymers that enter the drinking water system can be a nutrient source, and subsequently allow regrowth of microorganisms [5, 6]. Biopolymers can be degraded by a diverse spectrum of hydrolytic enzymes, and the resulting sugar or amino acid building blocks can be rapidly taken up by microorganisms. Unfortunately, chemical analyses of drinking water, such as through measurements and characterization of dissolved organic carbon (DOC; LC-OCD) or nitrogen content [7], provide only limited insights into the presence of specific biopolymers. The role of biopolymers has been discussed in terms of how they enter the water system, their impact on bacterial communities, and the implications for sanitation and disinfection procedures [8, 9]. However, the origin of these biopolymers is likely diverse and little understood. It is hypothesized that some of these biopolymers come from prokaryotes as well as small eukaryotes residing in the supplied drinking water, or proliferate in the drinking water network. Invertebrates, plants, and even fungi contain polymers such as chitin, cellulose, and chitosan in their exoskeletons and cell walls. These organisms release the biopolymers in the environment which provides a rich carbon and nitrogen source for the bacteria. However, these bacteria require a range of specialized enzymes to break down these polymers, enabling their uptake and incorporation into central carbon metabolism to support growth.

Aeromonas is a member of the genus *Gammaproteobacteria*, widely found in various water environments, including drinking water, where some representatives can cause diseases in both humans and animals [10, 11]. However, these pathogenic *Aeromonas* species are not found in drinking water in the Netherlands [12, 13]. *Aeromonas* bacteria are regulated in the Netherlands as an indicator strain to assess the regrowth conditions in non-

chlorinated drinking water distribution systems (DWDS). An interesting aspect of *Aeromonas* is that it can degrade and consume chitin as the sole carbon and nitrogen source [12]. A recent proteomic study of the secretome and intracellular proteome demonstrated that *Aeromonas* produces a range of specialized enzymes to degrade the polymer, while simultaneously taking up the resulting mono- and dimers and incorporating them into nitrogen and central carbon metabolism [14]. Also, the ability of *Aeromonas* to secret a range of other hydrolytic enzymes which allows to grow on different biopolymers found in nature has been demonstrated recently (Tugui et al., manuscript in preparation, Chapter 3, this thesis). Suspended competitive growth of *Aeromonas* in drinking water in the DWDS is not the cause of increased of nutrients in the water [8]. Moreover, the majority of these bacteria in the DWDS are found in loose deposits niche with high numbers in the invertebrate *Asellus aquaticus* [6]. This makes *Aeromonas* a generalist in the competitive and oligotrophic loose deposits niche like that is found in drinking water distribution systems.

Traditional microbial general (Heterotrophic Plate Count, HPC) as well selective (*Aeromonas)* cultivating methods on agar media have significant drawbacks when it comes to investigating the microbial diversity in drinking water. They are highly selective and time-consuming. Numerous other methods have been employed to unravel the microbial ecosystem and dynamics in drinking water environments [15-18]. However, ATP quantification and total cell count assays are faster and more reliable than HPCs. However, ATP quantification determines only active bacteria without capturing dormant cells [19]. Flow cytometry can determine between live and dead bacteria, it cannot accurately determine the number of bacteria present in loose particles, because multiple bacteria can be attached to a single particle [20].

Over the past decades, advancements in high throughput metagenomic sequencing advanced the investigation of complex microbial environments. These culture-independent methods helped identify the phylogenetic structure and functional potential of microbes present in the drinking water distribution systems [15, 21, 20, 22-24]. However, metagenomic methods are focused on the dominant microbial populations present. Therefore, these methods are not suited to detect accurately important drinking water species such as *Aeromonas* as well as opportunistic pathogens such as *Stenotrophomonas* and *Legionella* bacteria, since these species are only a very small part of the available microbial food web in the DWDS [18]. Still thanks to the recent advancements in sequencing techniques, a high sequence depth has been achieved that may allow the identification of low abundant bacteria and retrieval of MAGs [25, 26]. Furthermore and unfortunately , metagenomic studies do not provide information on microbial biomass abundance or the expressed metabolic functions but provides information on the relative abundance of the microbes and their potential function leaving room for speculation whether these functions are actually performed in the environment [27]. To achieve this, metaproteomic

analysis of expressed proteins can help determine the protein biomass contribution of individual community members and reveal their expressed enzymatic functions [28, 29].
The aim of the presented study is to employ whole metagenome sequencing and metaproteomics to study the microbial diversity and their metabolic and hydrolytic potential in the loose deposits of a drinking water distribution system. For this, we investigated the loose deposits from seven different locations in non-chlorinated distribution systems supplied from three different drinking water production plants in The Netherlands. This allowed us to explore the microbial diversity at different locations, and to investigate potential correlation with the presence of Aeromonas regrowth in these areas. While whole metagenome sequencing provided the metabolic potential to degrade and utilize different biopolymers, metaproteomics provided insights into the protein biomass composition of individual microbes and other protein sources.

Materials and Methods

Sampling of loose deposits from drinking water distribution systems. The sampling campaign occurred in the spring of 2023 from fire hydrants in the DWDS pipes before the water meters at the consumers, by the flushing method described by Ketelaars and co-workers [28]. The water was at the moment of sampling filtered to obtain >500 µm and 100–500 µm sediment fractions, where the latter was used for the subsequent whole metagenome and metaproteome sequencing analysis. In total, seven samples from production locations and their drinking water distribution systems (DWDS) were collected, with four samples from DWDS with and three samples from DWDS without *Aeromonas* regrowth conditions, as detailed in the table below.

Table 1. Sampling locations of loose deposits from DWDSs

Sampling location		Drinking water production[a]	*Aeromonas* Regrowth[b]
Delft	**DA4 Molenweide 54, 2614 LT**	**Mix of TP#1[a] + TP#3[a]**	**+**
Maassluis	**MsA3 Prins Mauritsstraat 29, 3143 LL;**	**TP#1 or mixed with TP#3**	**+**
Monster	**MA1 Duinvallei 5, 2681 XB**	**TP#1 or mixed with TP#3**	**+**
Naaldwijk	**NA2 Gerbrandyst raat 1, 2672 AH**	**TP#1 or mixed with TP#3**	**+**
Hoek	**HNA7 Molendreef 3, 4542 AA**	**TP#2[a]**	**-**
Philippine	**PNA6 Schorrenkuidelaan 22, 4553 BZ**	**TP#2**	**-**
Terneuzen	**GNA5 Geulstraat 88, 4535 CX;**	**TP#2**	**-**

[a]River Meuse water treated with reservoir storage, coagulation/filtration, UV disinfection granular activated carbon (GAC), GAC filtrate chlorination (Hijnen et al., 2024: TP#1, TP#2, TP#3 = Berenplaat, Braakman and Kralingen); [b]*Aeromonas*regrowth means elevated numbers above the regulated regrowth target level of 1000 colony forming unis per 100 mL of drinking water.

DNA extraction and sequencing. The DNA was extracted form the samples using the KingFischer instrument. Illumina MiSeq and NovaSeq 6000 systems were used for generating single-end and paired-end sequence reads. Metagenome assembly has been performed using MEGAHIT 1.2.9. Contigs smaller than 1000 base pairs have been removed from the final assembly. **Quality control of sequenced reads.** The quality of the sequenced raw reads was assessed by FastQC (version 0.11.7) with default parameters (Andrews, 2010) and visualized with MultiQC (version 1.0). Low-quality paired-end reads were trimmed and filtered by Trimmomatic version 0.39 on the paired-end mode [29]. **Microbiome profiling.** Taxonomic classification of raw reads and assembly was performed to profile the microbiome from each sample using the standard Kraken2 (version 2) database (uses all complete bacterial, archaeal, and viral genomes in NCBI Refseq database) complemented with a curated wastewater database [30] with default parameters [31]. The taxonomic classification outcomes from Kraken2.0 were converted into stats tables using the Pavian visualization tool [32] to explore metagenomics classification datasets. Biom files from kraken2 were generated with kraken-biom tool. The biom files were used as input datasets for generating a phyloseq file and further analysis. Figures were generated with the R

package "phyloseq" [33]. **Assembly of sequence reads.** Clean reads were assembled into contigs using MetaSPAdes (version 3.13.0) with default parameters [34]. **Binning of DNA contigs.** Contigs resulting from the sequencing of only the iDNA pools of bioaggregates were binned with MetaBAT version 2.2.15 [35] to reconstruct metagenome-assembled genomes (MAGs) on default parameters. **Functional annotation of the contigs** Possible proteins were extracted from the contigs using Prodigal 2.6.3 [36]. Mapping of proteins to function was performed with Eggnog-mapper2 version 2.1.12 [37] using the diamond method on the eggnog main database downloaded on 2024-10-03. **Taxonomical annotation of the contigs.** Mapping of contigs to likely taxonomical assignment was done with CAT_Pack [38] with GTDB r220 as reference proteome. **Taxonomical annotation of the bins.** Taxonomical assignment of the bins was performed using GTDBtk 2.4.0 using GTDB r220, using default parameters. **Protein extraction and proteolytic digestion.** 500 µL of the wastewater influent was taken and diluted with B-PER (175µl) and 50 mM TEAB buffer (175µl) and heated at 90 ℃, for 5 min under shaking at 300 rpm. Further, the sample was subjected to cell lysis using vortexing 3 times for 1 minute using a bench vortexing machine, sonication on a sonication bath for 15 minutes and one freeze/thaw cycle (frozen at -80 ℃, thawed in incubator at 40 ℃ for 5 minutes). The samples were then centrifuged and transferred to a 1.5 mL LoBind Eppendorf tube. TCA was added to the sample at a ratio 1:4 (v/v, TCA/sample), vortexed and incubated at 4 ℃ for 20 minutes. After centrifugation, the protein pellet was re-solubilized in 6 M urea and then reduced with DTT (dithiothreitol) and alkylated using IAA (iodoacetamide). After alkylation, the sample was transferred to a FASP filter (Millipore, MRCPRT010) which was previously conditioned by washing 2 times with 100 mM ABC buffer. The filters were centrifuged at 14K rpm in a bench top centrifuge, for 45 minutes, and then 2 times at 14K rpm for 40 minutes after adding 100 mM ABC buffer. Next, the proteins were proteolytically digested on the FASP filter by adding 100 µL trypsin solution, which was prepared by diluting 8 µL trypsin stock solution (0.1 ug/mL in 1 mM HCl, Promega, Cat No) in 100 µL 100 mM ABC. The FASP filters were incubated overnight for digestion, at 37 ℃, under gentle shaking at 300 rpm. The following day, the filters were centrifuged and then once washed with 100 mM ABC buffer followed by a second wash with 100µl of 10% ACN 0.1% FA/H2O collected the proteolytic peptides. The pooled fraction was then purified using an OASIS HLB well plate (Waters, UK) according to the manufacturer's protocol. The purified peptide fraction was speed-vac dried and stored at -20 ℃ until further analyzed. **Shotgun metaproteomics.** To the speed vac tried samples 20 µL of 3% acetonitrile and 0.01% trifluoroacetic acid in H_2O was added, vortexed, then left at room temperature for 30 min, and then once more vortexed. The peptide concentration was determined by measuring the absorbance at 280 nm using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Germany). Samples were diluted to a concentration of 0.5 mg/mL. Shotgun metaproteomics was performed as described previously[39], with a randomized sample order. Briefly, approximately 0.5µg protein digest was analyzed using a nano-liquid-

chromatography system consisting of an EASY nano-LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 25% solvent B over 90 min, from 25% to 55% over 60 min, followed by back equilibration to starting conditions. Solvent A was a 0.1% formic acid solution in water (FA), and solvent B consisted of 80% ACN in water and 0.1% FA. The Orbitrap was operated in data dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70 K resolution in full MS mode with a maximum ion injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD) at a normalized collision energy of 28. MS/MS scans were acquired at 17.5 K resolution with AGC target of 2E5 and IT of 75 ms, 1.2 m/z isolation width. **Data analysis.** The mass spectrometric raw data for each sample were de novo sequenced using PEAKS Studio X (Bioinformatics Solutions, Inc., Canada). *De novo* sequences with an ALC score >70 were subjected to taxonomic quality profiling using the NovoBridge pipeline as described previously. The metagenomic-based database was used to perform database searching using PEAKS Studio X (Bioinformatics Solutions, Inc., Canada). The search was performed allowing up to 3 missed cleavages, with carbamidomethylation as a fixed modification, and methionine oxidation and asparagine or glutamine deamidation as variable modifications, allowing 20 ppm precursor error and 0.02 Da fragment ion error. Peptide-spectrum matches from the second round were filtered to a 5% false discovery rate (FDR) at the PSM level, and protein identifications with ≥2 unique peptide sequences were considered significant. The complete dataset was combined using the PEAKSQ module allowing 10 minutes RT shifts and 10 ppm mass error. The taxonomic assignment was performed using the NCBI database and the GhostKOALA search engine [40].

Results

Loose deposits were obtained from seven different locations from drinking water distribution systems located in the south-west of The Netherlands and supplied by non-chlorinated drinking water from three different plants. Sampling was performed by flushing the water from hydrants, where the water was at the moment of sampling filtered to obtain > 500 µm and 100–500 µm sediment fractions as previous described by Ketelaars *et al*. [30] After sampling, the sediments were stored in sterile jars and transported in cooling compartments. The loose deposit samples were homogenized and subsequently concentrated through centrifugation in 50 mL Falcon tubes.

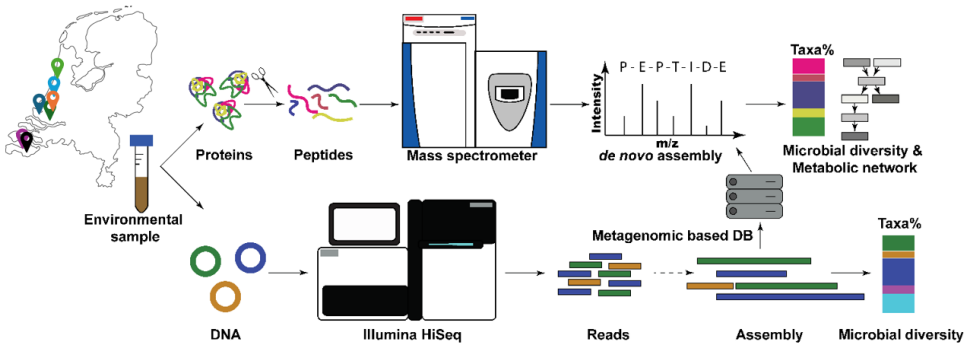


Figure 1: The multi-meta-omic workflow that was employed to study the sediments from seven different locations across a drinking water distribution system. First, metagenomic analysis using short-read sequencing was performed, which provided information about the taxonomies and metabolic potential present in each location. Furthermore, the proteins identified from the metagenomic analysis were used as a reference sequence database to annotate the metaproteomic data of the same sediments samples, which provided additional protein biomass composition for each location.

The main focuses of the study was on exploring the microbial community that is present in the sediments collected from the seven different locations. First, we performed a whole metagenome sequencing analysis and investigated the microbial diversity. Statistical differences in taxonomic abundance and occurrence between the sampled locations (**Figures 2B and 2D**) showed larger differences across the seven locations, albeit *Aeromonas* was identified only in very low abundance (**Figure 2C**). However, these are relative abundances, which do not correlate with the otherwise culture-based detection of *Aeromonas*. Also, the main fraction of *Aeromonas* is not expected in the loose deposits. For several samples, the relative abundance of *Aeromonas* was below the threshold that was applied for identification of bacterial genera.
A large spectrum (approx. 200) different genera were annotated based on the recovered reads, across all samples with *Streptomyces*, *Bradyrhizobium*, and *Pseudomonas* being the most abundant genera across all sampled locations. The only exception is sample MsA3 where *Candidatus* Rhabdochlamydia showed higher relative abundance compared to *Bradyrhizobium*. *Candidatus* Rhabdochlamydia appeared only in samples MsA3 and MA1. The presence of this genus may indicate the occurrence of invertebrates. The species *Candidatus* Rhabdochlamydia porcellionis belonging to the genus *Rhabdochlamydia*, is found in the hepatopancreas of the isopod Porcellio scaber [43], a wood louse which is not found in the sampled DWDSs. But the isopod A. aquaticus was found in these DWDS with *Aeromonas* regrowth occurring which may indicate that also in the hepatopancreas of this water isopod these bacteria can be found. Locations MsA3, MA1 and DA4 show the lowest

bacterial richness according to the Shannon diversity index (**Figure 2B**). A similar trend can be observed at the level of beta-diversity calculated through the Bray-Curtis dissimilarity index **(Figure 2D**). The samples formed one clear cluster: PNA6, GNA5, HNA7, and NA2, which indicate bacterial communities that were more closely related to each other, while the others MA1, DA4 and especially MsA3 were very different from the other locations. Interestingly, the cluster contains all samples without *Aeromonas* regrowth conditions including sampling location NA1 (with reported regrowth).
More *Aeromonas* regrowth and the presence of *A. aquaticus* means more biofouling in loose deposits which may be the cause of the lower bacterial richness in these samples. It is unclear why sample NA2 from a location with *Aeromonas* regrowth is located in the cluster of the sample locations without *Aeromonas* regrowth.
The relative abundance of *Aeromonas* reads in the loose deposits was very low and was only detected in trace levels in four locations out of seven, namely PNA6, GNA5, HNA7 and NA2 – which all clustered together. For all samples with *Aeromonas* regrowth in the bulk water (except NA2, **Table 1**), the relative abundance of *Aeromonas* in the loose deposits of MsA4, MA1 and DA4 was too low to pass the 0.2 % relative abundance threshold (**Figure 1C and SI Figure 1**). However, in another study where the bulk water and the loose deposits were sampled simultaneously no correlation was observed between the *Aeromonas* levels in the bulk water and the loose deposits [6]. These authors concluded that *Aeromonas* levels in the bulk water are high and above the regulated standard when *A. aquaticus* biomass, a large invertebrate visible with the naked eye, is present. This biomass is not homogeneously distributed over the loose deposit samples.

Other top 10 abundant bacteria identified in all samples were *Mycobacterium*, *Sphingomonas* and *Nocardia*. *Sphingomonas* is a well-known inhabitant of the drinking water system and is known for its ability to form biofilms [21]. The *Nocardia* genus generally inhabits the soil, and some members of this genus are considered pathogenic. However, these bacteria are not mentioned as problematic opportunistic pathogens for drinking water [44]. Its resistance to microbicides and its own production of antimicrobial agents can make it resistant to treatment and a good candidate to proliferate in complex ecosystems [45]. *Mycobacterium* is a well-known inhabitant of the drinking water system, especially its non-tuberculous species due to its resistance to disinfection [46].

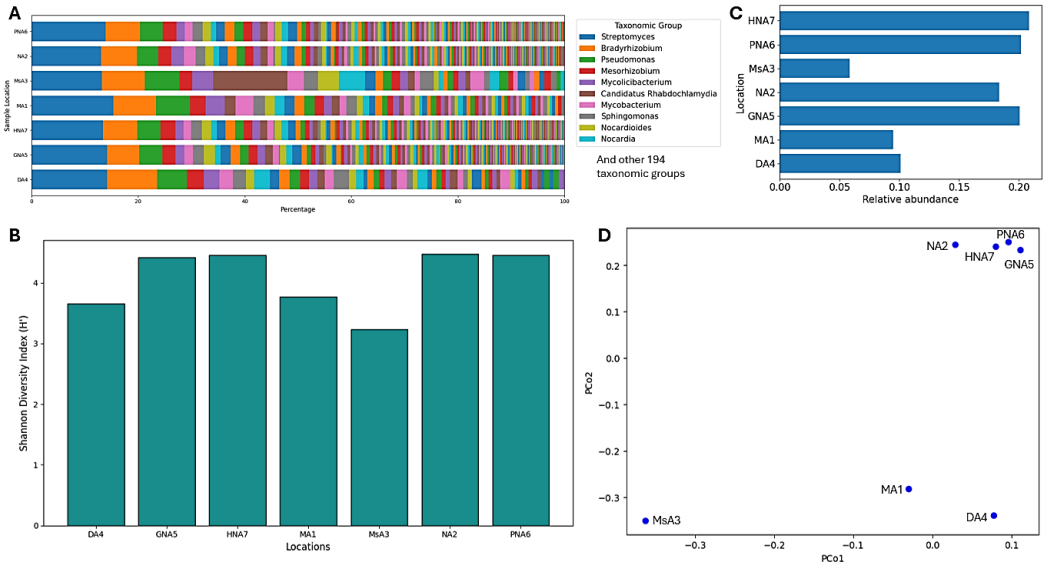


Figure 2. Microbial diversity, richness and abundance of the microbial community found in the sediments across seven sampling locations, as observed by whole metagenomic sequencing. **A**. Bacterial taxa that are present in the seven sampling locations and their relative abundance. The first 10 taxa are annotated in the legend. **B**. Alpha-diversity calculated through the Shannon diversity index of the bacteria genera across the seven sampling locations. **C**. Relative abundance of *Aeromonas* based on reads for all seven sampling locations. **D**. Beta-diversity from all seven sampling locations as determined by Bray-Curtis dissimilarity and plotted with PCoA.

Further, we investigated the microbial potential to express various glycoside hydrolases—enzymes necessary for degrading different biopolymers—enabling microbes to metabolize the hydrolysis products for biomass maintenance and growth. For this, we investigated the metabolic potential to degrade the four common biopolymers cellulose, chitin, xylan and rhamnogalacturonan. Interestingly, we found a wide range of glycoside hydrolases (GHs, or CAZy enzymes) encoded in the genomes of the identified microbes. One might

expect a trend where higher bacterial diversity corresponds to a greater fraction of retrieved GHs.

However, a Kendall Tau analysis showed a strong negative correlation between bacterial abundance and the fraction of glycoside hydrolases (GHs) in the samples (**SI Figure 2**). This indicates that bacterial abundance and richness are not correlated with the frequency of CAZy enzymes in these data. Several groups of GHs were identified, related to the degradation of different carbohydrate polymers. Chitinases (GH18, GH19, and GH20) and cellulases were detected in all samples, with the highest number of GH18 enzymes observed in sample MsA3. In addition to GH18, MsA3 also contained large numbers of GH13, GH31, and GH38 enzymes. The most abundant GH family overall was GH13, which belongs to the alpha-amylases—one of the largest glycoside hydrolase families [47]. The abundance of chitinase enzymes was the highest in the sample locations MsA3, DA4 and MA1 (**Figure 3B**) with *Aeromonas* regrowth which is related to the abundance of *Asellus aquaticus* a large invertebrate with a chitin exoskeleton. Samples GNA5, HNA7, NA2 and PNA6 show a similar profile of GH13, GH23, GH5, GH9, GH24 and GH84. GH5 and GH9 families are known to be involved in cellulose degradation, GH23 and GH24 are related to peptidoglycan degradation and putatively chitin (e.g. GH23) and GH84 is a hexosaminidase with a broad spectrum of functions (**Figure 3A**).

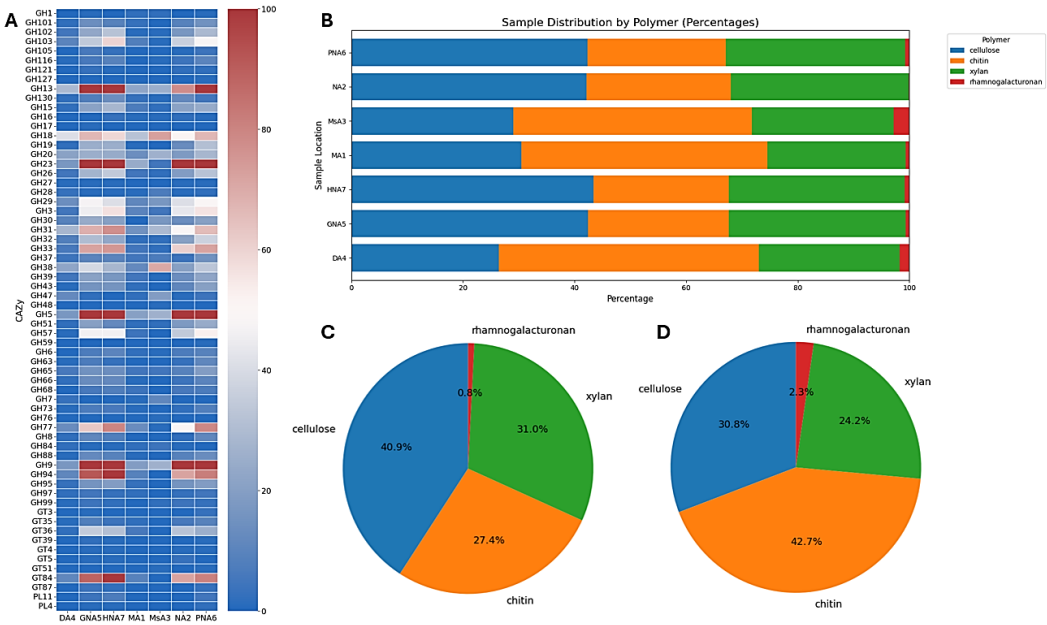


Figure 3. A. Frequency of CAZy enzymes detected in the whole metagenome sequencing data of the sediment community present in the seven different sampling locations. The color gradient is capped to be between 0 and 100. **B**. Fraction of CAZy enzymes (represented as percentage) involved in the degradation of cellulose, chitin, xylan and rhamnogalacturonan across all seven sampling locations. **C**. Fraction of the number of identified CAZy related to the degradation of selected biopolymers, across the seven sampling locations. **D**. Fraction of the relative abundance of the CAZy (calculated as the number of genes encoding for a GH corresponding to the degradation of chitin, cellulose, xylan or rhamnogalacturonan divided by the number of all the genes encoding for all GHs present across all samples) related to the degradation of selected biopolymers across the seven sampling locations.

Streptomyces, the most abundant genus in most loose deposits samples (according to the number of reads) possesses a rich repertoire of GHs. This genus has been also identified to alter the organoleptic properties of the drinking water when present in the drinking water reservoir by producing the metabolite geosmin [48]. The presence of GH13, GH18, GH20, GH9 and GH5 genes indicates the potential to degrade a wide range of polymers including chitin and cellulose, and indeed, members of the genus *Streptomyces* have been found to degrade and utilize chitin and cellulose [49- 52]. The second most abundant genus, *Bradyrhizobium*, possesses the potential to degrade cellulose [53]. Finally, also *Pseudomonas*

was found to be very abundant across all sampling locations, which possesses multiple CAZy enzymes related to degradation of biopolymers [54-56].

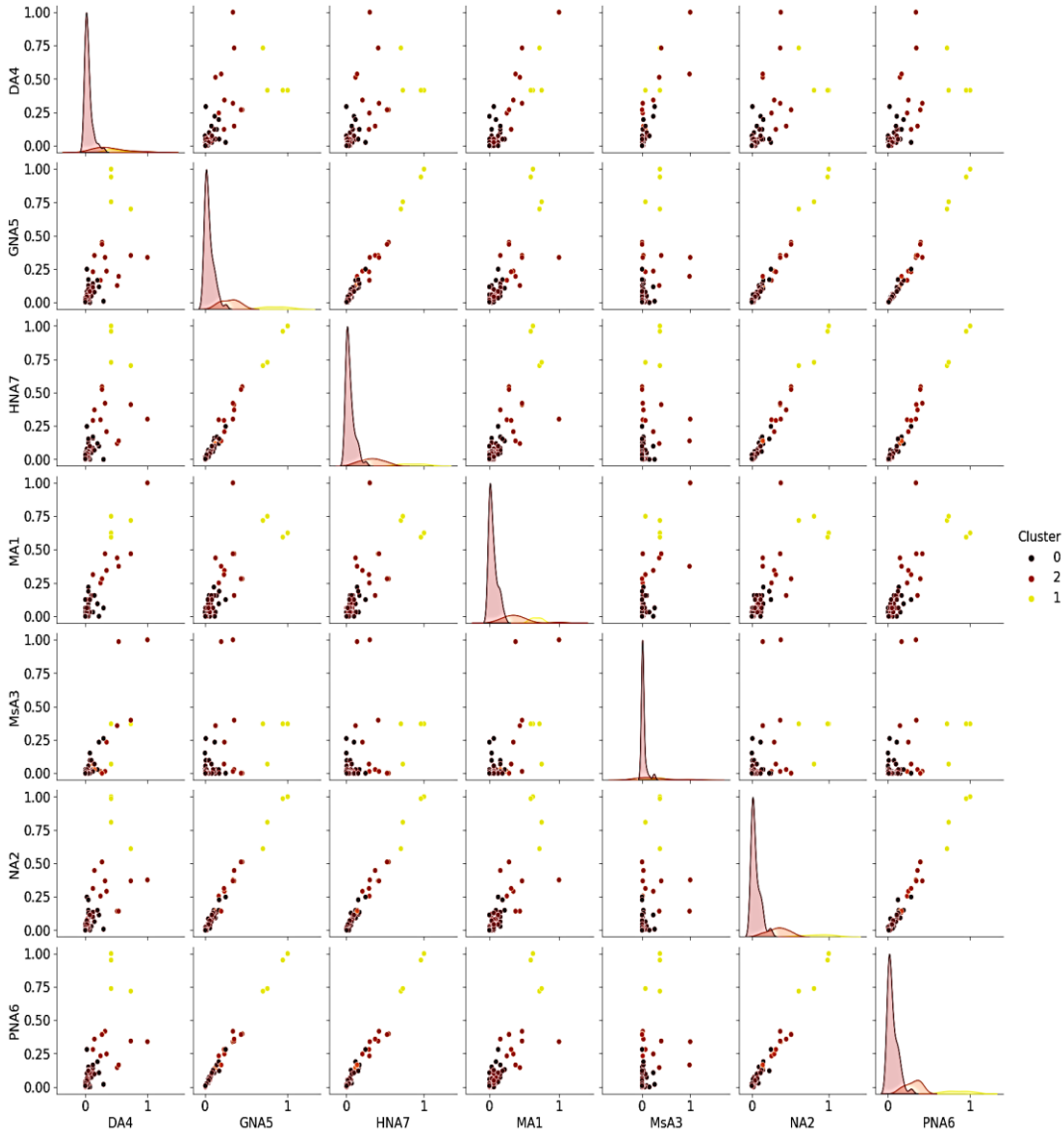


Figure 4: K-means clustering of glycoside hydrolase (GH) abundance across all seven sampling locations. For each location, the abundance of each GH family was calculated by dividing the count of a specific GH by the total number of identified genes encoding GHs in that sample. Using the Elbow method, the optimal number of clusters was determined to be three. The clusters appear to be fairly distinct although some overlap is observed. Cluster 1 provided the most distinctive cluster across all sample types, while clusters 0 and 2, overlaps are observed depending on the sample pair that is compared, The GH abundance was calculated as the number of each GH divided by the total

number of discovered GHs. The diagonal plots display kernel density estimates (KDE) for individual features while the scatterplots in the off-diagonal positions show pairwise relationships between different features. This result shows the degree of similarity between sample types at the level of identified glycoside hydrolases. The analysis of samples HNA7 and PNA6 show some overlap in groups 0 and 2, comparing the samples MsA3 and HNA7, all 3 clusters are more distinctively separated. At the same time sample MsA3 and DA4 show overlapping in clusters 0 and 2.

Across all sampling locations, the frequency and types of glycoside hydrolases varied, resulting in distinct clustering patterns when comparing the samples to one another. Samples HNA7, GNA5, PNA6, and NA2 (which also clustered based on the microbial diversity data) do not form clear clusters based on the glycoside hydrolases, and some overlap occurs. However, for MsA3, MA1, and DA4 (that also clustering based on microbial diversity data) stand out as clearly delineated when compared to the other samples. This indicates that not only the types of glycoside hydrolases present but also their relative abundance play an important role in identifying similarities between different sample types. We examined the distribution and prevalence of glycoside hydrolases in each sample to assess the potential for polymer degradation at each location. Four common biopolymers were selected for this analysis: cellulose, chitin, xylan, and rhamnogalacturonan. Cellulose and chitin are the two most abundant biopolymers in nature, while xylan is a plant cell wall-derived polymer commonly found in terrestrial environments. It is important to note that xylan degradation depends on its source and degree of solubility. Xylans from different plants may require distinct combinations of hydrolytic and non-hydrolytic enzymes for effective breakdown. Rhamnogalacturonan, another plant cell wall-derived polymer, is highly complex and often results from the degradation of pectin in terrestrial ecosystems.
The number of CAZy enzymes related to the degradation of each specified polymer is shown in **Figure 4**. The CAZy enzymes associated with cellulose degradation show the highest frequency, followed by the ones related to chitinolytic activity. The CAZy enzymes with the least frequency are related to degradation of rhamnogalacturonan. Exceptions from this trend were observed in samples MsA3, MA1 and DA4, where the glycoside hydrolases related to chitin degradation were the most frequent, indicating a high potential for chitin degradation. The number of enzymes associated with polymer degradation does not directly correspond to their relative abundance. Among all samples, cellulose-degrading enzymes are the most prevalent, accounting for 40.9% of the total, followed by chitin-degrading enzymes at 27.4%. However, in terms of relative abundance, chitin-degrading enzymes dominate at 42.7%, while cellulose-degrading enzymes represent 30.8% (**Figure 3C, D**). Across all sampling locations, a diverse range of bacterial genera exhibit the potential to degrade at least one polymer. Notably, more than two-thirds of bacteria can degrade multiple polymers, highlighting the functional versatility of microbial communities in the drinking water environment for biopolymers. Most bacteria

possess enzymes involved in the breakdown of cellulose and xylan, whereas rhamnogalacturonan is the least degradable polymer, with only five bacterial genera, *Yersinia*, *Cupriavidus*, *Acinetobacter*, *Bacteroides*, and *Burkholderia*, potentially capable of degrading such a polymer (**Figure 5**). For most bacterial genera have the potential to degrade at least one biopolymer, but the majority can likely degrade more than one biopolymer (**Figure 5**).

Finally, we wanted to investigate the protein biomass of the microbial community found in the sediments of the seven different locations. Interestingly, metaproteomic analysis revealed next to a large spectrum of bacteria, for many samples an abundant profile of proteins with non-bacterial origin (**Figure 6**). Metaproteomics detected proteins from plants, animals, protists, and fungi, which likely originate from the source water. In samples GNA5, PNA6, NA2, and HNA7, the largest fraction of proteins was derived from bacteria, with a large fraction of proteins remaining unassigned (**Figure 6B**). However, in samples MA1, MsA3, and NA2, animal-related proteins were the most abundant (**Figure 6B**). Although trypsin, introduced during sample preparation, was among the abundant proteins detected in MA1, neither trypsin nor albumin were identified in DA4 and MsA3, suggesting a different origin for the animal-related proteins in those samples.

Among the non-bacterial proteins, also proteins related to protists and more specifically amoeba were detected. Amoeba are known to be present in these environments, grazing on the biofilms, ingesting bacteria. Among these bacteria, some pathogenic ones have been detected to survive and proliferate inside an amoeba host [57, 58]. The close relationship between bacteria and amoeba in the drinking water has been described earlier [59]. It's worth mentioning that non-microbial species (in particular invertebrates) are not well represented in the employed database. As a result, proteins from these species may be assigned to related organisms, rather than the ones actually present. However, the high amount of non-bacterial biomass observed in several samples appears to be genuine.

Samples GNA5, PNA6 and HNA7 show similar trend in taxa composition compared with other samples **SI Figure 4**. However, regarding bacterial diversity, the Shannon diversity index shows high similarity between the metagenomic and the metaproteomic results (**Figure 2B** and **Figure 6D**). The samples MA1, DA4 and MsA3 show lower bacterial diversity than the other samples (**Figure 6C**). As in the case of the metagenomic results, *Bradyrhizobium* and *Pseudomonas* are abundant in samples GNA5, PNA6 and HNA7. Surprisingly, in the case of MsA3, *Bacillus* is the most abundant followed by *Clotridioides*, while *Hydrogenobacter* is highly abundant in MA1 (**Figure 6A**). However, there is a consensus regarding the differences and similarities in bacterial taxonomies between different samples for both omics datasets. Samples HNA7, PNA6, GNA5 and NA2 have relatively similar taxonomic composition and abundance. On the other hand, samples

MsA3, DA4 and MA1 are different from the other group and even different from each other.

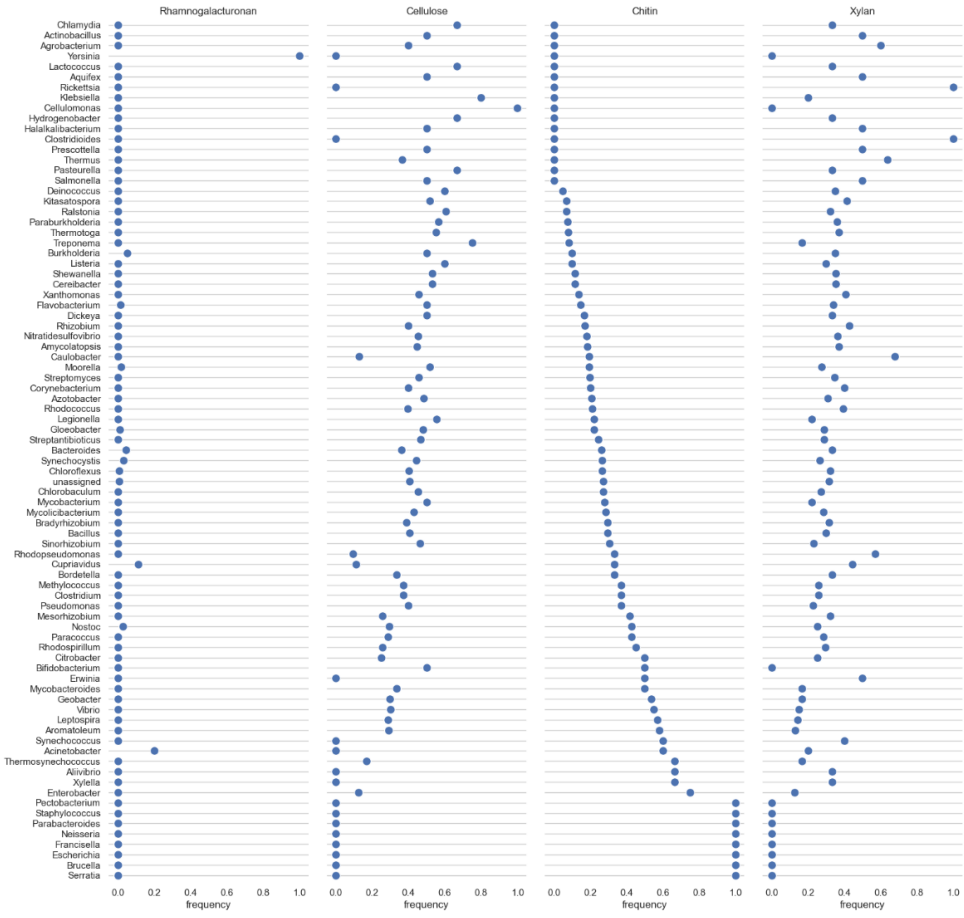


Figure 5: Detection of CAZy genes related to the degradation pathways for biopolymers chitin, cellulose, xylan and rhamno- galacturonan, by genus. The frequency corresponds to the number of CAZy genes per polymer degradation pathway and genus, relative to the total CAZy genes for all polymer degradation pathways and all genera. Interestingly, many genera show the potential to degrade 2 or more biopolymers.

The presence of certain Archaea may indicate turnover of organic and inorganic compounds, and the occurrence of anaerobic/anoxic conditions in these loose deposits despite the obvious oxic conditions in the bulk water. The presence of *Sulfolobus* indicates events of sulfur oxidation while the presence of *Methanothermobacter* [60], *Methanosarcina* [61], *Methanococcus* [62] and *Methanocaldococcus* [63] indicates the formation

of methane from carbon dioxide and hydrogen. *Nitrosopumilus* can oxidize ammonia in anoxic environments contributing to the nitrogen cycle [64].

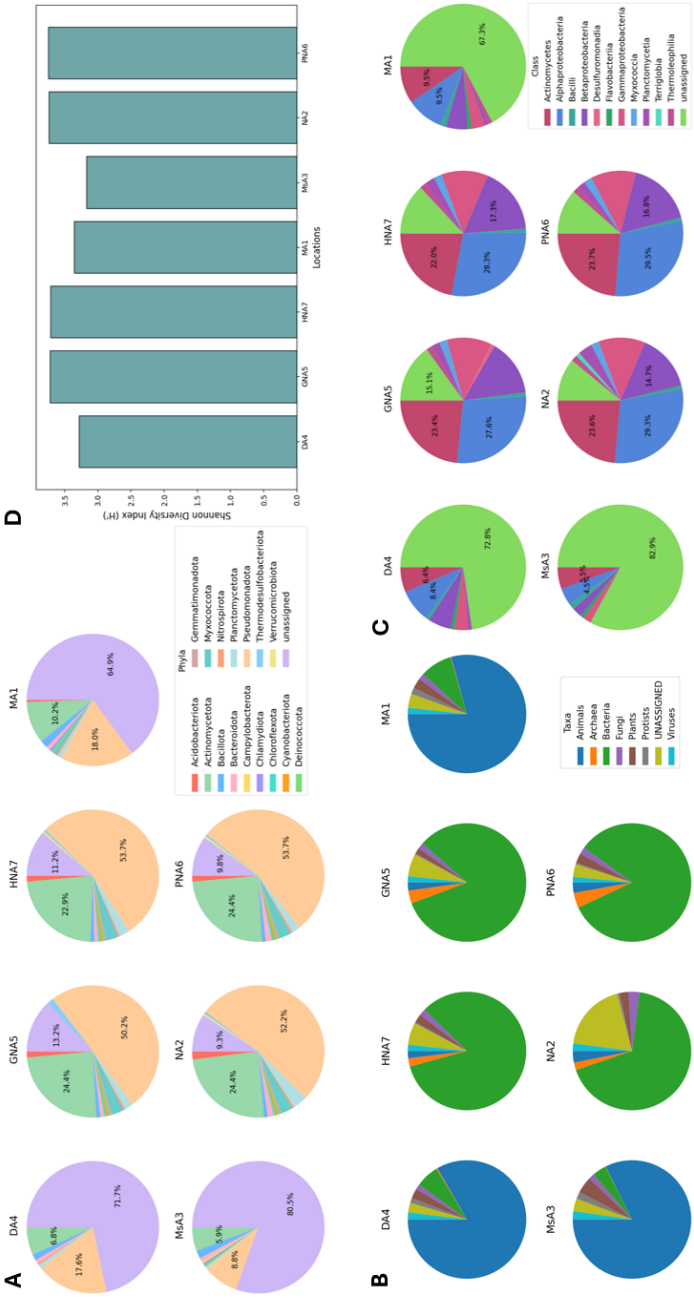


Figure 6: Community diversity based on metaproteomics and metagenomics. **A**. Taxonomic composition at the phylum level for all seven sampling locations based on read counts obtained from whole metagenome sequencing data. **B**. Taxonomic composition in all seven sampling locations based on protein counts from the metaproteomics data. **C.** Taxonomic composition at class level for all seven sampling locations based on read counts obtained from whole metagenome sequencing data. Only classes with more >1 count were taken into consideration for this graph. **D.** Shannon diversity index at genus level as determined based on protein counts from the metaproteomics data.

For *Proteobacteria*, the most abundant class is *Alphaproteobacteria*, followed by *Betaproteobacteria* and *Gammaproteobacteria*. Surprisingly, *Deltaproteobacteria* were not detected in any of the samples, although its presence in the drinking water has been confirmed in earlier studies [21, 65, 23]. However, a large fraction of reads in samples MA1, MsA3 and DA4 could not be assigned to a phylum (or other taxonomy).

Discussion and Conclusion

In the presented study the loose deposits were obtained from a drinking water distribution system that does not use residual chlorine. Most studies on drinking water microbiomes are associated with chlorinated drinking water. Chlorination widely impacts the drinking water community by selecting specific microbes. The lack of residual chlorine may select for a different bacterial community. The loose deposits (sediments in our case) contain a diverse and active bacterial community [21], which is indicative for the diversity present in the drinking water system. Bacteria present in drinking water sediments encompass a wide range of members originating from both aquatic and terrestrial ecosystems. As drinking water pipes are installed in the terrestrial environment, the pipes may be influenced by this environment during construction as well as during maintenance. Loose deposits in the DWDS harbor microorganisms from both sources. In addition, the sediments likely contain a diverse array of polymeric carbon sources, including cellulose and chitin.

Microbial ecology. Based on the CAZy genes identified at the different locations, polymers such as chitin, cellulose, and even more complex biopolymers can be degraded and utilized by the microbial communities. The three most abundant genera—*Streptomyces*, *Bradyrhizobium*, and Pseudomonas—are all known to possess enzymes involved in the breakdown of complex biopolymers (**Figure 5**). The metabolic versatility of these bacteria, with the ability to degrade multiple types of polymers, provides a clear competitive advantage, particularly in low-nutrient environments such as drinking water distribution systems.

Degradation of complex polymers in the environment by either specialists or generalists helps in maintaining a complex bacterial community that can thrive even in so called oligotrophic environments. In such environments the ability to recycle any of the (in)organic compounds from the accumulated biomass is essential to survival. At all locations, enzymes required for the degradation of chitin were detected in many microbial taxa, indicating a general potential to utilize this polymer as both a carbon and nitrogen source. Similarly, enzymes involved in cellulose degradation were also consistently observed, suggesting that cellulose is another readily available nutrient source for these microbial communities. Nevertheless, we found also indications for genes that encode for enzymes which are involved in the degradation of even more complex carbohydrate polymers, such as xylan, laminarin, pectin and pullulan.
Finally, if a more abundant or uniform carbon sources (e.g. biopolymer) becomes available, the increased regrowth potential could reduce overall microbial diversity and alter the ecosystem.

Methodology. The differences observed in microbial abundance and diversity between metagenomics and metaproteomics likely reflect the fundamental differences in how these methods capture and represent microbial communities (**Figure 2A** and **SI Figure 4**). For example, this has been recently investigated for metagenomic, 16S rRNA amplicon sequencing and metaproteomics [66]. Metagenomics usually provides a higher sensitivity toward lower abundant microbes, but it cannot discriminate between active and dormant cells or cells that may have a slow growth rate. On the other hand, active cells present in an environment are expected to produce more proteins and therefore will be well covered by metaproteomics studies. Also, some microbes produce larger cells, and therefore contain more protein than others, which makes them appear more abundant in the metaproteomics data. Furthermore, the chemistry of DNA and protein extraction from drinking water sediments is usually cumbersome due to the low amount of biological material and the interferences coming from salts and humic acids present. This further translates into low protein coverage or low taxonomic resolution. Another important factor to take into consideration are the cell lysis protocols, as these may be differently effective for different types microbes. For example, Gram+ and Gram- bacteria have different types of cell walls, harder to break for Gram+ bacteria which lead to low homogenization. This type of bias has been observed in the past for metagenomic sample preparation [67, 68], which can result in an underrepresentation of Gram+ bacteria in samples. The metaproteomic sample preparation involved harsher conditions, which may have reduced differences in protein extraction efficiency between different bacterial taxa. Finally, the choice of reference database for taxonomic classification of metagenomic and metaproteomic data has a large impact on the reported taxonomic composition. Smaller databases like GTDB may prove to be useful for fast annotation and lower the number of false positive or erroneous identifications. However, taxonomic resolution may be

impacted by the lack of coverage. On the other hand, the use of larger reference sequence databases may include high sequence redundancy, or poorly annotated data. Nevertheless, also the strategy of whether short-read or long read sequencing, or reads, contig or MAG based taxonomic classification is used, can significantly impact the accuracy of the taxonomic classification [69-71].

Regrowth assessment. The abundance differences between DNA based methods and metaproteomics have been observed in previous studies [66]. Nevertheless, as observed in the current study and also described in other studies, *Aeromonas* is only one of hundreds of microbes present in the drinking water distribution system and the culture-based assessment showed its low abundancy. This environment is a surprisingly complex ecosystem where *Aeromonas* is by far not the most abundant or prevalent bacterium. Apart from several studies where *Aeromonas* was detected in the drinking water [72], it was hardly detected with culture-independent techniques like 16S rRNA amplicon sequencing and even flow cytometry [15, 23]. In the current study – where we analyzed the loose deposits – *Aeromonas* was only detected in trace amounts in the whole metagenome sequencing data, and it was absent in the metaproteomic data. This did not allow to perform a more accurate quantification of the relative abundance of this microbe in this ecosystem. A similar observation was made in a previous study, where *Aeromonas* was not detect with 16S rRNA sequencing in unchlorinated drinking water [23]. The culture-based microbial analysis such as HPC and *Aeromonas* are highly selective but may also be not perfectly quantitative because bacteria are cultured on nutrient rich media. Thus, bacterial regrowth potential assessment as regulated in the Dutch Drinking Water Decree is based on detecting microbes growing under relatively nutrient rich conditions.

Regrowth and invertebrates. A recent study presented substantial data on the invertebrates in the loose deposits of a number of DWDS among which also the ones from the current study [30]. This study showed a good correlation with invertebrate biomass concentration with *A. aquaticus* as the dominant species and *Aeromonas* regrowth. Furthermore, elevated HPC and *Aeromonas* counts in the bulk water have been found in association with the abundance of the biomass of the large invertebrate *Asellus aquaticus* in the loose deposits [6, 73]. Moreover, *A. aquaticus* biomass contains high numbers of *Aeromonas* [6, 74]. Thus, where the dominant microbial populations in the DWDS are related to the more general ecological characteristics, the more selective microbial methods indicate the presence of selective nutrient rich environments. Consequently, the important ecological question is the association of *A. aquaticus* with the other food web contributors. This large invertebrate does not use the biopolymers directly from drinking water for their survival, but it can use the organic matter accumulated in loose deposits and biofilms from the DWDS for this purpose [13]. A current hypothesis says that biopolymers in the supplied drinking water is used for growth and maintenance of biofilms in loose deposits and biofilms [6], which is confirmed in the current study. *A. aquaticus*

populations in the DWDS feed on these biofilms for their survival and reproduction. The higher chitin enzymatic activity in the samples from *Aeromonas* regrowth conditions (**Figure 3B**) is an indication for the presence of *A. aquaticus* with chitin as part of its exoskeleton in these loose deposit samples from TP#1 and TP#3 and the absence of this invertebrate in the loose deposits from TP#2 as also observed by Hijnen *et al*. (2024) [6]. Though, chitinase was also observed in the latter samples, which shows that this biopolymer is also present in other and smaller invertebrates of the food web in the more oligotrophic environments.

Another example of bacterial association with invertebrates is the bacterium Legionella, which was detected in both metagenomic and metaproteomic data. It has been many times reported that several free-living amoeba (FLA) can harbor Legionella being part of their growth cycle. In DWDS with disinfectants this association offers protection against these disinfectants in the drinking water (75) and therefore can be observed with these techniques in these lower bacterial abundancy conditions due to the disinfection.

In the current study we did not detect *Asellus* related proteins with metaproteomics in the loose deposits, most likely because only 100 µm loose deposits were investigated. This fraction was chosen, because in this fraction most of the loose deposits including the invertebrate biomass showed a good homogeneity across all samples. In this 100 µm fraction we expect to find also major microbial communities related to the loose deposits, albeit missing some of the smaller invertebrates. This also enhanced the comparability of the data between the different sampling locations. However, Asellus biomass is usually only found in the larger fraction of 500 µm. Moreover, the presence and possible interaction between bacteria and larger organisms in the drinking water system is of crucial importance because these larger invertebrates have a dominant impact on the invertebrate biomass concentration in the DWDS. Not only may larger organisms provide shelter to microorganisms against environmental factors, helping the bacteria to overcome water treatment procedures, but they can also provide a food source for the bacteria helping them to proliferate as demonstrated by elevated HPC and *Aeromonas* counts in the bulk water [6, 73]. Finally, the total amount of obtained deposits and biomass (mg/m^3) per ample location should be included in the data interpretation in follow-up studies.

Biopolymers and ecology. The role of biopolymer concentrations in drinking water for biological stability of drinking water distribution was introduced before [5, 6, 76, 77]. The sources of biopolymers that can serve as nutrient are likely diverse. The presence of fungi in drinking water suggests an input of chitin and chitosan, as does the occurrence of insects such as coleopterans. Although not identified in the current study, crustaceans like *Asellus*, which have been observed in higher sediment fractions, are also potential sources of chitin due to their exoskeletons [78]. Cellulose and xylans, on the other hand, may come from plant material which may be present in the drinking water. *Sphingomonas* is a

bacterial genus commonly found in the drinking water system, especially in the fraction of loose deposits and suspended solids [21]. However, by metagenomics it is not the most abundant and in the metaproteomic data this genus was not observed. The presence of *Streptomyces* as the most abundant genus may influence the structure of the overall bacterial community. This genus is known for producing a wide array of antibiotic compounds, the most important being streptomycin [79] that can help in competing against other bacteria and only those that are resistant may have a chance to survive. Interestingly the second most abundant genera were *Bradyrhizobium* and *Pseudomonas*, both known to be resistant to antibiotics including streptomycin [80-83]. Nevertheless, also the drinking water disinfection method impacts the bacterial community. Since in The Netherlands there is no residual disinfectants released in the system, this allows to Archaea to survive in the drinking water [84]. Currently several Archaea genera were discovered to inhabit the drinking water system. Phylogenetic composition of the drinking water samples presented in this study is comparable with other studies. The abundance of *Proteobacteria* is similar to other sampled drinking water. In all sampled locations *Pseudomonadota* (*Proteobacteria*) is the most abundant. In sample PNA6, NA2, GNA5 and HNA7, Proteobacteria account for over 50% of all the phyla and is comparable to other studies [65]. Again, samples MA1, MsA3 and DA4 have a relatively different composition that the other samples at both phylum and class levels. The occurrence of *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* is comparable to other studies [21, 65] and they are usually dominating the drinking water system [85-87]. Overall, in the current study, the top three most abundant bacterial classes were *Alphaproteobacteria*, *Betaproteobacteria* and *Actinomycetes*. The presence of *Actinomycetes* in the top three most abundant classes has not been reported before. The presence of this bacterial class indicates possible degradation of the drinking water quality as it has been mentioned before [88]. These results are relatively different compared to previous studies that were performed on the bacterial community in unchlorinated drinking water systems. It is important to mention that previous studies were performed using low resolution and more targeted approaches such as 454 pyrosequencing or 16S rRNA amplicon sequencing susceptible to biases [89].

In summary, this study is the first to integrate metagenomic and metaproteomic analyses to investigate the microbes present in the sediments of a drinking water distribution system. The metagenomic data provided valuable insights into the overall taxonomic diversity and the metabolic potential for degrading a broad spectrum of biopolymers. Additionally, the metagenomic sequences served as a reference database for annotating proteins identified through metaproteomic analysis. The metaproteomic data, in turn, offered a detailed view of the protein biomass composition across the different sampling locations.

Acknowledgements: We would like to thank our colleagues from the department of biotechnology and the environment biotechnology section for valuable discussions and Dita Heikens from the mass spec facility for support in the lab. We appreciate the Hologenomix team, especially David Calderón and Eric van der Toorn, for their support with metagenomic sequencing and data processing. We would like to offer our special thanks to Evides Waterbedrijf for supporting this study and providing valuable feedback and samples. In particular, we are grateful to Julia Wurth, Leonie Marang, Nefs Roos and Professor Bert van der Wal for their contributions and collaboration. ChatGPT was used to assist with language editing and proofreading.

Competing interests: The authors declare no competing interests.

References

1. Van der Kooij D, Hein J, Van Lieverloo M, Schellart J, Hiemstra P. Maintaining quality without a disinfectant residual. Journal-American Water Works Association. 1999;91(1):55-64.
2. Escobar IC, Randall AA, Taylor JS. Bacterial Growth in Distribution Systems: Effect of Assimilable Organic Carbon and Biodegradable Dissolved Organic Carbon. Environmental Science & Technology. 2001;35(17):3442-7.
3. Riyadh A, Peleato N. Natural Organic Matter Character in Drinking Water Distribution Systems: A Review of Impacts on Water Quality and Characterization Techniques. Water. 2024;16:446.
4. van Lieverloo JHM, Hoogenboezem W, Veenendaal G, van der Kooij D. Variability of invertebrate abundance in drinking water distribution systems in the Netherlands in relation to biostability and sediment volumes. Water Research. 2012;46(16):4918-32.
5. Schurer R, Hijnen W, Van Der Wal A. The significance of the biomass subfraction of high-MW organic carbon for the microbial growth and maintenance potential of disinfectant-free drinking water produced from surface water. Water Research. 2022;209:117898.
6. Hijnen WA, Brouwer-Hanzens A, Schurer R, Wagenvoort AJ, van Lieverloo JHM, van der Wielen PW. Influence of biopolymers, iron, biofouling and Asellus aquaticus on Aeromonas regrowth in three non-chlorinated drinking water distribution systems. Journal of Water Process Engineering. 2024;61:105293.
7. Schurer R, Brouwer-Hanzens A, van der Wielen P, van Lieverloo J, Hijnen W. Quantification of high molecular weight organic carbon concentrations with LC-

OCD and PHMOC for biological stability investigation of drinking water produced from surface water. Water Research. 2025;271:122971.
8. Zhou Y, Li X, Zhou Z, Feng J, Sun Y, Ren J, et al. New insights into biopolymers: In situ collection and reuse for coagulation aiding in drinking water treatment plants and microbial mechanism. Separation and Purification Technology. 2024;337:126448.
9. Sack EL, van der Wielen PW, van der Kooij D. Polysaccharides and proteins added to flowing drinking water at microgram-per-liter levels promote the formation of biofilms predominated by bacteroidetes and proteobacteria. Appl Environ Microbiol. 2014;80(8):2360-71.
10. J. Holt NK, P. Sneath and J. Staley , Williams SJ. Genus Aeromonas1994. 190-1 p.
11. Janda JM, Abbott SL. Evolving concepts regarding the genus Aeromonas: an expanding Panorama of species, disease presentations, and unanswered questions. Clin Infect Dis. 1998;27(2):332-44.
12. van Bel N, van der Wielen P, Wullings B, van Rijn J, van der Mark E, Ketelaars H, et al. Aeromonas species from non-chlorinated distribution systems and their competitive planktonic growth in drinking water. Appl Environ Microbiol. 2021;87(5).
13. van Bel N, van Lieverloo JHM, Verschoor AM, Pap-Veldhuizen L, Hijnen WA, Peeters ET, et al. Survival and Growth of A. aquaticus on Different Food Sources from Drinking Water Distribution Systems. Arthropoda (2813-3323). 2024;2(3).
14. Tugui CG, Sorokin DY, Hijnen W, Wunderer J, Bout K, van Loosdrecht MCM, et al. Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin. RSC Chemical Biology. 2025;6(2):227-39.
15. Liu G, Van der Mark EJ, Verberk JQ, Van Dijk JC. Flow cytometry total cell counts: a field study assessing microbiological water quality and growth in unchlorinated drinking water distribution systems. Biomed Res Int. 2013;2013:595872.
16. Shen L, Zhang Z, Wang R, Wu S, Wang Y, Fu S. Metatranscriptomic data mining together with microfluidic card uncovered the potential pathogens and seasonal RNA viral ecology in a drinking water source. J Appl Microbiol. 2024;135(1).
17. Safford HR, Bischel HN. Flow cytometry applications in water treatment, distribution, and reuse: A review. Water Res. 2019;151:110-33.
18. Van der Wielen PW, van der Kooij D. Effect of water composition, distance and season on the adenosine triphosphate concentration in unchlorinated drinking water in the Netherlands. Water Research. 2010;44(17):4860-7.
19. Van Der Wielen PW, Lut MCJWS, Supply TW. Distribution of microbial activity and specific microorganisms across sediment size fractions and pipe wall biofilm in a drinking water distribution system. 2016;16(4):896-904.

20. Liu G, Ling FQ, van der Mark EJ, Zhang XD, Knezev A, Verberk JQJC, et al. Comparison of Particle-Associated Bacteria from a Drinking Water Treatment Plant and Distribution Reservoirs with Different Water Sources. Scientific Reports. 2016;6(1):20367.
21. Liu G, Bakker GL, Li S, Vreeburg JH, Verberk JQ, Medema GJ, et al. Pyrosequencing reveals bacterial communities in unchlorinated drinking water distribution system: an integral study of bulk water, suspended solids, loose deposits, and pipe wall biofilm. Environ Sci Technol. 2014;48(10):5467-76.
22. Brumfield KD, Hasan NA, Leddy MB, Cotruvo JA, Rashed SM, Colwell RR, et al. A comparative analysis of drinking water employing metagenomics. PLoS One. 2020;15(4):e0231210.
23. Vavourakis CD, Heijnen L, Peters M, Marang L, Ketelaars HAM, Hijnen WAM. Spatial and Temporal Dynamics in Attached and Suspended Bacterial Communities in Three Drinking Water Distribution Systems with Variable Biological Stability. Environ Sci Technol. 2020;54(22):14535-46.
24. Muth T, Benndorf D, Reichl U, Rapp E, Martens L. Searching for a needle in a stack of needles: challenges in metaproteomics data analysis. Molecular BioSystems. 2013;9(4):578-85.
25. Beach NK, Myers KS, Donohue TJ, Noguera DR. Metagenomes from 25 Low-Abundance Microbes in a Partial Nitritation Anammox Microbiome. 2022;11(6):e00212-22.
26. Jin H, You L, Zhao F, Li S, Ma T, Kwok LY, et al. Hybrid, ultra-deep metagenomic sequencing enables genomic and functional characterization of low-abundance species in the human gut microbiome. Gut microbes. 2022;14(1):2021790.
27. Kumar GC, Chaudhary J, Meena LK, Meena AL, Kumar A. 18 - Function-driven microbial genomics for ecofriendly agriculture. In: Singh JS, Tiwari S, Singh C, Singh AK, editors. Microbes in Land Use Change Management: Elsevier; 2021. p. 389-431.
28. Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass contributions in microbial communities via metaproteomics. Nature Communications. 2017;8(1):1558.
29. Shrestha HK, Appidi MR, Villalobos Solis MI, Wang J, Carper DL, Burdick L, et al. Metaproteomics reveals insights into microbial structure, interactions, and dynamic regulation in defined communities as they respond to environmental disturbance. BMC Microbiology. 2021;21(1):308.
30. Ketelaars HAM, Wagenvoort AJ, Peters M, Wunderer J, Hijnen WAM. Taxonomic diversity and biomass of the invertebrate fauna of nine drinking water treatment plants and their non-chlorinated distribution systems. Water Res. 2023;242:120269.

31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.
32. Calderón-Franco D, Sarelse R, Christou S, Pronk M, van Loosdrecht MCM, Abeel T, et al. Metagenomic profiling and transfer dynamics of antibiotic resistance determinants in a full-scale granular sludge wastewater treatment plant. Water Research. 2022;219:118571.
33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.
34. Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics. 2020;36(4):1303-4.
35. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.
36. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27(5):824-34.
37. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.
38. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11(1):119.
39. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825-9.
40. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biology. 2019;20(1):217.
41. Kleikamp HB, Grouzdev D, Schaasberg P, van Valderen R, van der Zwaan R, van de Wijgaart R, et al. Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. Water research. 2023;246:120700.
42. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428(4):726-31.
43. Kostanjšek R, Štrus J, Drobne D, Avguštin G. 'Candidatus Rhabdochlamydia porcellionis', an intracellular bacterium from the hepatopancreas of the terrestrial isopod Porcellio scaber (Crustacea: Isopoda). 2004;54(2):543-9.
44. Van der Wielen P, Italiaander R, Wullings B, Heijnen L, Van der Kooij D. Opportunistic pathogens in drinking water in the Netherlands. Microbial growth

in drinking-water supplies problems, causes, control and research needs. 2014:177-205.
45. Beaman BL, Beaman L. Nocardia species: host-parasite relationships. Clin Microbiol Rev. 1994;7(2):213-64.
46. Dowdell K, Haig SJ, Caverly LJ, Shen Y, LiPuma JJ, Raskin L. Nontuberculous mycobacteria in drinking water systems - the challenges of characterization and risk mitigation. Curr Opin Biotechnol. 2019;57:127-36.
47. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280 (Pt 2)(Pt 2):309-16.
48. Lylloff J, Mogensen MH, Burford M, Schlüter L, Jørgensen N. Detection of aquatic streptomycetes by quantitative PCR for prediction of taste-and-odour episodes in water reservoirs. Journal of Water Supply: Research and Technology-AQUA. 2012;61:272-82.
49. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Wendt-Pienkowski E, Doering DT, et al. Evolution of High Cellulolytic Activity in Symbiotic Streptomyces through Selection of Expanded Gene Content and Coordinated Gene Expression. PLoS Biol. 2016;14(6):e1002475.
50. Li X, Gao P. Isolation and partial characterization of cellulose-degrading strain of Streptomyces sp. LX from soil. Letters in Applied Microbiology. 1996;22(3):209-13.
51. Wang JL, Chen YC, Deng JJ, Mo ZQ, Zhang MS, Yang ZD, et al. Synergic chitin degradation by Streptomyces sp. SCUT-3 chitinases and their applications in chitinous waste recycling and pathogenic fungi biocontrol. Int J Biol Macromol. 2023;225:987-96.
52. Recognition and degradation of chitin by streptomycetes. Antonie van Leeuwenhoek. 2001;79:285-9.
53. Jimenéz-Zurdo J, Mateos PF, Dazzo FB, Martínez-Molina E. Cell-bound cellulase and polygalacturonase production by Rhizobium and Bradyrhizobium species. Soil Biology and Biochemistry. 1996;28(7):917-21.
54. Folders J, Algra J, Roelofs MS, van Loon LC, Tommassen J, Bitter W. Characterization of Pseudomonas aeruginosa chitinase, a gradually secreted protein. J Bacteriol. 2001;183(24):7044-52.
55. Neiendam Nielsen M, Sørensen J. Chitinolytic activity of Pseudomonas fluorescens isolates from barley and sugar beet rhizosphere. FEMS Microbiology Ecology. 1999;30(3):217-27.
56. Sun S, Zhang Y, Liu K, Chen X, Jiang C, Huang M, et al. Insight into biodegradation of cellulose by psychrotrophic bacterium Pseudomonas sp. LKR-1 from the cold region of China: optimization of cold-active cellulase production and the associated degradation pathways. Cellulose. 2020;27(1):315-33.

57. Delafont V, Brouke A, Bouchon D, Moulin L, Hechard Y. Microbiome of free-living amoebae isolated from drinking water. Water research. 2013;47.
58. Bichai F, Payment P, Barbeau B. Protection of waterborne pathogens by higher organisms in drinking water: a review. Can J Microbiol. 2008;54(7):509-24.
59. Corsaro D, Pages GS, Catalan V, Loret JF, Greub G. Biodiversity of amoebae and amoeba-associated bacteria in water treatment plants. Int J Hyg Environ Health. 2010;213(3):158-66.
60. Zeikus JG, Wolfe RS. Methanobacterium thermoautotrophicus sp. n., an anaerobic, autotrophic, extreme thermophile. J Bacteriol. 1972;109(2):707-15.
61. Welte C, Deppenmeier U. Chapter thirteen - Proton Translocation in Methanogens. In: Rosenzweig AC, Ragsdale SW, editors. Methods in Enzymology. 494: Academic Press; 2011. p. 257-80.
62. Li J, Akinyemi TS, Shao N, Chen C, Dong X, Liu Y, et al. Genetic and metabolic engineering of Methanococcus spp. Current Research in Biotechnology. 2023;5:100115.
63. Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS. Methanococcus jannaschii sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Archives of Microbiology. 1983;136(4):254-61.
64. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature. 2005;437(7058):543-6.
65. Liu G, Ling FQ, Magic-Knezev A, Liu WT, Verberk JQJC, Van Dijk JC. Quantification and identification of particle-associated bacteria in unchlorinated drinking water from three treatment plants by cultivation-independent methods. Water Research. 2013;47(10):3523-33.
66. Kleikamp HBC, Grouzdev D, Schaasberg P, van Valderen R, van der Zwaan R, Wijgaart Rvd, et al. Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. Water Research. 2023;246:120700.
67. Lima A, França A, Muzny CA, Taylor CM, Cerca N. DNA extraction leads to bias in bacterial quantification by qPCR. Appl Microbiol Biotechnol. 2022;106(24):7993-8006.
68. Nearing JT, Comeau AM, Langille MGI. Identifying biases and their potential solutions in human microbiome studies. Microbiome. 2021;9(1):113.
69. Chorlton SD. Ten common issues with reference sequence databases and how to mitigate them. Front Bioinform. 2024;4:1278228.
70. Portik DM, Brown CT, Pierce-Ward NT. Evaluation of taxonomic classification and profiling methods for long-read shotgun metagenomic sequencing datasets. BMC Bioinformatics. 2022;23(1):541.

71. Smith RH, Glendinning L, Walker AW, Watson M. Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. Animal Microbiome. 2022;4(1):57.
72. Blasco MD, Esteve C, Alcaide E. Multiresistant waterborne pathogens isolated from water reservoirs and cooling systems. Journal of Applied Microbiology. 2008;105(2):469-75.
73. Gunkel G, Michels U, Scheideler M. Water lice and other macroinvertebrates in drinking water pipes: Diversity, abundance and health risk. Water. 2021;13(3):276.
74. Mayer M. Zur Ernährungsweise von Isopoden in Trinkwasserverteilungssystemen. 2013.
75. Shaheen M, Scott C, Ashbolt NJ. Long-term persistence of infectious Legionella with free-living amoebae in drinking water biofilms. International Journal of Hygiene and Environmental Health. 2019;222(4):678-86.
76. Hijnen W, Schurer R, Bahlman J, Ketelaars H, Italiaander R, Van Der Wal A, et al. Slowly biodegradable organic compounds impact the biostability of non-chlorinated drinking water produced from surface water. Water Research. 2018;129:240-51.
77. Schurer R, Schippers J, Kennedy M, Cornelissen E, Salinas-Rodriguez S, Hijnen W, et al. Enhancing biological stability of disinfectant-free drinking water by reducing high molecular weight organic compounds with ultrafiltration posttreatment. Water Research. 2019;164:114927.
78. Iber BT, Kasan NA, Torsabo D, Omuwa JW. A Review of Various Sources of Chitin and Chitosan in Nature. Journal of Renewable Materials. 2021;10(4):1097-123.
79. de Lima Procópio RE, da Silva IR, Martins MK, de Azevedo JL, de Araújo JM. Antibiotics produced by Streptomyces. The Brazilian Journal of Infectious Diseases. 2012;16(5):466-71.
80. Ajayi OO, Nwodo DC, Adedeji T, Ogwugwa VH, Dianda M, Fagade OE. Antibiotic Resistance in Nitrogen-Fixing Rhizobial Strains: Implications for Agriculture. 2024;2024(1):9774054.
81. Mueller JG, Skipper HD, Shipe ER, Grimes LW, Wagner SC. Intrinsic antibiotic resistance in Bradyrhizobium japonicum. Soil Biology and Biochemistry. 1988;20(6):879-82.
82. Tseng JT, Bryan LE, Van den Elzen HM. Mechanisms and spectrum of streptomycin resistance in a natural population of Pseudomonas aeruginosa. Antimicrob Agents Chemother. 1972;2(3):136-41.
83. Pang Z, Raudonis R, Glick BR, Lin TJ, Cheng Z. Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies. Biotechnol Adv. 2019;37(1):177-92.

84. Inkinen J, Siponen S, Jayaprakash B, Tiwari A, Hokajärvi AM, Pursiainen A, et al. Diverse and active archaea communities occur in non-disinfected drinking water systems-Less activity revealed in disinfected and hot water systems. Water Res X. 2021;12:100101.
85. Williams MM, Domingo JW, Meckes MC, Kelty CA, Rochon HS. Phylogenetic diversity of drinking water bacteria in a distribution system simulator. J Appl Microbiol. 2004;96(5):954-64.
86. Revetta RP, Pemberton A, Lamendella R, Iker B, Santo Domingo JW. Identification of bacterial populations in drinking water using 16S rRNA-based sequence analyses. Water Res. 2010;44(5):1353-60.
87. Santo Domingo JW, Meckes MC, Simpson JM, Sloss B, Reasoner DJ. Molecular characterization of bacteria inhabiting a water distribution system simulator. Water Sci Technol. 2003;47(5):149-54.
88. Zaitlin B, Watson S. Actinomycetes in relation to taste and odour in drinking water: Myths, tenets and truths. Water research. 2006;40:1741-53.
89. Pinto AJ, Raskin L. PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. PLOS ONE. 2012;7(8):e43093.

Supplementary information material to:

Studying the microbiome of drinking water loose deposits using metagenomics and metaproteomics

TABLE OF CONTENTS

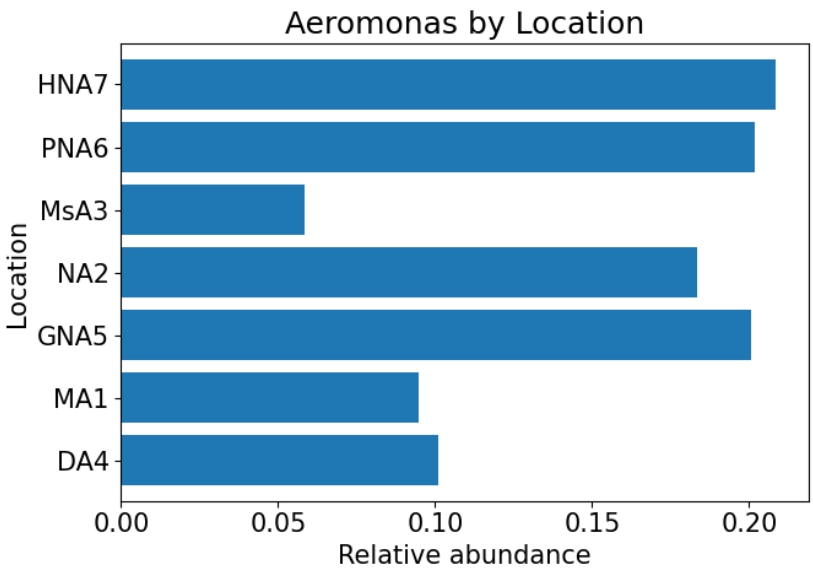


Figure 1. Relative abundance of the reads retrieved for *Aeromonas* across all samples, with a threshold cut of 0.2%

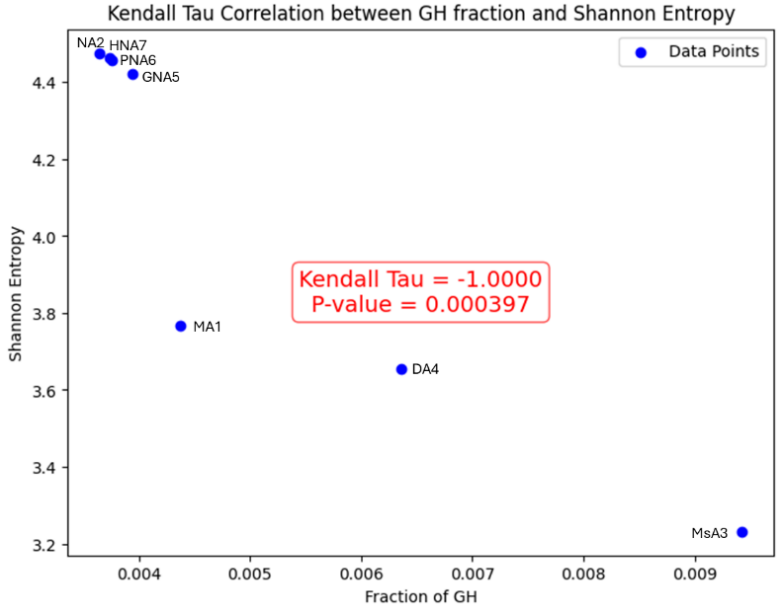


Figure 2. Kendall Tau analysis of the Shannon Entropy and Fraction of GH present in each sampled location. The graphic shows no correlation between the diversity in bacterial genera at a location and the abundance of GH that are present at that respective location

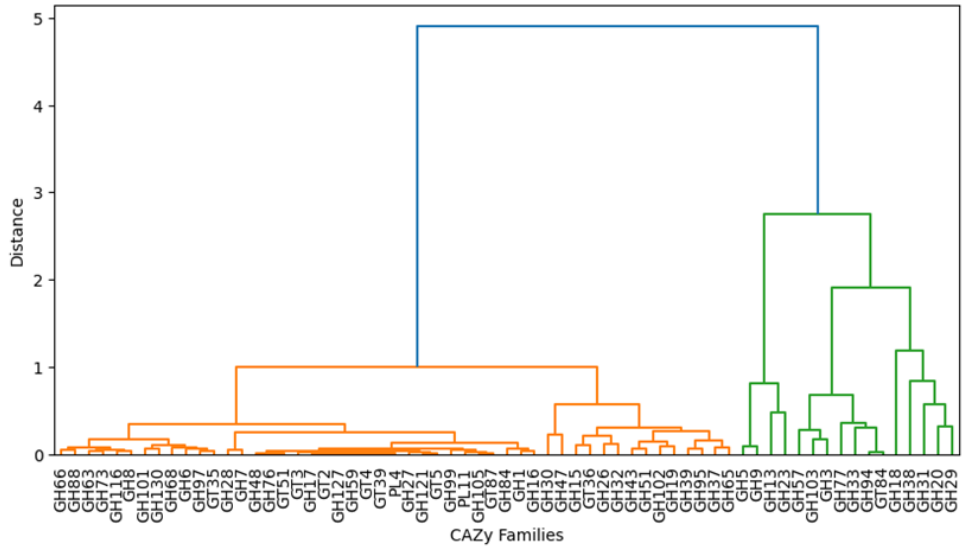


Figure 3. Clustering of the CAZy enzymes across all sample points. The clustering was performed taking into consideration the relative abundance of each enzyme across all sampled locations

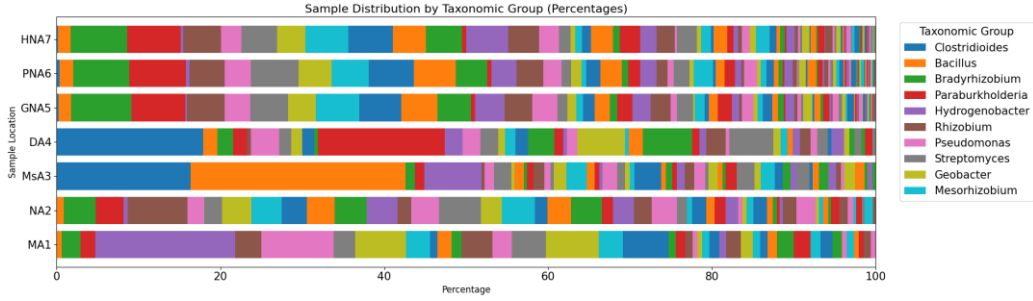
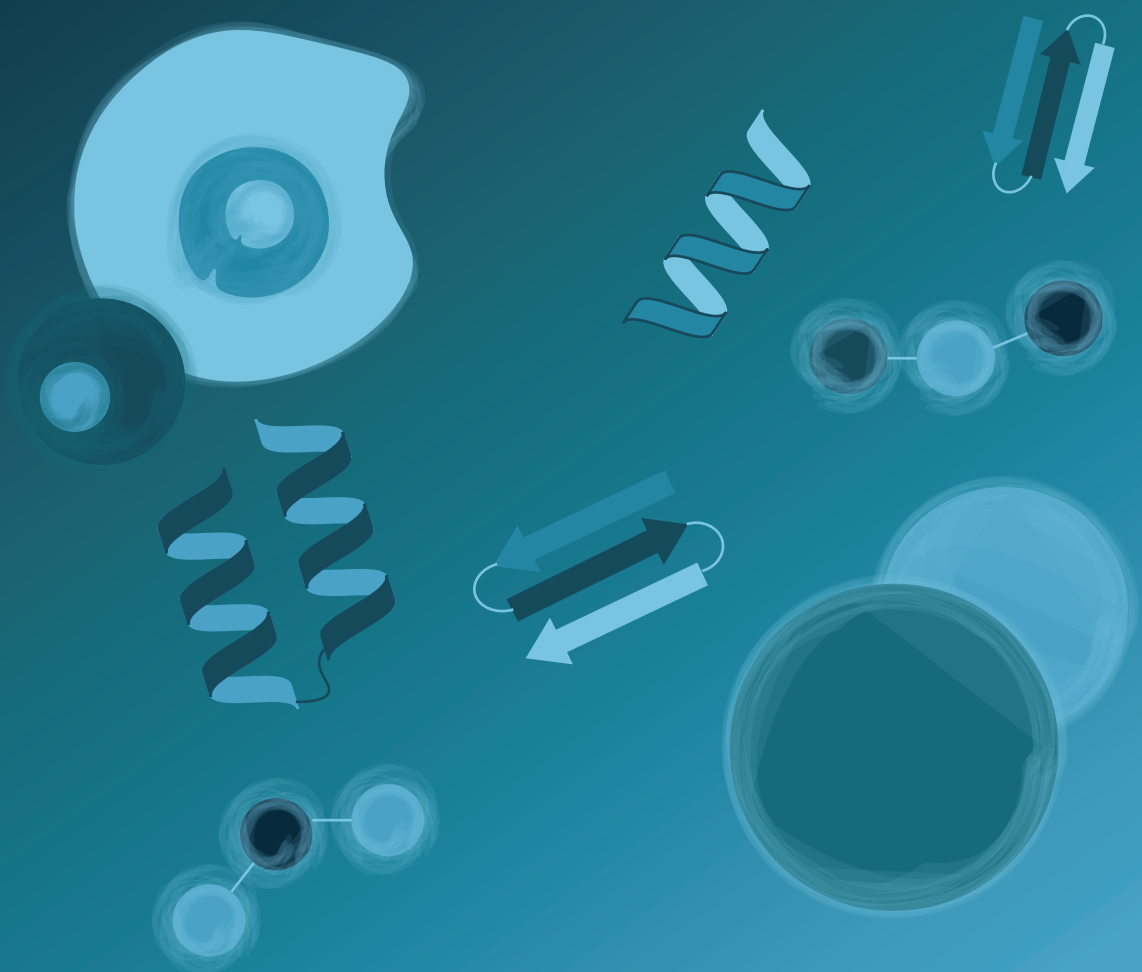


Figure 4. Bacterial taxa that are present in the seven sampling locations and their relative abundance according to the metaproteomic analysis. All the genera having at least 2 identified proteins were kept for analysis. The first 10 taxa in all the sampled locations are annotated in the legend

Chapter 5

Wastewater metaproteomics: tracking microbial and human protein biomarkers

Claudia G. Tugui, Filine Cordesius, Willem van Holthe, Mark C.M. van Loosdrecht and Martin Pabst



Chapter on bioRxiv (manuscript under review in Water Research, 2025):
Tugui, C. G., Cordesius, F., van Holthe, W., van Loosdrecht, M. C., & Pabst, M. (2025). Wastewater metaproteomics: tracking microbial and human protein biomarkers. bioRxiv, 2025-02. Supplementary material is available via https://-doi.org/10.1101/2025.02.08.637285.

Abstract

Wastewater-based surveillance has become a powerful tool for monitoring the spread of pathogens, antibiotic resistance genes, and measuring population-level exposure to pharmaceuticals and chemicals. While surveillance methods commonly target small molecules, DNA, or RNA, wastewater also contains a vast spectrum of proteins. However, despite recent advances in environmental proteomics, large-scale monitoring of protein biomarkers in wastewater is still far from routine. Analyzing raw wastewater presents a challenge due to its heterogeneous mixture of organic and inorganic substances, microorganisms, cellular debris, and various chemical pollutants. To overcome these obstacles, we developed a wastewater metaproteomics approach including efficient protein extraction and an optimized data-processing pipeline. The pipeline utilizes de novo sequencing to customize large public sequence databases to enable comprehensive metaproteomic coverage. Using this approach, we analyzed wastewater samples collected over three months from two urban locations. This revealed a core microbiome comprising a broad spectrum of microbes, gut bacteria and potential opportunistic pathogens. Additionally, we identified nearly 200 human proteins, including promising population-level health indicators, such as immunoglobulins, uromodulin, and cancer-associated proteins.

Key words: metaproteomics, wastewater, human proteins, potential pathogens

Introduction

Globally, approximately 380 trillion liters of wastewater are produced annually, and with the steadily growing world population, it is estimated to nearly double in the next 50 years[1]. Wastewater streams are a complex collection of chemicals, organic compounds, microorganisms, and biomolecules such as DNA and proteins, of which a large fraction originates from human activity. The analysis of wastewater for microbial pathogens, viruses, and substances such as pharmaceuticals, pesticides, and biomarkers of stress and diet has become a routine practice. This has been termed wastewater-based epidemiology (WBE) by Cristian G. Daughton in 2001[2-4]. Today, WBE includes various biological biomarkers, to assess the health status at a population level[5]. Wastewater-based epidemiology (WBE) has proven to be effective for identifying and monitoring epidemic outbreaks. For example, in the 1980s, wastewater surveillance in Finland and Israel provided insights into the spread of the poliovirus[6 7]. Furthermore, during the Coronavirus pandemic, various research groups and governments established COVID-19 surveillance programs[8 9 10]. This informed governmental bodies and the general public about the spread of SARS-CoV-2[11, 12]. Furthermore, the presence of certain bacteria also informs on the spread of antimicrobial resistance, and various diseases[13-17 18 19].

Apart from the advantage of anonymity, the collection of wastewaters is relatively cheap, and it can be applicable to a large population size. The detection of small molecules such as pharmaceuticals employs chromatographic separation combined with mass spectrometry[20]. The analysis of viruses, microbes or antimicrobial resistance genes commonly employs targeted approaches such as various nucleic acid–based polymerase chain reaction methods[21-26]. Recently, untargeted methods using next-generation sequencing methods have become more affordable and increasingly popular for studying water and wastewater environments[24, 27-30].

In addition to small molecules, microbes and viruses, wastewater also contains excreted human proteins and proteins from food waste or agricultural activities. Interestingly, many biomarkers that potentially contain information about population health are proteins, excreted through saliva, stool or urine. Currently, a transition toward precision medicine is underway, which prioritizes proactive, patient-centered approaches[31, 32]. One objective is to identify protein biomarkers that can improve early diagnosis[32]. For example, the proteins found in urine can indicate urogenital disorders, chronic conditions like cancer[33 34 35], autoimmune diseases[36], neurological disfunctions like Alzheimer[37] as well as diabetes[38]. However, currently, large volumes of clinical data from biofluids, such as urine and blood, must be collected and analyzed, ideally with minimal discomfort for patients[32]. Wastewater, on the other hand, is readily available and can be used by health professionals to assess the overall health of the population in a simple, non-invasive, and anonymous manner.

While the analysis of small molecules and the targeting of RNA and DNA have become routine, effective protocols for large-scale monitoring of macromolecules, such as proteins, are still lacking. Over the past decades, mass spectrometry-based proteomics has evolved from focusing on single species to encompassing the field of microbial ecology, known as metaproteomics. Metaproteomics enables the measurement of complex microbial mixtures, providing insights into the microbial composition and expressed microbial functions[39-41]. Furthermore, it allows to measure freely floating proteins, including those excreted by humans or released through industrial and agricultural activities. Therefore, metaproteomics can provide an alternative view on wastewater, which cannot be obtained by DNA-based approaches alone. Recent advancements in mass spectrometric instrumentation have significantly reduced measurement time, even for highly complex metaproteomic samples, while also enhancing sensitivity[42, 43]. This is a significant step toward establishing metaproteomics as a routine, untargeted wastewater surveillance approach. However, the heterogeneous nature of wastewater presents additional challenges. First, an effective sample preparation method is required to capture all present proteins. Second, data processing requires a reference sequence database that includes all proteins in the wastewater. While whole metagenome sequencing covers the microbial population, it does not capture freely floating proteins or those from food waste residues and agricultural activities[44, 45]. Additionally, although increasingly affordable, whole metagenome sequencing is time-consuming and prone to errors at various stages, including DNA extraction and data processing[45, 46].

The first metaproteomic study on wastewater, to the best of the authors' knowledge, was conducted by Carrascal and co-workers who used polymeric adsorbents immersed in the influent water of a wastewater treatment plant over several days[47]. The sorbed proteins allowed for the identification of 690 proteins from bacteria, plants, animals, and humans. In addition to the polymeric probe, the study utilized a large, generic database for database searching. This was later combined with the regions of interest multivariate curve resolution approach, to streamline data analysis[48]. Subsequent studies performed separate analysis of soluble and particulate fractions and concentrating of larger volumes of wastewater followed by SDS-PAGE gel electrophoresis and in-gel digestion to characterize the wastewater proteome[49, 50]. These studies identified various proteins, including potential human biomarkers, as well as a spectrum from various microbes. The proteomic profiles also provided insights into the presence of local industries, such as farming. However, polymeric probes may not capture all freely floating proteins, and separating fractions and concentrating large volumes of wastewater can be time-consuming. Additionally, the choice of reference sequence database affects the accuracy and comprehensiveness of the results. Using generic databases is computationally intensive and may reduce the sensitivity of the database search approach.

In this study, we demonstrate a streamlined metaproteomics approach which we applied to crude municipal wastewater samples collected over three months from two different locations. We developed an efficient sample preparation procedure that extracts proteins from both insoluble and soluble fractions starting with small volumes of wastewater, making it suitable for multiplexing. Additionally, we created a wastewater metaproteomics data processing pipeline that employs *de novo* sequencing to focus generic reference sequence databases in order to obtain a comprehensive metaproteomic coverage.

Material and Methods

Sampling. Samples were taken from two wastewater treatment plants over a period of 4 months, from November 2023 to February 2024. From the wastewater treatment plant Harnaschpolder[6] samples were taken on 29/11/23, 12/05/23, 24/01/24, 31/01/24, and 20/02/24, and from Utrecht (UT), samples were taken on 29/11/23, 05/12/23, 25/01/24, 23/02/24 and 27/02/24. The sampling was done from the influent, raw sewage, on the days with low precipitation. After sampling influent wastewater, samples were stored at -20 ◦C until further processed by a short sample preparation protocol (**Figure 1A**). **Protein extraction and proteolytic digestion.** 500 µL of the wastewater influent was taken and diluted with B-PER (175µl) and 50 mM TEAB buffer (175 µL) and heated at 90 ˚C, for 5 min under shaking at 300 rpm. Further, the sample was subjected to cell lysis using vortexing 3 times for 1 minute using a bench vortexing machine, sonication on a sonication bath for 15 minutes and one freeze/thaw cycle (frozen at -80 ˚C, thawed in incubator at 40˚C for 5 minutes). The samples were then centrifuged and transferred to a 1.5 mL LoBind Eppendorf tube. TCA was added to the sample at a ratio 1:4 (v/v, TCA/sample), vortexed and incubated at 4 ˚C for 20 minutes. After centrifugation, the protein pellet was re-solubilized in 6 M urea and then reduced with DTT (dithiothreitol) and alkylated using IAA (iodoacetamide). After alkylation, the sample was transferred to a FASP filter (Millipore, MRCPRT010) which was previously conditioned by washing 2 times with 100 mM ABC buffer. The filters were centrifuged at 14K rpm in a bench top centrifuge, for 45 minutes, and then 2 times at 14K rpm for 40 minutes after adding 100 mM ABC buffer. Next, the proteins were proteolytically digested on the FASP filter by adding 100 µL trypsin solution, which was prepared by diluting 8 µL trypsin stock solution (0.1 ug/mL in 1 mM HCl, Promega, Cat No) in 100 µL 100 mM ABC. The FASP filters were incubated over night for digestion, at 37 ˚C, under gentle shaking at 300 rpm. The following day, the filters were centrifuged and then once washed with 100 mM ABC buffer followed by a second wash with 100 µL of 10% ACN 0.1% FA/H_2O collected the proteolytic peptides. The pooled fraction was then purified using an OASIS HLB well plate (Waters, UK) according to the manufacturer's protocol. The purified peptide fraction was speed-vac dried and stored at -20 ˚C until further analyzed. **Shotgun metaproteomics.**

To the speed vac tried samples 20 µL of 3% acetonitrile and 0.01% trifluoroacetic acid in H_2O was added, vortexed, then left at room temperature for 30 min, and then once more vortexed. The peptide concentration was determined by measuring the absorbance at 280 nm using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). Samples were diluted to a concentration of approximately 0.5 µg/µL. Shotgun metaproteomics was performed as described previously[41], with a randomized sample order. Briefly, approximately 0.5 µg protein digest was analysed using a nano-liquid-chromatography system consisting of an EASY nano-LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 25% solvent B over 90 min, from 25% to 55% over 60 min, followed by back equilibration to starting conditions. Solvent A was a 0.1% formic acid solution in water (FA), and solvent B consisted of 80% ACN in water and 0.1% FA. The Orbitrap was operated in data dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70 K resolution in full MS mode with a maximum ion injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD) at a normalised collision energy of 28. MS/MS scans were acquired at 17.5 K resolution with AGC target of 2E5 and IT of 75 ms, 1.2 m/z isolation width. **Taxonomic profiling and database construction.** The mass spectrometric raw data for each sample were de novo sequenced using PEAKS Studio X (Bioinformatics Solutions, Inc., Canada). De novo sequences with an ALC score >70 were subjected to taxonomic profiling using the NovoBridge pipeline as described previsously[51]. An in-house constructed API sequence downloader "UniRefBuilder" was employed to construct a reference sequence database containing all UniRef90 entries of the identified families per sample. The NovoBridge+ pipeline and the UniRefBuilder are freely available via GitHub: https://github.com/hbckleikamp/NovoBridge_plus, and https://github.com/claudiatugui/UniRefBuilder (**Figure 1B**). **Database searching.** The focused UniRef90 database was used for database searching using PEAKS Studio X (Bioinformatics Solutions, Inc., Canada) employing a two-round search approach. The first round allowed for one missed cleavage and included carbamidomethylation as a fixed modification, allowing a 20 ppm precursor error and a 0.02 Da fragment ion error. From every sample, the matched proteins from the first round search (without score cut-offs) and the proteins from the human reference proteome (UP000005640) were combined into a new reference sequence database for the second-round search. The second-round search was performed allowing up to 3 missed cleavages, with carbamidomethylation as a fixed modification, and methionine oxidation and asparagine or glutamine deamidation as variable modifications, allowing 20 ppm precursor error and 0.02 Da fragment ion error. Peptide-spectrum matches from the second round were filtered to a 5% false discovery rate

(FDR) at the PSM level, and protein identifications with ≥2 unique peptide sequences were considered significant. The complete dataset was combined using the PEAKSQ module allowing 10 minutes RT shifts and 10 ppm mass error. All identified proteins were exported and further processed as described in the following. **Data analysis and visualization.** For further analysis, proteins were filtered for a minimum of 2 unique peptides and an overall top 3 peptide area (summed across all samples) of > 5E5, and finally only the top hit from every protein group was kept for further data visualization and interpretation. The protein identification table was further analyzed in Python. A taxonomic lineage based on NCBI taxonomy was assigned to every protein which was then used to prepare a taxonomic composition at different taxonomic levels. Two datasets were generated, the "microbial dataset" containing all proteins excluding those with "Eukaryota" and missing annotations at the superkingodm level, and the "human proteome" dataset which contains only protein identifications with the annotation "Homo" at the genus level. Finally, for both datasets, only one protein per protein group was retained ("one_per_group" datasets). Except stated otherwise, the taxonomic composition, protein abundance and diversity plots were determined by summing the top 3 peptide areas for each protein within the respective taxonomy. Principal coordinate analysis (PCoA) was performed using the MDS implementation from scikit-learn, after computing the Bray-Curtis dissimilarity matrix based on genus-level compositions from all time points and locations. The Shannon diversity index was calculated in Python using the formulae $H' = -\sum p_i \ln(p_i)$, where pi is the proportional abundance of each genus. The microbial proteins were categorized according to human gut bacteria and potential pathogens. Assigning potential pathogenic bacterial genera was based on the work by Bartlett et al.[52]. The assignment of human gut microbes was performed using the Human Gut Microbiome Atlas (www.microbiomeatlas.org) which was queried via an API. Annotation of potential pathogenic microbes was done according to the WHO report[53]. The human dataset was further analyzed for the enrichment of molecular and cellular functions, and pathways using STRING[54]. Potential cancer related protein biomarkers were taken from the Human Protein Atlas (www.proteinatlas.org)[55], which was queried via an API. Human proteins associated with coronary artery disease (CAD) were taken from the CAD biomarkers database[56], diabetic nephropathy (DN) from Zürbig *et al*. (2012)[57], breast cancer from Beretov *et al*. (2015)[58], urothelial cancer from Chen *et al*. (2021)[59], Abdominal-type Henoch-Schonlein purpura (HSP) from Jia *et al*. (2021)[60], prostate cancer from Fujita *et al*. (2018)[61], and for ovarian cancer and IBD from Owens *et al*. (2022)[62]. Parts of Figure 1A were generated with BioRender (Created in BioRender. Tugui, C. (2025) https://BioRender.com/v22q517). All mass spectrometric proteomics raw data are available via the ProteomeXchange consortium database, through the identifier PXD059455.

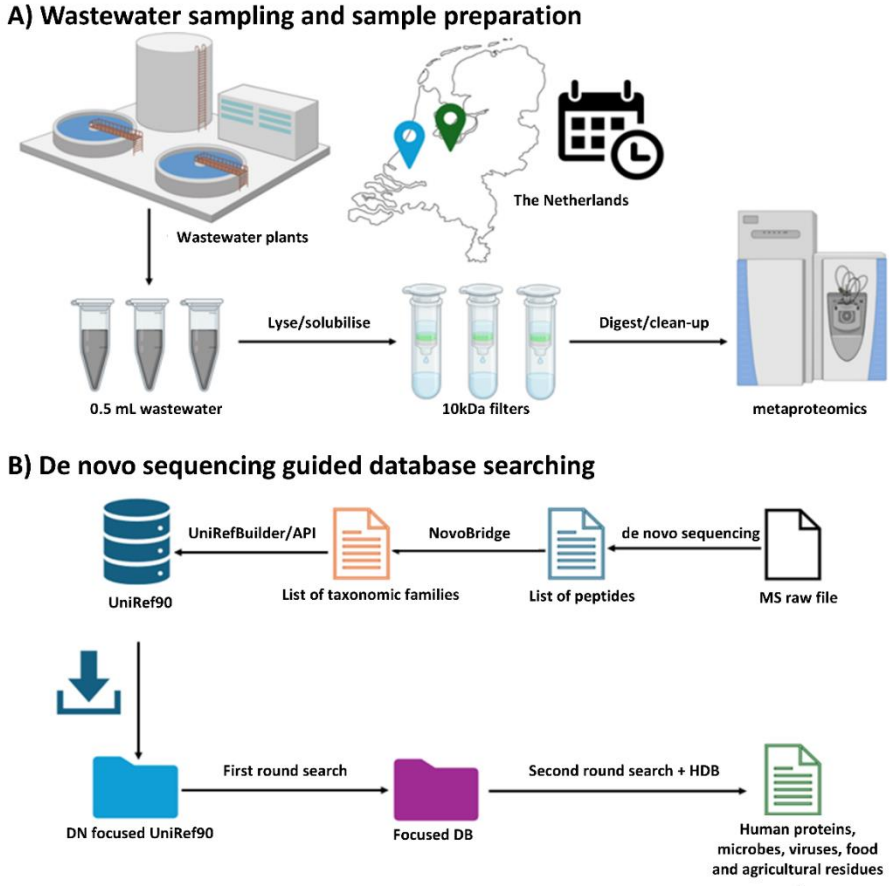


Figure 1. The wastewater metaproteomics workflow involved rapid protein extraction and de novo sequencing guided database searching to identify a broad spectrum of proteins across all domains of life. Influent samples were collected from two wastewater treatment plants, located in Utrecht and Harnaschpolder (The Netherlands), over a 3-month period, with five time points sampled from each plant. A 0.5 mL aliquot of influent was subjected to cell lysis, protein extraction, and proteolytic digestion using a filter-aided sample preparation (FASP) method. Shotgun metaproteomics was performed using a hybrid quadrupole-Orbitrap mass spectrometer. The resulting data were de novo sequenced to focus the global UniRef90 database. This enabled a rapid two-step database searching using the global Uniref90 database content to enhance protein identification.

The proteolytic peptides were finally analyzed with a shotgun metaproteomics experiment employing a one-dimensional chromatographic separation (with a 120 minutes gradient) and a hybrid quadrupole-Orbitrap mass spectrometer. This allowed to detect 4,919 proteins across 3,009 protein groups, including 249 human proteins from 198 protein groups, when considering proteins with ≥2 unique peptides. The number of proteins increased to >10,000

when considering proteins with only 1 unique peptide. The identified proteins could be assigned to a diverse array of taxonomies and sources, including environmental microbes, human microbiome microbes (both pathogenic and commensal bacteria), human and animal proteins, as well as agricultural waste and food residues.

Wastewater metaproteome biomass composition and diversity. The most abundant proteins in the wastewater across both treatment plants could be associated with human and animal feces, urine, and other animal-related sources. Although these proteins constituted only a minor fraction of the total protein fraction (approximately 10–15%), these accounted for more than 50% of the total protein abundance (considering the top 3 peptide area per protein) (**Figure 2A**).

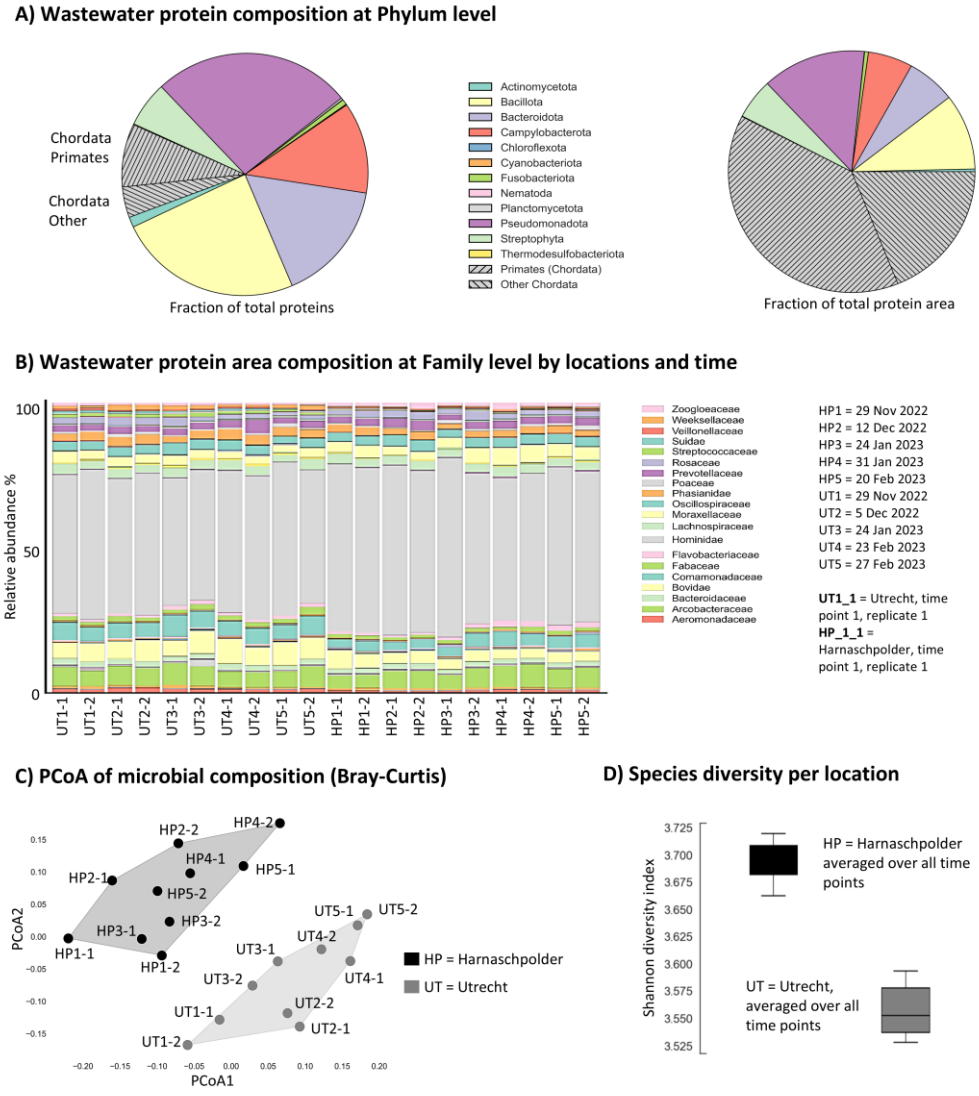


Figure 2: A. The graph displays the averaged phylum-level distribution from both locations, HP and UT. This is represented either by summing the top three peptide areas of all proteins ("Fraction of total protein") or by counting the number of identified proteins per phylum ("Fraction of total protein area"). B. The bar graph illustrates the protein area composition at the family level across 5 time points from Dec 2023 to February 2024, observed in the wastewater of Harnaschpolder [6] and Utrecht (UT), located in The Netherlands. C. The graph depicts alpha diversity, calculated for bacterial genera using the Shannon diversity index. D. The Principal Coordinate Analysis (PCoA) plot visualizes beta diversity, calculated by Bray-Curtis dissimilarity, using all microbial proteins detected in the samples.

Among the dominant microbial phyla, we identified *Pseudomonadota*, *Bacillota*, *Bacteroidota*, and *Campylobacterota*, which are typical wastewater-associated microbes or

microbes associated with various niches of the human microbiome. Minor other phyla that were consistently identified included *Actinomycetota*, *Fusobacteriota*, *Planctomycetota*, and *Cyanobacteriota*, which likely contribute to organic matter degradation and nutrient cycling in these environments. Additionally, significant amounts of *Streptophyta* were detected which are plant-derived residues, possibly from food processing activities. Finally, we also detected *Nematoda* which is a diverse phylum of worms commonly found in soil, and aquatic environments, and as parasites in plants and animals.

Interestingly, these Phyla could be further assigned to >60 taxonomic families, which indicates a diversity and complex ecosystem. Nevertheless, the individual sampling time points during the winter period, as well as the different locations (HP and UT), showed only comparatively small differences in their core taxonomic profiles (**Figure 2B**).

The most dominant families derived from fecal contamination such as *Streptococcaceae*, *Enterobacteriaceae*, *Bacteroidaceae*, *Veillonellaceae*, and *Bifidobacteriaceae*, which are indicative of human and animal gut microbiota and include potential pathogens like Streptococcus and *E. coli*[63]. These also included opportunistic pathogens, such as *Weeksellaceae*, *Leptotrichiaceae*, *Aeromonadaceae*, and *Arcobacteraceae*. Other families such as *Planctomycetaceae*, *Nitrobacteraceae*, *Comamonadaceae*, and *Rhodospirillaceae*, are rather related to biological processes including nitrogen cycling and organic matter degradation[64-66]. Additionally, biodegraders like *Pseudomonadaceae*, *Bacillaceae*, *Chitinophagaceae*, *Flavobacteriaceae*, *Burkholderiaceae*, and *Sphingomonadaceae* were detected which play crucial roles in the breakdown of organic material and xenobiotics[67-72]. Interestingly, also *Caldilineaceae* and *Nocardioidaceae* were detected which are frequently linked to operational challenges in wastewater treatment, such as sludge bulking and foaming[73, 74]. Also, *Zoogloeaceae* were detected *Zoogloeacease* is a major denitrifying bacterium which also produces extracellular polymeric substances relevant for floc formation[75]. Members of the families *Moraxellaceae* and *Burkholderiaceae* have been frequently linked to opportunistic infections[76-81].

Non-microbial taxa could be assigned to food (processing) residuals, livestock and agricultural runoff, which included families such as *Suidae*, *Bovidae*, *Phasianidae*, and *Callorhinchidae*. A broad range of different plant families were also detected including *Arecaceae*, *Asteraceae*, *Fabaceae*, *Solanaceae*, *Poaceae*, *Malvaceae*, *Zingiberaceae*, *Cucurbitaceae*, *Rosaceae*, *Anacardiaceae*, *Juglandaceae*, and *Pedaliaceae* which likely derive from food residuals, and agricultural activities, and which contribute organic material in wastewater environments. However, also proteins from potential pathogenic vectors such as *Nematoda* (e.g., *Steinernematidae*[82]) were detected, as well as from *Blastocystidae* which is the most prevalent gastrointestinal protist in humans and animals. While its clinical significance remains uncertain, it is increasingly regarded as a commensal component of the gut microbiome[83]. Overall, a large fraction (approx. 50%) of proteins derived from families that include potential pathogens, including several that are linked to the gut

microbiome. A comprehensive list of identified proteins, Phyla and Families is provided in the **SI EXCEL DOC**.

To further investigate the similarity and differences between the individual sampling time points and locations, we performed a principal coordinates analysis (PCoA) based on the Bray-Curtis dissimilarity index (**Figure 2C**). In general, in PCoA, points that are closer together on the plot represent taxonomic profiles with more similar compositions. The analysis demonstrated a clear clustering of replicate samples, and while the individual sampling time points showed distinct taxonomic profiles, the samples also clearly clustered by location. When further analyzing for differences between both locations, the UT wastewater generally showed a slightly lower alpha-diversity compared to the HP wastewater, with a Shannon index of 3.55 and 3.75, respectively (**Figure 2D**). The Shannon index is a metric that reflects both species richness (the total number of species) and the evenness of their abundances within a community. Higher values, such as detected for the analyzed wastewater, indicate a more diverse community with a larger number of species and relatively balanced abundances among them. However, considering that metaproteomics requires a minimum of protein biomass for a successful detection of a microbe, the true complexity is likely significantly larger.

We compared the average abundance of selected microbes between both locations, which showed that several bacterial families differed in their relative abundance (**Figure 3**). For example, in HP, *Aeromonadaceae* and *Weeksellaceae*, which are common in wastewater and are associated with potential pathogenic species[84] and are known to harbor antibiotic resistance genes or coexist with microbes involved in the spread of antimicrobial resistance[84-86], showed a much higher relative abundance in HP compared to UT. Additionally, *Streptococcaceae* and *Vibrionaceae* were more abundant in HP, both of which are fecal indicators and potential pathogens[87, 88]. In contrast, *Sphingomonadaceae*, which are well-known biodegraders of xenobiotics in wastewater[70, 72], were more abundant in the UT samples. On the other hand, *Nocardiaceae* and *Nocardioidaceae* were present at comparable levels in both locations. *Nocardiaceae* are known to cause foaming issues in wastewater treatment plants[89, 90], while the other includes members, such as *Nocardioides* which can degrade a variety of pollutants[91]. Overall, a significant proportion of the detected microbial families could be associated with pathogens and diseases. However, the unambiguous identification of pathogens requires species-level resolution, which could not be achieved using the generic reference sequence database.

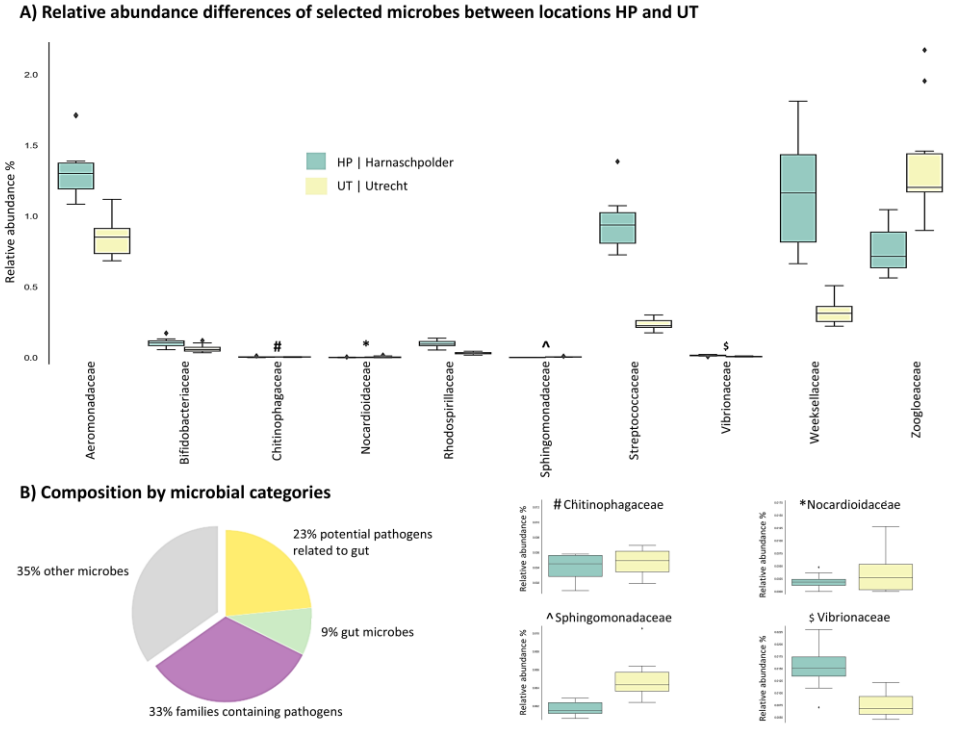


Figure 3: Graph A displays boxplots of the relative abundance of selected bacterial families across 5 time points from Dec 2023 to February 2024, observed in the wastewater of Harnaschpolder[6] and Utrecht (UT), located in The Netherlands. Graph B presents a pie chart illustrating the distribution of microbes observed in the wastewater, categorized into groups, by: i) those associated with pathogens, ii) gut microbes, iii) both and iv) other microbes.

The human wastewater proteome. Over 190 proteins groups from the human proteome were detected across all sampling time points and wastewater locations, which could be assigned to sources such as feces, urine, sweat, or saliva (**SI EXCEL DOC**). Interestingly, the human proteome profile was comparable across all sampling time points and locations. Furthermore, only a subset of these proteins exceeded 2.5% relative abundance, with chymotrypsin-like elastase family member 3A being the most abundant, which accounted for approximately 15% of the total abundance of all detected human proteins. Other abundant proteins did not exceed a relative abundance of 2.5%, including keratin type II cytoskeletal 1, chymotrypsin-C, keratin type I cytoskeletal 9, uromodulin, albumin, alpha-1-antitrypsin, immunoglobulin J chain, keratin type I cytoskeletal 10, and pancreatic alpha-amylase, while most others remained below 1% relative abundance (**Figure 4**, and **SI EXCEL DOC**). Nevertheless, many of these proteins are promising biomarkers for accessing the health status or detecting diseases within the general population. Multiple proteins could

be associated with breast cancer, intestinal diseases, pancreatic cancer, and gastrointestinal cancers, including stomach carcinoma and colon cancer. Cancer-related associations link to various types of carcinomas, such as adenocarcinoma, breast cancer, pancreatic carcinoma, and skin carcinoma. Additionally, this also includes associations with autoimmune diseases, genetic diseases, and metabolic disorders, including diabetes mellitus. Several diseases related to immune system dysfunction were also enriched. Furthermore, infectious diseases, including bacterial infectious disease and viral infectious disease, were also found, reflecting the role of these proteins in immune responses to infections. The complete tables of enriched terms and functions is provided in the **SI EXCEL DOC**. A more detailed study regarding proteins related to cancer with annotations from the Human Protein Atlas can be found in the SI DOC (**SI Figure 2 and SI Table 1**). An overview of potential biomarkers for widespread diseases is provided in **Figure 5C**, and **SI EXCEL DOC**, and **SI DOC** (**SI Table 1**).

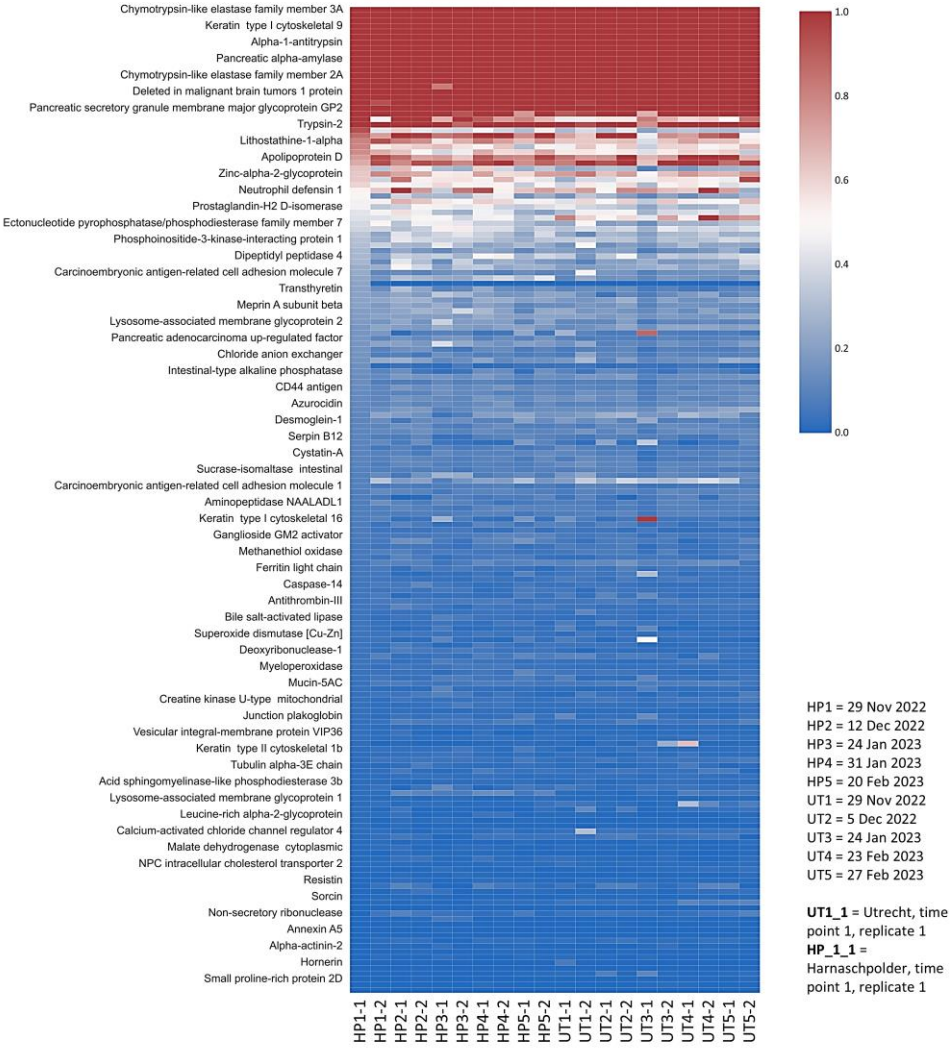


Figure 4: The heatmap provides an abundance profile of all human proteins identified in the wastewater, across 5 time points from Dec 2023 to February 2024, observed in the wastewater of Harnaschpolder[6] (HP) and Utrecht (UT), located in The Netherlands. The color gradient represents the relative abundance of each human protein, determined by the summed top-three peptide areas. The summed area of all proteins was normalized to 100. The proteins were further sorted by ascending abundance, based on sample HP1-1 (Harnaschpolder, time pointe 1 replicate 1). The y-axis annotation names every fourth protein.

Multiple proteins could be associated with breast cancer, followed by proteins linked to diabetic neuropathy, and some for inflammatory bowel disease (**Figure 5C**). Notable

proteins with clinical relevance include arginase-1 (P05089), neutrophil defensins 1 (P59665), Calmodulin-like protein (Q9NZT1) and a range of immune response-related proteins. Potential health indicators are also the various immunoglobulin chains (e.g., from IgG, IgM, and IgA), which reflect inflammation or immune responses to pathogens, including viruses.

To gain a more comprehensive understanding of the enriched functions, tissue expression profiles, cellular locations, and pathways associated with the human proteins identified in wastewater, we conducted an enrichment analysis using the STRING database[54] (**Figures 5A** and **5B**, and **SI EXCEL DOC**). The analysis for tissue expression terms showed enrichments for gastrointestinal tract, digestive glands, skin, liver, pancreas, salivary glands, and respiratory system, reflecting their presence in bodily fluids such as saliva, urine, bile, and tears (**Figure 5A**). Notably, proteins were also linked to tissues involved in immune responses, such as bone marrow, blood, and plasma cells, along with specific tissues like epidermis, epithelial cells, and intestinal epithelium. Additionally, certain proteins were associated with reproductive tissues, including the prostate gland, ovary, and seminal plasma, as well as specific cell types like keratinocytes and leukocytes. This underscores the wide range of biological sources contributing to the human proteome detected in wastewater. Similarly, the enriched cellular component Gene Ontology reflects their extracellular and membrane-associated roles. Key terms include extracellular space, extracellular exosome, secretory granule, and vesicle, indicating that these proteins are actively secreted or involved in extracellular processes. Proteins were also linked to extracellular matrix, suggesting their role in tissue structure and integrity, and collagen-containing extracellular matrix, which may point to their involvement in connective tissues. Several terms related to granules, such as azurophil granule and zymogen granule, include protein storage and secretion functions.

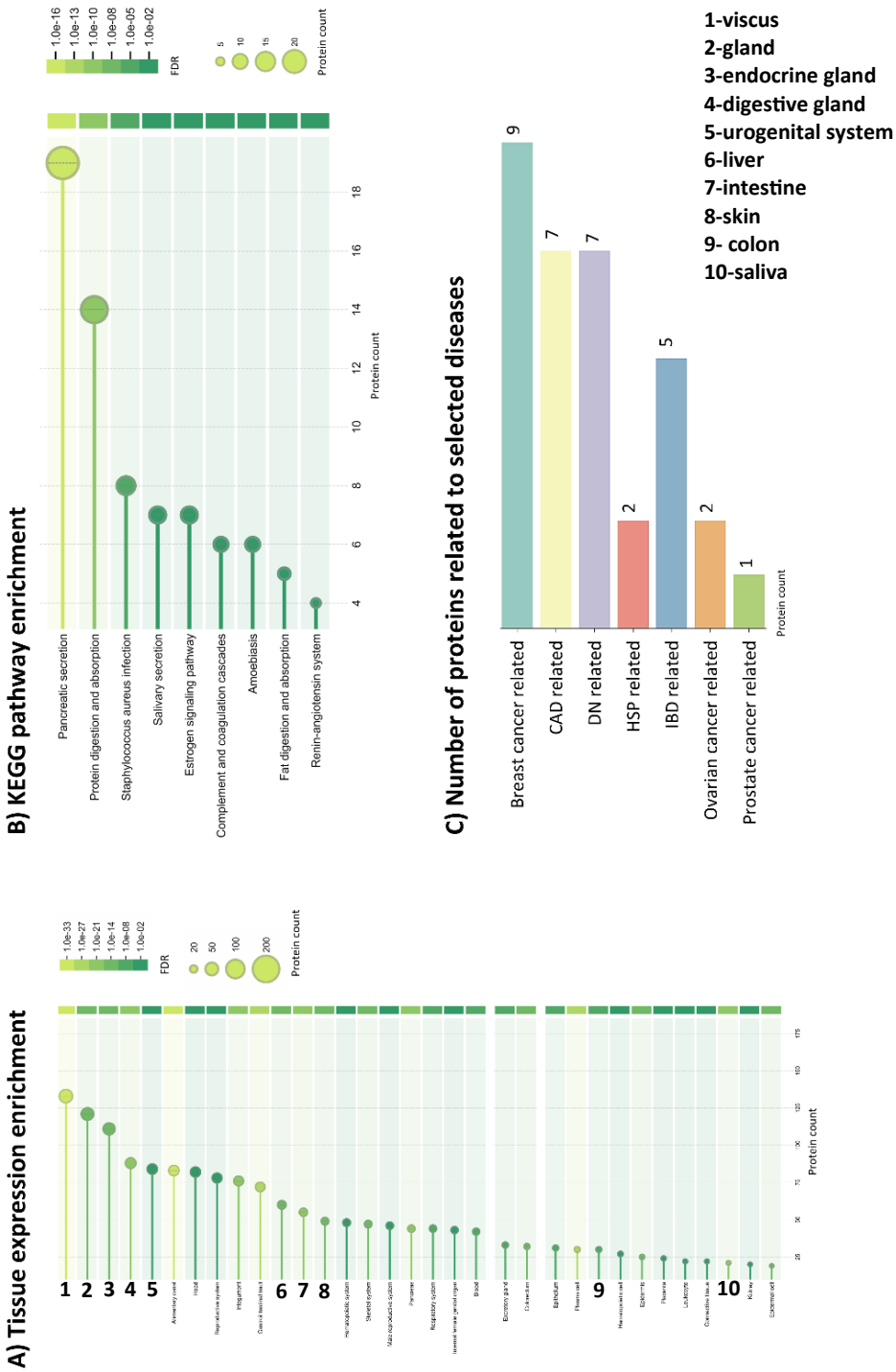

A) Tissue expression enrichment
FDR
Protein count
1 2 3 4 5
6 7 8
9
10
B) KEGG pathway enrichment
Pancreatic secretion
Protein digestion and absorption
Staphylococcus aureus infection
Salivary secretion
Estrogen signaling pathway
Complement and coagulation cascades
Amoebiasis
Fat digestion and absorption
Renin-angiotensin system
Protein count
C) Number of proteins related to selected diseases
Breast cancer related
CAD related
DN related
HSP related
IBD related
Ovarian cancer related
Prostate cancer related
Protein count
1-viscus
2-gland
3-endocrine gland
4-digestive gland
5-urogenital system
6-liver
7-intestine
8-skin
9- colon
10-saliva

Figure 5: Analysis of the identified human proteins for enriched terms and functions. A) The graph displays the enrichment of tissue expression terms. The length of each bar and the size of the corresponding circle represent the number of proteins assigned to each term. The shade of green reflects the false discovery rate (FDR), as calculated by the STRING database tool. B) The graph illustrates the enriched KEGG pathways, with the bar length and circle size indicating the number of proteins assigned to each pathway. The shade of green represents the FDR, as determined by the STRING database tool. C) The bar graph shows the number of proteins associated with selected diseases. Abbreviations: CAD (Coronary Artery Disease), DN (Diabetic Nephropathy), HSP (Henoch-Schönlein Purpura), and IBD (Inflammatory Bowel Disease). The enrichment output tables as obtained from the STRING database tool are available in the **SI EXCEL DOC**.

Additionally, intermediate filament and keratin filament show that several of these proteins are involved in maintaining cellular architecture. Other terms like lysosome and autolysosome suggest potential involvement in protein degradation pathways. Enriched KO pathways include pancreatic secretion and protein digestion and absorption which are linked to the digestive system (**Figure 5B**). The identification of associations with *Staphylococcus aureus* infection and amoebiasis suggests the potential involvement of these proteins in immune responses to bacterial and parasitic infections. Salivary secretion and fat digestion and absorption show their involvement in digestive and metabolic functions. Furthermore, the renin-angiotensin system, complement and coagulation cascades, and estrogen signaling pathway were enriched, which link to cardiovascular regulation, immune responses, and hormonal signaling. The most prevalent molecular function Gene Ontology terms associated with the human proteins were as expected related to enzymatic and structural roles. Key terms included endopeptidase inhibitor activity, peptidase activity, and serine-type peptidase activity, indicating the involvement of these proteins in proteolytic processes. Several structural roles were also identified, such as structural constituent of skin epidermis and the cytoskeleton, proteins which maintain cellular integrity. Other relevant activities include metal ion binding, antioxidant activity, immunoglobulin binding, enzyme inhibitor activity, and toxic substance binding, suggesting diverse physiological roles for these proteins in health and disease processes.

Most interestingly, the human proteins detected in wastewater can be linked to a wide range of clinically relevant disease gene associations, underscoring their potential relevance for public health monitoring (**Figure 5C**). Several skin diseases associations were enriched, including conditions such as keratosis, bullous skin disease, and skin cancer. These proteins were also linked to respiratory system diseases, with lung disease and lower respiratory tract disease. The identified proteins could also be linked to diseases such as pancreatitis. Proteins related to various types of cancer are further outlined in **SI Figure 2** and **SI Table 1**.

Discussion and Conclusions

The presented study demonstrates a streamlined wastewater metaproteomics approach which provides a comprehensive view of the wastewater metaproteome. This approach uses filter-aided sample preparation from small wastewater volumes and de novo sequencing to refine global reference sequence databases – which makes it suitable for multiplexing. We applied this approach to profile the wastewater metaproteome over three months from two different municipal locations. Interestingly, while clustering revealed distinct differences between individual time points and locations, a dominant core metaproteome remained consistent across all samples. This observation may be further amplified by the generic reference sequence database, which might not capture all differences. Additionally, such databases limit the exploration of microbial populations to the genus or family level. On the other hand, the core proteome is likely similar, as both municipal wastewater locations are in densely populated areas, only about 55 km apart, differing mainly in their sewer residence times. The identified proteins belong to environmental microbes as well as microbes form the human microbiome, such as the human gut. Among the many detected families, several contain pathogenic microbes and those known to spread antimicrobial resistance[53]. This may provide valuable insights into the spread of such genes in the environment. Since metaproteomics directly measures proteins, indicating that these microbes were either active or dormant, the information on the potential spread of antibiotic resistance genes would also be more relevant.
However, while metaproteomics identified a broad spectrum of microbes, the unambiguous identification of indicator strains or pathogens requires species-level resolution. This study used a public reference sequence database, which did not allow distinction between individual strains. Nevertheless, this could be achieved by utilizing metagenomic reference sequence databases in addition to the public UniRef database. Furthermore, the presented study did not include viruses, which are of great interest for controlling emerging pandemics. Viruses constitute only a tiny fraction of the protein biomass in such samples, and they typically produce only a few proteins. Consequently, a high viral load must be present in the wastewater for successful metaproteomic detection. Only a few studies employing targeted methods have reported the successful detection of SARS-CoV-2[92]. For example, Jagadeesan and co-workers demonstrated the simultaneous detection of SARS-CoV-2 proteins and the C-reactive protein, which next to the spread of the virus may also indicates inflammation responses[93]. Additionally, different antibody chains have been detected, such as the IgGFc-binding protein and the J chain from IgA, which links two monomer units of either IgM or IgA, and which may also indicate the presence of a viral infection in the population. Stephenson and colleagues explored the detection of antibodies in wastewater and evaluated their immunoaffinity against the SARS-CoV-2 spike protein[94].

They observed intact antibodies, predominantly of the IgG and IgA classes, which appeared to adhere to solid matter in the wastewater. IgA antibodies, critical for mucosal immunity, are detectable in the respiratory tract and serum within one to two weeks following infection or vaccination. In contrast, IgG antibodies typically emerge between two to three weeks and persist for several months. The study demonstrated that these IgG and IgA antibodies retained their immunoaffinity for SARS-CoV-2 spike protein antigens[94]. This suggests that active antibody fractions in wastewater could facilitate real-time monitoring of population immunity, vaccination coverage, and infection prevalence. This may help assess the effectiveness of vaccination campaigns and complement traditional seroprevalence surveys, which are time-intensive to conduct.

The detection of a wide spectrum of human proteins is another important aspect of wastewater metaproteomics. This data may support the evaluation of population health or signal the emergence of an epidemic. In this study, we identified approximately 200 human protein groups, including several potential disease-related proteins and biomarkers, which are promising indicators of population health. Recent studies have shown that population averages for different protein biomarkers can be estimated with good accuracy for several diseases[95] [96]. However, projecting the detected wastewater concentrations to individual averages requires consideration of dilution rates and the population size in the area. It has been suggested that using an abundant and easily detectable human protein, such as albumin, as a normalization factor could further improve protein quantification in wastewater[50]. It is worth mentioning that, in the case of the human proteome, diseases may also correlate with changes in protein modifications. For example, alterations in protein glycosylation have been extensively studied for their relevance as biomarkers[97] [98] [99]. However, no studies to date have investigated the fate of these modifications in the complex wastewater environment.

Besides the advantage of anonymity, wastewater collection is relatively inexpensive and can be applied to large population sizes. However, wastewater collected from extensive areas, including both urban and rural regions, tends to have a longer retention time in the pipes and likely a more uniform composition. For the implementation of wastewater surveillance, city-proximal sampling points may also allow to observe subtle changes in less prevalent pathogens or protein biomarkers. Finally, although mass spectrometry-based metaproteomics is a powerful approach for both fully untargeted screening but also sensitive targeting, this approach still shows some limitations towards throughput and sensitivity. Successful detection requires a minimum number of protein copies, making low-abundance microbial, viral and human proteins challenging to detect. At the same time, the developments in mass spectrometric instrumentation and methods has significantly improved sensitivity and throughput, which is highly relevant when monitoring complex and dynamic environments such as wastewater[42, 100, 101]. Furthermore, alternative technologies are currently being developed. For example, significant advancements have been made in

single-protein sequencing technologies, such as the use of nanopores, similar to DNA sequencing[102]. Progress has also been made in the development of single-protein fingerprinting approaches, which additionally may provide insights into protein modifications[103]. These advancements hold great promise for the integration of metaproteomics into routine wastewater monitoring in the near future.

Acknowledgements

The authors thank Dita Heikens for her support in the proteomics laboratory and acknowledge the NWO Spinoza Prize awarded to Mark van Loosdrecht for funding. They also thank Jelle Langedijk and Mario Pronk for their assistance in obtaining wastewater and extend special thanks to the engineers at the Utrecht and Harnaschpolder wastewater treatment plants for their help with sampling.

Conflict of interest

All authors declare that they have no conflicts of interest.

References

1. Qadir, M. et al. Global and regional potential of wastewater as a water, nutrient and energy source. **44**, 40-51 (2020).
2. Ryu, Y. et al. Increased levels of the oxidative stress biomarker 8-iso-prostaglandin F2α in wastewater associated with tobacco use. *Scientific Reports* **6**, 39055 (2016).
3. Ryu, Y. et al. Comparative measurement and quantitative risk assessment of alcohol consumption through wastewater-based epidemiology: An international study in 20 cities. *Science of The Total Environment* **565**, 977-983 (2016).
4. Daughton, C., Vol. 791 348-364 (2001).
5. O'Keeffe, J. Wastewater-based epidemiology: current uses and future opportunities as a public health surveillance tool. **64**, 44-52 (2021).
6. Hovi, T. et al. Role of environmental poliovirus surveillance in global polio eradication and beyond. *Epidemiology and infection* **140**, 1-13 (2012).
7. Roberts, L. Infectious disease. Israel's silent polio epidemic breaks all the rules. *Science (New York, N.Y.)* **342**, 679-680 (2013).
8. Spurbeck, R.R., Minard-Smith, A. & Catlin, L. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. *Science of The Total Environment* **789**, 147829 (2021).
9. Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. & Brouwer, A. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19

Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters* **7**, 511-516 (2020).

10. Daughton, C.G. Wastewater surveillance for population-wide Covid-19: The present and future. *The Science of the total environment* **736**, 139631 (2020).
11. Choi, P.M. et al. Wastewater-based epidemiology biomarkers: Past, present and future. *TrAC Trends in Analytical Chemistry* **105**, 453-469 (2018).
12. Rice, J. & Kasprzyk-Hordern, B. A new paradigm in public health assessment: Water fingerprinting for protein markers of public health using mass spectrometry. *TrAC Trends in Analytical Chemistry* **119**, 115621 (2019).
13. Clemente, J.C., Ursell, L.K., Parfrey, L.W. & Knight, R. The impact of the gut microbiota on human health: an integrative view. *Cell* **148**, 1258-1270 (2012).
14. Hassoun-Kheir, N. et al. Comparison of antibiotic-resistant bacteria and antibiotic resistance genes abundance in hospital and community wastewater: A systematic review. *Science of the Total Environment* **743**, 140804 (2020).
15. Łuczkiewicz, A., Jankowska, K., Fudala-Książek, S. & Olańczuk-Neyman, K. Antimicrobial resistance of fecal indicators in municipal wastewater treatment plant. *Water research* **44**, 5089-5097 (2010).
16. Novo, A., André, S., Viana, P., Nunes, O.C. & Manaia, C.M. Antibiotic resistance, antimicrobial residues and bacterial community composition in urban wastewater. *Water research* **47**, 1875-1887 (2013).
17. Gilbride, K., Lee, D.-Y. & Beaudette, L. Molecular techniques in wastewater: understanding microbial communities, detecting pathogens, and real-time process control. *Journal of microbiological methods* **66**, 1-20 (2006).
18. Kho, Z.Y. & Lal, S.K. The Human Gut Microbiome - A Potential Controller of Wellness and Disease. *Frontiers in microbiology* **9**, 1835 (2018).
19. Halfvarson, J. et al. Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology* **2**, 17004 (2017).
20. Picó, Y. & Barceló, D. Identification of biomarkers in wastewater-based epidemiology: Main approaches and analytical methods. *TrAC Trends in Analytical Chemistry* **145**, 116465 (2021).
21. Mao, K. et al. The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. *Current opinion in environmental science & health* **17**, 1-7 (2020).
22. Hillary, L.S., Malham, S.K., McDonald, J.E. & Jones, D.L. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Current opinion in environmental science & health* **17**, 14-20 (2020).
23. Hryniszyn, A., Skonieczna, M. & Wiszniowski, J. Methods for detection of viruses in water and wastewater. *Advances in Microbiology* **3**, 442-449 (2013).

24. Pruden, A., Vikesland, P.J., Davis, B.C. & de Roda Husman, A.M. Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Current Opinion in Microbiology* **64**, 91-99 (2021).
25. Parkins, M.D. et al. Wastewater-based surveillance as a tool for public health action: SARS-CoV-2 and beyond. *Clinical Microbiology Reviews* **37**, e00103-00122 (2024).
26. Toze, S. PCR and the detection of microbial pathogens in water and wastewater. *Water Research* **33**, 3545-3556 (1999).
27. Walden, C., Carbonero, F. & Zhang, W. Assessing impacts of DNA extraction methods on next generation sequencing of water and wastewater samples. *Journal of Microbiological Methods* **141**, 10-16 (2017).
28. Garner, E. et al. Next generation sequencing approaches to evaluate water and wastewater quality. *Water Research* **194**, 116907 (2021).
29. Munk, P. et al. (2022).
30. Nieuwenhuijse, D.F. et al. Setting a baseline for global urban virome surveillance in sewage. *Scientific Reports* **10**, 13748 (2020).
31. Kosorok, M.R. & Laber, E.B. Precision medicine. *Annual review of statistics and its application* **6**, 263-286 (2019).
32. Sequeira-Antunes, B. & Ferreira, H.A. Urinary biomarkers and point-of-care urinalysis devices for early diagnosis and management of disease: a review. *Biomedicines* **11**, 1051 (2023).
33. Yang, J. et al. Lipocalin 2 promotes breast cancer progression. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 3913-3918 (2009).
34. Feng, B., Yue, F. & Zheng, M.H. Urinary markers in colorectal cancer. *Advances in clinical chemistry* **47**, 45-57 (2009).
35. Marks, L.S. et al. PCA3 Molecular Urine Assay for Prostate Cancer in Men Undergoing Repeat Biopsy. *Urology* **69**, 532-535 (2007).
36. Song, H., Chan, J. & Rovin, B.H. Induction of chemokine expression by adiponectin in vitro is isoform dependent. *Translational Research* **154**, 18-26 (2009).
37. Ma, L. et al. Development of a Novel Urine Alzheimer-Associated Neuronal Thread Protein ELISA Kit and Its Potential Use in the Diagnosis of Alzheimer's Disease. *Journal of clinical laboratory analysis* **30**, 308-314 (2016).
38. Tashiro, K. et al. Urinary levels of monocyte chemoattractant protein-1 (MCP-1) and interleukin-8 (IL-8), and renal injuries in patients with type 2 diabetic nephropathy. *Journal of clinical laboratory analysis* **16**, 1-4 (2002).
39. Wilmes, P. & Bond, P.L. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends in microbiology* **14**, 92-97 (2006).

40. Kleiner, M. et al. Assessing species biomass contributions in microbial communities via metaproteomics. *Nature communications* **8**, 1558 (2017).
41. Kleikamp, H.B. et al. Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. *Water research* **246**, 120700 (2023).
42. Dumas, T. et al. The astounding exhaustiveness and speed of the Astral mass analyzer for highly complex samples is a quantum leap in the functional analysis of microbiomes. *Microbiome* **12**, 46 (2024).
43. Xian, F. et al. Ultra-sensitive metaproteomics (uMetaP) redefines the dark field of metaproteome, enables single-bacterium resolution, and discovers hidden functions in the gut microbiome. *bioRxiv*, 2024.2004. 2022.590295 (2024).
44. Blakeley-Ruiz, J.A. & Kleiner, M. Considerations for constructing a protein sequence database for metaproteomics. *Computational and structural biotechnology journal* **20**, 937-952 (2022).
45. Kleikamp, H.B. et al. NovoLign: metaproteomics by sequence alignment. *ISME communications* **4**, ycae121 (2024).
46. Kleikamp, H.B. et al. Database-independent de novo metaproteomics of complex microbial communities. *Cell Systems* **12**, 375-383. e375 (2021).
47. Carrascal, M., Abián, J., Ginebreda, A. & Barceló, D. Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. *Science of the Total Environment* **747**, 141145 (2020).
48. Perez-Lopez, C. et al. Non-target protein analysis of samples from wastewater treatment plants using the regions of interest-multivariate curve resolution (ROIMCR) chemometrics method. *Journal of Environmental Chemical Engineering* **9**, 105752 (2021).
49. Sánchez-Jiménez, E., Abian, J., Ginebreda, A., Barceló, D. & Carrascal, M. Shotgun proteomics to characterize wastewater proteins. *MethodsX* **11**, 102403 (2023).
50. Carrascal, M. et al. Sewage Protein Information Mining: Discovery of Large Biomolecules as Biomarkers of Population and Industrial Activities. *Environmental Science & Technology* **57**, 10929-10939 (2023).
51. Kleikamp, H.B.C. et al. Database-independent de novo metaproteomics of complex microbial communities. *Cell Systems* **12**, 375-383.e375 (2021).
52. Bartlett, A., Padfield, D., Lear, L., Bendall, R. & Vos, M. A comprehensive list of bacterial pathogens infecting humans. *Microbiology (Reading, England)* **168** (2022).
53. (Geneva: World Health Organization; 2024).
54. Szklarczyk, D. et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research* **39**, D561-D568 (2010).

55. Digre, A. & Lindskog, C. The human protein atlas—spatial localization of the human proteome in health and disease. *Protein Science* **30**, 218-233 (2021).
56. Perpétuo, L. et al. Coronary Artery Disease and Aortic Valve Stenosis: A Urine Proteomics Study. *International journal of molecular sciences* **23** (2022).
57. Zürbig, P. et al. Urinary Proteomics for Early Diagnosis in Diabetic Nephropathy. *Diabetes* **61**, 3304-3313 (2012).
58. Beretov, J. et al. Proteomic Analysis of Urine to Identify Breast Cancer Biomarker Candidates Using a Label-Free LC-MS/MS Approach. *PloS one* **10**, e0141876 (2015).
59. Chen, C.-J. et al. Discovery of Novel Protein Biomarkers in Urine for Diagnosis of Urothelial Cancer Using iTRAQ Proteomics. *Journal of Proteome Research* **20**, 2953-2963 (2021).
60. Jia, L. et al. Proteomic analysis of urine reveals biomarkers for the diagnosis and phenotyping of abdominal-type Henoch-Schonlein purpura. *Translational pediatrics* **10**, 510-524 (2021).
61. Fujita, K. & Nonomura, N. Urinary biomarkers of prostate cancer. *International journal of urology : official journal of the Japanese Urological Association* **25**, 770-779 (2018).
62. Owens, G.L., Barr, C.E., White, H., Njoku, K. & Crosbie, E.J. Urinary biomarkers for the detection of ovarian cancer: a systematic review. *Carcinogenesis* **43**, 311-320 (2022).
63. Tamburini, F.B. et al. Precision identification of diverse bloodstream pathogens in the gut microbiome. *Nature medicine* **24**, 1809-1814 (2018).
64. Watson, S.W., Valois, F.W. & Waterbury, J.B. in The prokaryotes: A handbook on habitats, isolation, and identification of bacteria 1005-1022 (Springer, 1981).
65. Lohmann, P. et al. Seasonal patterns of dominant microbes involved in central nutrient cycles in the subsurface. *Microorganisms* **8**, 1694 (2020).
66. Madigan, M., Cox, S.S. & Stegeman, R.A. Nitrogen fixation and nitrogenase activities in members of the family Rhodospirillaceae. *Journal of bacteriology* **157**, 73-78 (1984).
67. Goyal, P. & Basniwal, R.K. Environmental bioremediation: biodegradation of xenobiotic compounds. *Xenobiotics in the soil environment: monitoring, toxicity and management*, 347-371 (2017).
68. Gomez, N.C.F. & Onda, D.F.L. Potential of sediment bacterial communities from Manila Bay (Philippines) to degrade low-density polyethylene (LDPE). *Archives of Microbiology* **205**, 38 (2023).
69. Goff, K.L. & Hug, L.A. Environmental potential for microbial 1, 4-Dioxane degradation is sparse despite mobile elements playing a role in trait distribution. *Applied and Environmental Microbiology* **88**, e02091-02021 (2022).

70. García-García, R. et al. Assessment of the Microbial Communities in Soil Contaminated with Petroleum Using Next-Generation Sequencing Tools. *Applied Sciences* **13**, 6922 (2023).
71. Lünsmann, V. et al. In situ p rotein-SIP highlights Burkholderiaceae as key players degrading toluene by para ring hydroxylation in a constructed wetland model. *Environmental microbiology* **18**, 1176-1186 (2016).
72. Stolz, A. Molecular characteristics of xenobiotic-degrading sphingomonads. *Applied microbiology and biotechnology* **81**, 793-811 (2009).
73. Yu, L. et al. Effects of solids retention time on the performance and microbial community structures in membrane bioreactors treating synthetic oil refinery wastewater. *Chemical Engineering Journal* **344**, 462-468 (2018).
74. Bafghi, M.F. & Yousefi, N. Role of Nocardia in activated sludge. *The Malaysian journal of medical sciences: MJMS* **23**, 86 (2016).
75. An, W. et al. Comparative genomics analyses on EPS biosynthesis genes required for floc formation of Zoogloea resiniphila and other activated sludge bacteria. *Water Research* **102**, 494-504 (2016).
76. Rossau, R., Van Landschoot, A., Gillis, M. & De Ley, J. Taxonomy of Moraxellaceae fam. nov., a new bacterial family to accommodate the genera Moraxella, Acinetobacter, and Psychrobacter and related organisms. *International Journal of Systematic and Evolutionary Microbiology* **41**, 310-319 (1991).
77. Liu, H.-y. et al. The interactions of airway bacterial and fungal communities in clinically stable asthma. *Frontiers in microbiology* **11**, 1647 (2020).
78. Wong, D. et al. Clinical and pathophysiological overview of Acinetobacter infections: a century of challenges. *Clinical microbiology reviews* **30**, 409-447 (2017).
79. Munoz-Price, L.S. & Weinstein, R.A. Acinetobacter infection. *New England Journal of Medicine* **358**, 1271-1281 (2008).
80. Mali, S., Dash, L., Gautam, V., Shastri, J. & Kumar, S. An outbreak of Burkholderia cepacia complex in the paediatric unit of a tertiary care hospital. *Indian Journal of Medical Microbiology* **35**, 216-220 (2017).
81. Weinroth, M. et al. (2022).
82. Dillman, A.R. et al. Comparative genomics of Steinernema reveals deeply conserved gene regulatory networks. *Genome biology* **16**, 1-21 (2015).
83. Aykur, M. et al. Blastocystis: a mysterious member of the gut microbiome. *Microorganisms* **12**, 461 (2024).
84. Tugui, C.G. et al. Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin. *RSC Chemical Biology* (2025).

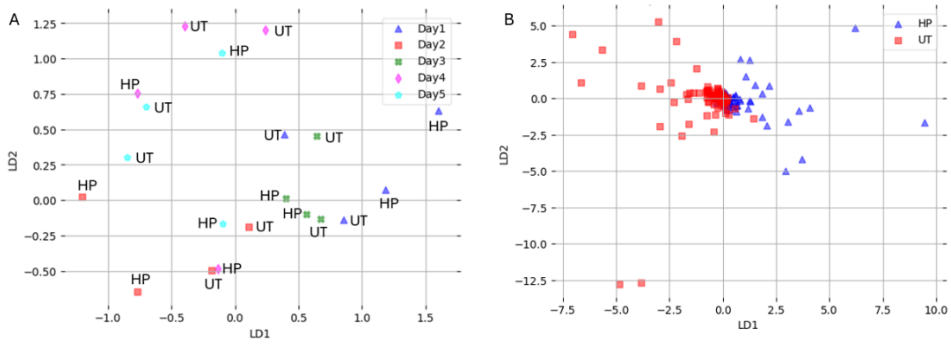
85. Grilo, M.L., Sousa-Santos, C., Robalo, J. & Oliveira, M. The potential of Aeromonas spp. from wildlife as antimicrobial resistance indicators in aquatic environments. *Ecological Indicators* **115**, 106396 (2020).
86. Fernandes, T., Vaz-Moreira, I. & Manaia, C.M. Neighbor urban wastewater treatment plants display distinct profiles of bacterial community and antibiotic resistance genes. *Environmental Science and Pollution Research* **26**, 11269-11278 (2019).
87. Sinton, L., Donnison, A. & Hastie, C. Faecal streptococci as faecal pollution indicators: a review. Part II: Sanitary significance, survival, and use. *New Zealand journal of marine and freshwater research* **27**, 117-137 (1993).
88. De, R., Mukhopadhyay, A.K. & Dutta, S. Metagenomic analysis of gut microbiome and resistome of diarrheal fecal samples from Kolkata, India, reveals the core and variable microbiota including signatures of microbial dark matter. *Gut Pathogens* **12**, 1-48 (2020).
89. Blackall, L.L. et al. Nocardia pinensis sp. nov., an actinomycete found in activated sludge foams in Australia. *Microbiology* **135**, 1547-1558 (1989).
90. Hao, O., * Strom, PF** & Wu, Y. A review of the role of Nocardia-like filaments in activated sludge foaming. *Water SA* **14**, 105-110 (1988).
91. Ma, Y. et al. Nocardioides:“specialists” for hard-to-degrade pollutants in the environment. *Molecules* **28**, 7433 (2023).
92. Lara-Jacobo, L.R., Islam, G., Desaulniers, J.-P., Kirkwood, A.E. & Simmons, D.B.D. Detection of SARS-CoV-2 Proteins in Wastewater Samples by Mass Spectrometry. *Environmental Science & Technology* **56**, 5062-5070 (2022).
93. Jagadeesan, K.K. et al. Wastewater-based proteomics: A proof-of-concept for advancing early warning system for infectious diseases and immune response monitoring. *Journal of Hazardous Materials Letters* **5**, 100108 (2024).
94. Stephenson, S. et al. Urban wastewater contains a functional human antibody repertoire of mucosal origin. *Water Research* **267** (2024).
95. Devianto, L.A. & Sano, D. Systematic review and meta-analysis of human health-related protein markers for realizing real-time wastewater-based epidemiology. *The Science of the total environment* **897**, 165304 (2023).
96. Amin, V., Bowes, D.A. & Halden, R.U. Systematic scoping review evaluating the potential of wastewater-based epidemiology for monitoring cardiovascular disease and cancer. *The Science of the total environment* **858**, 160103 (2023).
97. de Haan, N., Falck, D. & Wuhrer, M. Monitoring of immunoglobulin N- and O-glycosylation in health and disease. *Glycobiology* **30**, 226-240 (2020).
98. Gudelj, I., Lauc, G. & Pezer, M. Immunoglobulin G glycosylation in aging and diseases. *Cellular immunology* **333**, 65-79 (2018).

99. Reily, C., Stewart, T.J., Renfrow, M.B. & Novak, J. Glycosylation in health and disease. *Nature Reviews Nephrology* **15**, 346-366 (2019).
100. Vitko, D. et al. timsTOF HT improves protein identification and quantitative reproducibility for deep unbiased plasma protein biomarker discovery. *Journal of Proteome Research* **23**, 929-938 (2024).
101. Wang, J. et al. Evaluation of Protein Identification and Quantification by the diaPASEF Method on timsTOF SCP. *Journal of the American Society for Mass Spectrometry* (2024).
102. Yu, L. et al. Unidirectional single-file transport of full-length proteins through a nanopore. *Nature Biotechnology* **41**, 1130-1139 (2023).
103. Filius, M. et al. Full-length single-molecule protein fingerprinting. *Nature Nanotechnology* **19**, 652-659 (2024).

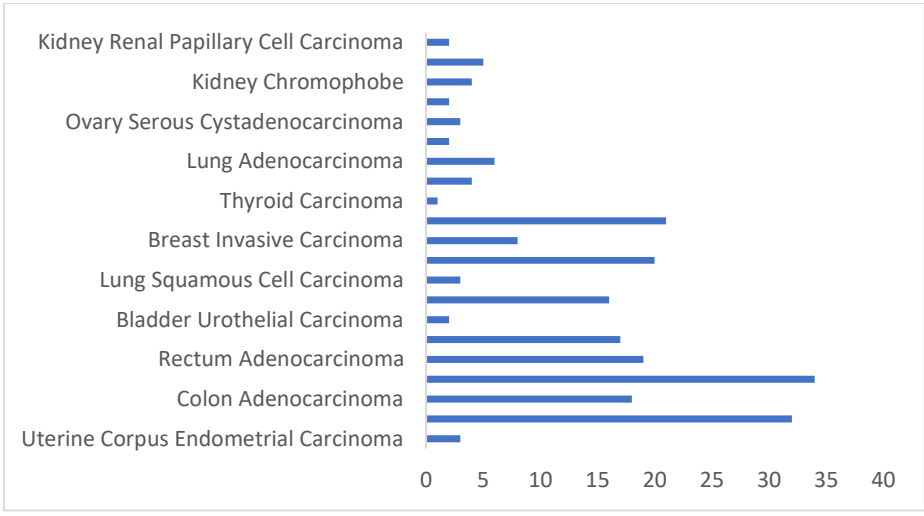
Supplementary information material to:

Wastewater metaproteomics: tracking microbial and human protein biomarkers

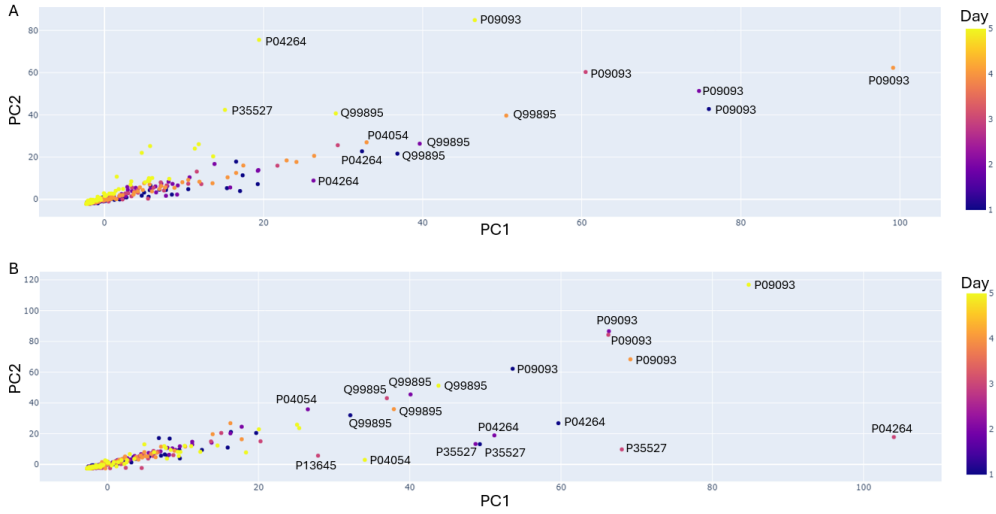
TABLE OF CONTENTS



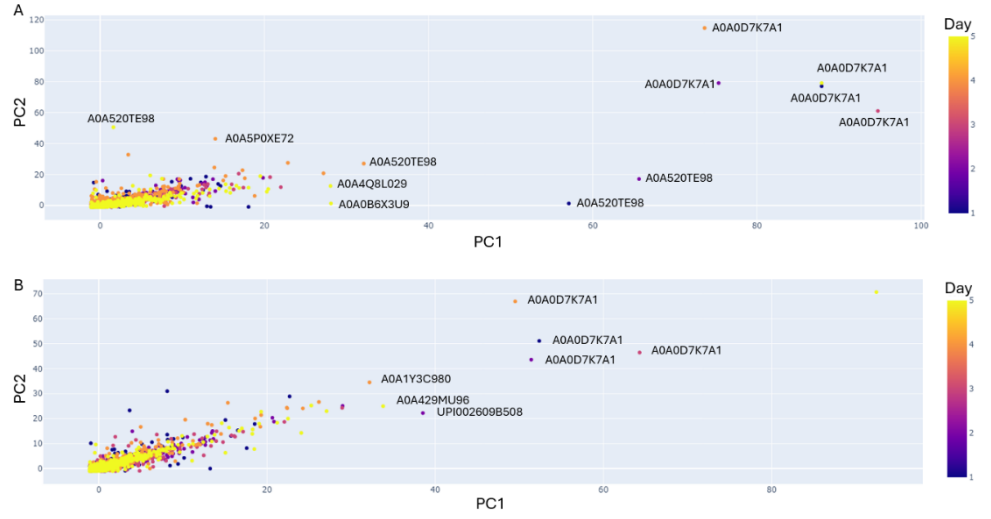
SI Figure 1. A) LDA analysis on potential pathogens (as defined by WHO report) discovered in wastewater samples based on sampling days. **B)** LDA analysis on the detected human proteome based on location (UT = Utrecht, HP = Harnaschpolder).



SI Figure 2. Number of human proteins with potential clinical relevance (e.g. biomarkers) for different types of cancer by both locations.



SI Figure 3. PCA analysis of the detected human proteome for HP (Graph A) and UT (Graph B). Outliers are annotated by their UniProtKB accession number. The color of the dot represents the day of sampling.



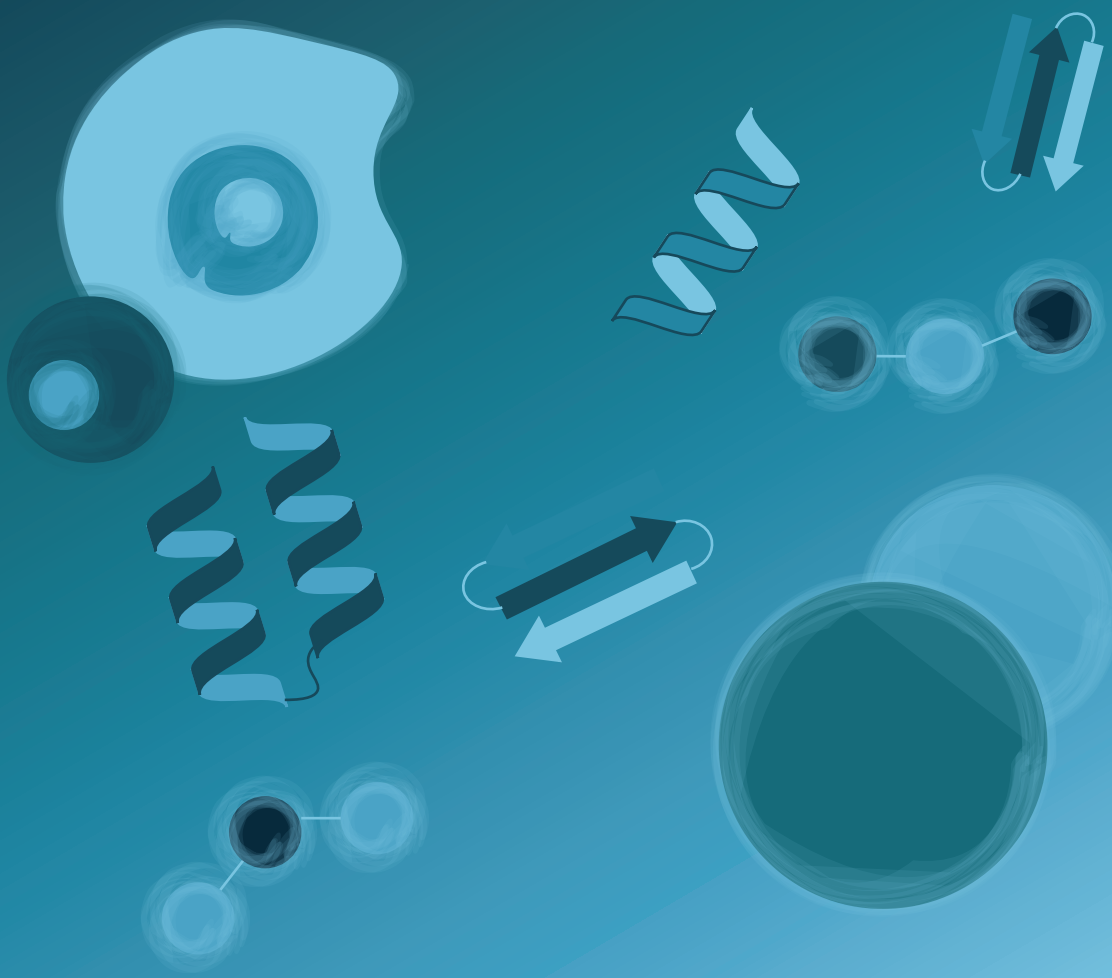
SI Figure 4. PCA analysis of the detected microbial metaproteome for HP (Graph A) and UT (Graph B). Outliers are annotated by their UniProtKB accession number. The color of the dot represents the day of sampling.

SI Table 1. List of proteins potentially related to different types of cancer as classified by the Human Protein Atlas (UT = Utrecht, HP = Harnaschpolder).

Cancer type	UniProtKB accessions of potentially related proteins
Uterine Corpus Endometrial Carcinoma	Q14508, P01036, P20160
Head and Neck Squamous Cell Carcinoma	P04264, P15924, P13645, P08779, P02533, P04259, P13647, Q04695, Q02413, P13646, P29508, Q8N1N4, P06702, Q01546, Q5T749, Q08554, Q08188, Q9NZT1, P04083, Q01469, P31944, P31151, Q96P63, P05109, P35609, Q13835, P22532, A8K2U0, P01040, Q15517, P63316, P55000
Colon Adenocarcinoma	Q9Y6R7, Q02817, A8K7I4, Q9UGM3, Q16819, Q8WWA0, P35030, P56470, O60844, P06731, P40879, P05451, P00915, Q14002, P13688, Q12864, Q9H3R2, Q6UX06
Pancreatic Adenocarcinoma	Q02817, P04746, P19961, A8K7I4, Q9UGM3, Q86UP6, P08861, P09093, P07477, P55259, P16233, P07478, Q6GPI1, P08217, P15085, P98088, Q8WWA0, Q99895, P48052, P04054, P98073, Q16820, P35030, P56470, P05451, Q6W4X9, P04118, Q03403, P19835, P02766, Q9H3R2, P31025, Q9HD89, Q6UX06
Rectum Adenocarcinoma	Q02817, A8K7I4, Q9UGM3, Q16819, Q8WWA0, Q8WWU7, P35030, P56470, O60844, P06731, P40879, P00915, Q14002, P13688, Q12864, P22748, P40199, Q9H3R2, Q6UX06
Stomach Adenocarcinoma	Q02817, Q9UGM3, P14410, P98088, Q8WWA0, P05164, P98073, P35030, P61626, P56470, P05451, P80188, Q12864, Q6W4X9, Q03403, Q9H3R2, Q6UX06
Bladder Urothelial Carcinoma	P13646, P31944
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma	P13647, Q04695, Q14CN2, P13646, P29508, P48594, Q8N1N4, P06702, Q9NZT1, Q01469, P31944, Q96P63, P05109, Q13835, P01040, Q9HCY8
Lung Squamous Cell Carcinoma	P13646, Q96P63, Q13835
Skin Cuteneous Melanoma	P04264, P35527, P35908, P13645, P04259, Q86YZ3, Q02413, Q5D862, Q8N1N4, P05090, Q5T749, Q08554, Q9NZT1, P31944, P31151, Q96P63, O14556, Q15517, P20930, P55000
Breast Invasive Carcinoma	P02788, P25311, P12273, Q96DA0, P00709, P01036, P81605, P02814
Liver Hepatocellular Carcinoma	P02768, P02787, P01009, P09923, O43895, P25311, Q6UWV6, P01011, P02760, P01008, P00738, P56470, P05089, P02763, O95497, P05154, P00734, P02790, P02766, P02750, P05155
Thyroid Carcinoma	P61916
Glioblastoma Multiforme	P01011, P35609, P10153, P59665
Lung Adenocarcinoma	P0DTE8, Q9UGM3, P98088, Q8WWA0, P40199, Q9HD89
Testicular Germ Cell Tumor	Q8WWU7, Q6PEY2
Ovary Serous Cystadenocarcinoma	P0DTE8, Q14508, P31025
Kidney Renal Clear Cell Carcinoma	P09923, P07911
Kidney Chromophobe	P07911, P06870, P24855, P01133
Prostate Adenocarcinoma	P27487, P25311, P15309, Q96DA0, P14555
Kidney Renal Papillary Cell Carcinoma	P15144, Q9BYE9

Outlook

Proteomes in the flow: proteomic insights into engineered water environments



Outlook

Microbial communities drive key processes in our global ecosystem, such as nutrient cycling and the turnover of organic matter. While culture-independent methods like metagenomics offer direct insights into microbial diversity and the potential functions of these communities, they do not reveal the actually expressed metabolic activities or how microbes respond to environmental changes.

Mass spectrometry-based microbial community proteomics, also known as metaproteomics, provides quantitative insights into the proteins, enzymes, and metabolic pathways expressed by microbial communities. Incorporating whole metagenome sequencing data as a reference for metaproteomic analysis improves proteome coverage and enables species-level, or in some cases even strain-level resolution. In addition, using stable isotope-labeled substrates allows to trace nutrient utilization across community members. Together, these tools provide a system-level understanding of complete microbial ecosystems and how they respond to environmental changes.

Using microbial proteomics, we demonstrated that *Aeromonas* can efficiently degrade and grow on chitin as its sole carbon and nitrogen source. *Aeromonas* not only breaks down chitin using a dedicated set of extracellular hydrolytic enzymes but also expresses a range of transporters and metabolic enzymes to funnel the degradation products into its central carbon metabolism.

In addition, we found that *Aeromonas* can degrade several other polymers, including pullulan, starch, and dextrin. A microbial proteomics study of the secretome revealed a diverse set of enzymes produced and secreted when the bacterium is grown in the presence of specific polymers.

However, conventional protein extraction protocols still recover membrane proteins poorly, meaning that valuable information about dedicated transporters is often lost. Furthermore, for approximately 20% of the secreted proteins, no functional classification could be assigned using existing databases and prediction tools. This highlights that a substantial fraction of proteins still lack characterized functional analogues. This is especially relevant for soil and sediment environments, which are highly complex and harbor a vast array of yet unknown microbes. In such environments, the proportion of proteins with unknown functions was found to be even higher.

In the whole metagenome sequencing data from the loose deposits microbiome, many proteins could not be assigned a taxonomic origin or functional classification. However, these so-called “unassigned” sequences may still hold valuable information about the structure and function of the organisms they originate from. None of the advanced algorithms or software tools currently available for function prediction based on amino acid sequence were able to reveal their functions [1-6]. More recently, however, advanced

machine learning algorithms such as AlphaFold, which integrates physical and biological knowledge about protein structure, and ESMFold, which leverages large protein language models trained on millions of sequences, have emerged [7-13]. When combined with structure-based search tools these approaches offer new opportunities to predict protein function based on their predicted 3D structures. Despite these advances, large-scale structure prediction, such as for complete metagenomic datasets, remains too time consuming for routine analysis. Looking ahead, improvements in model efficiency and computational power are expected to make structure-based function prediction a regular part of metagenomic and community proteomic workflows.

The metagenomic and metaproteomic analysis of the drinking water loose deposits revealed a highly diverse microbial community. Interestingly, a large fraction of the abundant members showed the potential to degrade a broad range of biopolymers, including cellulose, xylan, and chitin. However, while this indicates genetic potential, experimental confirmation that these microbes can actually utilize these biopolymers is still needed. A possible next step in confirming their ability to degrade and utilize different biopolymers would be to perform stable isotope probing experiments using either 13C or 15N labelled biopolymers. Protein stable isotope probing has already been applied to study nutrient fluxes in microbial enrichments [14] and complex microbial ecosystems [15, 16]. For the loose deposits, this could potentially be done in a bioreactor setup where labelled chitin is introduced, and the incorporation of heavy isotopes into proteins is tracked over time by metaproteomics. This approach would allow us to uncover the temporal sequence of polymer utilization by individual microbes within the community and shed light on food web relationships between microorganisms. However, the current metaproteomics data from the loose deposits sample identified only a few proteins or even just a handful of peptides for some of the lower abundant members. To better resolve these relationships, a deeper proteome coverage would be needed. The more recently released Astral Orbitrap mass spectrometer already provides significantly deeper coverage for highly complex communities compared to routine instruments like the quadrupole Orbitrap mass spectrometers as demonstrated recently [17]. Incorporating such advanced instrumentation would potentially reveal sufficient detail to track microbial activity and interactions of the major taxonomies in these systems.

Metagenomic and metaproteomic analyses of drinking water loose deposits only identified trace levels of *Aeromonas*. Several studies have reported that *Aeromonas* is often difficult to detect using culture-independent molecular techniques [18, 19]. Furthermore, *Aeromonas* is mostly found in the drinking water and only small quantities may be attached to the deposits. However, in order to quantify *Aeromonas* in these environments, a very sensitive, targeted proteomics approach could be developed. For example, parallel reaction monitoring has been successfully employed to detect SARS-CoV-2-specific peptides in

clinical samples with extremely low viral biomass [20]. Developing a targeted method using one of the most advanced mass spectrometers could allow quantifying *Aeromonas*, or other low-abundance indicator strains through a culture-dependent method.

Additional optimization of molecular techniques using a combination of in solution digestion and FASP allowed extraction of different types of proteins present in the wastewater. Novel approaches in cell lysis have been investigated and successfully applied for pure cultures [21]. This can be the start to adapt these types of protocols to more complex environments or environments with low bacterial load.

An ongoing challenge in metaproteomic analysis is the construction of a reference sequence database that enables deep proteome coverage and reliable taxonomic classification. In Chapter 5, we presented an alternative approach by tailoring public reference sequence databases using de novo sequencing. However, accessing large databases via an API can be time-consuming, and using a single, unified database for multiple conditions may introduce bias in the taxonomic profile, potentially leading to a false sense of homogeneity across different samples. Further advancements for database construction and the use of very large databases need to be developed. Employing HPC clusters for constructing and filtering public databases would already accelerate this procedure.

Wastewater based epidemiology proved to be an important tool for keeping the recent COVID pandemic under surveillance [22-24]. However, wastewater-based metaproteomics is still in its early stages. A large and highly informative portion of wastewater proteins are of human origin. While these human proteins hold potential as biomarkers, they are not yet utilized in wastewater-based epidemiology. Quantifying them in such a complex matrix remains a significant challenge, and the degradation dynamics of proteins in wastewater are still poorly studied and understood. For instance, protein glycosylation, especially in plasma proteins, could provide valuable insights into biomarker status. Although this is a well-established area of research in clinical proteomics, it has been largely overlooked in the context of wastewater. Moreover, human proteins are known to carry a wide range of post-translational modifications, which could offer further biological insights if systematically studied in wastewater samples. Although mass spectrometry-based metaproteomics has proven to be a valuable tool for profiling all proteins present in wastewater, a standardized, high-throughput workflow is still needed to establish it as a routine analytical method in wastewater-based epidemiology.

Technological advancements in mass spectrometry have transformed our ability to identify and quantify proteins, enabling large-scale proteomic analysis even from complex sample matrices such as cell lysates [25]. Over the last two decades the field has slowly transitioned into the field of microbial ecology [26]. However, several bottlenecks remain. Currently, mass spectrometry is limited in terms of throughput, as it sequences peptides one by one, and sensitivity, since proteomics-based techniques lack an amplification step, unlike genomic approaches. The most commonly used method for protein identification is bottom-up

proteomics, which involves digesting proteins into peptides and then sequencing those peptides individually. However, there are limitations towards proteome coverage for very complex samples, and the digestion into peptides introduced the problem of peptide inference [27, 28]. Recently, very sensitive and fast mass spectrometers were introduced, such as the Astral Orbitrap mass spectrometer and the TimsTOF mass spectrometer (including parallel accumulation-serial fragmentation PASEF) have emerged leading to very short measurement times and exceptional proteome coverage [29, 30]. Other technological advancements involve enhancements in sample separation and sensitivity, particularly within the domain of liquid chromatography [31]. The introduction of new chromatographic separation columns such as the micro-pillar array columns show very low backpressure and therefore allow to employ very long separation columns that ultimately provide high-resolution separation of complex samples [32, 33]. Acquisition strategies such as data-independent acquisition improved over the past few years, with multiple novel software tools being developed to aid in the analysis of DIA experiments from even complex metaproteomics samples [34, 35].

However, major bottlenecks remain unresolved, namely achieving ultra-high sensitivity, potentially down to single-protein detection, enabling parallel sequencing or identification, and accurately mapping post-translational modifications on intact proteins. As a result, alternative approaches to mass spectrometry are actively being explored, with the potential to transform the field of proteomics in the future. One such emerging approach is the application of nanopore sequencing to proteins, a significantly more complex challenge than DNA sequencing, which involves only four nucleotides. Despite the complexity, early studies have demonstrated promising proof-of-concept results, suggesting that nanopore-based protein sequencing could one day become a viable tool for large-scale proteomics [36, 37]. Still, substantial challenges remain, particularly in enhancing the accuracy of distinguishing individual amino acids. A key limitation lies in the subtle differences in ionic current (or gap voltage) generated by different amino acids, which makes precise identification difficult. Moreover, post-translational modifications and chemical alterations further complicate how amino acids behave in an electric field. Finally, in their natural state, proteins exist as folded 3D structures, not as linear chains, adding another layer of complexity to the application of this approach.

An alternative approach in single protein analysis is “single protein fingerprinting”. This technique enables the detection and targeting of specific proteins through fluorescent labeling of selective amino acids [38]. The resulting fluorescence fingerprint can be captured and compared to reference sequence databases for identification. Offering a complementary approach to both bottom-up and top-down proteomics, this method holds significant promise in microbial ecology, particularly for functional profiling in low-biomass samples. Its enhanced sensitivity enables protein detection in single cells, facilitating the study of rare and unculturable microorganisms. Currently, single protein fingerprinting

stands out as one of the most promising alternatives to mass spectrometry-based proteomics.

While mass spectrometry-based metaproteomics is still an emerging application and has seen significant advancements over the past years, other emerging techniques such as single protein sequencing or fingerprinting continue to evolve, holding the potential to provide even deeper and more comprehensive insights into complex ecosystems.

References

1. Eddy SR. Accelerated Profile HMM Searches. PLoS computational biology. 2011;7(10):e1002195.
2. Blum M, Andreeva A, Florentino Laise C, Chuguransky Sara R, Grego T, Hobbs E, et al. InterPro: the protein sequence classification resource in 2025. Nucleic acids research. 2024;53(D1):D444-D56.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403-10.
4. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic acids research. 2020;49(D1):D412-D9.
5. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic acids research. 2012;40(W1):W445-W51.
6. Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, et al. PROSITE: a documented database using patterns and profiles as motif descriptors. Briefings in bioinformatics. 2002;3(3):265-74.
7. Kulmanov M, Hoehndorf R. DeepGOPlus: improved protein function prediction from sequence. Bioinformatics. 2019;36(2):422-9.
8. Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, et al. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. IEEE transactions on pattern analysis and machine intelligence. 2022;44(10):7112-27.
9. Gligorijević V, Renfrew PD, Kosciolek T, Leman JK, Berenberg D, Vatanen T, et al. Structure-based protein function prediction using graph convolutional networks. Nature Communications. 2021;12(1):3168.
10. Asgari E, Mofrad MR. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. PLoS One. 2015;10(11):e0141287.
11. Brandes N, Ofer D, Peleg Y, Rappoport N, Linial M. ProteinBERT: a universal deep-learning model of protein sequence and function. Bioinformatics. 2022;38(8):2102-10.

12. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.
13. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. 2023;379(6637):1123-30.
14. Lawson CE, Nuijten GHL, de Graaf RM, Jacobson TB, Pabst M, Stevenson DM, et al. Autotrophic and mixotrophic metabolism of an anammox bacterium revealed by in vivo 13C and 2H metabolic network mapping. The ISME Journal. 2021;15(3):673-87.
15. Taubert M, Vogt C, Wubet T, Kleinsteuber S, Tarkka MT, Harms H, et al. Protein-SIP enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial consortium. 2012;6(12):2291-301.
16. Pan C, Fischer CR, Hyatt D, Bowen BP, Hettich RL, Banfield JF. Quantitative tracking of isotope flows in proteomes of microbial communities. Molecular & cellular proteomics : MCP. 2011;10(4):M110.006049.
17. Dumas T, Martinez Pinna R, Lozano C, Radau S, Pible O, Grenga L, et al. The astounding exhaustiveness and speed of the Astral mass analyzer for highly complex samples is a quantum leap in the functional analysis of microbiomes. Microbiome. 2024;12(1):46.
18. Liu G, Van der Mark EJ, Verberk JQJC, Van Dijk JC. Flow Cytometry Total Cell Counts: A Field Study Assessing Microbiological Water Quality and Growth in Unchlorinated Drinking Water Distribution Systems. 2013;2013(1):595872.
19. Vavourakis CD, Heijnen L, Peters M, Marang L, Ketelaars HAM, Hijnen WAM. Spatial and Temporal Dynamics in Attached and Suspended Bacterial Communities in Three Drinking Water Distribution Systems with Variable Biological Stability. Environmental science & technology. 2020;54(22):14535-46.
20. Bezstarosti K, Lamers MM, Doff WAS, Wever PC, Thai KTD, van Kampen JJA, et al. Targeted proteomics as a tool to detect SARS-CoV-2 proteins in clinical specimens. PLoS One. 2021;16(11):e0259165.
21. Abele M, Doll E, Bayer FP, Meng C, Lomp N, Neuhaus K, et al. Unified Workflow for the Rapid and In-Depth Characterization of Bacterial Proteomes. Molecular & Cellular Proteomics. 2023;22(8).
22. Betancourt WQ, Schmitz BW, Innes GK, Prasek SM, Pogreba Brown KM, Stark ER, et al. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. Science of The Total Environment. 2021;779:146408.
23. Spurbeck RR, Minard-Smith A, Catlin L. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. The Science of the total environment. 2021;789:147829.

24. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. The Science of the total environment. 2020;728:138764.
25. Makarov A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Analytical Chemistry. 2000;72(6):1156-62.
26. Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M, Schulze ED. A proteomic fingerprint of dissolved organic carbon and of soil particles. Oecologia. 2005;142(3):335-43.
27. Tsiatsiani L, Heck AJJTFj. Proteomics beyond trypsin. 2015;282(14):2612-26.
28. Miller RM, Smith LM. Overview and considerations in bottom-up proteomics. The Analyst. 2023;148(3):475-86.
29. Rethink what is possible Orbitrap Astral mass spectrometer [Available from: https://pragolab.cz/documents/br-001728-ms-orbitrap-astral-mass-spectrometer-br001728-en.pdf.
30. Koch HM, Kaspar-Schoenefeld S, Goedecke N, Raether O, Drechsler N, Krause M, et al., editors. PASEFTM on a timsTOF Pro defines new performance standards for shotgun proteomics with dramatic improvements in MS/MS data acquisition rates and sensitivity2018.
31. Broeckhoven K, Desmet G. Advances and Innovations in Liquid Chromatography Stationary Phase Supports. Analytical Chemistry. 2021;93(1):257-72.
32. Scientific T. Low-flow HPLC columns- Enabling high sensitivity LC-MS analysis for bottom-up and top-down proteomic research 2023 [Available from: https://assets.thermofisher.com/TFS-Assets/CMD/Flyers/fl-000478-ccs-lc-ms-low-flow-hplc-columns-portfolio-fl000478-en.pdf.
33. Rajczewski AT, Blakeley-Ruiz JA, Meyer A, Vintila S, McIlvin MR, Van Den Bossche T, et al. Data-Independent Acquisition Mass Spectrometry as a Tool for Metaproteomics: Interlaboratory Comparison Using a Model Microbiome. bioRxiv : the preprint server for biology. 2025.
34. Kitata RB, Yang J-C, Chen Y-J. Advances in data-independent acquisition mass spectrometry towards comprehensive digital proteome landscape. 2023;42(6):2324-48.
35. Barkovits K, Pacharra S, Pfeiffer K, Steinbach S, Eisenacher M, Marcus K, et al. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. Molecular & cellular proteomics : MCP. 2020;19(1):181-97.
36. Ouldali H, Sarthak K, Ensslen T, Piguet F, Manivet P, Pelta J, et al. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nature Biotechnology. 2020;38(2):176-81.

37. Yu L, Kang X, Li F, Mehrafrooz B, Makhamreh A, Fallahi A, et al. Unidirectional single-file transport of full-length proteins through a nanopore. Nature Biotechnology. 2023;41(8):1130-9.
38. Filius M, van Wee R, de Lannoy C, Westerlaken I, Li Z, Kim SH, et al. Full-length single-molecule protein fingerprinting. Nature Nanotechnology. 2024;19(5):652-9.

“Sometimes, when you've a very long street ahead of you, you think how terribly long it is and feel sure you'll never get it swept.'[...] 'And then you start to hurry,' he went on. 'You work faster and faster, and every time you look up there seems to be just as much left to sweep as before, and you try even harder, and you panic, and in the end you're out of breath and have to stop - and still the street stretches away in front of you. That's not the way to do it.' [...]'You must never think of the whole street at once, understand? You must only concentrate on the next step, the next breath, the next stroke of the broom, and the next, and the next. Nothing else.'[...] 'That way you enjoy your work, which is important, because then you make a good job of it. [...] 'And all at once, before you know it, you find you've swept the whole street clean, bit by bit. What's more, you aren't out of breath.' He nodded to himself. 'That's important, too,' he concluded.”

-Michael Ende, Momo-

Acknowledgements

The work presented in this dissertation would not have been possible without the support of many people, current and former employees and staff members. I would like to express my gratitude to all of you. Each of you chiseled the work presented here in various ways, be it with hands-on experiments, discussions, ideas or even emotional support.

First of all, I would like to thank **Martin** for offering me the chance to continue to work on this project that I started during my Master's. You helped me during my rough times and did not stop believing in me even when I started doubting myself. Your encouragement, patience and proficiency in mass spectrometry inspired me to learn more and surpass myself every time. You understood my need to explore and try out new things in the lab or in my analysis and you tried to meet me halfway. I will never forget your encouragement to explore and expand ideas, while at the same time, not letting myself get discouraged by the fear of possible feedback from reviewers. **Mark**, thank you for fulfilling the role of promotor and having faith in me to work on this project. Your expertise in biotechnology and microbiology was extremely helpful. Your insights during our meetings guided me into becoming a more well-rounded researcher.

I would like to give special thanks for my students, that worked with me during my PhD, **Willem**, **Kaatje** and **Filine**. Your contributions helped shape the research that I conducted. I am glad that I had the opportunity to be your supervisor and learn a lot from you, shaping me not only as a teacher but also as a person.

I would like to dedicate the following lines to my dear paranymphs to whom I am very grateful. **Maxime**, you were one of my first office mates, sitting next to me back when we were both in CSE. My PhD journey would not have been the same without you. Our discussions over different topics during our walks and your emotional support helped me overcome the harsh times of my PhD. I am very happy I met you and that I got to know you as a person. Your strong-willed, free-spirited nature inspire me to dare more in life and take on adventures with a smile on my face. **Carol**, you were the first person I worked with in the lab during my Master's project. You were the one delegated to supervise me and honestly, I could not ask for someone better. Your friendly personality and extensive knowledge of protein extraction built the pillars of my knowledge that I have today. Your discipline has always been inspiring, and I love that we share hobbies in common that we can freely discuss about.

The last 4 years of my life have been spent in the Mass-spectrometry group and the people that are part of this group deserve a huge shoutout for all their work and dedication. **Dita**, since you came into our lab, it looks better and more organized. Your diligence and willingness to help others is driving our projects further. **Hugo**, I always admired your skill in making very nice pictures and graphs, as well as your opennes to talk about anything. **Pim**, I always saw you as a very cheerful person and fun to talk to. **Jitske**, **Berdien** and

Ramon, I wish you good luck and all the best with your projects. **Shree,** your kindness and politeness are truly inspirational. I am happy I got to know you. **Yanfang**, thank you for helping me with the sugar measurements. I always admired your ambition and work ethic.
Further on I would like to thank my office mates that I had during my Master's and PhD. **Carla,** I always admired your resilience and your warm personality. Your work ethic and ambition is truly inspiring. **Agi,** it was very fun to share the office with you and listen to your stories from your PhD and all the sports you were doing. At the beginning of my PhD, I landed in the EBT group and for the last 2 years I shared the office with **Ali**, **Jin** and **Gonçalo**. It was fun to share the office with you guys and have interesting discussions about our projects.
Special thanks should be given to the backbone of every research group, the technicians. Either it was about bacteria culturing, microscopy or DNA extraction, **Ben** and **Zita** were there to guide. Therefore, I would like to thank them for their work and dedication. I would also like to thank **Apilena**, **Astrid** and **Jannie** for their help with the autoclave.
Further on I would like to thank my colleagues from the EBT group **Timmy**, **Francesc**, **Nina**, **Stefan**, **Bea**, **Marit**, **Siem**, **Natalia**, **Lemin**, **Puck, Linghang** and **Jelle** and all others for enjoying our lunch together on the "always too small" table in the coffee corner.
Dmitry, thank you for letting me work in your lab. You taught me what a real researcher is and what it takes to conduct high quality research. **Rebeca**, thank you for the discussions that we had about different strategies of substrate uptake.
Wim and **Julia** I would like to thank you both for offering me the opportunity to work together on the drinking water project. I would also like to thank the Evides team for providing insightful feedback and interesting discussions. **Wim** thank you for the contributions and discussions on Chapter 4.

Further, I would like to thank my friends and family that have been there for me and that have inspired and helped me in my work and endeavors.
Annie I loved sharing the house with you in The Hague. I loved that we can discuss freely about a lot of things and we vibe so well. **Andrea**, thank you very much for all the lessons you have given me. Your teaching skills and talent helped me tremendously and gave me courage to be more creative and courageous in expressing myself thorugh music.

Mi-aş dori să dedic următoarea secţiune prietenilor şi familiei care mi-au fost aproape. Ei sunt inspiraţia şi ambiţia mea de a deveni un om mai bun şi să excelez in tot ceea ce întreprind.
Evelyn, ne cunoaştem de la grădiniţă, si cumva, indiferent de distanţă, am orbitat mereu una în jurul celeilalte. Nu trebuie sa ne vorbim des dar ştim mereu ce vrem să ne spunem. Sunt foarte norocoasă că îmi eşti prietenă şi extrem de mândră de tot ceea ce ai realizat

până acum. Sunt convinsă că munca ta în analiza şi protejarea democraţiei ne vor aduce multe beneficii.

Cătă şi **Teo**, Aristotel spunea că un "prieten e un suflet împărţit în două corpuri". Sufletul nostru e prea mare aşa că a fost împărţit in 3. Sunteţi tot ce o persoana ar putea visa de la o prietenie ca a noastra. Sunteţi un cadou pe care simt că viaţa mi l-a dat nemeritat în doză dubla. Încă imi aduc aminte toate păţaniile din copilărie prin care am trecut, şi chiar şi acum, după atâţia ani, mă apucă râsul. Sunt extrem de norocoasă că v-am întâlnit şi sunteţi un element important al devenirii mele ca om şi ca cercetător.

Mihai, cuvintele sunt de prisos când mă gândesc ce ar trebui să-ţi spun. Îţi mulţumesc pentru tot ajutorul şi toată susţinerea, pe care mi le-ai oferit necondiţionat de-a lungul acestui doctorat. Pasiunea şi dedicarea în tot ceea ce faci au prezentat pentru mine fundamentele dupa care îmi ghidez mai departe deciziile. Restul lucrurilor le ştii deja. Mi-aş dori să le mulţumesc părinţilor tăi, **Gabriela** şi **Marian** pentru că m-au acceptat în familia lor.

Dragă **tata,** doresc să încep prin a-ţi mulţumi pentru disciplina şi ambiţia pe care mi le-ai insuflat. Îţi multumesc că, încă de când eram mică, petreceai timp cu mine şi îmi răspundeai la întrebări într-un mod matur şi concis, ca unui om matur, chiar dacă eu nu înţelegeam şi îţi puneam şi mai multe întrebări. Ție îţi mulţumesc pentru pasiunea pe care am dezvoltat-o pentru ştiinţă, când mă luai cu tine în bucătarie şi mă învăţai că sarea se dizolvă în apă şi că uleiul şi apa sunt nemiscibile. Îţi sunt recunoscătoare pentru serile în care dezbăteam evenimente istorice şi făceam schimb de idei pe marginea lor. Sunt încă miliarde de subiecte pe care mi-aş dori să le dezbat cu tine şi sper ca timpul să fie prietenos cu noi şi să ne faciliteze asta.

Curriculum vitae

Claudia Tugui was born on January 11, 1995, in Ploiești, Romania. Her passion for science led her to Cluj-Napoca in 2014, where she began her Bachelor's studies in Biochemistry at Babeș-Bolyai University. During this time, she gained her first hands-on research experience at the Institute of Biological Research, investigating the presence of antibiotic-resistant genes in manure and wastewater. In 2017, Claudia moved to the Netherlands to pursue a Master's degree in Microbiology at Radboud University. Her research journey there included two formative internships. The first, with the Environmental Microbiology group under Dr. Sebastian Luecker, focused on optimizing a CLICK-chemistry method for live staining of *commamox* bacteria. The second, carried out under the supervision of Dr. Martin Pabst at Delft University of Technology and Dr. Wim Hijnen from Evides Waterbedrijf N.V., explored the use of metaproteomics to study microbial communities in drinking water systems. In 2020, Claudia embarked on her PhD in Biotechnology at Delft University of Technology, under the supervision of Dr. Martin Pabst and Prof. Mark van Loosdrecht, in partnership with Evides Waterbedrijf N.V. Her research focuses on uncovering the drivers behind bacterial regrowth in drinking water. Alongside this, she is advancing the application of metaproteomics for monitoring wastewater and developing new bioinformatic tools to improve taxonomic assignment in metaproteomic studies. Her work bridges fundamental science with practical applications, aiming to better understand and manage microbial processes in water systems.

Key Publications and Conferences

- **C.G. Tugui**, F. Cordesius, W. van Holthe, M.C. van Loosdrecht, M. Pabst, bioRxiv, 2025-02.
- **C.G. Tugui**, D.Y. Sorokin, W. Hijnen, J. Wunderer, K. Bout, M.C.M. van Loosdrecht, M. Pabst, RSC Chemical Biology, 2025
- D.Y. Sorokin, A.Y. Merkel, E. Messina, **C. Tugui**, M. Pabst, P.N. Golyshin, M.M. Yakimov, The ISME Journal, 2022/6
- H.B.C. Kleikamp, M. Pronk, **C. Tugui**, L. Guedes da Silva, B. Abbas, Y.M. Lin, M C.M. van Loosdrecht, M. Pabst, Cell Systems, 2021/5/19

- 2024 – HUPO World Congress, The secret(ome) behind microbial survival in nutrient poor aquatic environments (Poster)
- 2023 - KNMV Spring Meeting, Unraveling the ability of Aeromonas *sp*. to degrade and utilize the biopolymer chitin (Oral presentation)
- 2022 – International Mass Spectrometry Conference (IMSC), Metaproteomics: a new tool in wastewater surveillance (Oral presentation)