ISME Communications, 2026, 6(1), ycaf243
https://doi.org/10.1093/ismeco/ycaf243
Advance access publication: 18 December 2025
Original Article

Wastewater metaproteomics: tracking microbial and human protein biomarkers

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

Department of Biotechnology, Delft University of Technology, Delft, HZ 2629, The Netherlands

*Corresponding author. Martin Pabst, Department of Biotechnology, Delft University of Technology, van der Maasweg 9, Delft, HZ 2629, The Netherlands. E-mail: m.pabst@tudelft.nl

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 approximately 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.

Keywords: metaproteomics; wastewater; wastewater-based epidemiology; biomarkers; gut microbes

Introduction

Globally, ~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]. 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 [8–10]. This informed governmental bodies and the general public about the spread of SARS-CoV-2. Several reviews built on the idea of WBE to apply different analytical methods that can target proteins in the wastewater like human biomarkers [11, 12]. Furthermore, the presence of certain bacteria also informs on the spread of antimicrobial resistance, and various diseases [13–19].

Apart from the advantage of anonymity, the collection of wastewater 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 [33–35], autoimmune diseases [36], neurological dysfunctions like Alzheimer's disease [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

Received: 22 June 2025. **Revised:** 9 November 2025. **Accepted:** 15 December 2025

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, noninvasive, 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 measurement of 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 proteins present. 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 proteomic 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 analyses of soluble and particulate fractions and concentrating 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 of proteins 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 that we applied to crude municipal wastewater samples collected over approximately 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 approximately three months, from November 2023 to February 2024. From the wastewater treatment plant Harnaschpolder [6] samples were taken on 29/11/23, 12/12/23, 24/01/24, 31/01/24, and 20/02/24, and from Utrecht (UT), samples were taken on 29/11/23, 05/12/23, 24/01/24, 23/02/24 and 27/02/24. The sampling was done from the influent, raw sewage, on days with low precipitation. After sampling influent wastewater, samples were stored at −20°C until further processed. Influent samples were collected from the buffer tank in 50 mL Falcon tubes, with sewer system retention times typically between 12 and 24 h. Before taking and processing 500 μL of wastewater in duplicate, for proteomic analysis, each sample was subjected to prolonged vortexing to ensure homogenization. Both sampling areas are highly urbanized where the Utrecht plant is located near the city center and serves only the city of Utrecht. In contrast, Harnaschpolder is the largest plant in the country, serving a broader region. Neither is significantly influenced by agricultural or industrial activities. The main difference in retention time is in the sewer, which is shorter in Utrecht compared to Harnaschpolder.

Protein extraction and proteolytic digestion

500 μL of the wastewater influent was taken and diluted with B-PER (175 μL) and 50 mM TEAB (triethylammonium bicarbonate) 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 three times for 1 min using a benchtop vortex mixer, sonication on a sonication bath for 15 min and one freeze/thaw cycle (frozen at −80°C, and thawed in an incubator at 40°C for 5 min). The samples were then Spun down on a benchtop centrifuge and the supernatant transferred to a 1.5 mL LoBind Eppendorf tube. TCA (trichloroacetic acid 100% stock) was added to the sample at a ratio of 1:4 (v/v, TCA/sample) to a final TCA concentration of 20%, vortexed and incubated at 4°C for 20 min. After centrifugation, the protein pellet was re-solubilized in 6 M urea and then reduced with DTT (dithiothreitol) and alkylated using IAA (iodoacetamide) [41, 51]. After alkylation, the sample was transferred to a FASP filter (Millipore, MRCPRT010) which was previously conditioned by washing 2 times with 100 mM ABC (ammonium bicarbonate) buffer. The filters were centrifuged at 14 K rpm in a bench top centrifuge, for 45 min, and then 2 times at 14 K rpm for 40 min 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 μg/μL in 1 mM HCl, Promega, Cat No) in 100 μL 100 mM ABC. The FASP filters were incubated overnight 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 (acetonitrile) in H_2O containing 0.1% FA (formic acid) in order to collected the proteolytic peptides. The pooled peptide 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 dried 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 ~0.5 μg/μL. Shotgun metaproteomics was performed as described previously [41], with a randomized sample order. Briefly, ~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 × 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 35% solvent B over 90 min, from 35% to 65% over 30 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 to 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 100 ms, 2.5 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 (average local confidence) score ≥70 were subjected to taxonomic profiling using the NovoBridge pipeline as described previously [46]. 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.

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 (peptide spectrum matches) level, and protein identifications with ≥2 unique peptide sequences were considered significant. Evaluation of the employed two-round search for sensitivity and specificity is outlined in the SI EXCEL DOC, where entrapment experiments showed that the number of additional false positives was very low and within the FDR range estimated by PEAKS. The complete dataset was combined using the PEAKSQ module allowing 10 min RT (retention time) shifts and 10 ppm mass error [41, 45]. All identified proteins were exported and further processed as described in the following. Analysis for the presence of oxonium ions was performed as described previously [52]. N-glycosylation profile analysis was carried out manually and by using Byonic v3.3.9 [53], searching the data against the human proteome with a 20 ppm precursor mass tolerance, a 0.02 Da fragment mass tolerance, carbamidomethylation as a fixed modification, and 50 common biantennary N-glycans as variable modifications. The protein FDR was set to 2%, and the resulting search outputs were manually reviewed. Glycopeptides with a PEP 2D score < 0.01 (a Byonic defined metric) were used to generate N-glycan modification frequency histogram.

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 superkingdom 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 where 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 $H' = -\sum pi \ln(pi)$, 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.* [54]. 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 [55]. The human dataset was further analyzed for the enrichment of molecular and cellular functions, and pathways using STRING [56]. Potential cancer related protein biomarkers were taken from the Human Protein Atlas (www.proteinatlas.org) [57], which was queried via an API. Human proteins associated with coronary artery disease (CAD) were taken from the CAD biomarkers database [58], diabetic nephropathy (DN) from Zürbig *et al.* (2012) [59], breast cancer from Beretov *et al.* (2015) [60], urothelial cancer from Chen *et al.* (2021) [61], Abdominal-type Henoch–Schönlein purpura (HSP) from Jia *et al.* (2021) [62], prostate cancer from Fujita *et al.* (2018) [63], and for ovarian cancer and inflammatory bowel disease (IBD) from Owens *et al.* (2022) [64]. Parts of Fig. 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 identifier PXD059455.

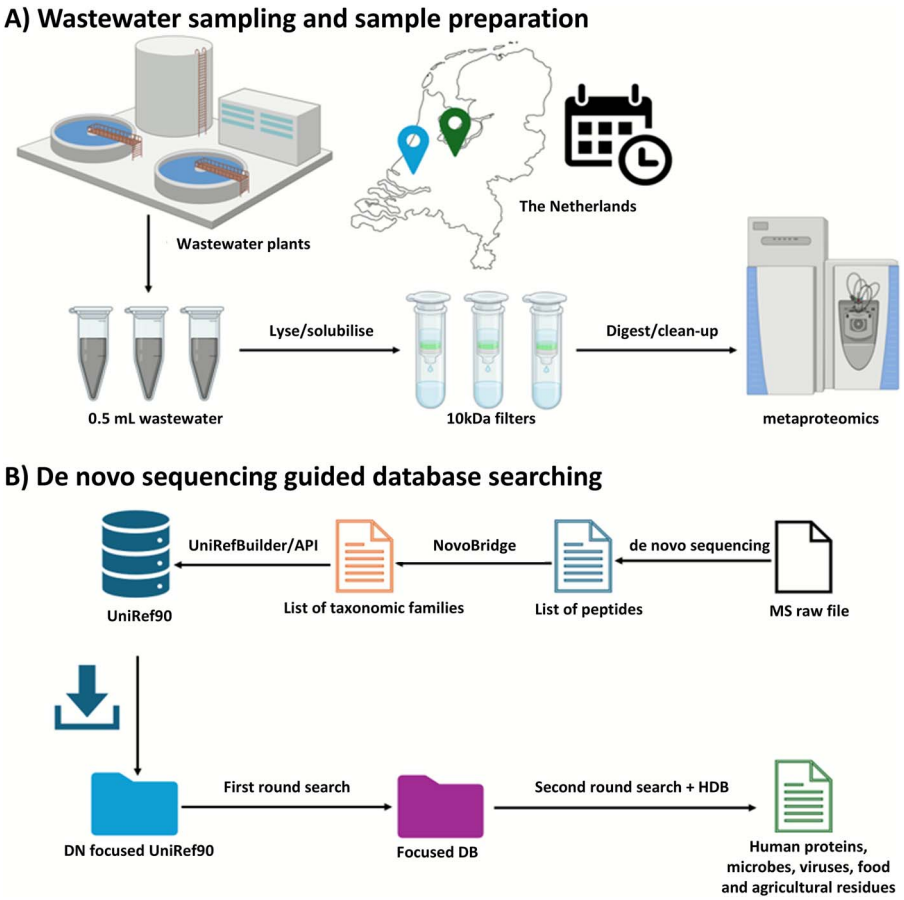

Figure 1. The wastewater metaproteomics workflow involved rapid protein extraction and de novo sequencing guided database searching to identify a broad spectrum of proteins across all domains of life. Influent samples were collected from two wastewater treatment plants, located in Utrecht and Harnaschpolder, 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 emplying a filter-aided sample preparation (FASP) approach. 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-round database searching using a global reference database to enhance protein identification.

Results

Wastewater samples from two different wastewater treatment plants were sampled during the winter period over approximately 3 months (Nov 2023–Feb 2024). The wastewater treatment plant in Harnaschpolder is the biggest wastewater treatment plant in the Netherlands, serving over 1 million inhabitants and 40 000 companies in the The Hague area. The wastewater treatment plant located very close to the city of Utrecht serves the city and the surrounding area of around 375 000 people. The wastewater treatment plant located in Utrecht is close to the city serving a highly urbanized area, with relatively low sewer residence times. Harnaschpolder, on the other hand, is located between The Hague and Delft, in a peri-urban zone including different larger and smaller industries. The wastewater is transported over longer distances by pressure mains. To capture a comprehensive spectrum of taxonomies from the sampled wastewater, we developed a streamlined wastewater metaproteomics workflow. This included the extraction and analysis of proteins from both the liquid and solid components starting from small volumes of influent wastewater (raw sewage). For this, 0.5 mL of wastewater was directly mixed with a lysis buffer, followed by solubilization and cell lysis using a combination of sonication and freeze–thaw cycles. Proteins were then extracted and digested employing a filter-aided sample preparation (FASP) workflow. The proteolytic peptides were finally analyzed with a shotgun metaproteomics experiment employing a one-dimensional chromatographic separation (with a 120 min gradient) and a hybrid quadrupole-Orbitrap mass spectrometer. This allowed the detection of 4919 proteins across 3009 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 (~10%–15%), these accounted for more than 50% of the total protein abundance (considering the top 3 peptide area per protein) (Fig. 2A). 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, dietary consumption by humans or animals or naturally occurring algae belonging to *Streptophyta*. 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 diverse 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 (Fig. 2B).

The most dominant families were 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 some *Escherichia coli* [65]. These also included opportunistic pathogens, such as *Weeksellaceae*, *Leptotrichiaceae*, *Aeromonadaceae*, and *Arcobacteraceae*. Other families such as *Planctomycetaceae*, *Nitrobacteraceae*, *Comamonadaceae*, and *Rhodospirillaceae* are often associated with biological processes including nitrogen cycling and organic matter degradation [66–68]. Additionally, biodegraders like *Pseudomonadaceae*, *Bacillaceae*, *Chitinophagaceae*, *Flavobacteriaceae*, *Burkholderiaceae*, and *Sphingomonadaceae* were detected which play crucial roles in the breakdown of organic material and xenobiotics [69–74]. Interestingly, also *Caldilineaceae* and *Nocardioidaceae* were detected which are frequently linked to operational challenges in wastewater treatment, such as sludge bulking and foaming [75, 76]. Also, *Zoogloeaceae* were detected, which are major denitrifying bacteria that also produce extracellular polymeric substances relevant for floc formation [77]. Members of the families *Moraxellaceae* and *Burkholderiaceae* have been frequently linked to opportunistic infections [78–83].

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. Previous studies have demonstrated that protein and DNA residues derived from human dietary intake can be detected using omics-based approaches such as metagenomics and metaproteomics [84, 85]. This opens new perspectives and enables the assessment of general dietary habits within the human population in an unbiased manner through the analysis of wastewater composition. However, also proteins from potential pathogenic vectors such as *Nematoda* (e.g. *Steinernematidae* [86]) 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 [87]. 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 (Fig. 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 (Fig. 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 (Fig. 3). For example, in HP, *Aeromonadaceae* and *Weeksellaceae*, which are common in wastewater and are associated with potential pathogenic species [51] and are known to harbor antibiotic resistance genes or coexist with microbes involved in the spread of antimicrobial resistance [51, 88, 89], showed a much higher relative abundance compared to UT. Additionally, *Streptococcaceae* and *Vibrionaceae* were more abundant in HP, both of which are fecal indicators and potential pathogens [90, 91]. In contrast, *Sphingomonadaceae*, which are well-known biodegraders of xenobiotics in wastewater [72, 74], 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 [92, 93], while the other includes members, such as *Nocardioides* which can degrade a variety of pollutants [94]. 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.

The human wastewater proteome

Over 190 protein 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 ~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 (Fig. 4 and SI EXCEL DOC). Nevertheless, many of these proteins are promising biomarkers for accessing the

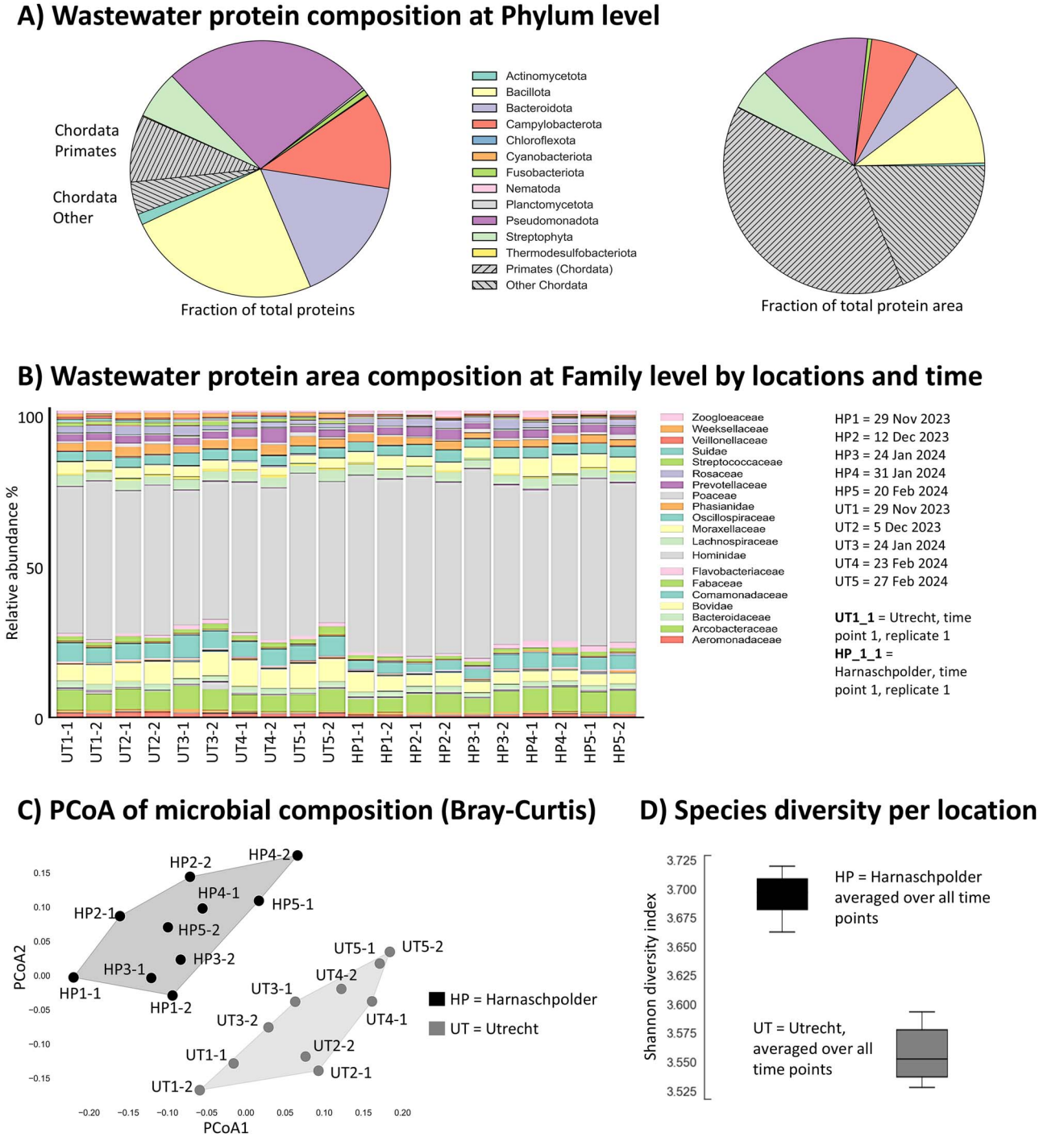

Figure 2. (A) The graph displays the averaged phylum-level distribution from both locations, HP and UT. This is represented either by summing the top three peptide areas of all proteins ("Fraction of total protein area") or by counting the number of identified proteins per phylum ("Fraction of total protein"). (B) The bar graph illustrates the protein area composition at the family level across 5 time points from December 2023 to February 2024, observed in the wastewater of Harnaschpolder (HP) [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, based on all microbial proteins detected in the samples.

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. Tables listing all enriched terms and functions are provided in the SI EXCEL DOC. A more

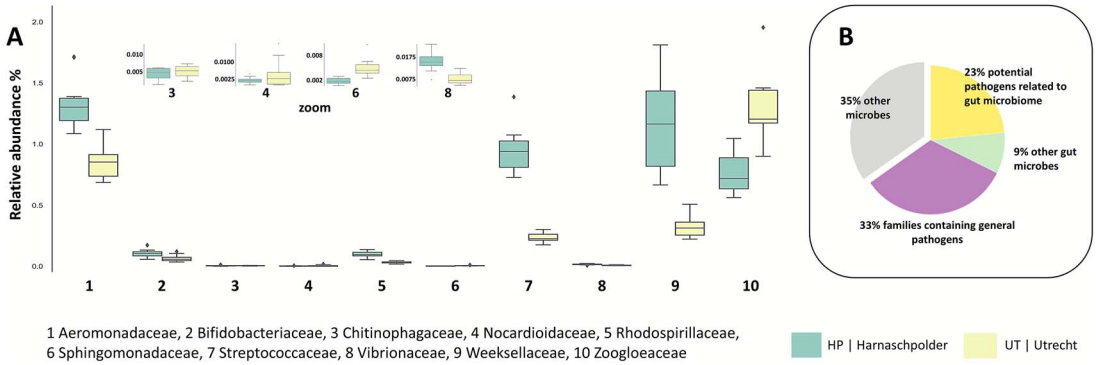

Figure 3. Graph A displays boxplots of the relative abundance of selected bacterial families across 5 time points from Dec 2023 to February 2024, observed in the wastewater of Harnaschpolder [6] and Utrecht (UT), located in the Netherlands. Graph B presents a pie chart illustrating the distribution of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

detailed study regarding pathogens and proteins related to cancer with annotations from the Human Protein Atlas can be found in the SI DOC (Figs S1–S4 and Table S1).

Multiple proteins could be associated with breast cancer, followed by proteins linked to diabetic nephropathy, and some for inflammatory bowel disease (Fig. 5). 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 [56] (Fig. 5 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 (Fig. 5). 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 [95, 96] 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 [97, 98].

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 (Fig. 5). 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 potentially be linked to a wide range of clinically relevant disease gene associations, underscoring their potential relevance for public health monitoring (Fig. 5). Proteins related to various types of cancer are further outlined in Fig. S2 and Table S1.

An additional feature of human proteins is their post-translational modification state, which can reflect health and disease, including on antibodies and other glycoproteins. We therefore examined the N-glycosylation profile on the detected human proteins. Although we observed abundant oxonium ions indicative of N-glycopeptides, the resulting N-glycan mass-modification profile showed predominantly trimmed core structures, including single HexNAc or HexNAc-Fuc residues, consistent with exo- and endoglycosidase activity in wastewater (Fig. S5). Therefore, the original glycosylation patterns were largely degraded and could not be interpreted from wastewater samples.

Finally, several human proteins showed significant abundance differences between Harnaschpolder and Utrecht. For example, carcinoembryonic antigen-related cell adhesion molecule 5, S100-A7, and myeloperoxidase were more abundant in Harnaschpolder, whereas hemopexin and aminopeptidase N were more abundant in Utrecht. Clear temporal abundance trends were difficult to resolve with the current shotgun approach,

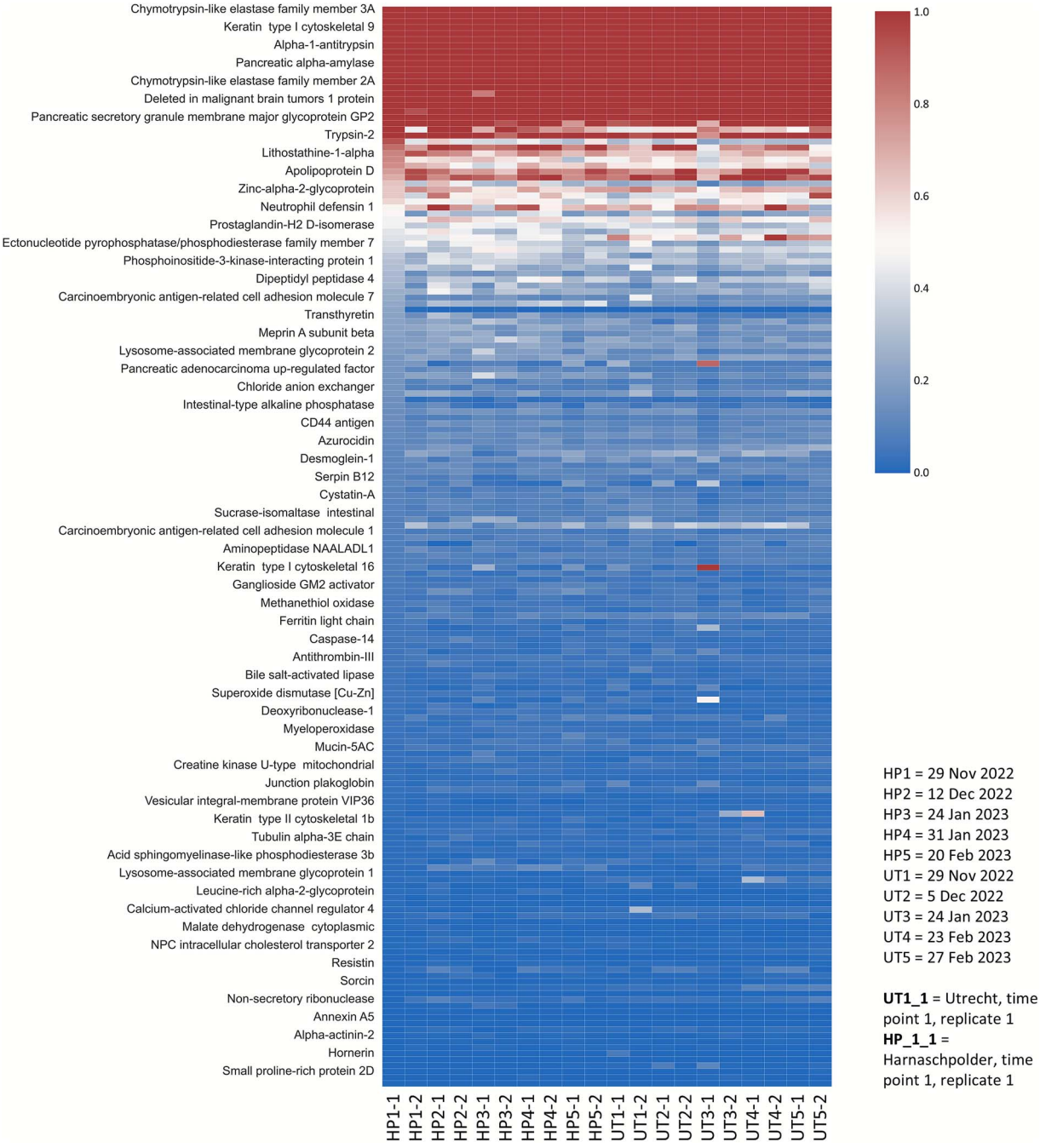

Figure 4. The heatmap provides an abundance profile of all human proteins identified in the wastewater, across 5 time points from Dec 2023 to February 2024, observed in the wastewater of Harnaschpolder [6] and Utrecht (UT), located in the Netherlands. The color gradient represents the relative abundance 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 point 1 replicate 1). The y-axis annotation names every fourth protein. The full heatmap dataset is provided in the SI Excel DOC, worksheet "HS heatmap raw data".

except for a few proteins from the Utrecht wastewater location that showed a consistent increase from Dec to March (e.g. lactotransferrin, S100-A9; SI Excel, worksheet "HS relative abundances", and Fig. S6 and S7). A more robust assessment of temporal trends would require a targeted approach with internal standards to minimize sampling and preparation biases.

Finally, many of the detected tissue-enriched and disease-associated proteins are not unique to a single tissue and may also be present under healthy conditions, with changes in abundance across tissues during disease.

Discussion and conclusions

The presented study demonstrates a streamlined wastewater metaproteomics approach that provides a comprehensive view of the wastewater metaproteome. The approach performs filter-aided sample preparation of small wastewater volumes and de novo peptide sequencing to refine global reference sequence databases, which makes it suitable for large sample cohorts and multiplexing. We applied this approach to profile the wastewater metaproteome from two different municipal locations over approximately 3 months. For sampling, we chose the winter

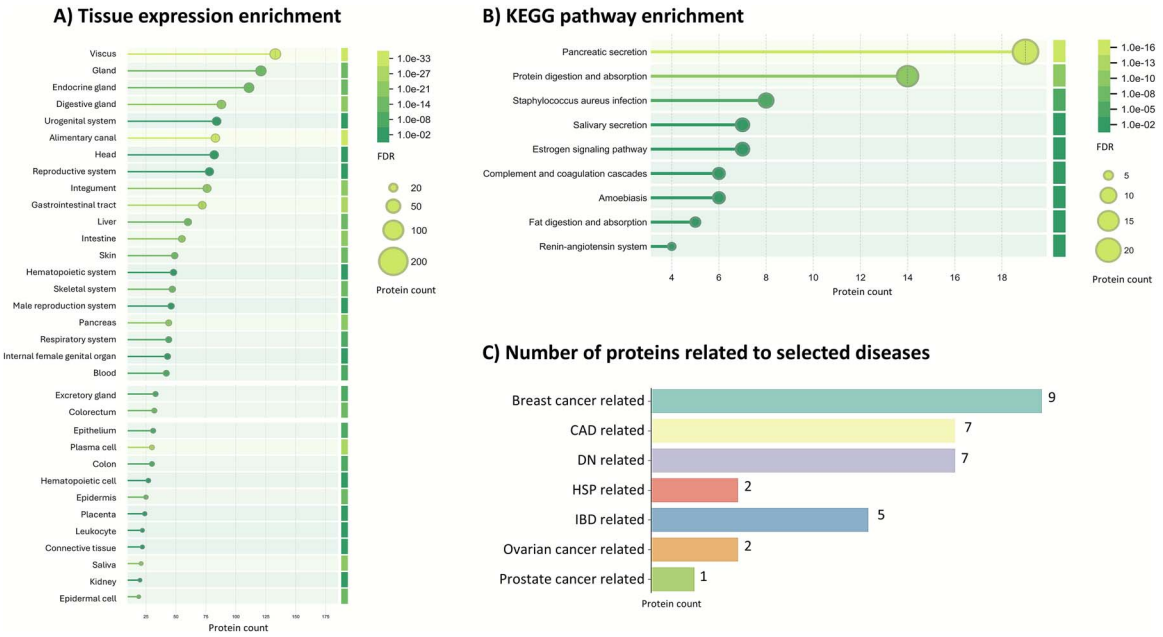

Figure 5. Analysis of the identified human proteins as obtained from both locations, for enriched terms and functions. (A) The graph displays the enrichment of tissue expression terms. The length of each bar and the size of the corresponding circle represent the number of proteins assigned to each term. 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.

season, as this period typically is associated with higher infection rates, and lower precipitation, which is particularly interesting for biomarker analysis.

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 minor differences. Additionally, such generic databases limit the exploration of microbial populations to the genus or family level. However, the core proteome of both wastewater locations is likely similar, as both serve densely populated areas, only ~55 km apart, differing mainly in their sewer residence times. The identified proteins belong to environmental microbes as well as microbes from the human microbiome, including the human gut. Among the many detected families, several contain pathogenic microbes and those known to spread antimicrobial resistance [55]. This may provide valuable insights into the spread of such genes in the environment. Since metaproteomics directly measures the proteins, this confirms that the respective microbes are either active or dormant.

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. Higher resolution could be also achieved by incorporating more specific databases, e.g. through selected genomes or pathogen-unique peptides [99]. 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 [100]. 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 [101]. For example, Jagadeesan and coworkers 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 [102]. 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 [103]. They observed intact antibodies, predominantly of the IgG and IgA classes, which appeared to adhere to solid matter in the wastewater. The study demonstrated that these IgG and IgA antibodies retained their immunoaffinity for SARS-CoV-2 spike protein antigens [103]. 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 may support the evaluation of population health or signal the emergence of an epidemic. In this study, we identified ~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 [104, 105]. 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]. Finally, evaluating the potential of individual proteins as wastewater-derived biomarkers will require sampling campaigns across multiple locations and seasons, in close collaboration with clinicians and health specialists to correlate protein levels with real health data. 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 [106–108]. However, no studies to date have investigated the fate of these modifications in the complex wastewater environment. Here, we show that protein N-glycosylation on human proteins can still be detected. However, we observed strong enzymatic trimming, therefore, only single HexNAc or HexNAc-Fuc core residues remained in most cases, meaning that the original glycoform information was not preserved.

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 for monitoring complex and dynamic environments such as wastewater [42, 109, 110]. 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 [111]. Progress has also been made in the development of single-protein fingerprinting approaches, which additionally may provide insights into protein modifications [112]. These advancements hold great promise for the integration of metaproteomics into routine wastewater monitoring in the near future.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT to assist with language editing and proofreading. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgements

The authors thank Dita Heikens for her support in the proteomics laboratory, 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.

Supplementary material

Supplementary material is available at *ISME Communications* online.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Funding

This research was supported by TU Delft and, in part, by the NWO Spinoza Prize awarded to Mark van Loosdrecht by the Netherlands Organisation for Scientific Research (NWO).

Data availability

The datasets generated during and/or analyzed during the current study are available via the ProteomeXchange repository, under the project code PXD059455.

References

1. Qadir M, Drechsel P, Jiménez Cisneros B. *et al.* Global and regional potential of wastewater as a water, nutrient and energy source. *Nat Res Forum* 2020;**44**:40–51. https://doi.org/10.1111/1477-8947.12187
2. Ryu Y, Gracia-Lor E, Bade R. *et al.* Increased levels of the oxidative stress biomarker 8-iso-prostaglandin F2α in wastewater associated with tobacco use. *Sci Rep* 2016;**6**:39055. https://doi.org/10.1038/srep39055
3. Ryu Y, Barceló D, Barron LP. *et al.* Comparative measurement and quantitative risk assessment of alcohol consumption through wastewater-based epidemiology: An international study in 20 cities. *Sci Total Environ* 2016;**565**:977–83. https://doi.org/10.1016/j.scitotenv.2016.04.138
4. Daughton C. *Illicit Drugs in Municipal Sewage: Proposed New Non-intrusive Tool to Heighten Public Awareness of Societal Use of Illicit/Abused Drugs and their Potential for Ecological Consequences*, Daughton CG, Jones-Lepp TL, (eds.), Washington, D.C.: American Chemical Society, 2001, 348–64.
5. O'Keeffe J. Wastewater-based epidemiology: current uses and future opportunities as a public health surveillance tool. *Environ Health Rev* 2021;**64**:44–52. https://doi.org/10.5864/d2021-015
6. Hovi T, Shulman LM, van der Avoort H. *et al.* Role of environmental poliovirus surveillance in global polio eradication and beyond. *Epidemiol Infect* 2012;**140**:1–13. https://doi.org/10.1017/S095026881000316X
7. Roberts L. Infectious disease. Israel's silent polio epidemic breaks all the rules. *Science* 2013;**342**:679–80. https://doi.org/10.1126/science.342.6159.679
8. Spurbeck RR, Minard-Smith A, Catlin L. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. *Sci Total Environ* 2021;**789**:147829. https://doi.org/10.1016/j.scitotenv.2021.147829
9. Medema G, Heijnen L, Elsinga G. *et al.* 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. *Environ Sci Technol Lett* 2020;**7**:511–6. https://doi.org/10.1021/acs.estlett.0c00357
10. Daughton CG. Wastewater surveillance for population-wide Covid-19: the present and future. *Sci Total Environ* 2020;**736**:139631. https://doi.org/10.1016/j.scitotenv.2020.139631
11. Choi PM, Tscharke BJ, Donner E. *et al.* Wastewater-based epidemiology biomarkers: past, present and future. *TrAC Trends Anal Chem* 2018;**105**:453–69. https://doi.org/10.1016/j.trac.2018.06.004
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 Anal Chem* 2019;**119**:15621. https://doi.org/10.1016/j.trac.2019.115621
13. Clemente JC, Ursell LK, Parfrey LW. *et al.* The impact of the gut microbiota on human health: an integrative view. *Cell* 2012;**148**:1258–70. https://doi.org/10.1016/j.cell.2012.01.035
14. Hassoun-Kheir N, Stabholz Y, Kreft J-U. *et al.* Comparison of antibiotic-resistant bacteria and antibiotic resistance genes abundance in hospital and community wastewater: a systematic review. *Sci Total Environ* 2020;**743**:140804. https://doi.org/10.1016/j.scitotenv.2020.140804
15. Łuczkiewicz A, Jankowska K, Fudala-Książek S. *et al.* Antimicrobial resistance of fecal indicators in municipal wastewater treatment plant. *Water Res* 2010;**44**:5089–97. https://doi.org/10.1016/j.watres.2010.08.007
16. Novo A, André S, Viana P. *et al.* Antibiotic resistance, antimicrobial residues and bacterial community composition in urban wastewater. *Water Res* 2013;**47**:1875–87. https://doi.org/10.1016/j.watres.2013.01.010
17. Gilbride K, Lee D-Y, Beaudette L. Molecular techniques in wastewater: understanding microbial communities, detecting pathogens, and real-time process control. *J Microbiol Methods* 2006;**66**:1–20. https://doi.org/10.1016/j.mimet.2006.02.016
18. Kho ZY, Lal SK. The human gut microbiome - a potential controller of wellness and disease. *Front Microbiol* 2018;**9**:1835. https://doi.org/10.3389/fmicb.2018.01835
19. Halfvarson J, Brislawn CJ, Lamendella R. *et al.* Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol* 2017;**2**:17004. https://doi.org/10.1038/nmicrobiol.2017.4
20. Picó Y, Barceló D. Identification of biomarkers in wastewater-based epidemiology: main approaches and analytical methods. *TrAC Trends Anal Chem* 2021;**145**:116465. https://doi.org/10.1016/j.trac.2021.116465
21. Mao K, Zhang K, Du W. *et al.* The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. *Curr Opin Environ Sci Health* 2020;**17**:1–7. https://doi.org/10.1016/j.coesh.2020.04.006
22. Hillary LS, Malham SK, McDonald JE. *et al.* Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Current Opinion in Environmental Science & Health* 2020;**17**:14–20. https://doi.org/10.1016/j.coesh.2020.06.001
23. Hryniszyn A, Skonieczna M, Wiszniowski J. Methods for detection of viruses in water and wastewater. *Adv Microbiol* 2013;**03**:442–9. https://doi.org/10.4236/aim.2013.35060
24. Pruden A, Vikesland PJ, Davis BC. *et al.* Seizing the moment: now is the time for integrated global surveillance of antimicrobial resistance in wastewater environments. *Curr Opin Microbiol* 2021;**64**:91–9. https://doi.org/10.1016/j.mib.2021.09.013
25. Parkins MD, Lee BE, Acosta N. *et al.* Wastewater-based surveillance as a tool for public health action: SARS-CoV-2 and beyond. *Clin Microbiol Rev* 2024;**37**:e00103–22. https://doi.org/10.1128/cmr.00103-22
26. Toze S. PCR and the detection of microbial pathogens in water and wastewater. *Water Res* 1999;**33**:3545–56. https://doi.org/10.1016/S0043-1354(99)00071-8
27. Walden C, Carbonero F, Zhang W. Assessing impacts of DNA extraction methods on next generation sequencing of water and wastewater samples. *J Microbiol Methods* 2017;**141**:10–6. https://doi.org/10.1016/j.mimet.2017.07.007
28. Garner E, Davis BC, Milligan E. *et al.* Next generation sequencing approaches to evaluate water and wastewater quality. *Water Res* 2021;**194**:116907. https://doi.org/10.1016/j.watres.2021.116907
29. Munk P, Brinch C, Møller F. *et al.* Genomic analysis of sewage from 101 countries reveals global landscape of antimicrobial resistance. *Nat Commun* 2022;**13**:7251. https://doi.org/10.1038/s41467-022-34312-7
30. Nieuwenhuijse DF, Oude Munnink BB, Phan MV. *et al.* Setting a baseline for global urban virome surveillance in sewage. *Sci Rep* 2020;**10**:13748. https://doi.org/10.1038/s41598-020-69869-0
31. Kosorok MR, Laber EB. Precision medicine. *Annual review of statistics and its application* 2019;**6**:263–86. https://doi.org/10.1146/annurev-statistics-030718-105251
32. Sequeira-Antunes B, Ferreira HA. Urinary biomarkers and point-of-care urinalysis devices for early diagnosis and management of disease: a review. *Biomedicine* 2023;**11**:1051. https://doi.org/10.3390/biomedicines11041051
33. Yang J, Bielenberg DR, Rodig SJ. *et al.* Lipocalin 2 promotes breast cancer progression. *Proc Natl Acad Sci USA* 2009;**106**:3913–8. https://doi.org/10.1073/pnas.0810617106
34. Feng B, Yue F, Zheng MH. Urinary markers in colorectal cancer. *Adv Clin Chem* 2009;**47**:45–57.
35. Marks LS, Fradet Y, Lim Deras I. *et al.* PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. *Urology* 2007;**69**:532–5. https://doi.org/10.1016/j.urology.2006.12.014
36. Song H, Chan J, Rovin BH. Induction of chemokine expression by adiponectin in vitro is isoform dependent. *Transl Res* 2009;**154**:18–26. https://doi.org/10.1016/j.trsl.2009.04.003
37. Ma L, Wang R, Han Y. *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. *J Clin Lab Anal* 2016;**30**:308–14. https://doi.org/10.1002/jcla.21856
38. Tashiro K, Koyanagi I, Saitoh A. *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. *J Clin Lab Anal* 2002;**16**:1–4. https://doi.org/10.1002/jcla.2057
39. Wilmes P, Bond PL. Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* 2006;**14**:92–7. https://doi.org/10.1016/j.tim.2005.12.006
40. Kleiner M, Thorson E, Sharp CE. *et al.* Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* 2017;**8**:1558. https://doi.org/10.1038/s41467-017-01544-x
41. Kleikamp HB, Grouzdev D, Schaasberg P. *et al.* Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. *Water Res* 2023;**246**:120700. https://doi.org/10.1016/j.watres.2023.120700
42. Dumas T, Martinez Pinna R, Lozano C. *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**:46. https://doi.org/10.1186/s40168-024-01766-4

43. Xian F, Brenek M, Krisp C. *et al.* Ultra-sensitive metaproteomics redefines the dark metaproteome, uncovering host-microbiome interactions and drug targets in intestinal diseases. *Nature Communications* 2025;**16**:6644.
44. Blakeley-Ruiz JA, Kleiner M. Considerations for constructing a protein sequence database for metaproteomics. *Comput Struct Biotechnol J* 2022;**20**:937–52. https://doi.org/10.1016/j.csbj.2022.01.018
45. Kleikamp HB, Van der Zwaan R, Van Valderen R. *et al.* NovoLign: metaproteomics by sequence alignment. *ISME communications* 2024;**4**:ycae121. https://doi.org/10.1093/ismeco/ycae121
46. Kleikamp HB, Pronk M, Tugui C. *et al.* Database-independent de novo metaproteomics of complex microbial communities. *Cell Systems* 2021;**12**:375–383.e5. https://doi.org/10.1016/j.cels.2021.04.003
47. Carrascal M, Abián J, Ginebreda A. *et al.* Discovery of large molecules as new biomarkers in wastewater using environmental proteomics and suitable polymer probes. *Sci Total Environ* 2020;**747**:141145. https://doi.org/10.1016/j.scitotenv.2020.141145
48. Perez-Lopez C, Ginebreda A, Carrascal M. *et al.* Non-target protein analysis of samples from wastewater treatment plants using the regions of interest-multivariate curve resolution (ROIMCR) chemometrics method. *J Environ Chem Eng* 2021;**9**:105752. https://doi.org/10.1016/j.jece.2021.105752
49. Sánchez-Jiménez E, Abian J, Ginebreda A. *et al.* Shotgun proteomics to characterize wastewater proteins. *MethodsX* 2023;**11**:102403. https://doi.org/10.1016/j.mex.2023.102403
50. Carrascal M, Sánchez-Jiménez E, Fang J. *et al.* Sewage protein information mining: discovery of large biomolecules as biomarkers of population and industrial activities. *Environ Sci Technol* 2023;**57**:10929–39. https://doi.org/10.1021/acs.est.3c00535
51. Tugui CG, Sorokin DY, Hijnen W. *et al.* Exploring the metabolic potential of Aeromonas to utilise the carbohydrate polymer chitin. *RSC Chem Biol* 2025;**6**:227–39. https://doi.org/10.1039/D4CB00200H
52. Pabst M, Grouzdev DS, Lawson CE. *et al.* A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. *ISME J* 2022;**16**:346–57. https://doi.org/10.1038/s41396-021-01073-y
53. Bern M, Kil YJ, Becker C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics* 2012;**40**:13.20.1–14. https://doi.org/10.1002/0471250953.bi1320s40
54. Bartlett A, Padfield D, Lear L. *et al.* A comprehensive list of bacterial pathogens infecting humans. *Microbiology (Reading)* 2022;**168**. https://doi.org/10.1099/mic.0.001269
55. *WHO Bacterial Priority Pathogens List, 2024: Bacterial Pathogens of Public Health Importance to Guide Research, Development and Strategies to Prevent and Control Antimicrobial Resistance*, Geneva: World Health Organization;2024, Available from: https://iris.who.int/bitstream/handle/10665/376776/9789240093461-eng.pdf?sequence=1.
56. Szklarczyk D, Franceschini A, Kuhn M. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2010;**39**:D561–8. https://doi.org/10.1093/nar/gkq973
57. Digre A, Lindskog C. The human protein atlas—spatial localization of the human proteome in health and disease. *Protein Sci* 2021;**30**:218–33. https://doi.org/10.1002/pro.3987
58. Perpétuo L, Barros AS, Dalsuco J. *et al.* Coronary artery disease and aortic valve stenosis: a urine proteomics study. *Int J Mol Sci* 2022;**23**. https://doi.org/10.3390/ijms232113579
59. Zürbig P, Jerums G, Hovind P. *et al.* Urinary proteomics for early diagnosis in diabetic nephropathy. *Diabetes* 2012;**61**:3304–13. https://doi.org/10.2337/db12-0348
60. Beretov J, Wasinger VC, Millar EK. *et al.* Proteomic analysis of urine to identify breast cancer biomarker candidates using a label-free LC-MS/MS approach. *PLoS One* 2015;**10**:e0141876. https://doi.org/10.1371/journal.pone.0141876
61. Chen C-J, Chou C-Y, Shu K-H. *et al.* Discovery of novel protein biomarkers in urine for diagnosis of urothelial cancer using iTRAQ proteomics. *J Proteome Res* 2021;**20**:2953–63. https://doi.org/10.1021/acs.jproteome.1c00164
62. Jia L, Wu J, Wei J. *et al.* Proteomic analysis of urine reveals biomarkers for the diagnosis and phenotyping of abdominal-type Henoch-Schonlein purpura. *Transl Pediatr* 2021;**10**:510–24. https://doi.org/10.21037/tp-20-317
63. Fujita K, Nonomura N. Urinary biomarkers of prostate cancer. *Int J Urol* 2018;**25**:770–9. https://doi.org/10.1111/iju.13734
64. Owens GL, Barr CE, White H. *et al.* Urinary biomarkers for the detection of ovarian cancer: a systematic review. *Carcinogenesis* 2022;**43**:311–20. https://doi.org/10.1093/carcin/bgac016
65. Tamburini FB, Andermann TM, Tkachenko E. *et al.* Precision identification of diverse bloodstream pathogens in the gut microbiome. *Nat Med* 2018;**24**:1809–14. https://doi.org/10.1038/s41591-018-0202-8
66. Watson SW, Valois FW, Waterbury JB. *The family nitrobacteraceae*. In: Starr Mortimer P, Heinz Stolp, Hans G Trüper, Albert Balows, Hans G Schlegel, (eds.), *The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria*. Heidelberg: Springer, 1981, 1005–22.
67. Lohmann P, Benk S, Gleixner G. *et al.* Seasonal patterns of dominant microbes involved in central nutrient cycles in the subsurface. *Microorganisms* 2020;**8**:1694. https://doi.org/10.3390/microorganisms8111694
68. Madigan M, Cox SS, Stegeman RA. Nitrogen fixation and nitrogenase activities in members of the family Rhodospirillaceae. *J Bacteriol* 1984;**157**:73–8. https://doi.org/10.1128/jb.157.1.73-78.1984
69. Goyal P, Basniwal RK. Environmental bioremediation: Biodegradation of xenobiotic compounds. In: Hashmi M, Kumar V, Varma A. (eds.), *Xenobiotics in the Soil Environment: Monitoring, Toxicity and Management*, Cham: Springer, 2017, 347–71.
70. Gomez NCF, Onda DFL. Potential of sediment bacterial communities from Manila Bay (Philippines) to degrade low-density polyethylene (LDPE). *Arch Microbiol* 2023;**205**:38. https://doi.org/10.1007/s00203-022-03366-y
71. Goff KL, Hug LA. Environmental potential for microbial 1, 4-dioxane degradation is sparse despite mobile elements playing a role in trait distribution. *Appl Environ Microbiol* 2022;**88**:e02091–21. https://doi.org/10.1128/aem.02091-21
72. García-García R, Bocanegra-García V, Vital-López L. *et al.* Assessment of the microbial communities in soil contaminated with petroleum using next-generation sequencing tools. *Appl Sci* 2023;**13**:6922. https://doi.org/10.3390/app13126922
73. Lünsmann V, Kappelmeyer U, Benndorf R. *et al.* In situ p rotein-SIP highlights Burkholderiaceae as key players degrading toluene by Para ring hydroxylation in a constructed wetland model. *Environ Microbiol* 2016;**18**:1176–86. https://doi.org/10.1111/1462-2920.13133
74. Stolz A. Molecular characteristics of xenobiotic-degrading sphingomonads. *Appl Microbiol Biotechnol* 2009;**81**:793–811. https://doi.org/10.1007/s00253-008-1752-3
75. Yu L, Yang Y, Yang B. *et al.* Effects of solids retention time on the performance and microbial community structures in

membrane bioreactors treating synthetic oil refinery wastewater. *Chem Eng J* 2018;**344**:462–8. https://doi.org/10.1016/j.cej.2018.03.073
76. Bafghi MF, Yousefi N. Role of nocardia in activated sludge. *The Malaysian journal of medical sciences: MJMS* 2016;**23**:86–8.
77. An W, Guo F, Song Y. *et al.* Comparative genomics analyses on EPS biosynthesis genes required for floc formation of Zoogloea resiniphila and other activated sludge bacteria. *Water Res* 2016;**102**:494–504. https://doi.org/10.1016/j.watres.2016.06.058
78. Rossau R, Van Landschoot A, Gillis M. *et al.* Taxonomy of Moraxellaceae fam. Nov., a new bacterial family to accommodate the genera Moraxella, Acinetobacter, and Psychrobacter and related organisms. *Int J Syst Evol Microbiol* 1991;**41**:310–9.
79. Liu H-Y, Li C-x, Liang Z-Y. *et al.* The interactions of airway bacterial and fungal communities in clinically stable asthma. *Front Microbiol* 2020;**11**:1647. https://doi.org/10.3389/fmicb.2020.01647
80. Wong D, Nielsen TB, Bonomo RA. *et al.* Clinical and pathophysiological overview of Acinetobacter infections: a century of challenges. *Clin Microbiol Rev* 2017;**30**:409–47. https://doi.org/10.1128/CMR.00058-16
81. Munoz-Price LS, Weinstein RA. Acinetobacter infection. *N Engl J Med* 2008;**358**:1271–81. https://doi.org/10.1056/NEJMra070741
82. Mali S, Dash L, Gautam V. *et al.* An outbreak of Burkholderia cepacia complex in the paediatric unit of a tertiary care hospital. *Indian J Med Microbiol* 2017;**35**:216–20. https://doi.org/10.4103/ijmm.IJMM_16_258
83. Weinroth M, Thomas K, Doster E. *et al.* Resistomes and microbiome of meat trimmings and colon content from culled cows raised in conventional and organic production systems. *Anim Microbiome* 2022;**4**:21. https://doi.org/10.1186/s42523-022-00166-z
84. Diener C, Holscher HD, Filek K. *et al.* Metagenomic estimation of dietary intake from human stool. *Nat Metab* 2025;**7**:1–14.
85. Petrone BL, Bartlett A, Jiang S. *et al.* A pilot study of metaproteomics and DNA metabarcoding as tools to assess dietary intake in humans. *Food Funct* 2025;**16**:282–96. https://doi.org/10.1039/D4FO02656J
86. Dillman AR, Macchietto M, Porter CF. *et al.* Comparative genomics of Steinernema reveals deeply conserved gene regulatory networks. *Genome Biol* 2015;**16**:1–21. https://doi.org/10.1186/s13059-015-0746-6
87. Aykur M, Malatyalı E, Demirel F. *et al.* Blastocystis: a mysterious member of the gut microbiome. *Microorganisms* 2024;**12**:461. https://doi.org/10.3390/microorganisms12030461
88. Grilo ML, Sousa-Santos C, Robalo J. *et al.* The potential of Aeromonas spp. from wildlife as antimicrobial resistance indicators in aquatic environments. *Ecol Indic* 2020;**115**:106396. https://doi.org/10.1016/j.ecolind.2020.106396
89. Fernandes T, Vaz-Moreira I, Manaia CM. Neighbor urban wastewater treatment plants display distinct profiles of bacterial community and antibiotic resistance genes. *Environ Sci Pollut Res* 2019;**26**:11269–78. https://doi.org/10.1007/s11356-019-04546-y
90. Sinton L, Donnison A, Hastie C. Faecal streptococci as faecal pollution indicators: a review. Part II: sanitary significance, survival, and use. *N Z J Mar Freshw Res* 1993;**27**:117–37. https://doi.org/10.1080/00288330.1993.9516550
91. De R, Mukhopadhyay AK, 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* 2020;**12**:1–48. https://doi.org/10.1186/s13099-020-00371-8
92. Blackall LL, Parlett J, Hayward A. *et al.* Nocardia pinensis sp. nov., an actinomycete found in activated sludge foams in Australia. *Microbiology* 1989;**135**:1547–58. https://doi.org/10.1099/00221287-135-6-1547
93. Hao O, Strom PF, Wu Y. A review of the role of nocardia-like filaments in activated sludge foaming. *Water SA* 1988;**14**:105–10.
94. Ma Y, Wang J, Liu Y. *et al.* Nocardioides:"specialists" for hard-to-degrade pollutants in the environment. *Molecules* 2023;**28**:7433. https://doi.org/10.3390/molecules28217433
95. Ashburner M, Ball CA, Blake JA. *et al.* Gene ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–9. https://doi.org/10.1038/75556
96. Consortium TGO, Aleksander SA, Balhoff J. *et al.* The gene ontology knowledgebase in 2023. *Genetics* 2023;**224**.
97. Cieutat A-M, Lobel P, August JT. *et al.* Azurophilic granules of human neutrophilic leukocytes are deficient in lysosome-associated membrane proteins but retain the mannose 6-phosphate recognition marker. *Blood* 1998;**91**:1044–58. https://doi.org/10.1182/blood.V91.3.1044
98. Gómez-Lázaro M, Rinn C, Aroso M. *et al.* Proteomic analysis of zymogen granules. *Expert Rev Proteomics* 2010;**7**:735–47. https://doi.org/10.1586/epr.10.31
99. Muth T, Karlsson R, Yu Y-K. *et al.* From identification to insight: making full use of the diagnostic potential of MS/MS Proteotyping in clinical microbiology using efficient bioinformatics. *J Proteome Res* 2025;**24**:5336–51. https://doi.org/10.1021/acs.jproteome.5c00733
100. Tisza M, Javornik Cregeen S, Avadhanula V. *et al.* Wastewater sequencing reveals community and variant dynamics of the collective human virome. *Nat Commun* 2023;**14**:6878. https://doi.org/10.1038/s41467-023-42064-1
101. Lara-Jacobo LR, Islam G, Desaulniers J-P. *et al.* Detection of SARS-CoV-2 proteins in wastewater samples by mass spectrometry. *Environ Sci Technol* 2022;**56**:5062–70. https://doi.org/10.1021/acs.est.1c04705
102. Jagadeesan KK, Elliss H, Standerwick R. *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* 2024;**5**:100108. https://doi.org/10.1016/j.hazl.2024.100108
103. Stephenson S, Eid W, Wong CH. *et al.* Urban wastewater contains a functional human antibody repertoire of mucosal origin. *Water Res* 2024;**267**:122532. https://doi.org/10.1016/j.watres.2024.122532
104. Devianto LA, Sano D. Systematic review and meta-analysis of human health-related protein markers for realizing real-time wastewater-based epidemiology. *Sci Total Environ* 2023;**897**:165304. https://doi.org/10.1016/j.scitotenv.2023.165304
105. Amin V, Bowes DA, Halden RU. Systematic scoping review evaluating the potential of wastewater-based epidemiology for monitoring cardiovascular disease and cancer. *Sci Total Environ* 2023;**858**:160103. https://doi.org/10.1016/j.scitotenv.2022.160103
106. de Haan N, Falck D, Wuhrer M. Monitoring of immunoglobulin N- and O-glycosylation in health and disease. *Glycobiology* 2020;**30**:226–40. https://doi.org/10.1093/glycob/cwz048

107. Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. *Cell Immunol* 2018;**333**:65–79. https://doi.org/10.1016/j.cellimm.2018.07.009
108. Reily C, Stewart TJ, Renfrow MB. *et al.* Glycosylation in health and disease. *Nat Rev Nephrol* 2019;**15**:346–66. https://doi.org/10.1038/s41581-019-0129-4
109. Vitko D, Chou W-F, Nouri Golmaei S. *et al.* timsTOF HT improves protein identification and quantitative reproducibility for deep unbiased plasma protein biomarker discovery. *J Proteome Res* 2024;**23**:929–38. https://doi.org/10.1021/acs.jproteome.3c00646
110. Wang J, Tan H, Fu Y. *et al.* Evaluation of protein identification and quantification by the diaPASEF method on timsTOF SCP. *J Am Soc Mass Spectrom* 2024;**35**:1253–60. https://doi.org/10.1021/jasms.4c00067
111. Yu L, Kang X, Li F. *et al.* Unidirectional single-file transport of full-length proteins through a nanopore. *Nat Biotechnol* 2023;**41**:1130–9. https://doi.org/10.1038/s41587-022-01598-3
112. Filius M, van Wee R, de Lannoy C. *et al.* Full-length single-molecule protein fingerprinting. *Nat Nanotechnol* 2024;**19**:652–9. https://doi.org/10.1038/s41565-023-01598-7

ISME Communications, 2026, **6(1)**, ycaf243
https://doi.org/10.1093/ismeco/ycaf243
Original Article