nature water

Article

https://doi.org/10.1038/s44221-024-00345-z

Daily sampling reveals household-specific water microbiome signatures and shared antimicrobial resistomes in premise plumbing

Received: 11 January 2024

Accepted: 24 October 2024

Published online: 10 December 2024

Check for updates

A list of authors and their affiliations appears at the end of the paper

Stagnation in premise plumbing can lead to the degradation of drinking water quality, yet the variability of microbiomes and resistomes in these systems at fine spatiotemporal scales remains poorly understood. Here we track the water microbiome daily across households in St. Louis, Missouri, alongside functional gene profiles and antimicrobial resistomes. Our results show substantial differences in species composition between households, with household identity, instead of temporal fluctuations or specific water-use devices, emerging as the dominant variable shaping microbiome composition. Using LASSO regression models, we identified informative taxa for each household, achieving an average accuracy of 97.5% in estimating a sample's household origin. Notably, distinct profiles of opportunistic premise plumbing pathogens (OPPPs) were detected, with *Mycobacterium gordonae* being twice as prevalent as *M. chelonae*. Community assembly simulations indicated that stochastic processes primarily drive household-level taxonomic variation. In contrast, antimicrobial resistomes and functional gene repertoires were similar across households, likely influenced by common chloramine residuals applied throughout the local water distribution systems. Genes conferring resistance to beta-lactams were prevalent in bathtub faucet water across all households. These results highlight the need to incorporate household-level species variation when assessing health risks from OPPPs and monitoring antimicrobial resistance. These findings also pave the way for new research to better understand plumbing environments as potential routes for the transmission of resistant bacteria and their genes.

Drinking water harbours a diverse collection of microorganisms, exhibiting spatiotemporal variations in abundance[1], community composition[2–5] and metabolic functions[6,7] from source to tap. The vast majority of these microorganisms are not pathogenic to humans. Some of them even contribute positively, in such ways as by seeding human commensals[8–10], protecting infrastructure from corrosion[11] and degrading disinfection byproducts[12]. Nonetheless, overgrowth of indicator microorganisms (*Escherichia coli* and total coliforms), proliferation of opportunistic premise plumbing pathogens (OPPPs), nitrification and microbially mediated corrosion can lead to public health concerns, regulatory compliance challenges and operational issues[13,14]. To ensure the microbiological safety of drinking water, a better fundamental understanding of the diversity, variation and dynamics of the drinking water microbiome is crucial[15]. The urgency of this task is underscored by the World Health Organization's guideline that "water entering drinking water distribution systems must be microbially safe and ideally should also be biologically stable", highlighting the imperative of understanding the spatiotemporal variations in drinking water microbiomes[16,17].

e-mail: fangqiong@wustl.edu

Premise plumbing encompasses the water supply systems within the property boundaries. The unique engineering characteristics of premise plumbing, for example, a high surface area to volume ratio, low concentrations of residual disinfectant, high temperatures and intermittent stagnation, lead to an environment distinct from that of the distribution system[18]. Therefore, premise plumbing is usually associated with opportunistic pathogen proliferation and overgrowth of indicator organisms[19,20]. In addition, microbial communities of premise plumbing, as represented by stagnant water, are often distinct from those of distribution systems, as represented by flushed water samples from the same building[20–22]. Furthermore, premise plumbing has been observed to induce changes in water microbiome composition within hours[21] and to exhibit spatial variations between volumes as small as 1 litre (ref. 22). These findings highlight the critical need to characterize premise plumbing microbiomes at various spatial and temporal resolutions. Because premise plumbing is currently beyond the scope of routine sampling conducted by public water utilities, understanding the spatiotemporal variations in premise plumbing microbiomes can provide insights for future monitoring strategy designs and risk assessments[23].

Our current understanding of spatiotemporal variations in premise plumbing microbiomes relies primarily on studies using large buildings, such as hospitals and university buildings, as model systems[24,25]. In these settings, community compositions are influenced primarily by factors such as stagnation and the plumbing configuration; moreover, residual disinfectant levels and water temperature are often affected by these factors[24–26], which probably act as environmental stressors on the microbial communities. Our knowledge of the premise plumbing microbiome in household settings remains limited, even though they represent the primary setting where drinking water quality is experienced by individual consumers as well as the main location for aerosolized exposure[27]. Experimental studies have shown that variations in community composition, including the occurrence of opportunistic pathogens[28,29], can arise from various system ages and water heater materials. For example, Mathys and colleagues showed that relatively new plumbing systems were less likely to be colonized by *Legionella* and that plumbing systems in houses using on-demand water heaters had lower occurrences of *Legionella* than households using a storage tank and recirculation[30]. These results underscore the need to investigate premise plumbing microbiome variations in actual households. Although variations in the occurrence and abundance of opportunistic pathogens in households have been reported[29,31,32], species-level characterization of OPPPs in real-world premise plumbing water remains a critical knowledge gap in risk mitigation[33]. Furthermore, we still lack a clear understanding of how premise plumbing microbiomes vary by household, especially at finer taxonomic resolutions.

Recent studies have also indicated that drinking water is an overlooked reservoir of antimicrobial resistance. An international antimicrobial resistance survey across multiple cities in Hong Kong, Singapore and Mainland China revealed the widespread prevalence of multidrug, bacitracin and aminoglycoside resistance genes in various size fractions of the drinking water microbiome[34,35]. These findings underscore the necessity of monitoring antimicrobial resistance in drinking water. Furthermore, studies in China and Brazil have shown greater resistance gene loadings in distribution systems than in finished water[36,37], emphasizing the need to study the spatiotemporal variation in resistomes (all acquired and intrinsic resistance genes, their precursors and resistance mechanisms) during water distribution. Nevertheless, previous works on antimicrobial resistance in drinking water distribution systems have focussed mainly on spatiotemporal variations at coarse granularities (for example, across cities or regions), and whether within-city variations in drinking water microbiomes play a role in shaping resistome remains an open question.

In this study, we are particularly interested in the following questions: (1) How much do the water microbiomes in premise plumbing vary within and between households? (2) Are these variations, if any, robust to the daily dynamics of water microbiomes and water use devices? (3) Can cross-validated models identifying individual households be built on the basis of the microbial species profiles in bathtub faucet water? To address these questions, we recruited volunteers in St. Louis to collect daily water samples over a week as well as samples from different water usage devices from their homes. We examined the water microbiome and resistome variations within and between individual households via complementary 16S ribosomal RNA (rRNA) gene amplicon sequencing and shotgun metagenomics. In particular, daily variations in the microbiomes and resistomes of household bathtub faucet water were examined. In addition, within-household variations were examined by comparing three types of water: stagnant water from bathtub faucets, stagnant water from kitchen faucets and fresh water from kitchen faucets. We found that there was strong between-household variation in taxonomic compositions but that the resistomes were more similar across households. This finding motivated us to use null-model-based simulations to investigate the community assembly mechanisms and explore the interactions between taxonomic compositions, metabolic functions and resistomes.

Results

Households had distinctive bath faucet water microbiomes

To capture the dynamic fluctuations in the compositions of premise plumbing microbiomes, we collected 56 water samples from eight households over a span of seven consecutive days (Fig. 1a). All samples were taken from bathtub faucets. During the week-long sampling, the microbial taxon relative abundances, as shown by metagenomic taxonomic profiling, varied from day to day (Fig. 2). The variations in the microbial communities at each home did not significantly correlate with differences in sample temperature ($R^2 = 0.04$, $P = 0.141$) but were moderately correlated with differences in total chlorine concentration ($R^2 = 0.36$, $P = 9.79 \times 10^{-7}$; Supplementary Fig. 1). Despite the day-to-day variations, the species profiles in an individual household were strongly correlated (Pearson ρ: mean of 0.91, s.d. of 0.13). This observation emphasized the resilience of bathtub faucet water microbiomes to perturbations in physicochemical water properties.

Each household harboured distinct taxa, in particular in the finer taxonomic ranks (Fig. 2). Principal coordinate analysis (PCoA) based on the Bray–Curtis distance of the species profiles clearly grouped by household (Fig. 3a; permutational multivariate analysis of variance (PERMANOVA) $P = 0.001$; pairwise PERMANOVA P values after multiple test adjustments are shown in Supplementary Table 1). The non-significant differences between the dispersions of the Bray–Curtis dissimilarities (permutational analysis of multivariate dispersions (PERMDISP) $P = 0.176$) further support the significant differences in centroids between households. A consistent PCoA pattern emerged when the same species profile data were analysed via the Jaccard index (Supplementary Fig. 2), indicating that community composition similarities and differences were driven by both the occurrence and relative abundance of taxa.

Household individuality based on LASSO-selected taxa

Upon detecting microbial community variations among households (Fig. 3a), we asked whether microbial species can serve as distinctive signatures of the plumbing system of a particular household. To address this question, leave-one-out least absolute shrinkage and selection operator (LASSO) models were built to identify a minimal set of species sufficient to distinguish each household uniquely (Fig. 3b). Remarkably, models estimating the household origin of a water sample on the basis of the selected microbial taxa achieved an average accuracy of 97.5% (see Supplementary Table 1 for the accuracy for each household), indicating the persistence of household-specific signatures across daily sampling of bathtub faucet water microbiomes.

https://doi.org/10.1038/s44221-024-00345-z

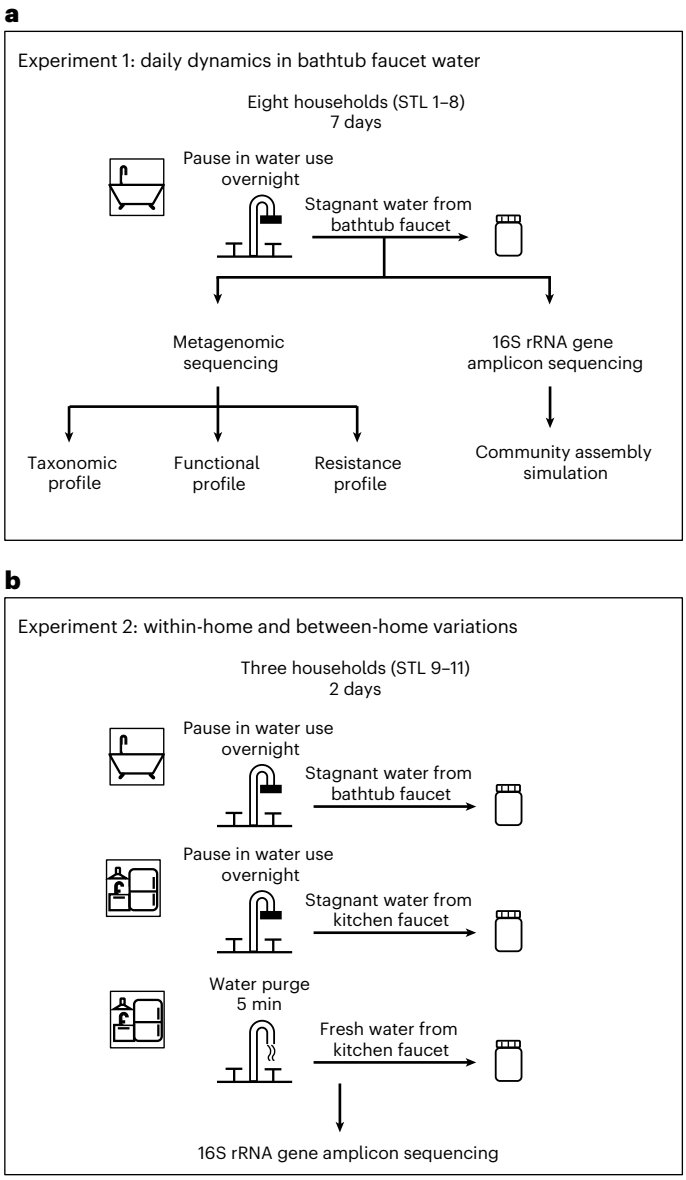


Fig. 1 | Schematic of the two sampling campaigns and analyses performed. **a**, Bathtub faucet water samples were collected from 8 households across 7 days to examine daily dynamics in bathtub faucet water micrbiome. **b**, Three types of water samples were collected from 3 households on 2 days to examine within- and between-home water microbiome variations.

Closer examinations of the household-indicative taxa revealed associations with biofilm-forming species, such as *Mycobacterium* and *Sphingomonas*[38,39]. Furthermore, distinct species within the same genus can signify separate households. For example, *Sphingomonas ursincola* was associated with household STL-7, whereas *Sphingomonas hankookensis* and *Sphingomonas koreensis* were associated with STL-1 and *S. hankookensis* was detected only in STL-1. Moreover, among *Mycobacterium* spp., *M. arupense*, *M. mucogenicum* and *M. gordonae* were associated with households STL-3, STL-4 and STL-6, respectively (Fig. 3b).

While not specifically looking for OPPPs, the model selected the aforementioned non-tuberculous mycobacteria (NTMs) as well as an *Acinetobacter* species, suggesting that household-level individuality was contributed by differences in relative abundances of OPPPs (for extended reports on OPPPs, see 'Presence and prevalence of OPPPs vary by household' section). It is worth noting that the selection of a taxon by LASSO does not necessarily mean that it is unique to the household. Despite this, the relative abundance of these taxa showed strong associations with specific households.

Household identity explains more variation than devices

To investigate how various water use devices and volumetric segments of samples affect our ability to detect between-household differences in microbiomes, we performed a follow-up experiment (Fig. 1b). A total of 54 samples were collected over 2 days from three households, including overnight-stagnant water from a bathtub faucet (hot side), overnight-stagnant water from a kitchen faucet (cold side) and fresh tap water from a kitchen faucet (cold side) following a 5 min faucet flush (Fig. 1b). For each water type, three volumetric segments were taken consecutively: the first, the second and the third litre coming out of the faucet. These samples were analysed via 16S rRNA gene amplicon sequencing. When water samples across households, sample types and sample volumes were analysed together, the PCoA of Bray–Curtis dissimilarities exhibited the most marked clustering by households (Fig. 4). This distinction was supported by tests on the group centroids (PERMANOVA $P = 0.001$; Extended Data Table 1), but it was not explained by differences in the degrees of dispersion (PERMDISP $P = 0.061$).

To what extent can the large heterogeneity be explained by household identity, sample type and volumetric segment, respectively? Household identities significantly affected the group centroids (PERMANOVA $P = 0.001$) and explained the highest proportion of variations (52%) (Extended Data Table 1). Sample type had a significant effect on the microbiome composition (PERMANOVA $P = 0.001$; Fig. 4), but the proportion of variance explained was much smaller than that accounted for by household identities ($R^2 = 0.11$; Extended Data Table 1). Specifically, stagnant bathtub faucet water microbiomes differed significantly from those of both stagnant (PERMANOVA $P = 0.001$) and flushed kitchen faucet water (PERMANOVA $P = 0.001$), whereas the stagnant and flushed kitchen faucet water microbiomes did not significantly differ (PERMANOVA $P = 0.163$; Extended Data Table 1). Sample volumetric segment did not have a significant effect on the water microbiome composition when all samples were analysed together (PERMANOVA $P = 0.291$; Extended Data Table 1) or when each type of sample was analysed separately (PERMANOVA $P > 0.937$; Supplementary Table 2), although this could be affected by the limited statistical power caused by the small sample size. A closer examination of the Bray–Curtis dissimilarities within each household showed that, among the bathtub faucet water microbiomes, the first litre was more different from the other two volumetric segments (Extended Data Fig. 1a). A plausible explanation is that volumetric segments affect water microbiome compositions, yet their effect can be blurred by daily dynamics and household-level differences (Extended Data Fig. 1b).

Presence and prevalence of OPPPs vary by household

To characterize the presence of OPPPs and their relative abundances, we applied MetaPhlAn4 to generate species profiles from the metagenomes of daily bathtub faucet water samples (experiment 1). Diverse NTMs, including *M. arupense*, *M. chelonae*, *M. frederiksbergense*, *M. gadium*, *M. gilvum*, *M. gordonae*, *M. llatzerense*, *M. mucogenicum* and an uncultured species *M.* sp YC RL4, were detected in bathtub faucet water (Fig. 5). These organisms have various degrees of involvement in clinical cases. Among the NTM species, *M. chelonae* is known as one of the most pathogenic rapidly growing mycobacteria, and it is commonly associated with skin and soft tissue infections[40]. *M. arupense* has been associated with a variety of clinical presentations[41]. *M. mucogenicum* is one of the most common causes of bloodstream infection in patients with catheters in place[42,43]. *M. frederiksbergense*, *M. llatzerense* and *M. gordonae* have been associated with rare cases of human infection[44–48]. *M. gilvum* and *M. gadium* have not been reported to cause any human infections. Notably, individual households varied in terms of the NTM diversity and species composition. For example, at a detection

https://doi.org/10.1038/s44221-024-00345-z

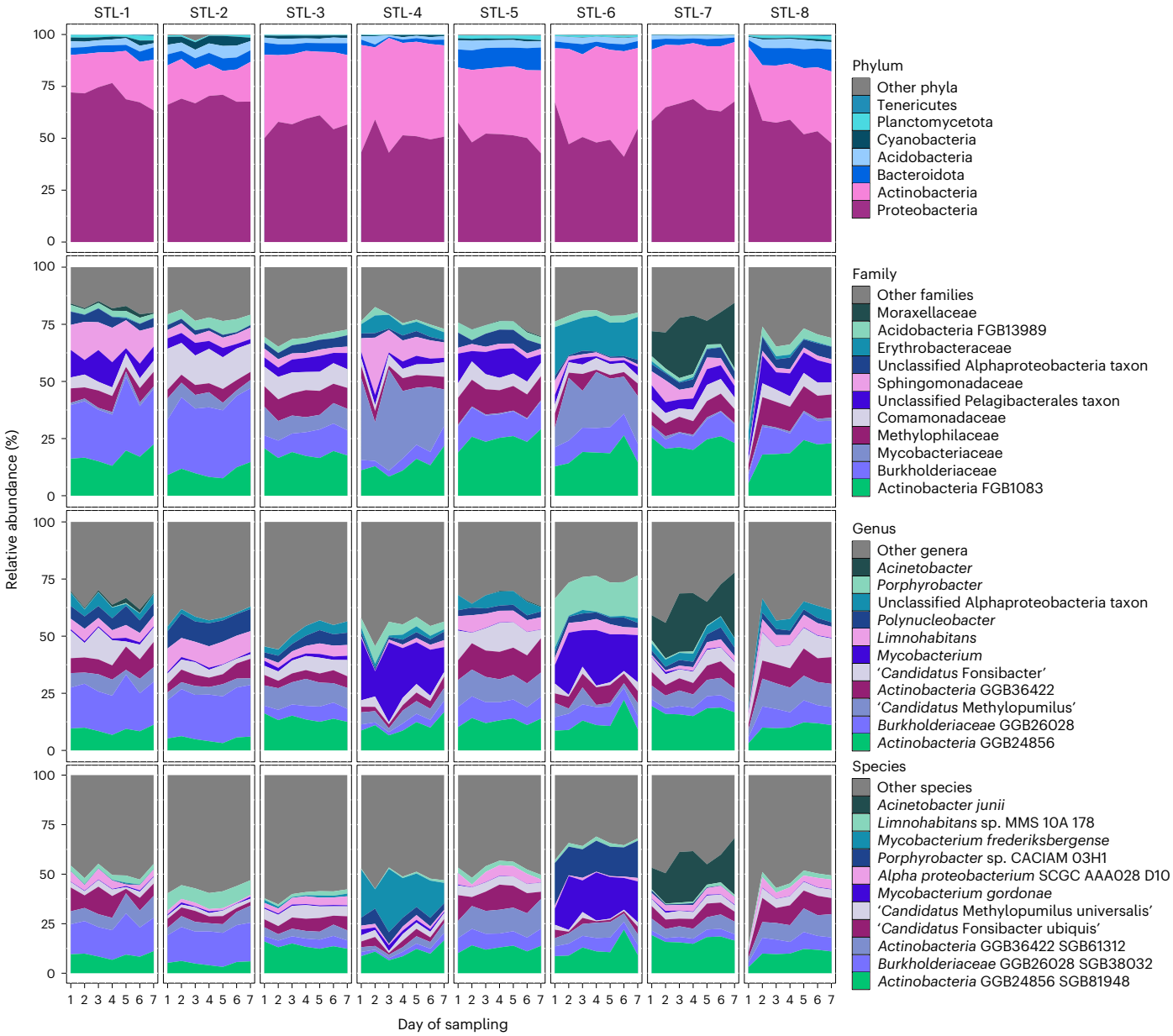


Fig. 2 | Area charts showing the relative abundance of taxa reported by MetaPhlAn4. Taxa with relative abundances lower than 0.1%, 2%, 2% and 2% were grouped as 'Other phyla', 'Other families', 'Other genera' and 'Other species', respectively. FGB, GGB and SGB refer to family-, genus- and species-level genome bins, respectively, as defined in MetaPhlAn4.

limit (DL) of 0.1%, *M. gordonae* was found in all households, whereas *M. chelonae* was found in 50% of the households.

With respect to the other OPPPs, *Legionella pneumophila* was below the DL (0.1%) in all the samples. If the reports below DL were considered, there were extremely low relative abundances in several samples (0.0007–0.006%; Supplementary Table 4). *Mycobacterium avium*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii* were not detected. Among the organisms related to OPPPs at the genus level, *A. junii* and *A. parvus* were above the DL in two and three households, respectively (Fig. 5). Although these two species are rarely associated with infections, *A. junii* can affect patients who have had prior antimicrobial therapy, invasive procedures or malignancies[49,50].

Two *Pseudomonas* clades, *P. chengduensis* and *P. fluvialis*, were detected in multiple households. Both species were recently discovered with isolates from landfill leachate and river sediment, respectively[51,52]. There have not been reports of their associations with infections. A list of all clades related to OPPPs as well as their relative abundances, including both those above and those below the DL, is provided in Supplementary Table 4.

Among the opportunistic pathogen-related species, four out of the nine *Mycobacterium* species (*M.* sp YC RL, *M. mucogenicum*, *M. gordonae* and *M. arupense*) as well as *A. junii* were selected by LASSO as taxa associated with household identity ('Household individuality based on LASSO-selected taxa' section). Their environmental associations were investigated, with regression models built to estimate their relative abundances with total chlorine levels and water temperature as the explanatory variables. Significant counter-correlations with total chlorine levels were detected in three of the four NTM groups (*M. gordonae*, *M. mucogenicum* and *M. arupense*) as well as *A. junii*, indicating a possible positive role of residual disinfectants in controlling these groups (Supplementary Table 5). With respect to the temperature association, only *M.* sp YC RL4 and *A. junii* presented significant

https://doi.org/10.1038/s44221-024-00345-z

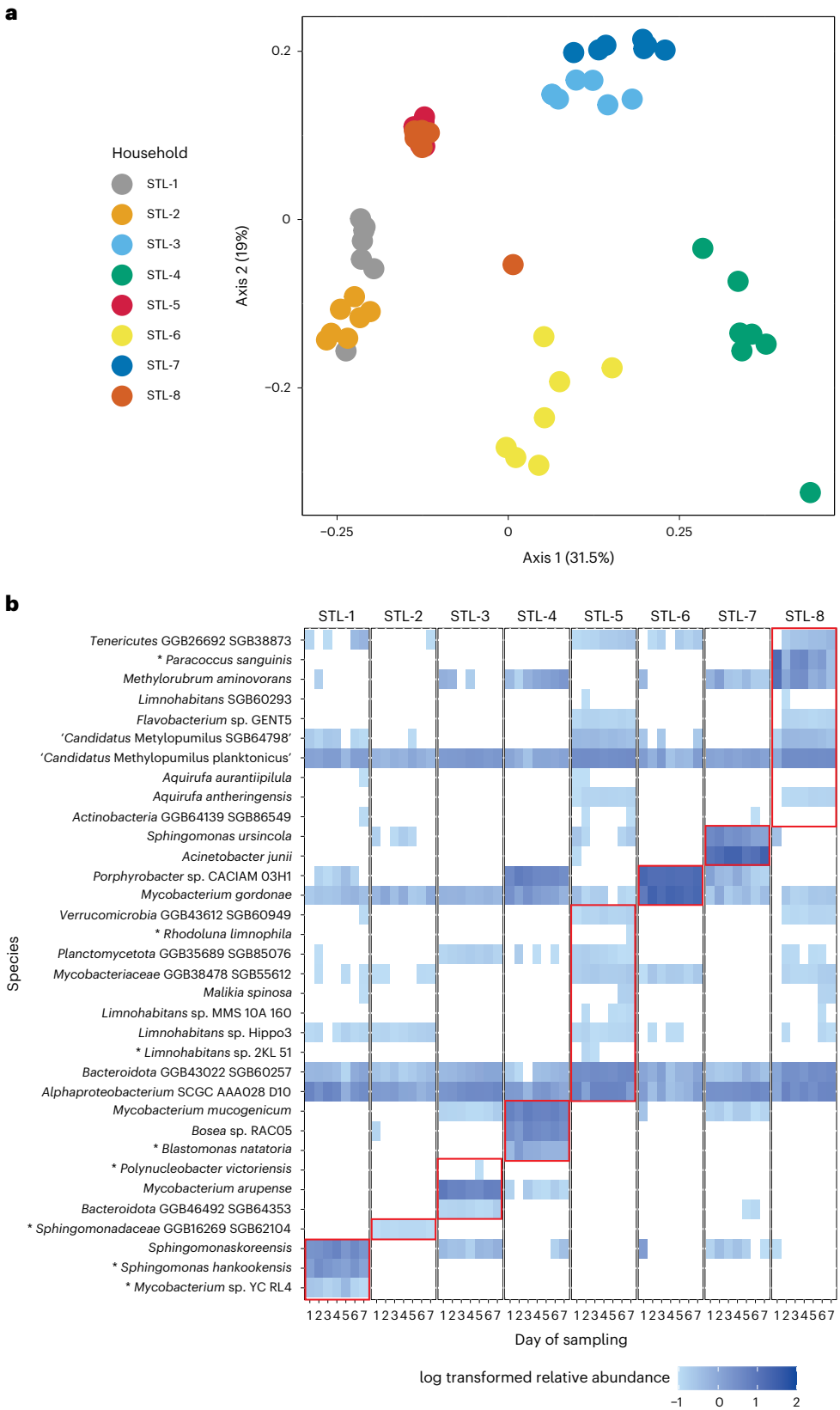


Fig. 3 | Variation in microbial community composition across households. **a**, A PCoA plot of Bray–Curtis distances based on species profiles reported by MetaPhlAn4. Households formed distinctive groups (PERMANOVA $P = 0.001$, $R^2 = 0.84$). PERMDISP indicated non-significant differences in dispersions ($P = 0.176$). *P* values were computed on the basis of two-sided tests. Between-household comparisons yielded adjusted *P* values below 0.05 for all comparisons after the Benjamin–Hochberg and Holm corrections. **b**, LASSO regression models identified informative species for each household. The heatmap was plotted using log transformed relative abundances, where darker colour indicates higher relative abundance, and white indicates the absence of detection or a value lower than the detection limit (0.1%). Red rectangles from left to right enclose the bacterial species whose combinations were indicative of households STL-1 to STL-8, respectively. Taxa with asterisks are those that were uniquely detected in only one household.

https://doi.org/10.1038/s44221-024-00345-z

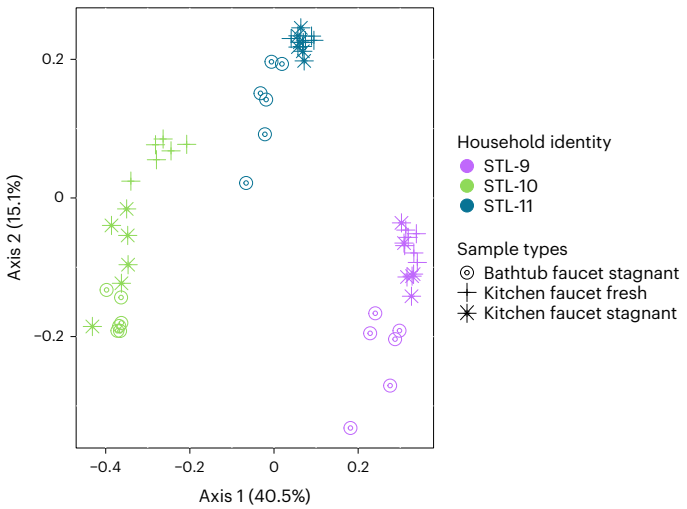


Fig. 4 | Variation in microbial community composition within- and between-households. PCoA plot of Bray–Curtis distances between water microbiome samples collected from multiple types of water across three households. Water microbiomes are primarily clustered by household (PERMANOVA $P = 0.001$, $R^2 = 0.52$) despite water use devices. Within each household, the water microbiomes were clustered by sample type (PERMANOVA $P = 0.001$, $R^2 = 0.11$), that is, fresh water collected from kitchen cold faucets (crosses), overnight-stagnant water from kitchen cold faucets (asterisks) and overnight-stagnant water from bathtub hot faucets (circles). Pairwise comparisons after Benjamin–Hochberg and Holm corrections are shown in Extended Data Table 1.

associations, yet the directions of the effects differed, suggesting a more nuanced role of temperature in OPPP control.

Bath faucet water across households has similar resistomes

To explore the antimicrobial resistome in bathtub faucet water, we mapped the metagenomic reads to MEGARes 2.0, a database of antimicrobial resistance genes (ARGs), using the AMR++ 2.0 pipeline[53]. A total of 162 ARGs were detected across 56 water metagenomes. The most abundant traits in the resistomes were multidrug resistance, macrolide–lincosamide–streptogamine resistance and beta-lactam resistance (Fig. 6a). These traits have also been reported in other drinking water systems[35,36,54]. Multidrug efflux pumps have been associated with increased microbial survival in response to water disinfectants and other environmental stressors[55–57]. Because chloramines are routinely used for disinfecting the drinking water supply, the enrichment of these genetic determinants in the resistome is anticipated. ARGs conferring resistance to critically important antimicrobials, such as Guiana extended-spectrum β-lactamase (bla_{GES}), $bla_{FAR\text{-}1}$, $bla_{BJP\text{-}1}$ and Adelaide imipenemase gene (bla_{AIM}), were also detected (Fig. 6b). These genes can confer resistance to high-priority critically important antimicrobials such as carbapenems and monobactams, which are used for the treatment of recalcitrant multidrug-resistant infections[58–62].

It is noteworthy that, while the microbial community compositions were household specific, the antimicrobial resistance profiles did not show strong clustering by home (Extended Data Fig. 2; PERMANOVA P values after adjustment are shown in Supplementary Table 6). In addition, the abundant ARGs, with an average reads per kilobase of transcript per million reads mapped (RPKM) ranking in the top 20, were detected in all the bathtub faucet water samples for at least 1 day. The detection frequency of those genes ranged from 0.59 to 1 (Supplementary Table 7). It is possible that the similarities between ARG profiles from different households and multiple days within the same household arise from the use of a common disinfectant, chloramines, in the municipality. There are slight variations in ARG detection within households (Fig. 6b), yet with the ARGs detected at small RPKMs, these variations are potentially related to the daily dynamics of the rare or less abundant microbes.

Similar coarse-grained functions and distinct gene profiles

To explore the functional profile of bathtub faucet water microbiomes, we annotated the functional pathways from their metagenomes. In total, 9,650 Kyoto Encyclopedia of Genes and Genomes orthology (KO) functions were detected. The functional profiles were very similar between homes at coarse functional groups, specifically at the Functional Ontology Assignments for Metagenomes (FOAM) levels 1 (Fig. 7a) and 2 (Supplementary Figs. 3–22). For example, in terms of the nitrogen cycle, genes related to nitrogenated compound reduction, nitrogen fixation, nitrification and ammonium assimilation were detected in all the bathtub faucet water samples (Fig. 7b). Similarly, with respect to methylotrophy, genes related to methane oxidation to CO_2, methanol oxidation to formaldehyde, methylamine to formaldehyde and various pathways involved in the transformation of formaldehyde were detected in all households (Extended Data Fig. 3). Not all functions were well resolved at the finer levels, for example, FOAM level 3. However, for those functions that were well classified at FOAM level 3, such as the nitrogen cycle, the profiles were highly similar across homes (Extended Data Fig. 4). Curiously, at the gene level, variations across households emerged. Distinct clustering by homes was observed in the PCoA of the Bray–Curtis dissimilarity of the KO profiles (Extended Data Fig. 5; PERMANOVA $P = 0.001$; PERMANOVA P values after adjustment are shown in Supplementary Table 8), suggesting household individuality.

We further asked whether the cross-household similarities in coarse functional groups (FOAM Levels 1 and 2) were specific to the St. Louis drinking water or common for other premise plumbing water microbiomes. Premise plumbing water metagenomes from a pilot-scale hot water plumbing rig study in Blacksburg, Virginia, were analysed[63]. Similar FOAM Level 1 profiles were detected between the St. Louis water and pipe rig water samples (Extended Data Fig. 6). At FOAM level 2, when the nitrogen cycle was used as an example, functions involved in the nitrogen cycle were present in most of the samples (Extended Data Fig. 7), although the detection of specific KOs varied (Extended Data Fig. 8). Thus, the coarse-level functional similarity probably arises from adaptation to residual disinfectant-carrying drinking water systems as a kind of habitat.

Correlation between taxonomic and functional profiles

When the similarity of the functional profiles and that of the taxonomic profiles were compared, the similarity of the species profiles was well correlated with that of the KO profiles ($R^2 = 0.6$, $P = 2.2 \times 10^{-16}$; Fig. 7c). However, the similarity in ARG profiles was less well explained by the similarity in the KO profiles ($R^2 = 0.06$, $P = 7.51 \times 10^{-12}$; Fig. 7d) and not well explained by that of the species profiles ($R^2 = 0.001$, $P = 0.097$). The dissociation between the ARG and KO profiles further highlights that ARGs are highly mobile.

The contrast between taxonomic diversity and functional diversity has been used to indicate functional redundancy, that is, taxonomically distinct community assemblages carrying similar functions[64]. In this context, a complete lack of correlation indicates functional redundancy, whereas a very high correlation indicates otherwise. In our bathtub faucet water metagenomes, the correlation between taxonomic diversity and functional diversity was moderate (Fig. 7e; $R^2 = 0.18$, $P = 0.0006$) in comparison with reports of other microbial systems ($R^2 = 0.76$ in soil systems and $R^2 = 0.69$ in a marine system)[65,66]. The moderate correlation suggests partial functional redundancy. The variation in functional diversity was also affected by household identity (Supplementary Fig. 23). Nevertheless, the Bray–Curtis dissimilarities between water microbiomes from different homes were significantly greater for the species profile (mean of 0.47, s.d. of 0.16) than for the KO profile (mean of 0.17, s.d. of 0.05), suggesting that taxonomically

https://doi.org/10.1038/s44221-024-00345-z

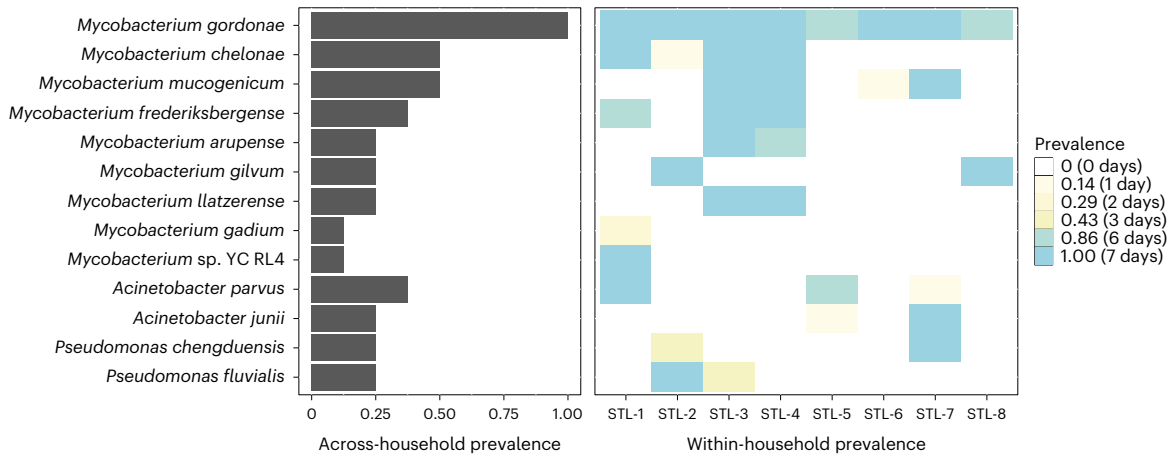


Fig. 5 | Presence and prevalence of OPPPs. Prevalences of *Pseudomonas*-, *Mycobacterium*- and *Acinetobacter*-related species across and within households. Left: a bar plot showing the prevalence across households, where presence is defined as detection on 1 or more days (detection limit of 0.1%). Right: a heatmap indicating the prevalence within each household over a 7 day sampling period, with white indicating values below the detection limit, shades of yellow representing intermediate prevalences and shades of blue indicating high prevalences.

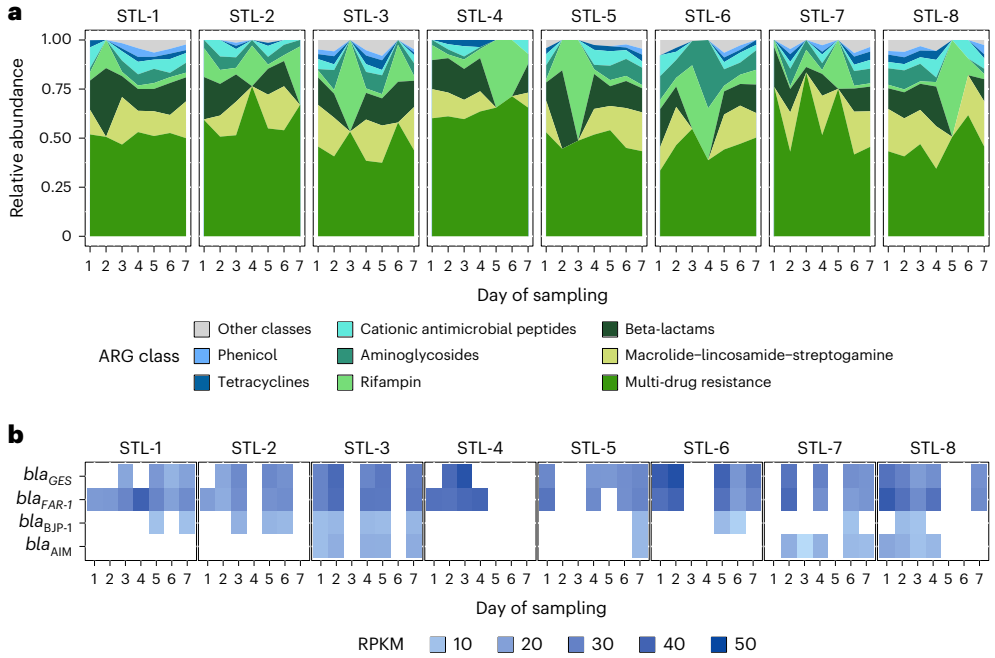


Fig. 6 | Antimicrobial resistance across households. a, An area chart showing the relative abundances of ARGs associated with resistance to different drug classes. Relative abundances were computed on RPKMs. **b**, A heatmap showing RPKMs of four extended-spectrum beta-lactamases that confer resistance to last-resort antibiotics.

distinct microbes in the bathtub faucet water microbiome may carry similar functions (Fig. 7f). Functional redundancy has been examined in other microbial systems as a mechanism to maintain community resilience. This mechanism potentially contributes to the maintenance of the drinking water microbiome.

Community assembly in the bath faucet water microbiome

We used a null-model-based approach to further investigate the relative importance of stochastic and deterministic assembly processes in shaping the bathtub faucet water microbial communities. In this approach, the stochastic processes include 'drift' (DR), dispersal limitation and homogenizing dispersal (HD), and the deterministic processes include homogeneous selection (HoS) and heterogeneous selection (HeS). The household premise plumbing water microbial community could be seeded by two possible sources: the fresh supply from the distribution system and the biofilm in the premise plumbing system. The process of migration from a common public water supply to individual houses can be conceived of as homogenizing dispersal, which usually results in a similar community structure in the local community in each house (Fig. 8a). On the other hand, the plumbing biofilms in each house could harbour distinct microbial communities, depending on the specific plumbing environment conditions in each house. The migration of microbes from biofilms to the water in each house could be considered as dispersal limitation, since those microbes are isolated from one house to another. In addition, the premise plumbing microbiome is constantly under pressure from disinfectant residuals and warm temperatures, which might cause environmental filtering for particular microorganisms. The structure of the microbial community

https://doi.org/10.1038/s44221-024-00345-z

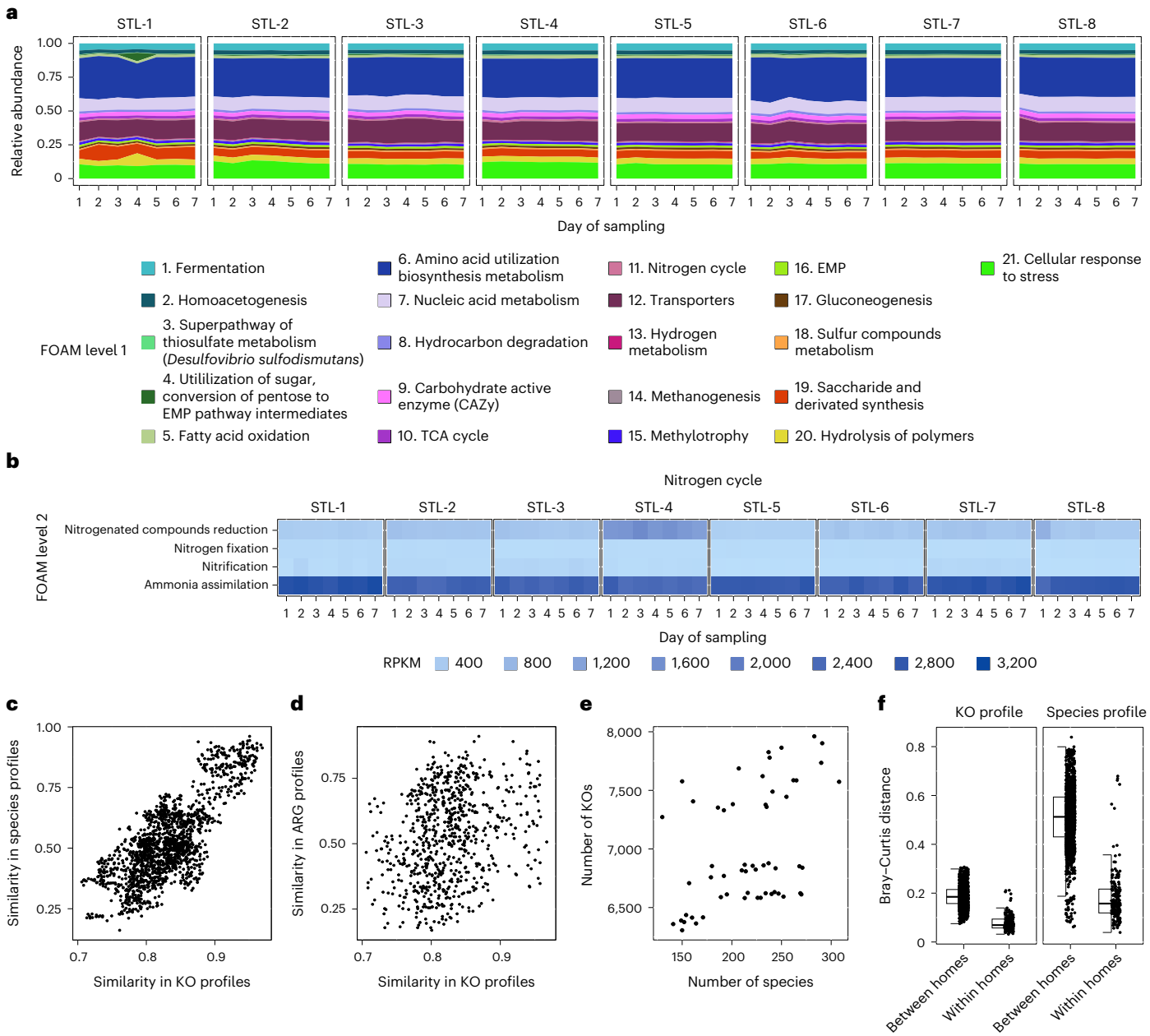
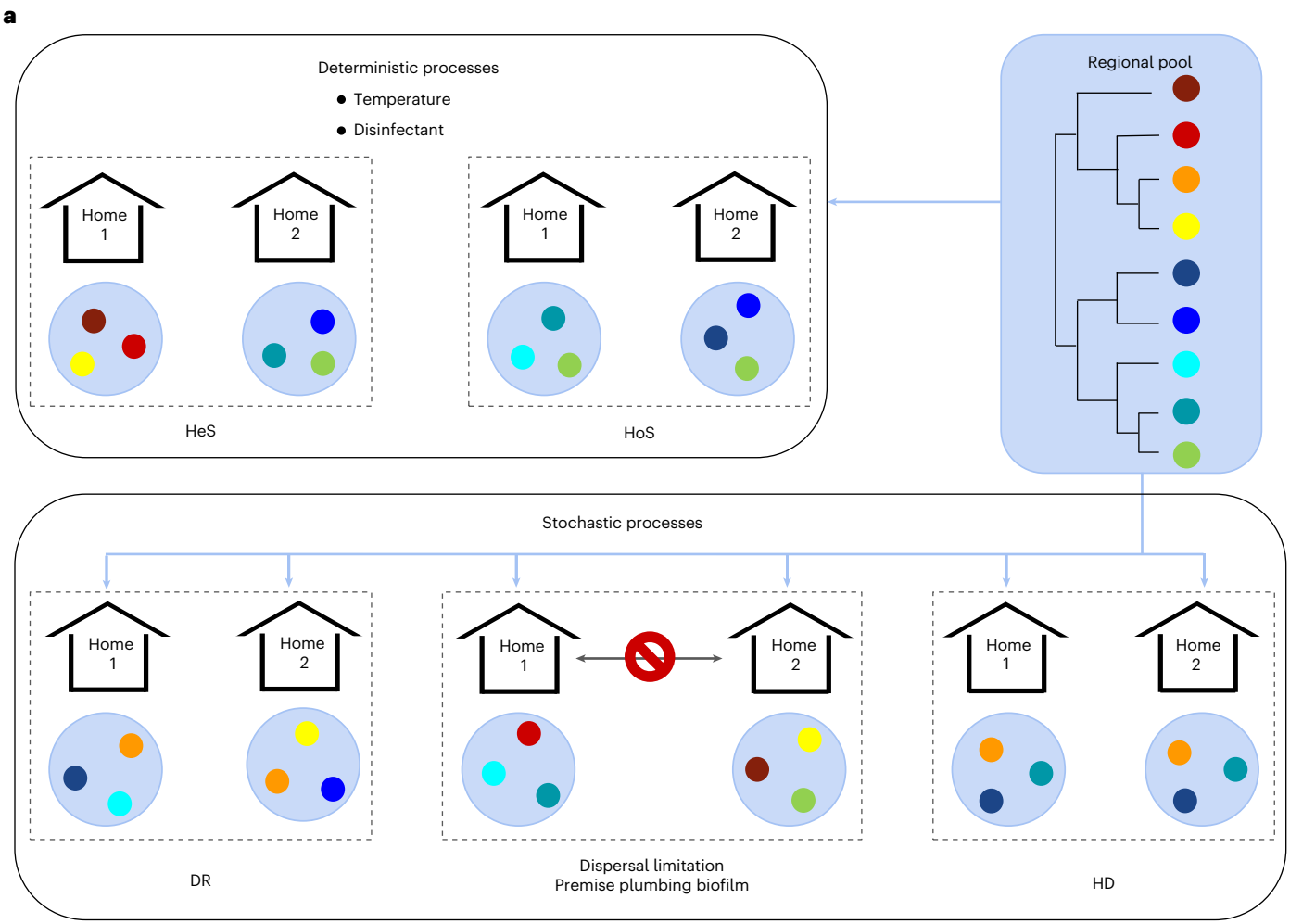


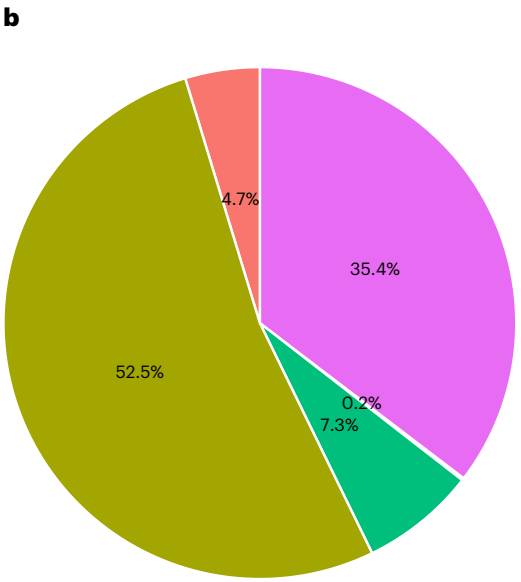
Fig. 7 | Functional profile and its correlation to taxonomic profile. a, An area chart showing the relative abundances of environmentally relevant functional groups (FOAM level 1) in bathtub faucet water microbiomes. Overall similarity in functional potentials emerged at broad functional groups. Relative abundances were computed on RPKMs. EMP, Embden Meyerhof–Parnas; TCA, citric acid cycle. **b**, A heatmap showing RPKMs of FOAM level 2 functions in the nitrogen cycle. **c**–**e**, Linear regression models performed to examine the relationship between variable and response (*P* values for coefficients computed based on two-sided tests), revealing that similarity in species profiles was highly correlated with similarity in KO profiles (**c**, $R^2 = 0.6$, $P = 2.2 \times 10^{-16}$), similarity in ARG profiles was weakly correlated with similarity in KO profiles (**d**, $R^2 = 0.06$, $P = 7.51 \times 10^{-12}$) and functional diversity was moderately correlated with taxonomic diversity (**e**, $R^2 = 0.18$, $P = 0.0006$). **f**, There were greater similarities between households in the KO profiles than in the taxonomic profiles. In the box plot, there were 168 within-home comparisons and 1,372 between-home comparisons. The box shows the interquartile range (IQR), which spans from the 25th percentile (Q1) to the 75th percentile (Q3) of the data. The thick line inside the box represents the median of the data. The lower whisker extends from Q1 to the smallest value in the dataset that is greater than or equal to Q1 − 1.5 × IQR (minima); the upper whisker extends from Q3 to the largest value in the dataset that is less than or equal to Q3 + 1.5 × IQR (maxima).

in the premise plumbing could be a joint result of all the aforementioned ecological processes.

The 16S amplicon data from daily sampling across eight homes were examined for the community assembly processes. Stochastic processes contributed greater proportions of the community assembly forces than did deterministic processes (64.48% versus 35.52%). Among stochastic processes, 'drift' was the most dominant, contributing an average relative importance of 52.51% across homes (s.d. of 10.93%, $n_{home} = 8$). Homogenizing dispersal and dispersal limitations contributed an average of 7.27% (s.d. of 5.35%, $n_{home} = 8$) and 4.71% (s.d. of 3.07%, $n_{home} = 8$; Fig. 8b), respectively. Between the two deterministic processes, HoS was dominant, contributing to an average of 35.37% (s.d. of 9.09%, $n_{home} = 8$). The overall dominant role of stochastic processes is in agreement with reports of community assembly in

https://doi.org/10.1038/s44221-024-00345-z





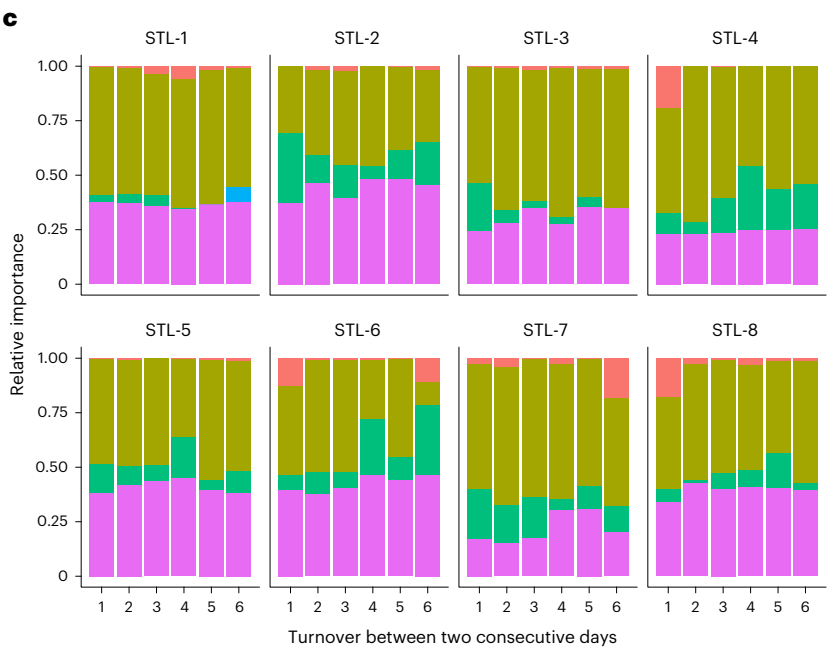


Fig. 8 | Community assembly in drinking water microbiome. a, Schematics of the community assembly mechanism in drinking water. **b**, The relative importance of each process in between-household variations in community composition. The '0.2%' label represents HeS. **c**, The relative importance of each ecological process in the daily turnover of the whole microbial community in each household.

environmental systems elsewhere[67,68]. The greater importance of HoS indicates that the selection may take place among taxa that are closely related, for example, different amplicon sequence variant (ASVs) within the same genus.

Furthermore, we investigated the relative importance of each ecological process in the daily turnover of the microbial communities within the same home. Stochastic processes still governed the turnover between communities on consecutive days in each home (Fig. 8c). HoS, the most important deterministic process, can be attributed to the fact that the water in each home was exposed to similar environmental filtering from day to day, for example, disinfectant residuals and warm temperatures. As a result, the microbial communities were stable under similar environmental pressures. In addition, the relative importance of stochastic and deterministic processes remained fairly stable in the daily turnover of the communities in each home.

To discern the relative importance of the measured environmental variables contributing to the variation in microbial community composition, we applied multiple regression on distance matrices (MRM), with the Bray–Curtis dissimilarity as the response distance matrix and total chlorine concentration, temperature, building age and pedestrian distance as explanatory distance matrices. The model outputs revealed significant effects of total chlorine concentration ($P = 0.001$) and pedestrian distance ($P = 0.008$) but not temperature ($P = 0.142$) or building age ($P = 0.275$) (Extended Data Table 2). Upon backwards selection, a final model consisting of total chlorine concentrations and pedestrian distances explained 23.17% of the variance. When each variable was examined independently, total chlorine concentrations explained the highest variance ($R^2 = 0.21$; Supplementary Table 9). The MRM analysis showed that the majority of the community variation (76.83%) could not be explained by the measured environmental variables, which is consistent with the dominant role of stochastic processes inferred from the null-model-based approach.

Discussion

Implications for monitoring antimicrobial resistance risks

The antimicrobial resistance profiles were highly similar across households despite differences in taxonomic compositions (Fig. 6), potentially driven by the common chloramine residuals applied throughout local drinking water distribution systems. Notably, we detected the presence of genes encoding extended-spectrum beta-lactamases in multiple households, such as bla_{GES}, $bla_{FAR\text{-}1}$, $bla_{BJP\text{-}1}$ and bla_{AIM}, which are genes that can confer resistance to aztreonam and meropenem, two last-resort antibiotics. The presence of extended-spectrum beta-lactamase genes in the premise plumbing microbiome is alarming because clinically important resistance, even when carried by non-pathogenic environmental organisms, can be horizontally transferred to opportunistic pathogens under selective pressures[58]. Following the resistome analysis, we performed follow-up analyses using metagenome assembly and binning to link the resistance determinants to specific taxa (Supplementary Method 1). This analysis revealed the presence of genome bins related to *Pseudoxanthomonas mexicana* carrying the bla_{AIM} gene (Supplementary Results). Recently, it has been hypothesized that *Pseudomonas aeruginosa* with newly discovered beta-lactam resistance could have acquired the bla_{AIM} gene from *P. mexicana* through horizontal gene transfer; however, the exact mechanisms involved remain to be elucidated[58]. With *P. aeruginosa* being a group of opportunistic premise plumbing pathogens and our findings of *P. mexicana* carrying the bla_{AIM} gene in premise plumbing, we hypothesize that potable water uses (for example, shower and wound cleaning) could act as routes where the human microbiome interacts with bla_{AIM}-carrying *P. mexicana*. Future work to identify and confirm the host range for resistance determinants, using techniques such as long-read sequencing or emulsion, paired isolation and concatenation polymerase chain reaction[69–71], will probably be a promising avenue, despite the challenges when applying these techniques in low-biomass drinking water systems. More broadly, strategies to monitor antimicrobial resistance in drinking water deserve further attention. Liguori and colleagues suggested discussions on prioritized targets and suitable methodologies for antimicrobial resistance monitoring in various aquatic environments[72]. Proctor and colleagues advocated for a holistic approach to address antimicrobial resistance and OPPP challenges[73]. With respect to the premise plumbing environment, our findings suggest the importance of broadly monitoring the resistome, as well as gaining a more specific understanding about non-pathogenic environmental organisms with potential horizontal gene transfer mechanisms to OPPPs.

Implications for monitoring opportunistic pathogen risks

Our study detected diverse species of *Mycobacterium*, *Pseudomonas* and *Acinetobacter* in household bathtub faucet water. Notably, the diversity of OPPPs at the species level can vary highly across households; however, once OPPPs are detected, they can be robustly detected across multiple days. While a lot of attention has been given to the public health risks associated with OPPPs, a quantitative risk assessment framework for NTM is currently available for only the *M. avium* complex[74]. The varied prevalences of NTM species across households are especially relevant when considering future developments in dose–response models. Doses estimated from representative samples from water facilities or distribution systems may not be sufficient to infer the health risk faced by each resident, since every home has a unique species profile. Therefore, a more intensive and individual household-based sampling plan, for example, an environmental exposome approach that considers multiple pathogens and/or contaminants, exposure points and spatiotemporal variations[73,75], should be considered when evaluating the health risk from waterborne opportunistic pathogens.

Implications for mechanism-based biological quality control

We showed that variations in water microbiome taxonomic compositions, both the between household variation and the daily turnover, were largely driven by stochastic processes (Fig. 8b,c). MRM analysis revealed that the majority of the variation in the microbial community was not explained by the measured environmental variables (Extended Data Table 2), which is consistent with the result inferred from the null-model-based approach (Fig. 8), although uncharacterized environmental variables could also contribute to the unexplained variance. HoS came second. The dominance of stochasticity and HoS appeared counterintuitive given the observed household individuality in community compositions, yet multiple ecological processes, such as drift, HoS and HeS, can each independently lead to household individuality (as detailed in Supplementary Discussion).

The new understandings of the community assembly mechanisms in premise plumbing could shed light on more mechanism-informed biological quality control strategies. For example, stochastic assembly does not mean that the community assembly is not controllable, given that stochastic processes have been found to be highly related to resources[76–78] or stress levels[79]. In the context of the drinking water microbiome, this provides a justification for strategies to reduce the total cell count, such as limiting the overall nutrient input to the distribution system by biofiltration and reducing stagnation at the point of use. The multitude of community-level assembly mechanisms shown here provide a justification for future work to examine specific mechanisms driving the distribution of organisms of interest to develop mechanism-informed control strategies. For example, species governed by selection need to be controlled via environmental conditions, for example, temperature and disinfectants, whereas those governed by dispersal will require the intervention of dispersal or source communities, for example, the control of hydraulic conditions. Computational tool development to link community assembly mechanisms to species-level taxa still presents challenges, yet their

successful development and implementation will probably generate valuable new insights.

Conclusion

This study examined the within- and between-household variations in the premise plumbing microbiomes as well as the daily variations in the microbiomes and resistomes of household bathtub faucet water. We found that households had distinctive water microbiomes and that household-specific signatures persisted throughout the week-long sampling of bathtub faucet water. Households differ in their species-level compositions of NTMs as well as the prevalence and relative abundances of detected species, with *M. gordonae* being twice as prevalent as *M. chelonae*. Despite differences in microbial community composition, broad categories of antimicrobial resistomes were shared across households, and ARGs conferring resistance to critically important antimicrobials were detected in the bathtub faucet water from each household. Community assembly simulations revealed that variations in water microbiome taxonomic compositions, both between-home variations and the daily turnover, were largely driven by stochastic processes. Our results highlight the importance of species- and household-level data for assessing health risks associated with NTMs. The detection of clinically important resistance genes also calls for further monitoring of antimicrobial resistance and its putative relationship with opportunistic pathogens in premise plumbing.

Methods

Background on distribution systems

In this study, water microbiomes were sampled from 11 households in the St. Louis, Missouri, area between 2018 and 2021 (details in 'Sampling design' section). St. Louis City received drinking water from the St. Louis City Water Division, whereas St. Louis County received water from Missouri American Water. Depending on water pressure, households near the border between the city and county could receive water from either source or receive a mixture. Because the objective of this study did not involve examining distribution system factors, volunteer recruitment was agnostic to water sources.

Sampling design

Two sampling campaigns examined the daily water microbiome dynamics ('Sampling to examine daily water microbiome dynamics' section; Fig. 1a) and within-household water microbiome variations ('Sampling to examine within-household variations' section; Fig. 1b).

Sampling to examine daily water microbiome dynamics. Following protocol training provided by the core research team, the occupants of eight households sampled water from a bathtub hot-side faucet on seven consecutive days in October 2018 and October 2019. In this study, water was collected from bathtub faucets to accurately control the sampling volume and avoid any loss of water to splashing, which are crucial factors in retrieving samples representative of the premise plumbing water. All samples were taken in the morning, immediately after overnight stagnation, to minimize the impacts of same-day variation in the water microbiomes[80,81]. To collect the samples, participants used a provided sampling kit containing three steam-sterilized 1 litre polypropylene containers, three 50 ml Falcon tubes and a rack to hold them, a thermometer, a data sheet, gloves, alcohol pads and an insulated bag. Before any water usage in the morning, 3 liters of water samples from the bathtub faucet were collected in the sterilized polypropylene containers. The participants were trained to wear sterile nitrile gloves during sampling. The water stagnation times (that is, the time between the use of the shower or bathtub faucet, typically overnight) were recorded. Within 2 h after collection, the water samples were transported in an insulated bag to the laboratory for biomass collection and water chemistry analyses.

Sampling to examine within-household variations. To examine the effects of water usage scenarios on the water microbiome compositions, a separate sampling effort was conducted. Three types of water samples were collected: (1) overnight stagnant water from a bath hot-side faucet, (2) overnight stagnant water from a kitchen sink cold-side faucet and (3) fresh water from the kitchen sink cold-side faucet, which was collected after allowing the faucet to run cold water for 5 min. For each type of water, three 1 litre water samples were collected and labelled sequentially, with the tap closed between samples. On two consecutive days in June 2021, water samples were collected from three households. Within 2 h after collection, the samples were transported to the laboratory in an insulated bag for biomass collection and water chemistry analyses.

Water temperature and total chlorine measurements

Water temperatures were measured on site by trained participants using thermometers provided by the research group. To protect the integrity of the water samples in the 1 litre containers for later microbiological analyses, for immediate temperature measurement, participants were trained to pour a small sample from each 1 litre container directly into one of the provided 50 ml tubes immediately after sample collection. The total chlorine concentration was measured in the laboratory on the same day as the samples were received, using the *N*,*N*-diethyl-*p*-phynylenediamine colorimetric method (Standard Method 4500-Cl G). The temperature and total chlorine concentration data are provided in Supplementary Figs. 26 and 27.

Total organic carbon and nitrate measurements

Total organic carbon (TOC) and nitrate concentrations were measured for four households in the first campaign: STL-1, 2, 6 and 8. The other households did not participate in the chemical analyses because they had moved out. TOC was measured following the non-purgeable organic carbon method using a TOC-L analyser (Shimadzu, Japan). Nitrate concentration was measured using ion chromatography (Dionex, USA) according to Environmental Protection Agency method 300.0. The nitrate and TOC results are shown in Supplementary Fig. 28.

Other metadata

We collected building age data and renovation records from real estate brokerage websites, for example, LoopNet and Redfin (Supplementary Table 10). Street addresses were used to calculate pedestrian distances between households from the Google Maps application programming interface. The actual addresses were excluded from the reported metadata to protect the participants' privacy. We chose to use pedestrian distances as a proxy for pipe distances between households, which are more relevant for the potential of microbial dispersal.

Biomass collection and genomic DNA extraction

Biomass was collected by filtering each water sample through a sterile 0.22 µm mixed cellulose ester membrane filter (Millipore Sigma; catalogue information provided in Supplementary Table 11) to collect the retained cells. Genomic DNA from the water samples was extracted using a protocol previously demonstrated to have high yield and low bias towards the drinking water microbiome[82,83]. Briefly, the membrane filters were dissected and transferred into sterile screw-cap centrifuge tubes with sterile 0.1 mm glass beads. The samples were subjected to bead beating at 4.5 m s^{-1}, enzymatic treatment with lysozyme (Millipore Sigma) and achromopeptidase (Millipore Sigma), followed by two incubations. The first incubation was with a mixture of proteinase K (New England Biolabs) and sodium dodecyl sulfate (Millipore Sigma), and the second was with a solution of hexadecyltrimethylammonium bromide (Millipore Sigma) and sodium chloride (Millipore Sigma). Following these treatments, DNA was extracted with a premixed phenol–chloroform–isoamyl alcohol solution (25:24:1) (Millipore Sigma), and precipitated overnight in an isopropanol solution (Millipore Sigma).

The extracted DNA pellets were resuspended in molecular biology grade water (Sigma Aldrich) and then purified with the Wizard DNA Cleanup System (Promega). Two types of blank controls were included in the analysis. One was a membrane filter autoclaved under the same conditions as the filters for biomass collection from water samples and extracted using the same protocol, and the other contained only DNA extraction reagents. The blank samples were used in sequencing analyses to minimize noise from reagents. Purified DNA samples were quantified with an AccuBlue High Sensitivity dsDNA Quantification Kit (Biotium, USA) and then stored at −80 °C until sequencing library preparations.

16S rRNA gene amplicon library preparation and sequencing

The 16S rRNA libraries were constructed through two steps of PCR[84]. First, the extracted DNA was amplified using a universal primer set 515F(GTGCCAGCMGCCGCGGTAA)/806R(GGACTACHVGGGTWTCTAAT), which targets the V4 region of the 16S rRNA gene. Next, unique indexing sequences were added to each library. Technical replicates were performed for every sample by preparing two libraries with different barcodes. The final products were purified with the Wizard SV Gel and PCR Clean-Up System. Multiplexed sequencing was performed on an Illumina MiSeq platform (2×250) at the Edison Family Center for Genome Sciences and Systems Biology at Washington University in St. Louis.

Shotgun metagenomic sequencing

Metagenomic sequencing was performed on daily water samples collected as described in 'Sampling to examine daily water microbiome dynamics' section. Shotgun genomic libraries were prepared using the Nextera DNA Flex Library Prep Kit, with an average library insert size of 437 bp (ranging from 200 to 2,375 bp), and sequenced on a NovaSeq 6000 platform at the Genome Technology Access Center at the McDonnell Genome Institute.

16S rRNA gene sequence analysis

Raw sequencing reads were demultiplexed using the Illumina bcltofastq2 program at the Edison sequencing facility. Upon receiving demultiplexed raw fastq paired-end reads, we completed quality filtering and denoising and then generated ASV tables using the QIIME 2 platform[85]. For quality control, the forward reads were trimmed at 10 bp and truncated at 180 bp; the reverse reads were trimmed at 10 bp and truncated at 150 bp. This procedure resulted in a quality score above 30 for all the nucleotides. The trimmed reads were denoised and merged using the DADA2 algorithm as implemented in the q2-dada2 plugin[86], which uses a quality-aware model of Illumina amplicon errors and resolves differences as subtle as one nucleotide. Sequencing reads with more than two expected errors (that is, the average number of errors under the error probability distribution) and detected chimeras were discarded. An ASV table was thus generated. Taxonomy classification of the resulting ASVs was performed using a multinomial naive Bayes classifier, which was trained on a SILVA 138 reference database with the confidence threshold set at 0.7 (refs. 87,88).

The QIIME 2 objects were imported into R (version 3.6.1) using the qiime2R package (https://github.com/jbisanz/qiime2R) for diversity analysis, statistical analysis and visualization. Details of the preprocessing of technical replicates and blank control samples are provided in Supplementary Method 2. To compare microbial diversity across samples, the obtained sequences were first rarefied to 70,399 sequences per sample for the experiment 1 data and 4,434 sequences per sample for the experiment 2 data (determined by the sample with the fewest sequences). Alpha and beta diversity analyses were conducted with the R packages phyloseq v1.30.0 and vegan v2.5-6 (refs. 89,90). Community composition differences between samples were computed via the Bray–Curtis and Jaccard distances and visualized by PCoA. To test the significance of between-group differences in community structure and dispersion (within-group variance), PERMANOVA and PERMDISP were performed using the 'adonis2' and 'betadisper' functions in the vegan package[89]. In cases where multiple comparisons occurred, the 'p.adjust' function was implemented to report *P* values after both Benjamini–Hochberg and Holm corrections.

Metagenomic sequence quality control

The shotgun metagenomic sequencing data were demultiplexed at the sequencing center using Illumina's BCL Convert software. PCR duplicates were removed from paired-end reads using FastUniq v1.1 (ref. 91). Quality trimming of the raw reads was performed using Trimmomatic v0.39 with a sliding window of 6 bp and quality cutoff of 30 (ref. 92). The average number of reads in each metagenome was 55,774,223 (s.d. of 24,673,160, $n = 56$) after quality filtering.

Species profiling from metagenomic data

MetaPhlAn4 was used for taxonomic classification of 56 water metagenomes[93]. We applied MetaPhlAn4 because it is a marker-based profiler that directly outputs relative taxonomic abundance representing the fraction of each detected taxon[94]. The MetaPhlAn4 database incorporates approximately 5.1 million unique clade-specific marker genes identified from 26,970 species-level genome bins[93]. The presence and relative abundances of species related to *Legionella*, *Mycobacterium*, *Pseudomonas* and *Acinetobacter* were retrieved. Because of the recent taxonomic debate around the *Mycobacterium* genus, manual curation of MetaPhlAn4 outputs was performed to reflect the latest consensus[95,96]. Specifically, the genera *Mycolicibacterium*, *Mycolicibater*, *Mycolicibacillus* and *Mycobacteroides* were updated to *Mycobacterium* according to the recommendations of Meehan and colleagues[95]. To prevent false positives in metagenomics species detection, outputs below 0.1% were excluded from opportunistic pathogen and household indicator analyses[97,98].

LASSO regression

Informative species for each home were identified by performing LASSO regression using the caret package v6.0-94 (ref. 99). The metagenomic data from the daily sampling campaign (Fig. 1a) were used for this analysis. For each household, a LASSO regression model was built to separate the home from the rest of the homes, with microbial taxon abundances as inputs. For each home, a logistic approach was applied by labelling samples from this home as 1 and those that were not from this home as 0 (equation (1)):

$$\log\left(\frac{p(y=1)}{1-p(y=1)}\right) = \beta_0 + \sum_{i=1}^{n}\beta_i x_i, \quad (1)$$

where $i = 1, 2, \ldots, n$, with n the number of taxa, β_i is the coefficient of the ith taxon and x_i is the relative abundance of the ith taxon.

The dataset was first randomly split into a training/validation set (80%) and a testing set (20%). The hyperparameter in the LASSO model (λ) for each household was tuned via leave-one-out cross-validation. The optimal model was subsequently used to estimate the household origin of the samples in the testing set. The classification accuracy was computed as the percentage of samples whose household origin was correctly estimated.

Regression models on selected taxa

Linear regression models were built to examine the environmental variables (total chlorine level and temperature) as sources of fixed effects on the relative abundances of LASSO-selected bacterial species related to opportunistic premise plumbing pathogens. The 'lm' function in the R stats package was used to build the models[100]. A linear model was used to fit the log transformed relative abundances of each species. A pseudo relative abundance of 0.1% was used to avoid taking logarithms

of zeros. The model formula was as follows: $\log_{10}$(species relative abundance + 0.001) ~ total chlorine level + temperature + intercept.

Functional profiling

Each metagenome was assembled individually with SPAdes v3.14.0, using the metaSPAdes pipeline with default settings (*k*-mer sizes of 21, 33 and 55)[101,102]. After removing contigs with length <500 bp, the remaining contigs were annotated by Prokka 1.12 (ref. 103). All the prediction transcripts from each single home were clustered at the nucleotide level by using CD-HIT 4.8.1, with the following parameters: -G 0, -aS 0.9, -c 0.95, -M 30000, -T 30 and -d 0, where genes sharing >95% identity were considered as redundant[104]. After clustering, an average of 663,628 (s.d. of 226,902) sequences were detected across eight homes. The clustered genes were further annotated with eggNOG 5.0.2 ($e = 0.001$)[105]. KO information was extracted from the eggNOG annotation, resulting in 9,650 KOs in total. To generate a count matrix of the number of reads mapped to each gene, the metagenomic reads from each sample were mapped to the clustered genes using Bowtie2 (-sensitive-local)[106]. The numbers of mapped reads were normalized by the number of bacterial reads reported by Bracken in each sample and the gene length to obtain RPKMs, which represented the relative abundance of each gene. The RPKM matrix was rarefied to 632,529 so that each sample had the same sum of RPKM values. The clustered genes were further annotated using FOAM by searching each KO reported by eggNOG against the FOAM database[107,108]. Finally, for each FOAM level, a count matrix was created. Out of the 9,650 identified KOs, 2,162 were classified at both FOAM level 1 and level 2. Among those, 1,050 were classified at level 3, and 445 were classified at level 4.

Analysis of ARGs

To generate the abundance profiles of ARGs, we followed the AmrPlus-Plus v2.0 pipeline. Quality-filtered metagenomic reads were mapped to the MEGARes 2.0 database using BWA MEM with default parameters (-k 19, -w 100, -d 100, -r 1.5 and -c 10000). The alignments were then analysed by ResistomeAnalyzer to generate profiles of the numbers of aligned reads for each ARG, class, mechanism and group. The numbers of aligned reads were normalized by the number of bacterial reads in each sample and the gene length to obtain RPKMs, which represented the relative abundance of each ARG.

Community assembly mechanism

To quantitatively estimate the relative importance of different ecological processes in the bathtub faucet water microbial community assembly, we applied iCAMP v1.6.5 (Inferring Community Assembly Mechanisms by Phylogenetic-bin-based null model) to the 16S rRNA gene amplicon sequencing data[67]. The estimation was performed as follows: the observed ASVs were divided into 100 phylogenetic bins. For each bin, if the between-sample turnover showed a significantly greater (β net relatedness index of a bin, βNRIbin > 1.96) or smaller (βNRIbin < -1.96) phylogenetic dissimilarity than the null expectation, the turnover was considered to be governed by HeS or HoS, respectively. If the turnover shows a non-significant phylogenetic dissimilarity ($-2 \leq$ βNRIbin ≤ 2) but a significantly greater (Raup-Crick metric, RCbin > 0.95) or smaller (RCbin < -0.95) taxonomic dissimilarity than the null expectation, it was regarded as governed by dispersal limitation or homogenizing dispersal, respectively. Otherwise, ($-2 \leq$ βNRIbin ≤ 2 and $-0.95 \leq$ RCbin ≤ 0.95), the turnover represented the influence of 'drift' (including stochastic drift, diversification, weak selection and dispersal). The relative importance of an assembly process in each turnover was assessed as the relative abundance sum of the bins governed by the process. In this method, dispersal limitation, homogenizing dispersal and 'drift' are deemed stochastic processes. Attention was given to the regional species pool assumption (metagroup setting) in iCAMP to reflect varied water sources and sampling times (Supplementary Table 12). Notably, iCAMP does not rely on environmental variables as inputs.

Multiple regression on distance matrices

To explore the relationships between environmental variables and microbial community compositions, we applied MRM using the MRM function in the R package ecodist v2.1.3. The Bray–Curtis dissimilarity based on the 16S rRNA gene sequencing data was used as the response distance matrix, and the total chlorine concentration, temperature, building age and pedestrian distance between households were used as explanatory distance matrices. The 16S rRNA gene sequencing data were used here to estimate the Bray–Curtis dissimilarities because they provide a better representation of the diversity of the community. Nitrate and TOC were excluded from MRM because of the limited variability in nitrate concentrations and the non-significant correlation between the Bray–Curtis dissimilarities of the microbial communities and the Euclidean distances of the TOC concentrations. The final model was selected via backwards selection, where we started with a full model using all four variables and then deleted the least significant one at a time until all variables remaining in the model were significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw DNA sequences from this study are available on NCBI under Bioproject PRJNA1066374.

Code availability

The custom code used in this study is available on GitHub(https://github.com/linglab-washu/DW_daily_dynamics).

References

1. Lautenschlager, K. et al. A microbiology-based multi-parametric approach towards assessing biological stability in drinking water distribution networks. *Water Res.* **47**, 3015–3025 (2013).
2. Hwang, C., Ling, F., Andersen, G. L., LeChevallier, M. W. & Liu, W.-T. Microbial community dynamics of an urban drinking water distribution system subjected to phases of chloramination and chlorination treatments. *Appl. Environ. Microbiol.* **78**, 7856–7865 (2012).
3. Pinto, A. J., Schroeder, J., Lunn, M., Sloan, W. & Raskin, L. Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiome. *mBio* **5**, e01135–14 (2014).
4. Hull, N. M. et al. Longitudinal and source-to-tap New Orleans, LA, U.S.A. Drinking water microbiology. *Environ. Sci. Technol.* **51**, 4220–4229 (2017).
5. Perrin, Y., Bouchon, D., Delafont, V., Moulin, L. & Héchard, Y. Microbiome of drinking water: a full-scale spatio-temporal study to monitor water quality in the Paris distribution system. *Water Res.* **149**, 375–385 (2019).
6. Dai, Z. et al. Disinfection exhibits systematic impacts on the drinking water microbiome. *Microbiome* **8**, 42 (2020).
7. Zhang, Y., Oh, S. & Liu, W. Impact of drinking water treatment and distribution on the microbiome continuum: an ecological disturbance's perspective. *Environ. Microbiol.* **19**, 3163–3174 (2017).
8. Di Rienzi, S. C. et al. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife* **2**, e01102 (2013).
9. Cremers, G. et al. Draft genome sequence of a novel Methylobacterium brachiatum strain isolated from human skin.*Microbiol. Resour. Announc.* **9**, e01093–20 (2020).
10. Gallego, V., García, M. T. & Ventosa, A. *Methylobacterium adhaesivum* sp. nov., a methylotrophic bacterium isolated from drinking water. *Int. J. Syst. Evol. Microbiol.* **56**, 339–342 (2006).

11. Kip, N. & van Veen, J. A. The dual role of microbes in corrosion. *ISME J.* **9**, 542–551 (2015).
12. Tung, H. & Xie, Y. F. Association between haloacetic acid degradation and heterotrophic bacteria in water distribution systems. *Water Res.* **43**, 971–978 (2009).
13. Berry, D., Xi, C. & Raskin, L. Microbial ecology of drinking water distribution systems. *Curr. Opin. Biotechnol.* **17**, 297–302 (2006).
14. Vacs Renwick, D., Heinrich, A., Weisman, R., Arvanaghi, H. & Rotert, K. Potential public health impacts of deteriorating distribution system infrastructure. *J. Am. Water Works Assoc.* **111**, 42–53 (2019).
15. Hull, N. M. et al. Drinking water microbiome project: is it time? *Trends Microbiol.* **27**, 670–677 (2019).
16. Favere, J. et al. Safeguarding the microbial water quality from source to tap. *npj Clean Water* **4**, 28 (2021).
17. Prest, E. I., Hammes, F., van Loosdrecht, M. C. M. & Vrouwenvelder, J. S. Biological stability of drinking water: controlling factors, methods, and challenges. *Front. Microbiol.* **7**, 45 (2016).
18. *Drinking Water Distribution Systems: Assessing and Reducing Risks* (National Academies Press, 2006); https://doi.org/10.17226/11728
19. Rhoads, W. J., Pruden, A. & Edwards, M. A. Survey of green building water systems reveals elevated water age and water quality concerns. *Environ. Sci. Water Res. Technol.* **2**, 164–173 (2016).
20. Wang, H., Edwards, M., Falkinham, J. O. & Pruden, A. Molecular survey of the occurrence of *Legionella* spp., *Mycobacterium* spp., *Pseudomonas aeruginosa*, and amoeba hosts in two chloraminated drinking water distribution systems. *Appl. Environ. Microbiol.* **78**, 6285–6294 (2012).
21. Lautenschlager, K., Boon, N., Wang, Y., Egli, T. & Hammes, F. Overnight stagnation of drinking water in household taps induces microbial growth and changes in community composition. *Water Res.* **44**, 4868–4877 (2010).
22. Ling, F., Whitaker, R., LeChevallier, M. W. & Liu, W.-T. Drinking water microbiome assembly induced by water stagnation. *ISME J.* **12**, 1520–1531 (2018).
23. Wang, H. et al. Methodological approaches for monitoring opportunistic pathogens in premise plumbing: a review. *Water Res.* **117**, 68–86 (2017).
24. Vosloo, S. et al. Gradual recovery of building plumbing-associated microbial communities after extended periods of altered water demand during the COVID-19 pandemic. *Environ. Sci. Technol.* **57**, 3248–3259 (2023).
25. Yao, M. et al. Building water quality deterioration during water supply restoration after interruption: influences of premise plumbing configuration. *Water Res.* **241**, 120149 (2023).
26. Dowdell, K. S. et al. *Legionella pneumophila* occurrence in reduced-occupancy buildings in 11 cities during the COVID-19 pandemic. *Environ. Sci.-Wat. Res.* **9**, 2847–2865 (2023).
27. Logan-Jackson, A. R. et al. A critical review on the factors that influence opportunistic premise plumbing pathogens: from building entry to fixtures in residences. *Environ. Sci. Technol.* **57**, 6360–6372 (2023).
28. Song, Y., Finkelstein, R., Rhoads, W., Edwards, M. A. & Pruden, A. Shotgun metagenomics reveals impacts of copper and water heater anodes on pathogens and microbiomes in hot water plumbing systems. *Environ. Sci. Technol.* **57**, 13612–13624 (2023).
29. Rhoads, W. J. et al. Residential water heater cleaning and occurrence of *Legionella* in Flint, MI. *Water Res.* **171**, 115439 (2020).
30. Mathys, W., Stanke, J., Harmuth, M. & Junge-Mathys, E. Occurrence of *Legionella* in hot water systems of single-family residences in suburbs of two German cities with special reference to solar and district heating. *Int. J. Hyg. Environ. Health* **211**, 179–185 (2008).
31. Haig, S.-J., Kotlarz, N., LiPuma, J. J. & Raskin, L. A high-throughput approach for identification of nontuberculous mycobacteria in drinking water reveals relationship between water age and *Mycobacterium avium*. *mBio* **9**, e02354-17 (2018).
32. Haig, S.-J. et al. Emerging investigator series: bacterial opportunistic pathogen gene markers in municipal drinking water are associated with distribution system and household plumbing characteristics. *Environ. Sci. Water Res. Technol.* **6**, 3032–3043 (2020).
33. Dowdell, K. et al. Nontuberculous *Mycobacteria* in drinking water systems – the challenges of characterization and risk mitigation. *Curr. Opin. Biotechnol.* **57**, 127–136 (2019).
34. Ma, L. et al. Catalogue of antibiotic resistome and host-tracking in drinking water deciphered by a large scale survey. *Microbiome* **5**, 154 (2017).
35. Ma, L., Li, B. & Zhang, T. New insights into antibiotic resistome in drinking water and management perspectives: a metagenomic based study of small-sized microbes. *Water Res.* **152**, 191–201 (2019).
36. Xu, L. et al. High-throughput profiling of antibiotic resistance genes in drinking water treatment plants and distribution systems. *Environ. Pollut.* **213**, 119–126 (2016).
37. Dias, M. F. et al. Exploring the resistome, virulome and microbiome of drinking water in environmental and clinical settings. *Water Res.* **174**, 115630 (2020).
38. Gulati, P. & Ghosh, M. Biofilm forming ability of *Sphingomonas paucimobilis* isolated from community drinking water systems on plumbing materials used in water distribution. *J. Water Health* **15**, 942–954 (2017).
39. Vega-Dominguez, P. et al. Biofilms of the non-tuberculous *Mycobacterium* chelonae form an extracellular matrix and display distinct expression patterns. *Cell Surf.* **6**, 100043 (2020).
40. Akram, S. M., Rathish, B. & Saleh, D. *Mycobacterium chelonae* infection. in *StatPearls* (StatPearls, 2022).
41. Abudaff, N. N. & Beam, E. *Mycobacterium arupense*: a review article on an emerging potential pathogen in the *Mycobacterium terrae complex*. *J. Clin. Tuberc. Mycobact. Dis.* **10**, 1–5 (2017).
42. Beydoun, N., Wiley, Z. & Rouphael, N. *Mycobacterium mucogenicum* bacteremia in an immunocompetent host: a case report and concise review. *IDCases* **23**, e01032 (2020).
43. Pradier, M. et al. *Mycobacterium mucogenicum* bacteremia: major role of clinical microbiologists. *BMC Infect. Dis.* **18**, 646 (2018).
44. Chang, H.-Y., Tsai, W.-C., Lee, T.-F. & Sheng, W.-H. *Mycobacterium gordonae* infection in immunocompromised and immunocompetent hosts: a series of seven cases and literature review. *J. Formos. Med. Assoc.* **120**, 524–532 (2021).
45. Freyne, B. & Curtis, N. *Mycobacterium gordonae* skin infection in an immunocompetent child. *Pediatr. Infect. Dis. J.* **36**, 523–525 (2017).
46. Mazumder, S. A., Hicks, A. & Norwood, J. *Mycobacterium gordonae* pulmonary infection in an immunocompetent adult. *North Am. J. Med. Sci.* **2**, 205–207 (2010).
47. Senozan, E. A., Adams, D. J., Giamanco, N. M., Warwick, A. B. & Eberly, M. D. A catheter-related bloodstream infection with *Mycobacterium frederiksbergense* in an immunocompromised child. *Pediatr. Infect. Dis. J.* **34**, 445–447 (2015).
48. Torosian, A., Ly, L., Murina, A. & Bitar, C. A case of cutaneous *Mycobacterium llatzerense*. *JAAD Case Rep.* **31**, 53–55 (2022).
49. Hung, Y.-T. et al. Clinical characteristics of patients with *Acinetobacter junii* infection. *J. Microbiol. Immunol. Infect. Wei Mian Yu Gan Ran Za Zhi* **42**, 47–53 (2009).
50. Abo-Zed, A., Yassin, M. & Phan, T. Acinetobacter junii as a rare pathogen of urinary tract infection. *Urol. Case Rep.* **32**, 101209 (2020).

51. Sudan, S. K. et al. *Pseudomonas fluvialis* sp. nov., a novel member of the genus *Pseudomonas* isolated from the river Ganges, India. *Int. J. Syst. Evol. Microbiol.* **68**, 402–408 (2018).
52. Tao, Y., Zhou, Y., He, X., Hu, X. & Li, D. *Pseudomonas chengduensis* sp. nov., isolated from landfill leachate. *Int. J. Syst. Evol. Microbiol.* **64**, 95–100 (2014).
53. E, D. et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res.* **48**, D561–D569 (2020).
54. Xi, C. et al. Prevalence of antibiotic resistance in drinking water treatment and distribution systems. *Appl. Environ. Microbiol.* **75**, 5714–5718 (2009).
55. Hou, A.-M. et al. Chlorine injury enhances antibiotic resistance in *Pseudomonas aeruginosa* through over expression of drug efflux pumps. *Water Res.* **156**, 366–371 (2019).
56. Martinez, J. L. et al. Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol. Rev.* **33**, 430–449 (2009).
57. Blanco, P. et al. Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. *Microorganisms* **4**, 14 (2016).
58. Amsalu, A. et al. Worldwide distribution and environmental origin of the Adelaide imipenemase (AIM-1), a potent carbapenemase in *Pseudomonas aeruginosa*. *Microb. Genomics* **7**, 000715 (2021).
59. Laurent, F. et al. Biochemical-genetic analysis and distribution of FAR-1, a class A β-lactamase from *Nocardia farcinica*. *Antimicrob. Agents Chemother.* **43**, 1644–1650 (1999).
60. Meletis, G. Carbapenem resistance: overview of the problem and future perspective. *Ther. Adv. Infect. Dis.* **3**, 15–21 (2016).
61. Stoczko, M., Frère, J.-M., Rossolini, G. M. & Docquier, J.-D. Postgenomic scan of metallo-β-lactamase homologues in Rhizobacteria: identification and characterization of BJP-1, a subclass B3 ortholog from *Bradyrhizobium japonicum*. *Antimicrob. Agents Chemother.* **50**, 1973–1981 (2006).
62. *Critically Important Antimicrobials for Human Medicine* (World Health Organization, 2019).
63. Dai, D., Rhoads, W. J., Edwards, M. A. & Pruden, A. Shotgun metagenomics reveals taxonomic and functional shifts in hot water microbiome due to temperature setting and stagnation. *Front. Microbiol.* **9**, 2695 (2018).
64. Tian, L. et al. Deciphering functional redundancy in the human microbiome. *Nat. Commun.* **11**, 6217 (2020).
65. Fierer, N. et al. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science* **342**, 621–624 (2013).
66. Galand, P. E., Pereira, O., Hochart, C., Auguet, J. C. & Debroas, D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* **12**, 2470–2478 (2018).
67. Ning, D. et al. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4717 (2020).
68. Sun, C. et al. Seasonal dynamics of the microbial community in two full-scale wastewater treatment plants: diversity, composition, phylogenetic group based assembly and co-occurrence pattern. *Water Res.* **200**, 117295 (2021).
69. Lou, E. G., Fu, Y., Wang, Q., Treangen, T. J. & Stadler, L. B. Sensitivity and consistency of long- and short-read metagenomics and epicPCR for the detection of antibiotic resistance genes and their bacterial hosts in wastewater. *J. Hazard. Mater.* **469**, 133939 (2024).
70. Arango-Argoty, G. A. et al. NanoARG: a web service for detecting and contextualizing antimicrobial resistance genes from nanopore-derived metagenomes. *Microbiome* **7**, 88 (2019).
71. Wu, Z. et al. Nanopore-based long-read metagenomics uncover the resistome intrusion by antibiotic resistant bacteria from treated wastewater in receiving water body. *Water Res.* **226**, 119282 (2022).
72. Liguori, K. et al. Antimicrobial resistance monitoring of water environments: a framework for standardized methods and quality control. *Environ. Sci. Technol.* **56**, 9149–9160 (2022).
73. Proctor, C. et al. Tenets of a holistic approach to drinking water-associated pathogen research, management, and communication. *Water Res.* **211**, 117997 (2022).
74. Hamilton, K. A., Weir, M. H. & Haas, C. N. Dose response models and a quantitative microbial risk assessment framework for the *Mycobacterium avium* complex that account for recent developments in molecular biology, taxonomy, and epidemiology. *Water Res.* **109**, 310–326 (2017).
75. Garner, E., Zhu, N., Strom, L., Edwards, M. & Pruden, A. A human exposome framework for guiding risk management and holistic assessment of recycled water quality. *Environ. Sci. Water Res. Technol.* **2**, 580–598 (2016).
76. Chase, J. M. Stochastic community assembly causes higher biodiversity in more productive environments. *Science* **328**, 1388–1391 (2010).
77. Ning, D., Deng, Y., Tiedje, J. M. & Zhou, J. A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl Acad. Sci. USA* **116**, 16892–16898 (2019).
78. Zhou, J. et al. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc. Natl Acad. Sci. USA* **111**, E836–E845 (2014).
79. Ning, D. et al. Environmental stress mediates groundwater microbial community assembly. *Nat. Microbiol.* **9**, 490–501 (2024).
80. Bautista-de los Santos, Q. M. et al. The impact of sampling, PCR, and sequencing replication on discerning changes in drinking water bacterial community over diurnal time-scales. *Water Res.* **90**, 216–224 (2016).
81. Gabrielli, M., Turolla, A. & Antonelli, M. Bacterial dynamics in drinking water distribution systems and flow cytometry monitoring scheme optimization. *J. Environ. Manage.* **286**, 112151 (2021).
82. Schmidt, T. M., DeLong, E. F. & Pace, N. R. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**, 4371–4378 (1991).
83. Hwang, C., Ling, F., Andersen, G. L., LeChevallier, M. W. & Liu, W.-T. Evaluation of methods for the extraction of DNA from drinking water distribution system biofilms. *Microbes Environ.* **27**, 9–18 (2012).
84. Preheim, S. P. et al. Computational methods for high-throughput comparative analyses of natural microbial communitie. *Methods Enzymol.* **531**, 353–370 (2013).
85. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
86. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
87. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
88. Bokulich, N. A. et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
89. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
90. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).

91. Xu, H. et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS ONE* **7**, e52249 (2012).
92. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
93. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
94. Sun, Z. et al. Challenges in benchmarking metagenomic profilers. *Nat. Methods* **18**, 618–626 (2021).
95. Meehan, C. J., Barco, R. A., Loh, Y.-H. E., Cogneau, S. & Rigouts, L. Reconstituting the genus *Mycobacterium*. *Int. J. Syst. Evol. Microbiol.* **71**, 004922 (2021).
96. Zhang, L., Lin, T.-Y., Liu, W.-T. & Ling, F. Toward characterizing environmental sources of non-tuberculous *Mycobacteria* (NTM) at the species level: a tutorial review of NTM phylogeny and phylogenetic classification. *ACS Environ. Au* **4**, 127–141 (2024).
97. Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G. & Warinner, C. Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research. *mSystems* https://doi.org/10.1128/msystems.00080-18 (2018).
98. Wirbel, J. et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019).
99. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
100. *R: The R project for Statistical Computing* (R Project, 2024).
101. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
102. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
103. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
104. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinforma. Oxf. Engl.* **28**, 3150–3152 (2012).
105. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
106. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
107. Prestat, E. et al. FOAM (Functional Ontology Assignments for Metagenomes): a hidden Markov model (HMM) database with environmental focus. *Nucleic Acids Res.* **42**, e145 (2014).
108. White III, R. A. & Figueroa, J. Functional Ontology Assignments for Metagenomes (FOAM) database. *Open Science Framework* https://osf.io/muan4 (2023).

Acknowledgements

This work was supported by a McKelvey School of Engineering Startup Fund and a Ralph E. Powe Junior Faculty Enhancement Award by the Oak Ridge Associated Universities to F.L. This research was also partially supported by the Division of Chemical, Bioengineering, Environmental and Transport Systems (CBET) of the National Science Foundation under award 2047470 to F.L. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the sponsoring organizations, agencies or the US government. We thank D. Giammar, J. Ballard, K. Andres, T.-Y. Lin, and K. Chibwe for valuable discussions. We thank A. Dang, P. Liu, Y. Liu, P. Prathibha, J. Wei, H. Zhang and Z. Zou for collecting the water samples.

Author contributions

L.Z.: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing. D.N.: formal analysis, methodology, software, writing—review and editing. D.M.-C.: writing—original draft, writing—review and editing. Y.X.: investigation, methodology. B.L.: investigation, methodology. W.C.: investigation, writing—review and editing. J.G.: investigation, writing—review and editing. K.A.H.: writing—review andediting. J.L.: investigation, methodology, writing—review and editing. J.Z.: writing—review and editing. F.L.: conceptualization, funding acquisition, project administration, supervision, writing—original draft, writing—review and editing.

Competing interests

The Washington University Human Research Protection Office reviewed this project and determined that it did not involve activities that are subject to Institutional Review Board oversight. All participants provided informed consent at enrollment. The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s44221-024-00345-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44221-024-00345-z.

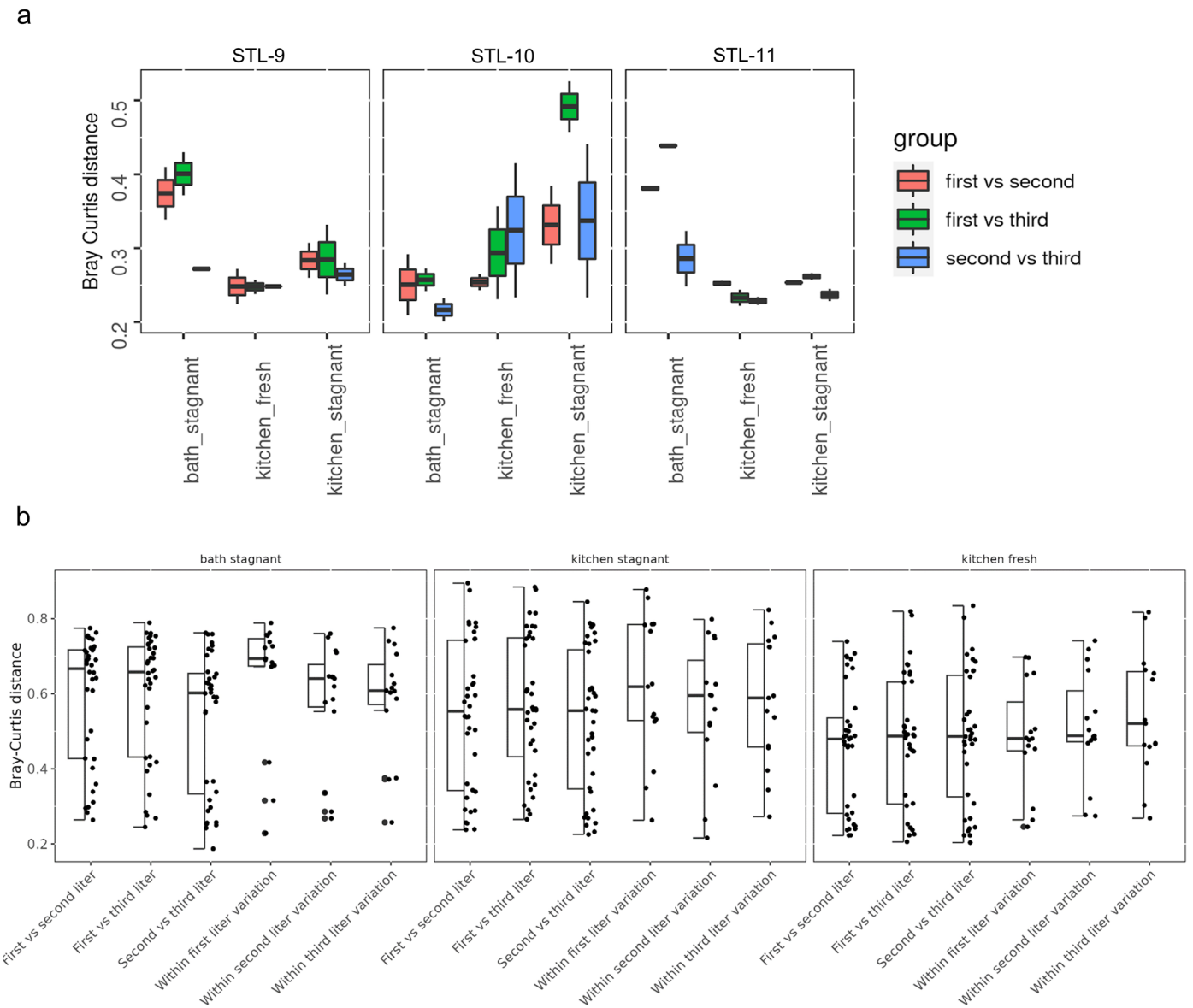
Correspondence and requests for materials should be addressed to Fangqiong Ling.

Peer review information *Nature Water* thanks Xianghua Wen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Lin Zhang[1], Daliang Ning[2,3], David Mantilla-Calderon[1], Yirong Xu[4], Bingdi Liu[1], Winston Chen[1,11], Jinyu Gao[5], Kerry A. Hamilton[6,7], Jinyong Liu[5], Jizhong Zhou[2,3,8,9,10] & Fangqiong Ling[1,4]

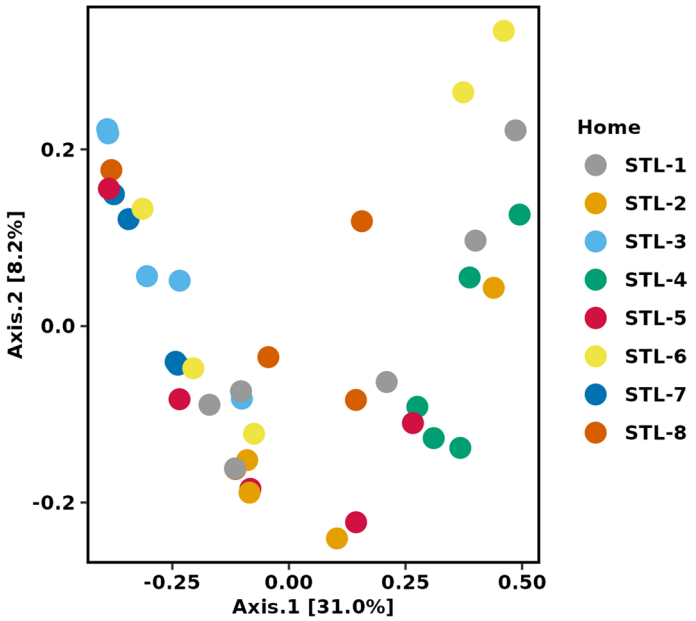
[1]Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, MO, USA. [2]Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA. [3]School of Biological Sciences, University of Oklahoma, Norman, OK, USA. [4]Division of Computational and Data Sciences, Washington University in St. Louis, St. Louis, MO, USA. [5]Department of Chemical and Environmental Engineering, University of California, Riverside, CA, USA. [6]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA. [7]The Biodesign Center for Environmental Health Engineering, Arizona State University, Tempe, AZ, USA. [8]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA. [9]School of Computer Science, University of Oklahoma, Norman, OK, USA. [10]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [11]Present address: Department of Earth and Environmental Science, University of Pennsylvania, Philadelphia, PA, USA. e-mail: fangqiong@wustl.edu

https://doi.org/10.1038/s44221-024-00345-z



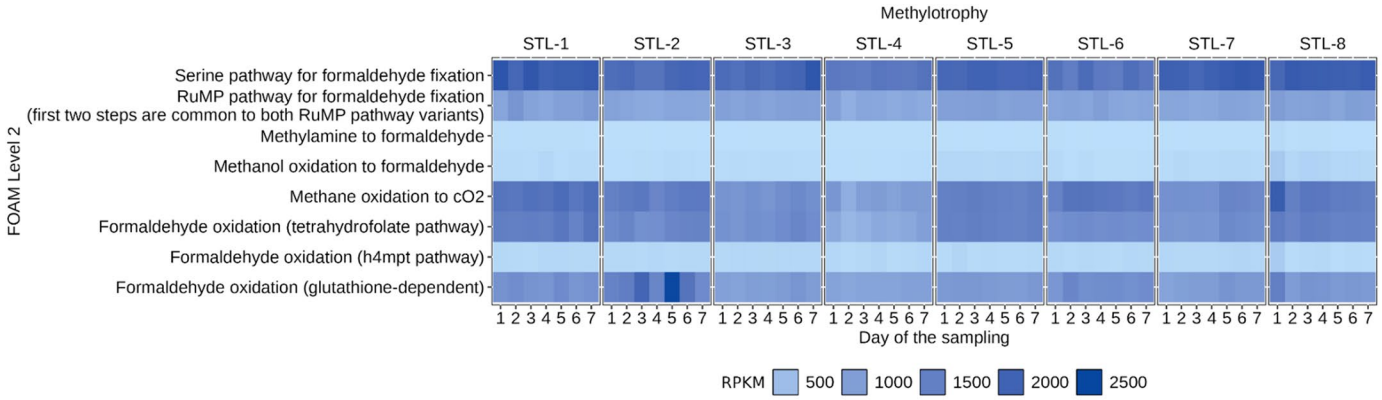
Extended Data Fig. 1 | Bray Curtis dissimilarities relevant for sample volumetric segments comparisons. a, The between-group Bray Curtis dissimilarities between each volumetric segments in each households. Each box represents four pairs of comparisons. **b**, The between and within-group Bray Curtis dissimilarities grouped by sample types. Each box comparing Bray Curtis dissimilarities between different volumetric segments represents 36 pairs of comparisons. Each box comparing Bray Curtis dissimilarities within the same volumetric segment represents 15 pairs of comparisons. In the box plot, the box shows the interquartile range (IQR), which spans from the 25th percentile (Q1) to the 75th percentile (Q3) of the data. The thick line inside the box represents the median of the data. The lower whisker extends from Q1 to the smallest value in the dataset that is greater than or equal to Q1 - 1.5 × IQR (minima); the upper whisker extends from Q3 to the largest value in the dataset that is less than or equal to Q3 + 1.5 × IQR (maxima).

https://doi.org/10.1038/s44221-024-00345-z



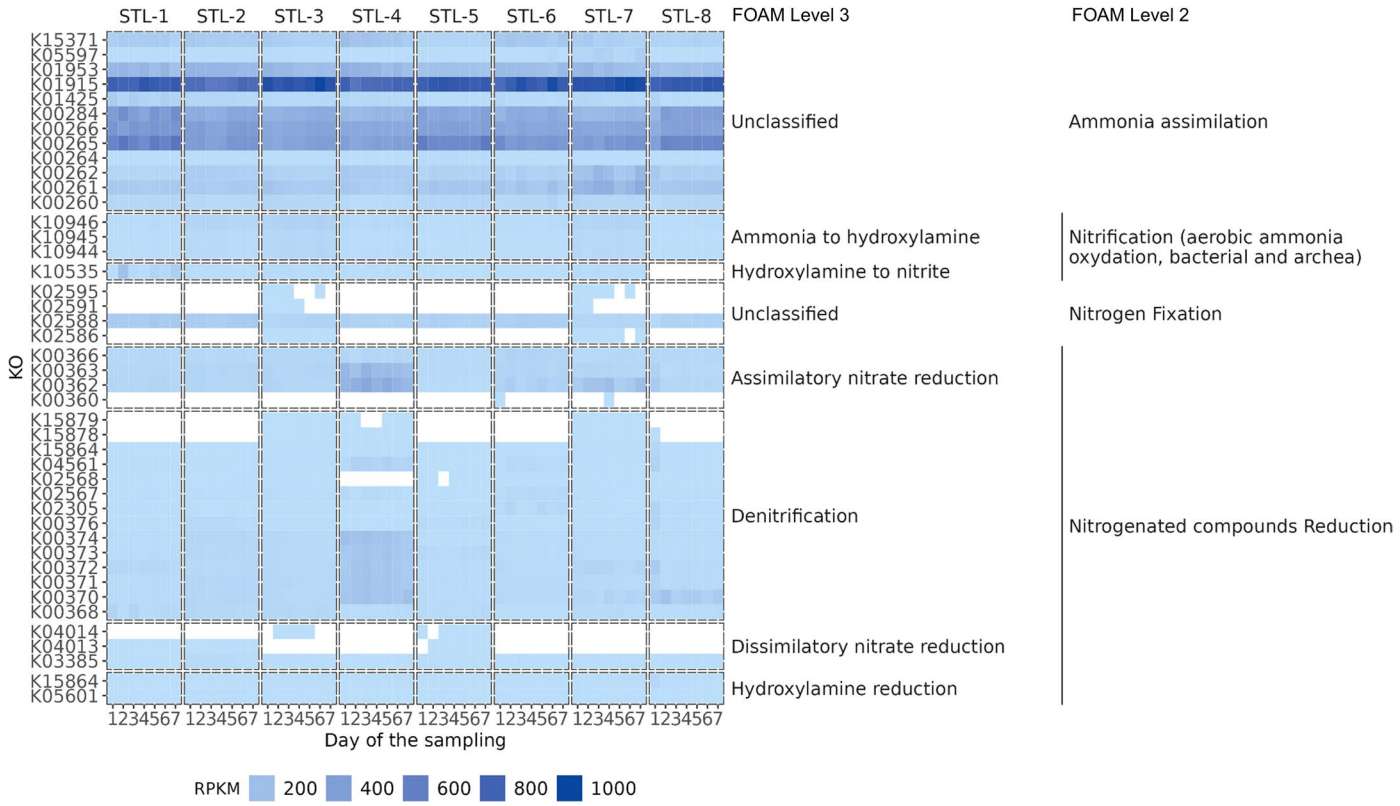
Extended Data Fig. 2 | PCoA plot on ARG profiles. PCoA plot on Hellinger distance between ARG profiles from different homes. Distance was computed on number of reads (each sample rarefied to 18,783 reads, which got 17 samples excluded).

https://doi.org/10.1038/s44221-024-00345-z



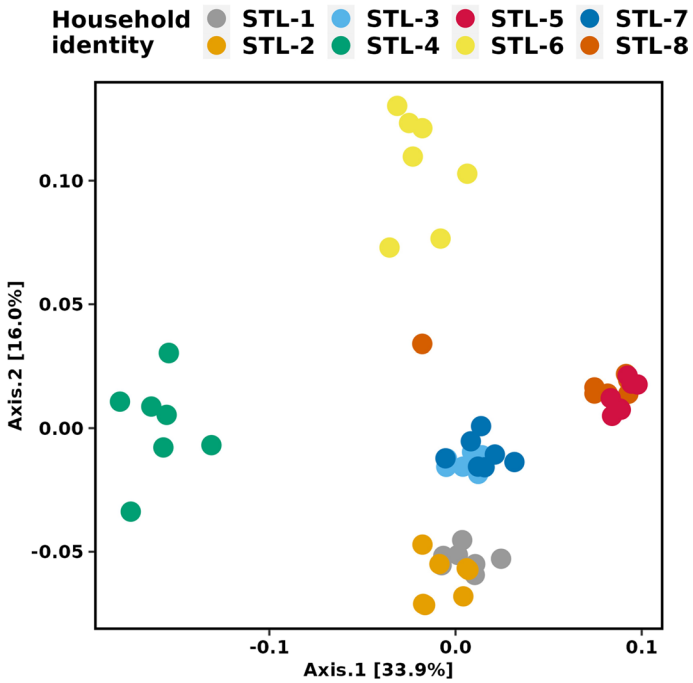
Extended Data Fig. 3 | RPKMs of FOAM level 2 functions in methylotrophy. A heatmap showing RPKMs of FOAM level 2 functions in methylotrophy.

https://doi.org/10.1038/s44221-024-00345-z



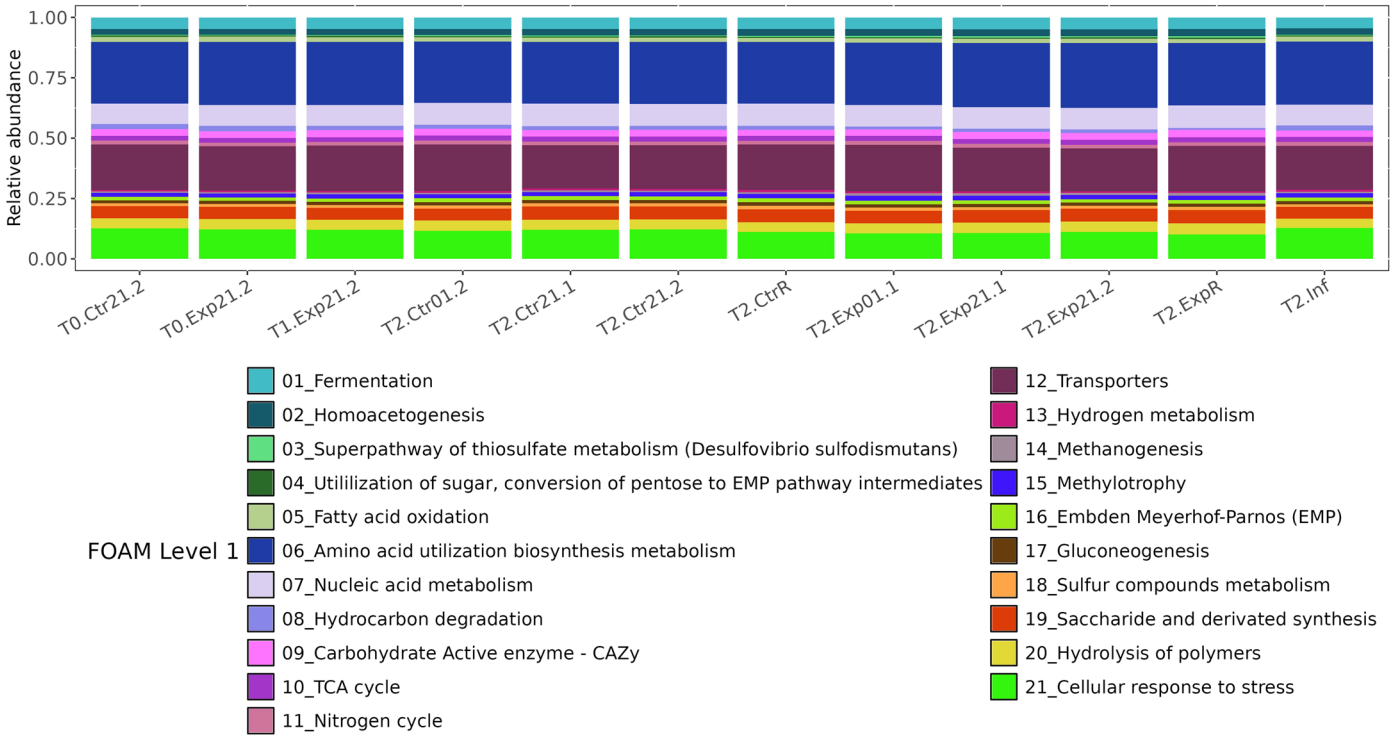
Extended Data Fig. 4 | RPKMs of KOs involved in the nitrogen cycle. A heatmap showing RPKMs of KOs involved in the nitrogen cycle.

https://doi.org/10.1038/s44221-024-00345-z



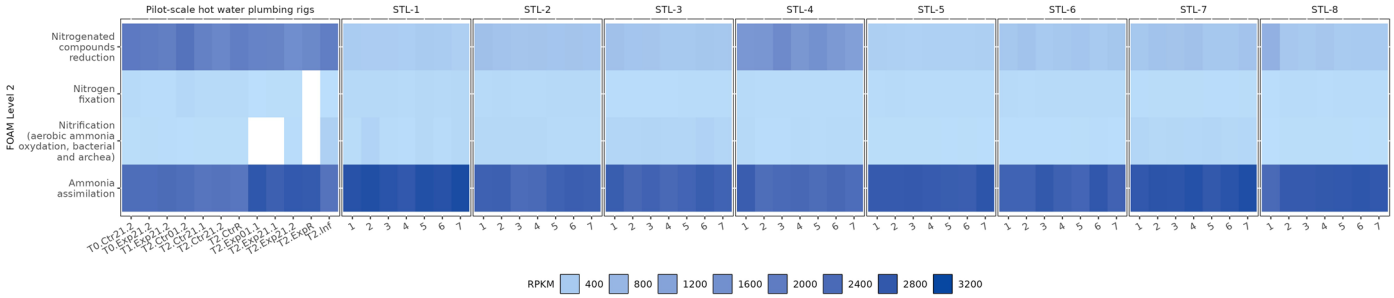
Extended Data Fig. 5 | PCoA plot of Bray Curtis distances computed on KO profiles. KO profiles clustered by households (PERMANOVA $p = 0.001$ $R^2 = 0.81$). P-value was computed based on two-sided tests.

https://doi.org/10.1038/s44221-024-00345-z



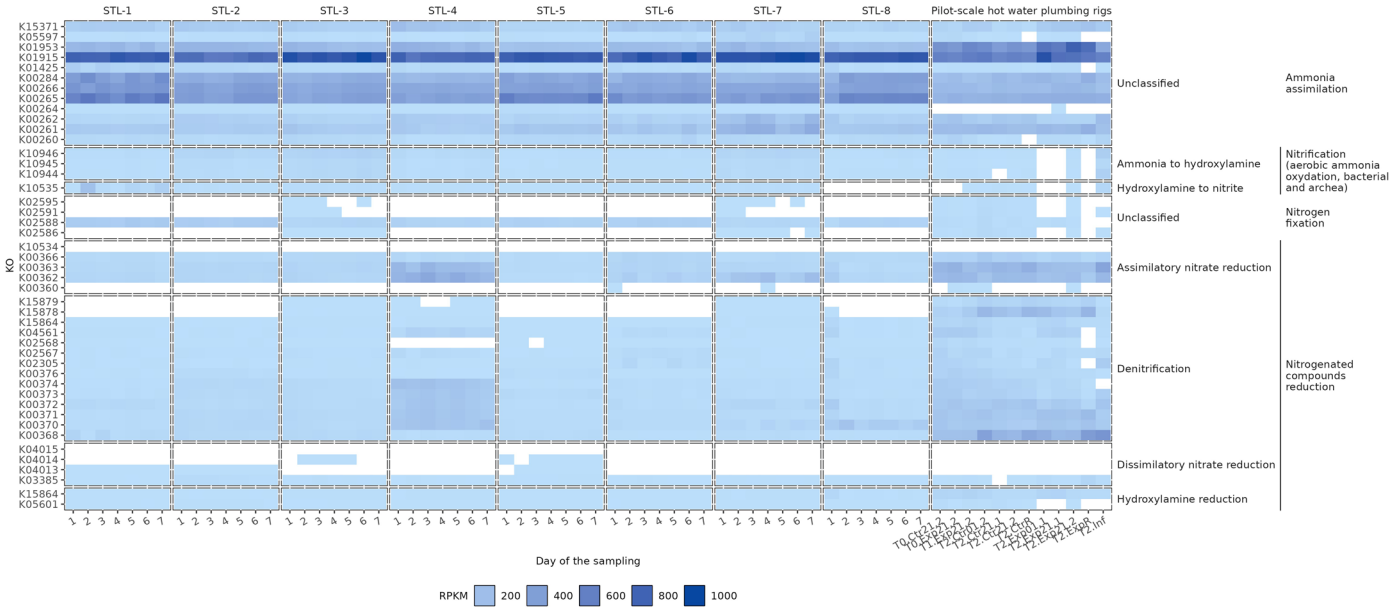
Extended Data Fig. 6 | Relative abundances of FOAM level 1 functions of water samples from a pilot-scale hot water plumbing rig study. A stacked bar chart showing relative abundances of environmentally relevant functional groups at FOAM level 1 of water samples from a pilot-scale hot water plumbing rig study (Dai et al., 2018). The sample labels were kept consistent with the original publication.

https://doi.org/10.1038/s44221-024-00345-z



Extended Data Fig. 7 | RPKMs of level 2 functions in the nitrogen cycle of St. Louis premise plumbing water in comparison to a pilot-scale hot water plumbing rig study. A heatmap showing RPKMs of level 2 functions in the nitrogen cycle of St. Louis premise plumbing water from this study (STL-1 through 8) in comparison to a previously published pilot-scale hot water plumbing rig study in Blacksburg, VA (those samples of which the sample names starting with "T", Dai et al., 2018).

https://doi.org/10.1038/s44221-024-00345-z



Extended Data Fig. 8 | RPKMs of KOs involved in the nitrogen cycle of St. Louis premise plumbing water in comparison to a pilot-scale hot water plumbing rig study. A heatmap showing RPKMs of KOs involved in the nitrogen cycle in bathtub faucet water samples from this study (STL-1 through 8) in comparison to those from a published pilot-scale hot water plumbing rig study in Blacksburg, VA (those samples of which the sample names starting with "T", Dai et al., 2018).

https://doi.org/10.1038/s44221-024-00345-z

Extended Data Table 1 | PERMANOVA p values between different households, sample types, and volumetric segments

Variables	p values	R^2	Pairwise comparison	R^2	$p_{unadjusted}$	$p_{Benjamini-Hochberg}$	p_{Holm}
Household N_{STL-9}=18 N_{STL-10}=18 N_{STL-11}=18	0.001	0.52	STL 9 vs 10	0.54	0.001	0.0018	0.009
			STL 9 vs 11	0.38	0.001	0.0018	0.009
			STL 10 vs 11	0.39	0.001	0.0018	0.009
Sample type N_{SB}=18 N_{SK}=18 N_{FK}=18	0.001	0.11	SB vs SK	0.09	0.001	0.0018	0.009
			SB vs FK	0.13	0.001	0.0018	0.009
			SK vs FK	0.03	0.163	1	1
Volumetric segment $N_{1st-liter}$=18 $N_{2nd-liter}$=18 $N_{3rd-liter}$=18	0.291	0.02	1st vs 2nd liter	0.01	0.925	1	1
			1st vs 3rd liter	0.02	0.764	1	1
			2nd vs 3rd liter	0.01	0.974	1	1

Note: SB represents overnight-stagnant water from bathtub hot faucets; SK represents overnight-stagnant water from kitchen cold faucets; FK represents fresh water from kitchen cold faucets.

PERMANOVA p values between different households, sample types, and volumetric segments after Benjamini-Hochberg and Holm corrections. P-values were computed based on two-sided tests.

https://doi.org/10.1038/s44221-024-00345-z

Extended Data Table 2 | Coefficients and p values of MRM models

Variables	Coefficients	p values
Full model: $R^2 = 0.238$		
Total chlorine concentration	0.051	0.001
Pedestrian distance	0.003	0.079
Temperature	0.002	0.142
Building age	0.0003	0.275
Final model: $R^2 = 0.232$		
Total chlorine concentration	0.052	0.001
Pedestrian distance	0.005	0.008

Coefficients and p values of MRM models built on distance matrices of environmental variables against the Bray Curtis dissimilarity of ASV profiles. P-values were computed based on two-sided tests.

nature portfolio

Corresponding author(s): Fangqiong Ling

Last updated by author(s): Oct 17, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom code used in this study will be made available on GitHub upon publication of the manuscript.
Data analysis	Custom code used in this study will be made available on GitHub upon publication of the manuscript.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Raw DNA sequences from this study will be made available on NCBI under Bioproject PRJNA1066374 upon publication of the manuscript.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This item does not apply to this study because this study does not involve research on human populations.
Reporting on race, ethnicity, or other socially relevant groupings	This item does not apply to this study because this study does not involve research on human populations.
Population characteristics	This item does not apply to this study because this study does not involve research on human populations.
Recruitment	Participants in this study were volunteers from the Department of Energy, Environmental and Chemical Engineering at Washington University in St Louis. They were recruited to collect water samples from their residences.
Ethics oversight	The Washington University Human Research Protection Office reviewed this project and determined that it did not involve activities that are subject to Institutional Review Board (IRB) oversight.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This study examined the microbiome and resistome of drinking water samples collected by volunteers from their residences.
Research sample	The samples in this study were drinking water collected from bathtub faucets and kitchen sink faucets. We were interested in the premise plumbing microbiome because it represents the primary setting where drinking water quality is experienced by customers. To represent the premise plumbing environment, samples were collected after overnight stagnation.
Sampling strategy	This study examined drinking water microbiomes in different households on the city scale and no such data were available in literature. We made our best effort to recruit volunteers to collect samples from their residences and the sample size was mostly restrained by logistic challenges.
Data collection	Sample temperatures were recorded by volunteers on site using the sampling kits we provided. Other water chemistry data were measured in the laboratory. Sequencing libraries were prepared in Ling Lab and sequencing data were provided by the Genome Technology Access Center at the McDonnell Genome Institute and the Edison Family Center for Genome Sciences & Systems Biology at Washington University in St. Louis.
Timing and spatial scale	The first experiment examining daily water microbiome dynamics was conducted in October 2018 and October 2019. Each volunteer collected bathtub faucet water samples from their homes on seven consecutive days. The second experiment examining within-household water microbiome variations was conducted in June 2021. Each volunteer collected water samples from bathtub faucet and kitchen sink faucet in their homes on two consecutive days.
Data exclusions	No data were excluded from the analyses.
Reproducibility	Bathtub faucet water samples were collected on seven days and data were reproducible as revealed by high Pearson correlations between microbial community compositions collected on different days.
Randomization	This item does not apply to this study because this study is not a randomized controlled trial and does not involve research on human populations.
Blinding	This item does not apply to this study because this study is not a randomized controlled trial and does not involve research on human populations.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	This item does not apply to this study because the sampling was conducted in built environments (households).
Location	Samples were collected from 11 households in St Louis, Missouri, USA.
Access & import/export	The Washington University Human Research Protection Office reviewed this project and determined that it did not involve activities that are subject to Institutional Review Board (IRB) oversight. The samples were transported to the laboratory by volunteers within two hours of collection.
Disturbance	The volunteers were aware of the time demands (e.g., time cost to collect and transport water samples) before implementing this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	This study does not involve plant material.
Novel plant genotypes	This study does not involve plant material.
Authentication	This study does not involve plant material.