nature water

Supplementary information

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

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

In the format provided by the authors and unedited

SUPPLEMENTARY TEXTS

Supplementary Method 1. Identification of potential hosts of antimicrobial resistance genes (ARGs).

Each sample was assembled individually via SPAdes v3.14.0 using the metaSPAdes pipeline with default settings (k-mer sizes of 21, 33, and 55)[1]. After removing contigs with lengths shorter than 1.5 kb, the assemblies were binned into draft genomes via MetaBAT v2.12.1[2] and Maxbin v2.2.7[3]. DAS Tool v1.1.1 was then used to integrate MAGs produced by these two tools with the option '-score_threshold 0'[4]. The quality of the MAGs was inspected using CheckM v1.0.18[5]. MAGPurify was used to identify and remove contamination from MAGs on the basis of the taxonomic annotation of contigs, outlier nucleotide composition, and outlier sequencing read depth[6]. These processes generated 3,296 MAGs in total across all samples. Among those, 308 MAGs met the MIMAG high-quality criterion of ≥90% completeness with ≤5% contamination[7]. Taxonomy assignment for the MAGs was performed using GTDB-Tk based on the Genome Taxonomy Database (GTDB)[8]. The ARGs carried by those high-quality MAGs were identified using BLASTN against MEGARes v3.0 with an e-value cutoff of 10^{-10}, 80% similarity, and 70% gene coverage.

Supplementary Method 2. Details on the preprocessing of 16S rRNA gene amplicon sequencing data.

Sequences assigned to eukaryotes, mitochondria, or chloroplasts were removed. Sequences with fewer than three reads across all samples were also excluded. For quality control, the blank control samples were also sequenced. The number of reads in the blank samples after quality filtering was between 6 and 1,874 sequences, with more than 96.3% of them assigned to a single *Lysobacter* ASV. This particular *Lysobacter* ASV, which accounted for 0.72% of the total sequence reads, was removed from our samples. Technical replicates of the same water sample were examined via Pearson correlation. Upon observing the high reproducibility of the technical replicates in every sample (Pearson rho >0.95, Supplementary Fig. 24), we merged the technical replicates of the same samples to achieve deeper sequencing depths.

Supplementary Result. Microbial hosts of ARGs.

Among the 308 high-quality MAGs, a total of 21 MAGs were identified as carrying ARGs. These 21 MAGs belong to the genera *Acinetobacter*, *Aquabacterium*, *Bosea*, *Hylemonella*, *Microbacterium*, *Mycobacterium*, *Pseudomonas*, *Pseudoxanthomonas*, and *Serpentinomonas*. The 21 MAGs carried mainly metal resistance and multi-compound resistance. We detected a betalactamase AIM, which was carried by *Pseudoxanthomonas mexicana*. AIM can hydrolyze a broad spectrum of betalactams, including carbapenems, which are reserved for the treatment of infections caused by multidrug-resistant and ESBL-producing bacteria. This gene has been found to be carried by clinical isolates of *Pseudomonas aeruginosa*, making this opportunistic pathogen resistant to carbapenem, which is the leading cause of mortality in critically ill and immunocompromised populations. More importantly, a recent study reported that the potential origin of the bla_{AIM} gene in *P. aeruginosa* might be the environmental organism, *Pseudoxanthomonas mexicana*[9]. The detection of bla_{AIM} carried by *P. mexicana* in drinking water is alarming, revealing the potential of this gene being horizontally transferred to *P. aeruginosa* under particular selection pressures, e.g., a patient infected with *P. aeruginosa* on carbapenem medication.

Broad metal resistance was detected, motivating future studies examining the interactions between corrosion and the selection of antimicrobial resistance. Previously, Kimbell and colleagues conducted laboratory experiments, showing that the addition of zinc orthophosphate can lead to an increased abundance of antimicrobial-resistant bacteria resistant to ciprofloxacin, sulfonamides, trimethoprim, and vancomycin, as well as the genes *sul1* and *qac*EΔ1, indicating resistance to quaternary ammonium compounds[10]. The detection of multiple efflux pumps and metal resistance genes revealed that the potential mechanisms may be more complex in the real drinking water environment.

Supplementary Discussion. Community assembly mechanisms.

We sought to use three examples to show that each of the following processes, heterogeneous selection (HeS), homogenous selection (HoS), and drift, can lead to a 'location-specific' pattern independently.

Example 1: Consider species A and B which have optimum growth temperatures of 20°C and 40°C, respectively. The temperatures at locations X and Y are right at 20°C and 40°C, respectively. Species A then predominates at X, and B predominates at Y. This is a location-specific pattern driven by HeS.

Example 2: Consider three species A, B, and C, which belong to a closely related phylogenetic clade. Although they evolved slightly different traits, their optimum growth temperature remains the same, i.e., 40 degrees Celsius. In contrast, another species, D, has an optimum temperature of 20 degrees. Assume a total of 100 individuals consisting of species A, B, C, and D migrating from the same source community to several final locations, namely locations X, Y, and Z. Among the 100 traveling individuals, there are 97 individuals of species D and only 1 individual of each of species A, B, and C. We consider a scenario where the environmental temperatures at locations X, Y, and Z are all 40°C, and each location can hold only 10 individuals.

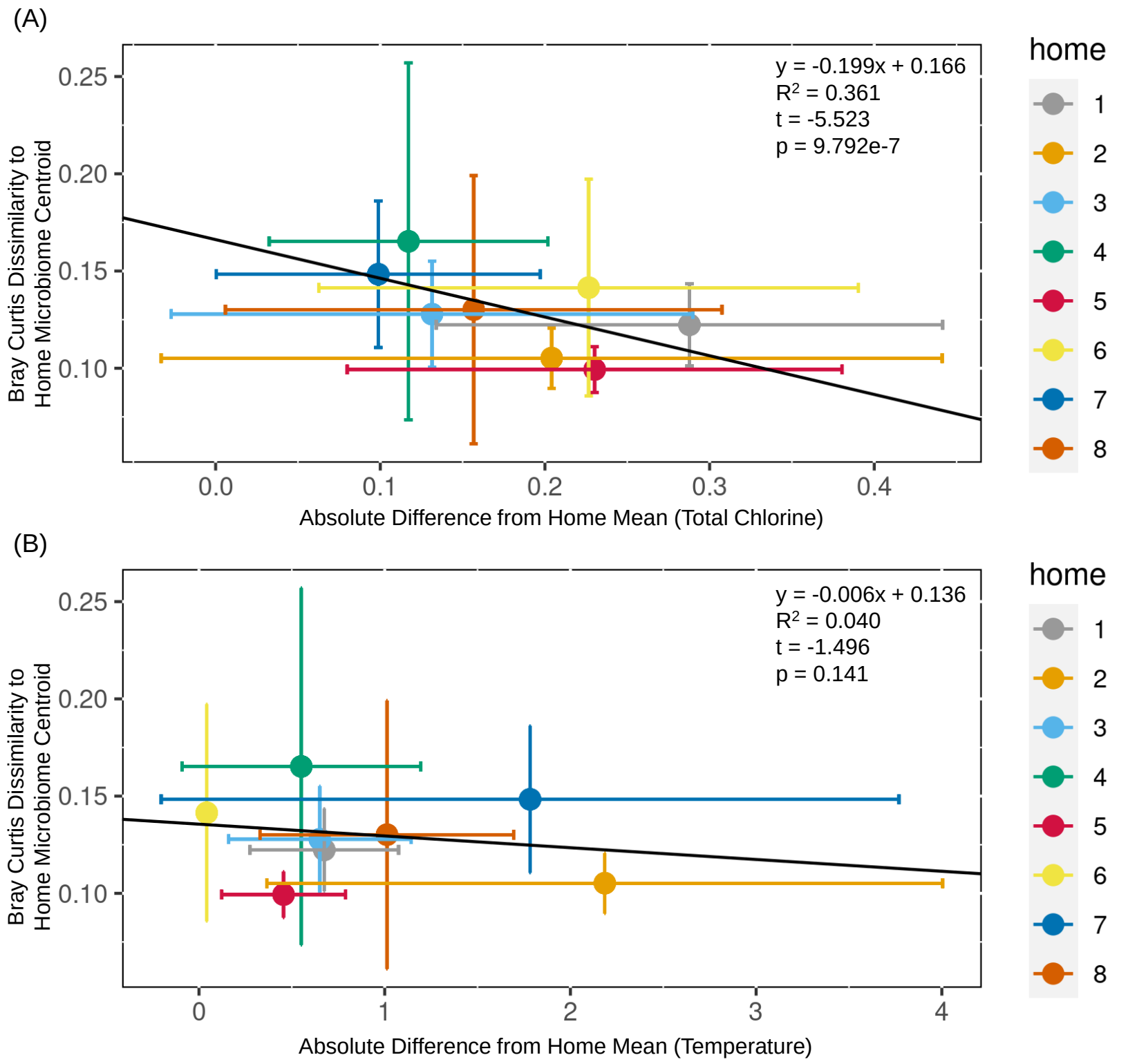
Consider the following intermediate state caused by migration only: location X received 9 individuals of species D and 1 individual of species A, location Y received 9 individuals of species D and 1 individual of species B, and location Z received 9 individuals of species D and 1 individual of species C. Due to temperature selection (where the species with the optimum growth temperature at 40°C are favored at all locations), the following scenario is possible: species A predominates in location X, species B predominates in location Y, and species C predominates in location Z. In all locations, species D can become a rare species. This is a location-specific pattern driven by HoS. Recall that species A, B, and C are closely related; thus, homogeneous selection that drives location specificity may result in a community composition pattern where many homes share common genera, but each home is dominated by different species of those genera. We speculate that the results in our LASSO analysis, where the indicator taxa in various households were mostly different species yet similar genera (Fig. 3b), could be explained by homogenous selection. Similarly, opportunistic pathogen visualization also revealed the presence of various species yet similar genera across households (Fig. 5).

Example 3: Species A and B have very similar niche preferences, and locations X and Y have the same environmental conditions. A and B have no obvious interactions. The population sizes of A and B in the regional pool (or the source community) are the same. Each location can hold only 20 individuals. At the start time point, each location hosts 10 individuals of species A and 10 individuals of species B. At each

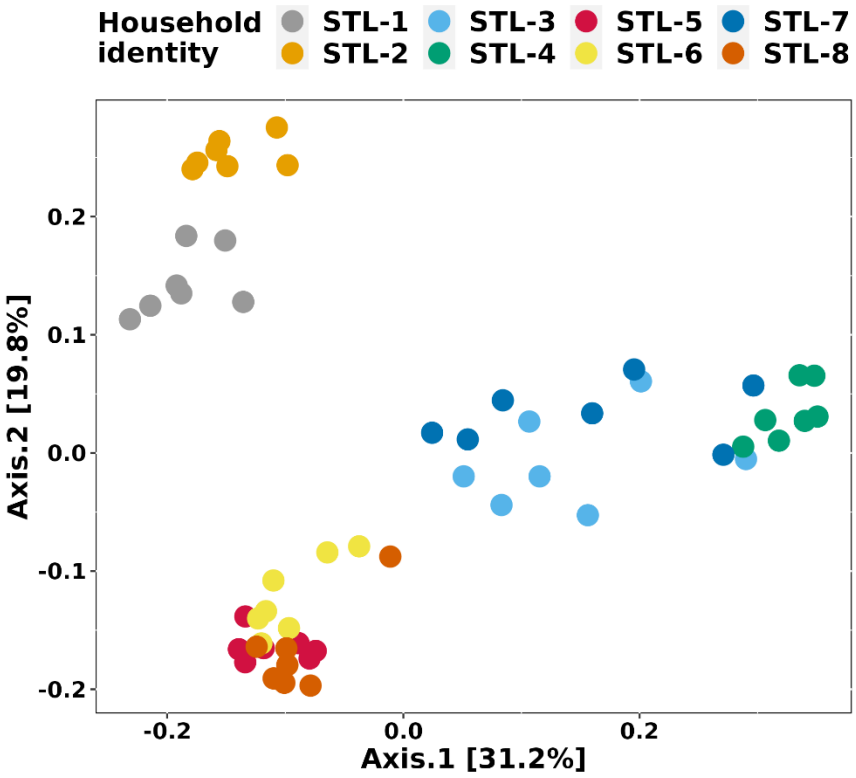
time step, one individual dies, which can randomly be species A or B; the niche is replaced by a newly born individual, which can randomly be species A or B. After 200 time steps, one location can be predominated by species A, and the other can be predominated by species B. Supplementary Fig. 25 shows the simulation results of each time step. This is a location-specific pattern driven by drift.

References

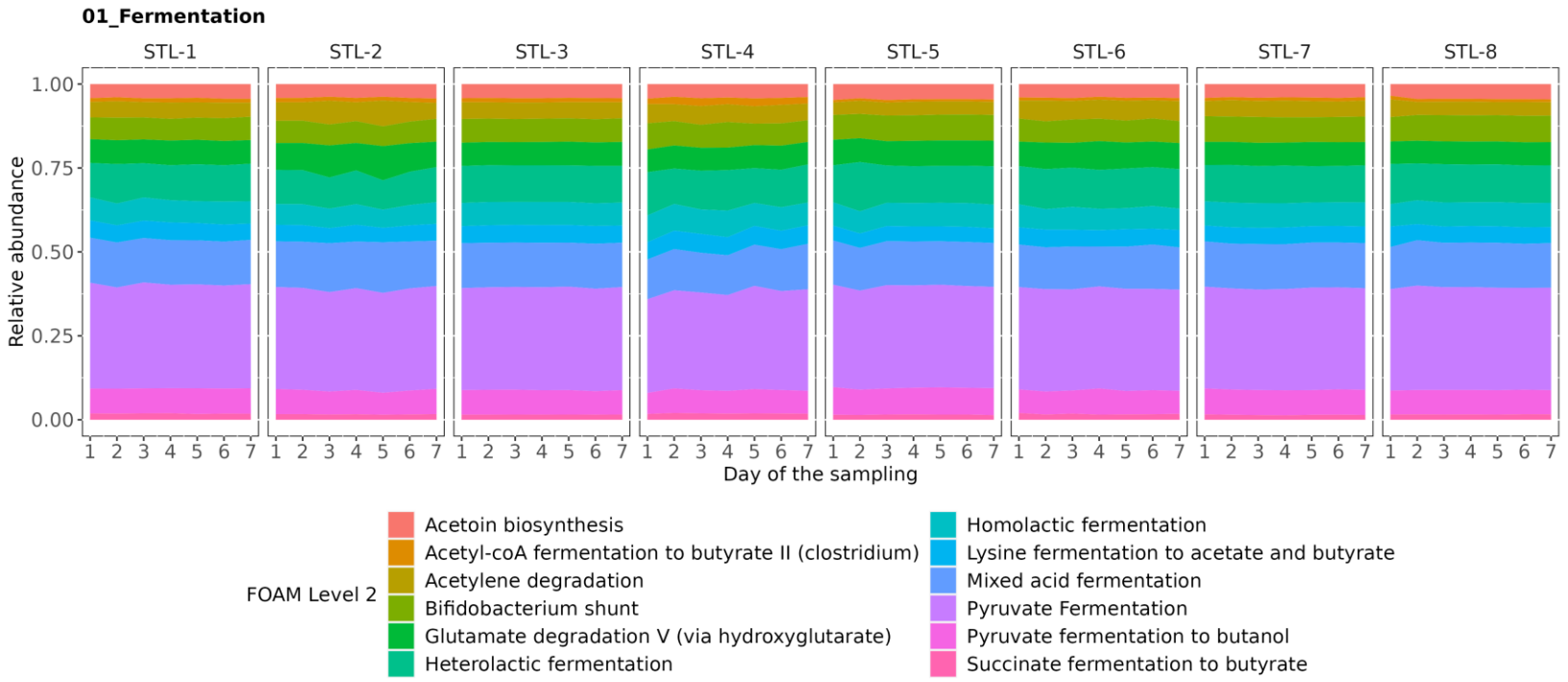
1. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
2. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
3. Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 26 (2014).
4. Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation, and scoring strategy. *Nat Microbiol* **3**, 836–843 (2018).
5. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
6. Nayfach, S., Shi, Z. J., Seshadri, R., Pollard, K. S. & Kyrpides, N. C. New insights from uncultivated genomes of the global human gut microbiome. *Nature* **568**, 505–510 (2019).
7. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**, 725–731 (2017).
8. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* btz848 (2019) doi:10.1093/bioinformatics/btz848.
9. Amsalu, A. *et al.* Worldwide distribution and environmental origin of the Adelaide imipenemase (AIM-1), a potent carbapenemase in Pseudomonas aeruginosa. *Microbial Genomics* **7**, (2021).
10. Kimbell, L. K. *et al.* Impact of corrosion inhibitors on antibiotic resistance, metal resistance, and microbial communities in drinking water. *mSphere* **8**, e00307-23 (2023).



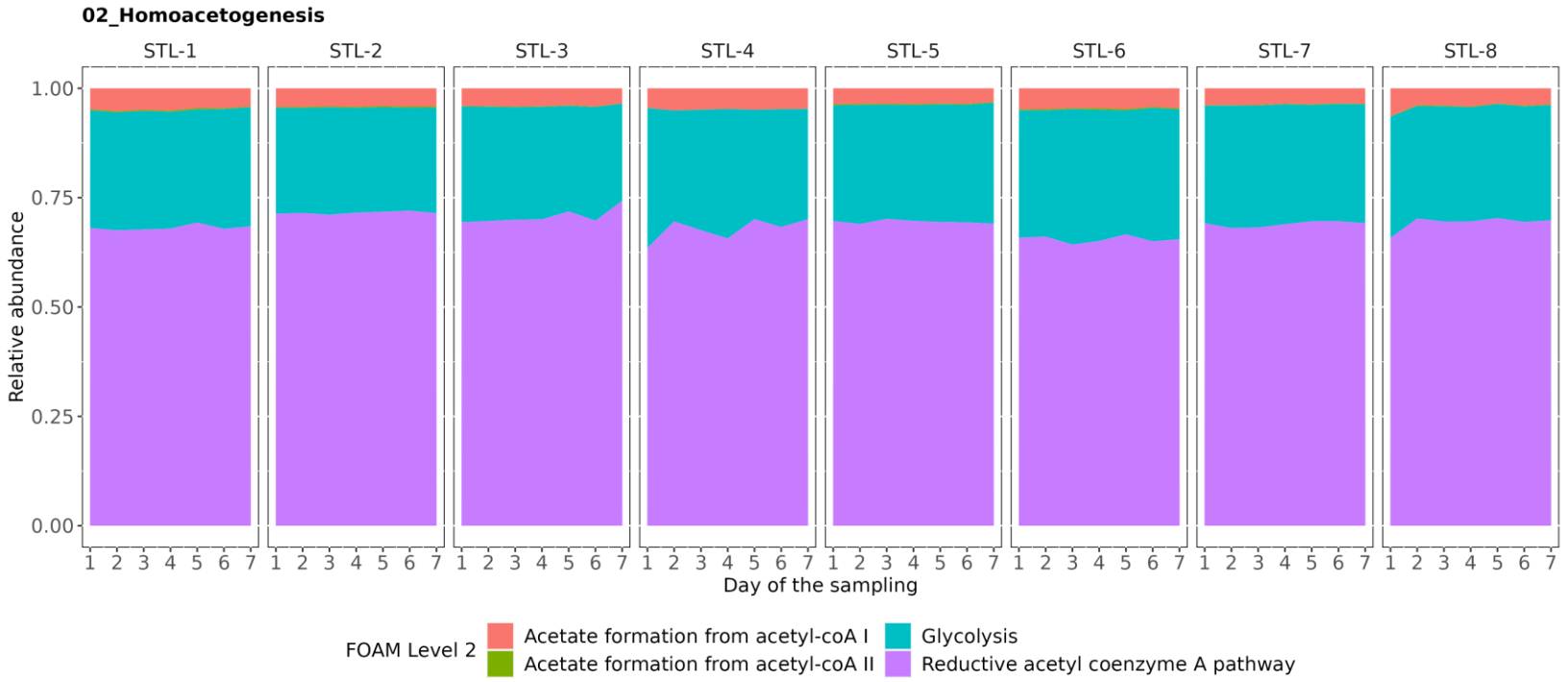
Supplementary Fig. 1: Relationships of household-level bathtub faucet water microbiome variability with environmental variables. Each dot represents data from an individual home, with colors distinguishing different homes. The X-axis shows the average absolute deviation of daily measurements for total chlorine (Panel A) or temperature (Panel B) from the home-specific mean. For both temperature and total chlorine, each home has 21 measurements (the first, second, and third liter of bathtub faucet water were measured separately on seven days). Error bars on the X-axis indicate the standard deviations of these mean deviations. The Y-axis represents the average Bray-Curtis dissimilarity (B-C dissimilarity) between daily microbiome observations (ASV profiles) and the household centroid. Error bars on the Y-axis denote the standard deviations of the mean dissimilarities. The within-home variability in water microbiome associated significantly with the total chlorine levels but not with the temperature. P-values for the coefficient were computed based on two-sided tests.



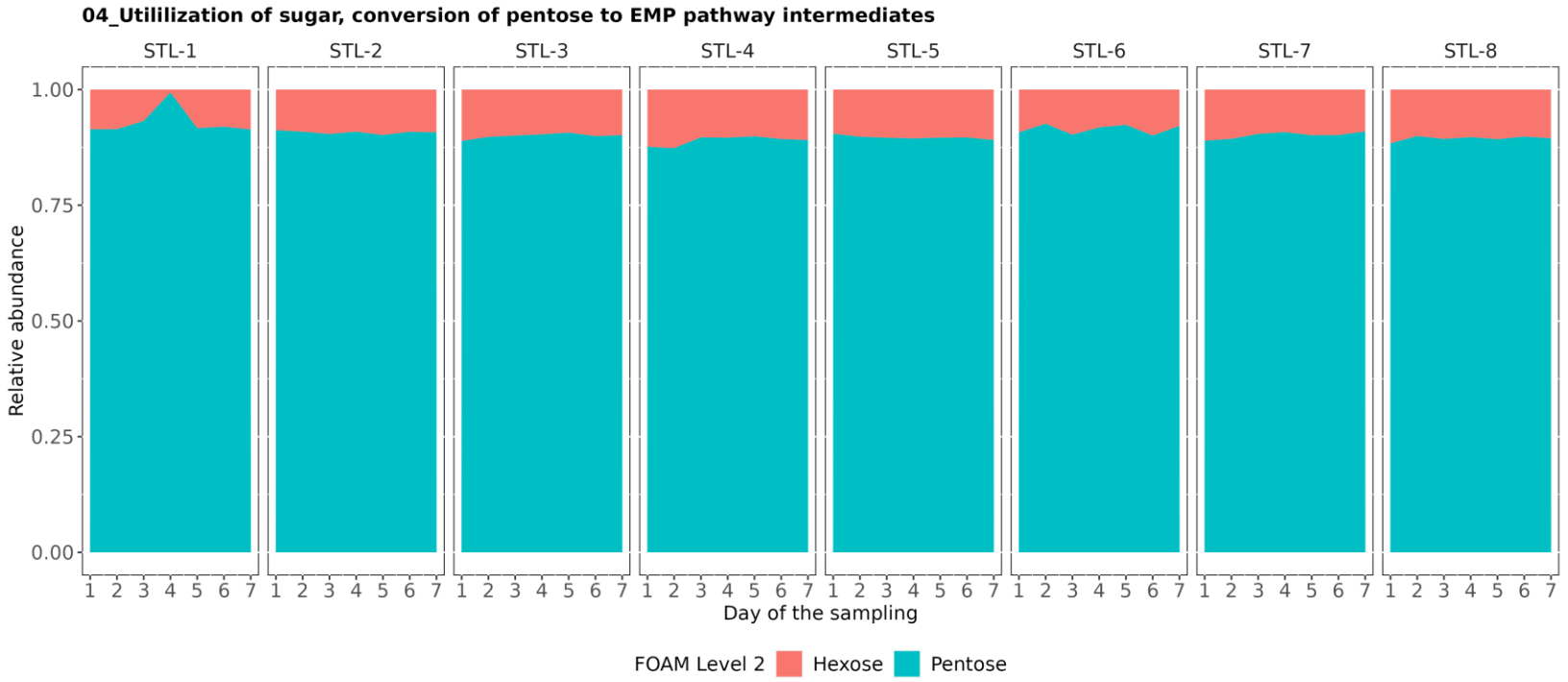
Supplementary Fig. 2: PCoA plot of Jaccard distances computed on species profile from MetaPhlAn4. Homes formed distinctive groups (PERMANOVA p=0.001 R^2=0.73). P-value was computed based on two-sided tests.



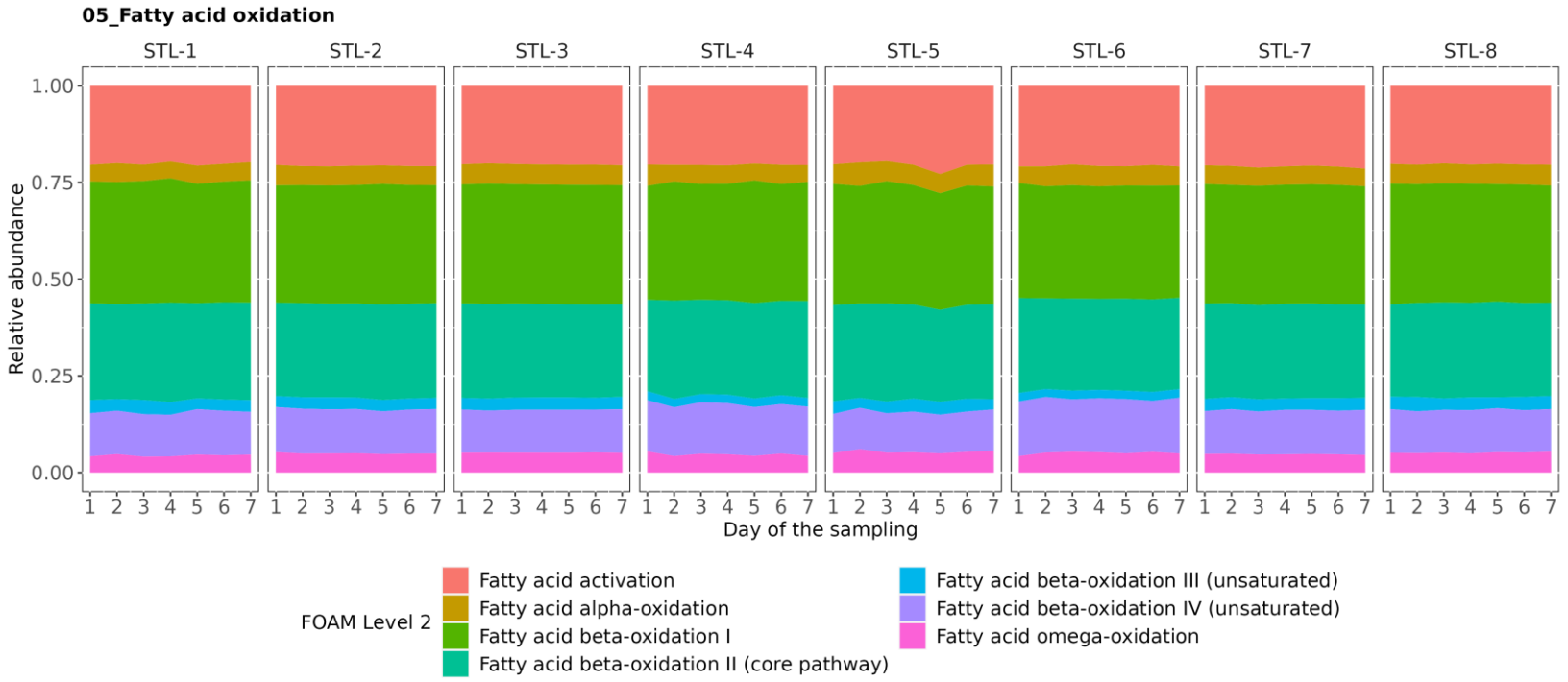
Supplementary Fig. 3: An area chart showing relative abundances of FOAM level 2 functions in fermentation. Relative abundances were computed on RPKMs.



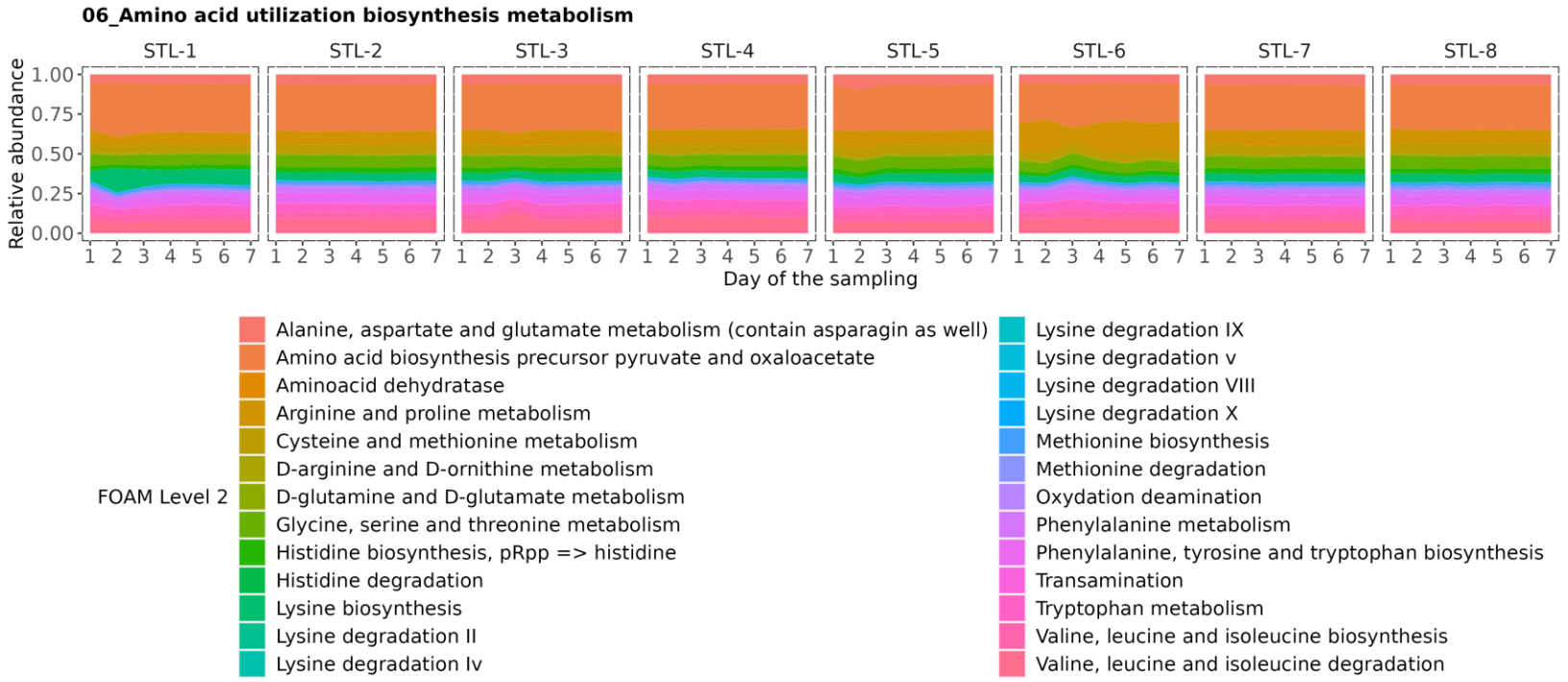
Supplementary Fig. 4: An area chart showing relative abundances of FOAM level 2 functions in homoacetogenesis. Relative abundances were computed on RPKMs.



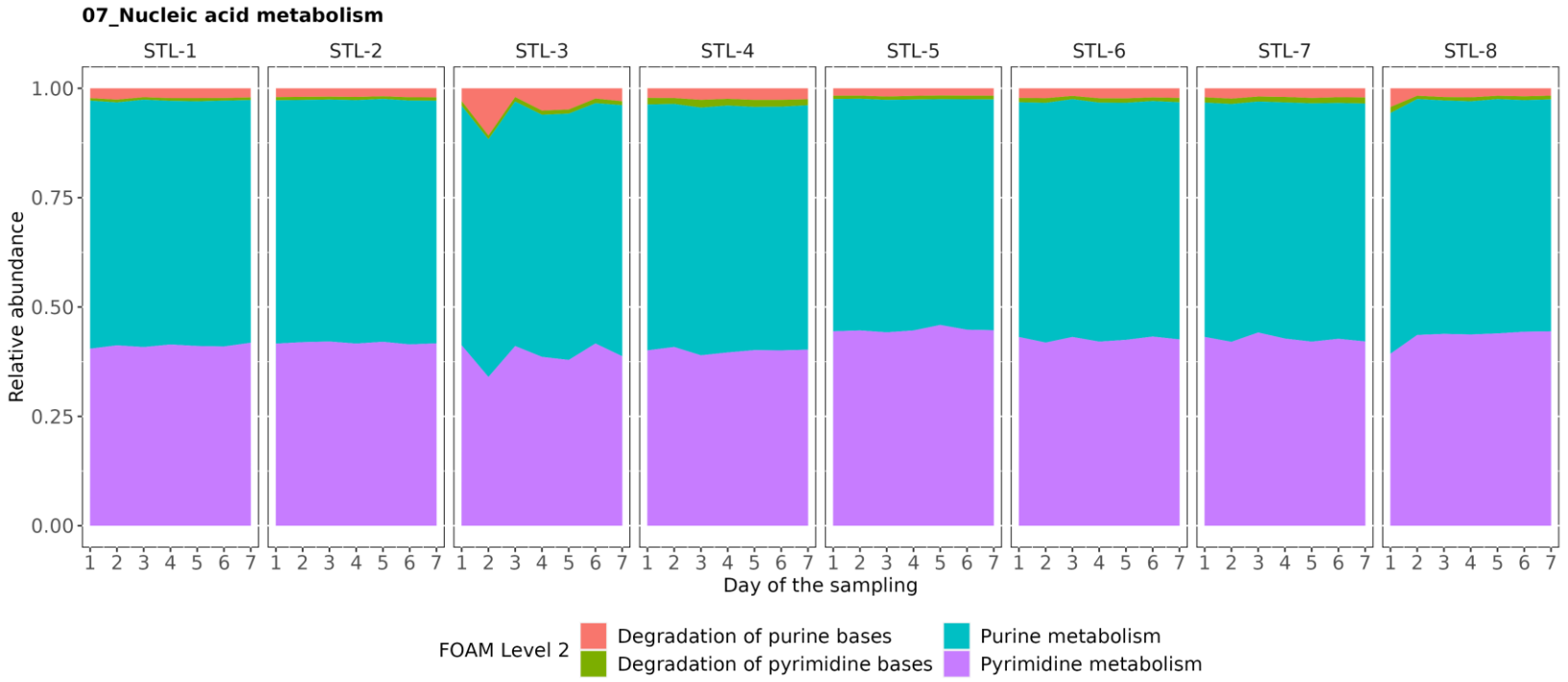
Supplementary Fig. 5: An area chart showing relative abundances of FOAM level 2 functions in untilization of sugar. Relative abundances were computed on RPKMs.



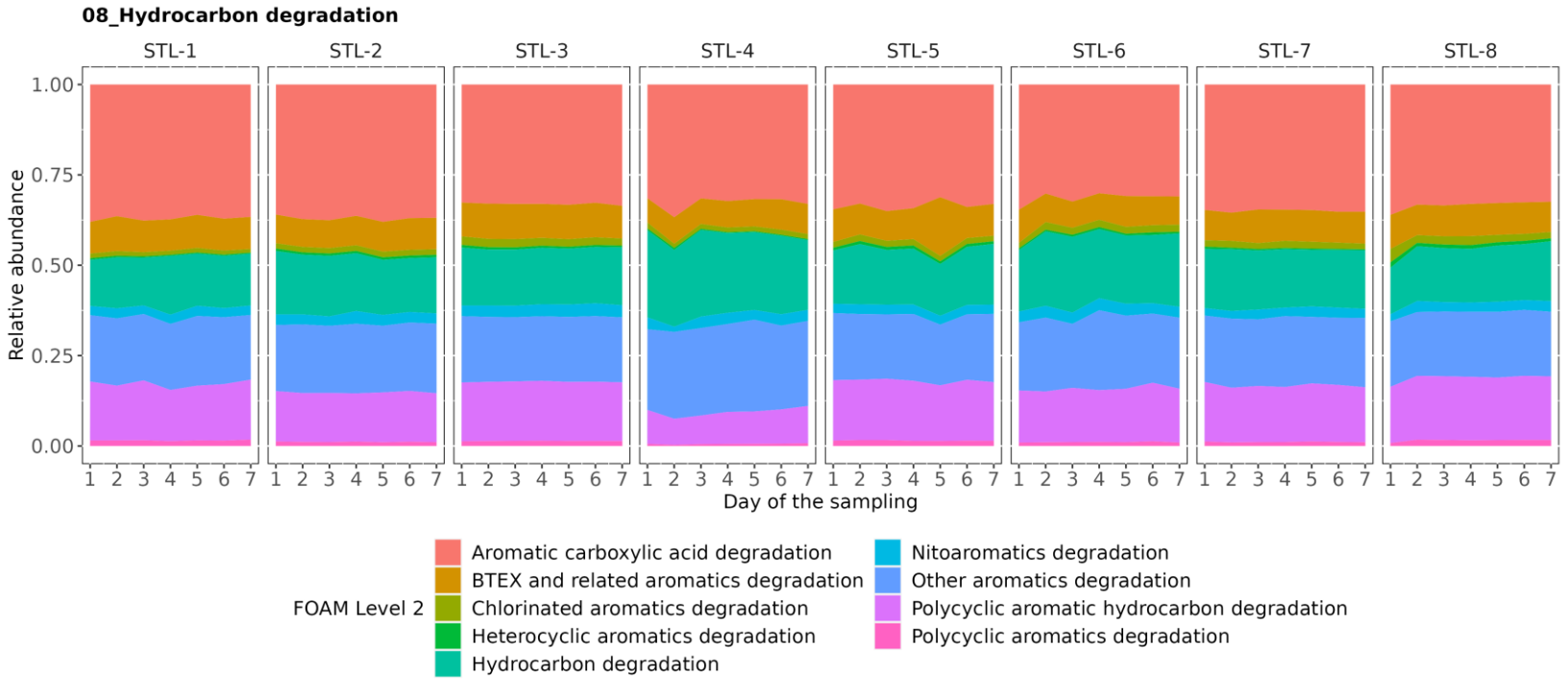
Supplementary Fig. 6: An area chart showing relative abundances of FOAM level 2 functions in fatty acid oxidation. Relative abundances were computed on RPKMs.



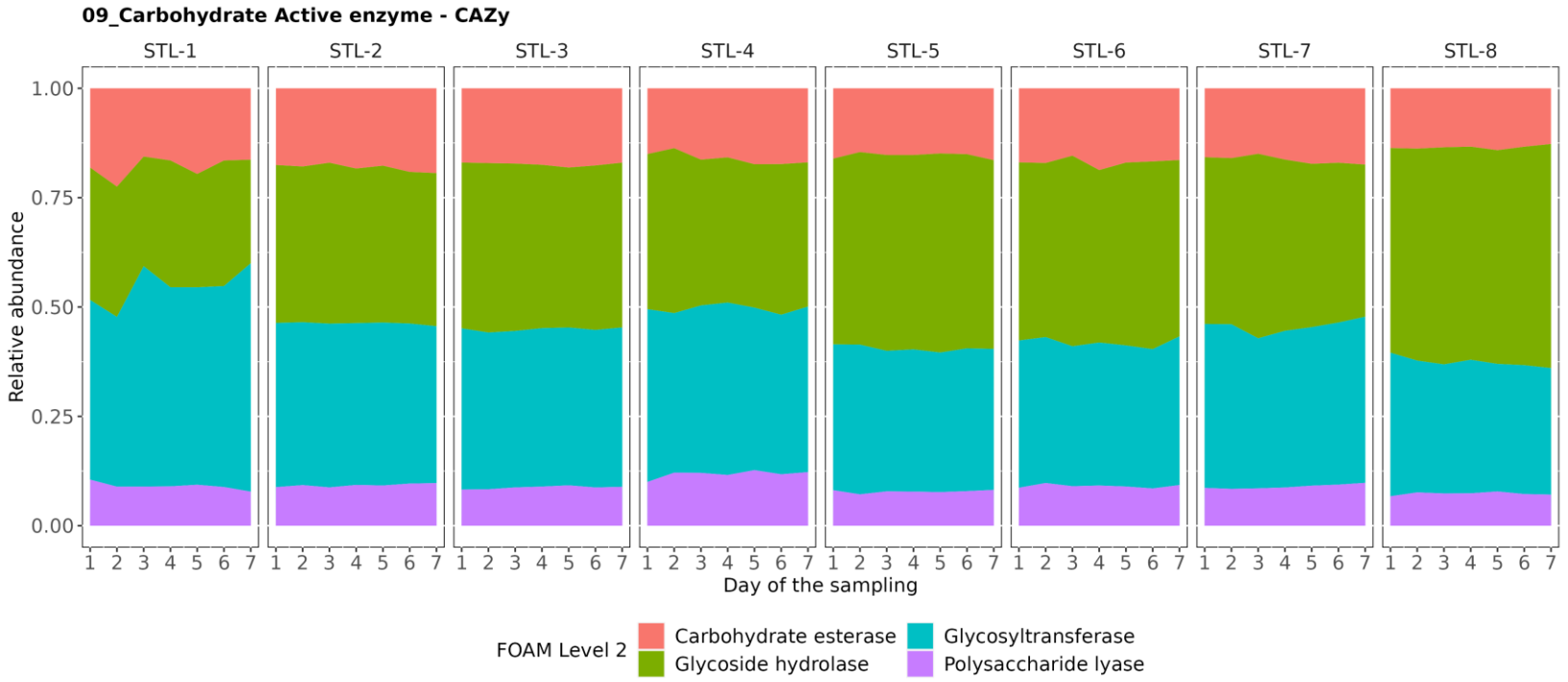
Supplementary Fig. 7: An area chart showing relative abundances of FOAM level 2 functions in amino acid utilization. Relative abundances were computed on RPKMs.



Supplementary Fig. 8: An area chart showing relative abundances of FOAM level 2 functions in nucleic acid metabolism. Relative abundances were computed on RPKMs.

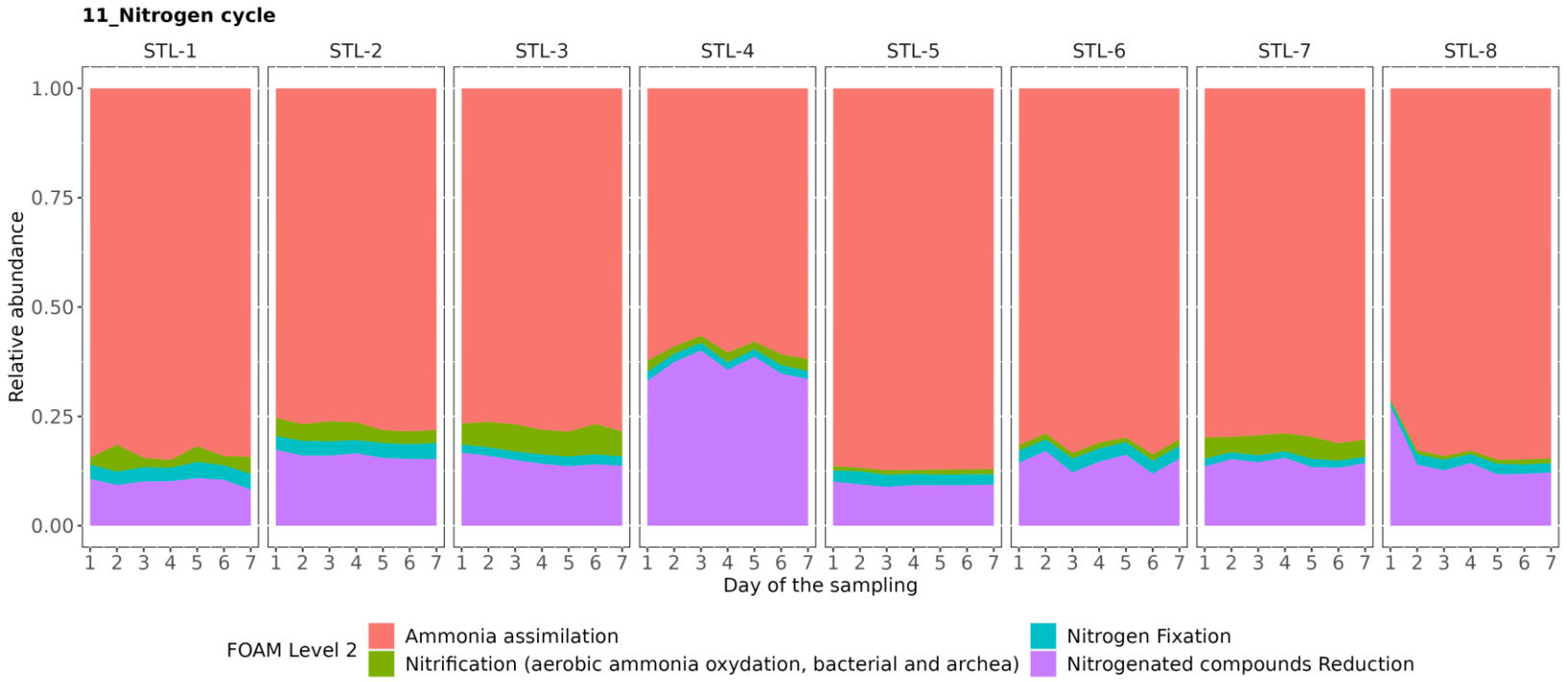


Supplementary Fig. 9: An area chart showing relative abundances of FOAM level 2 functions in hydrocarbon degradation. Relative abundances were computed on RPKMs.

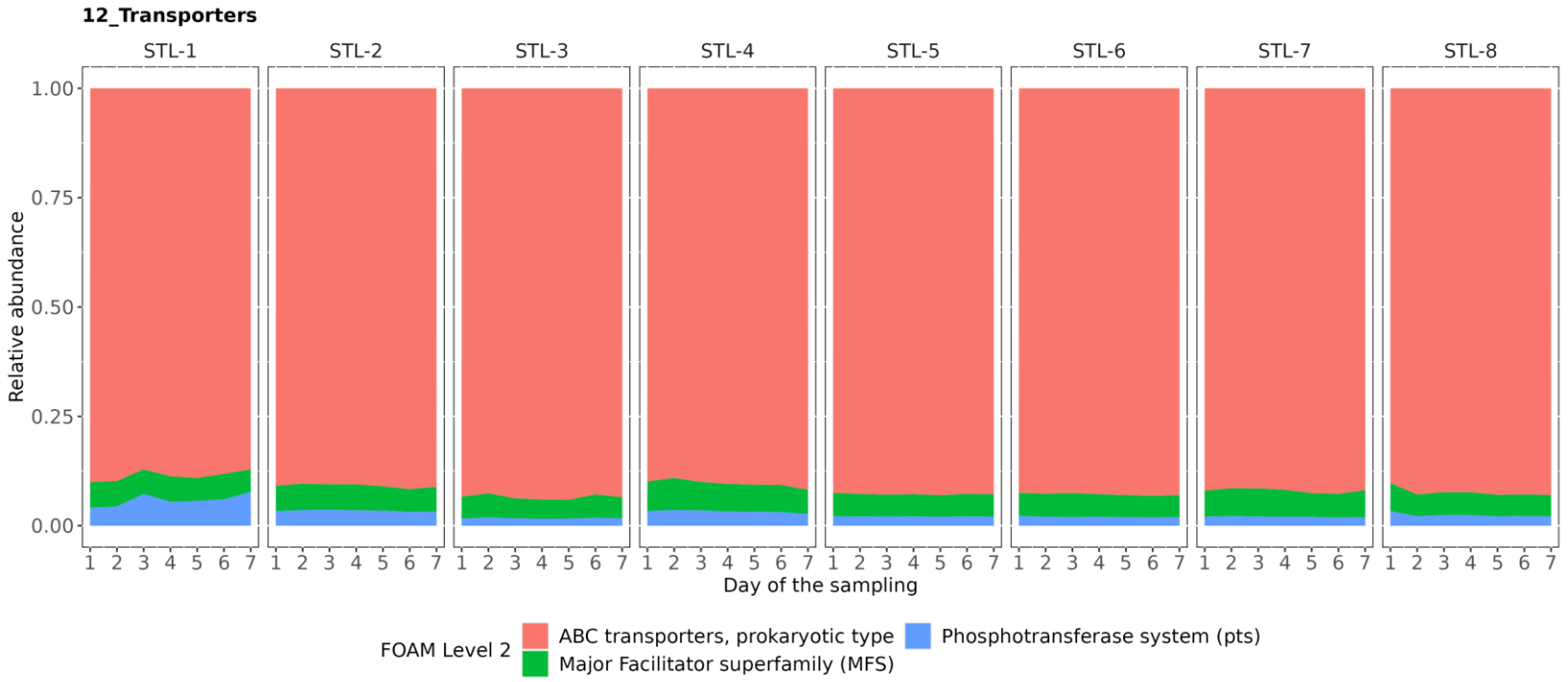


Supplementary Fig. 10: An area chart showing relative abundances of FOAM level 2 functions in carbohydrate active enzyme. Relative abundances were computed on RPKMs.

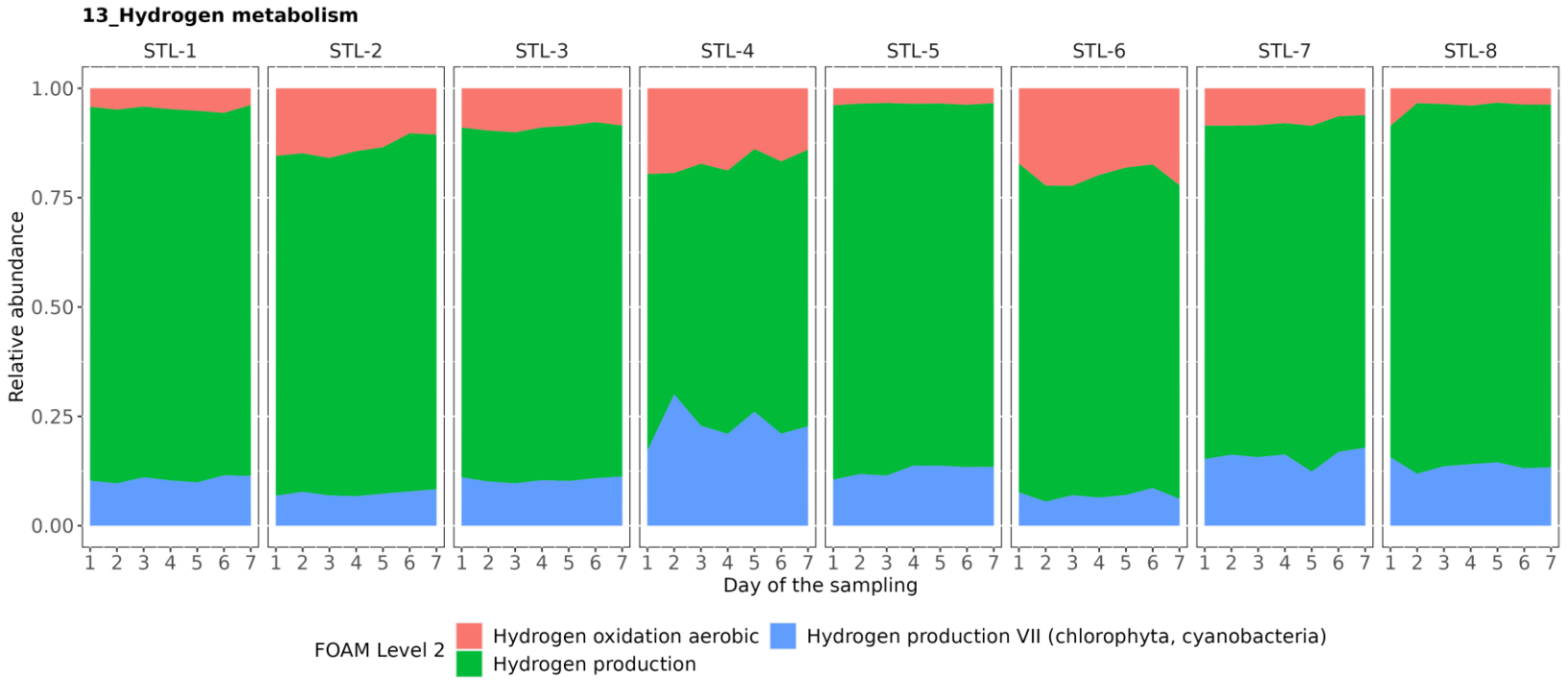
Supplementary Fig. 11: An area chart showing relative abundances of FOAM level 2 functions in TCA cycle. Relative abundances were computed on RPKMs.



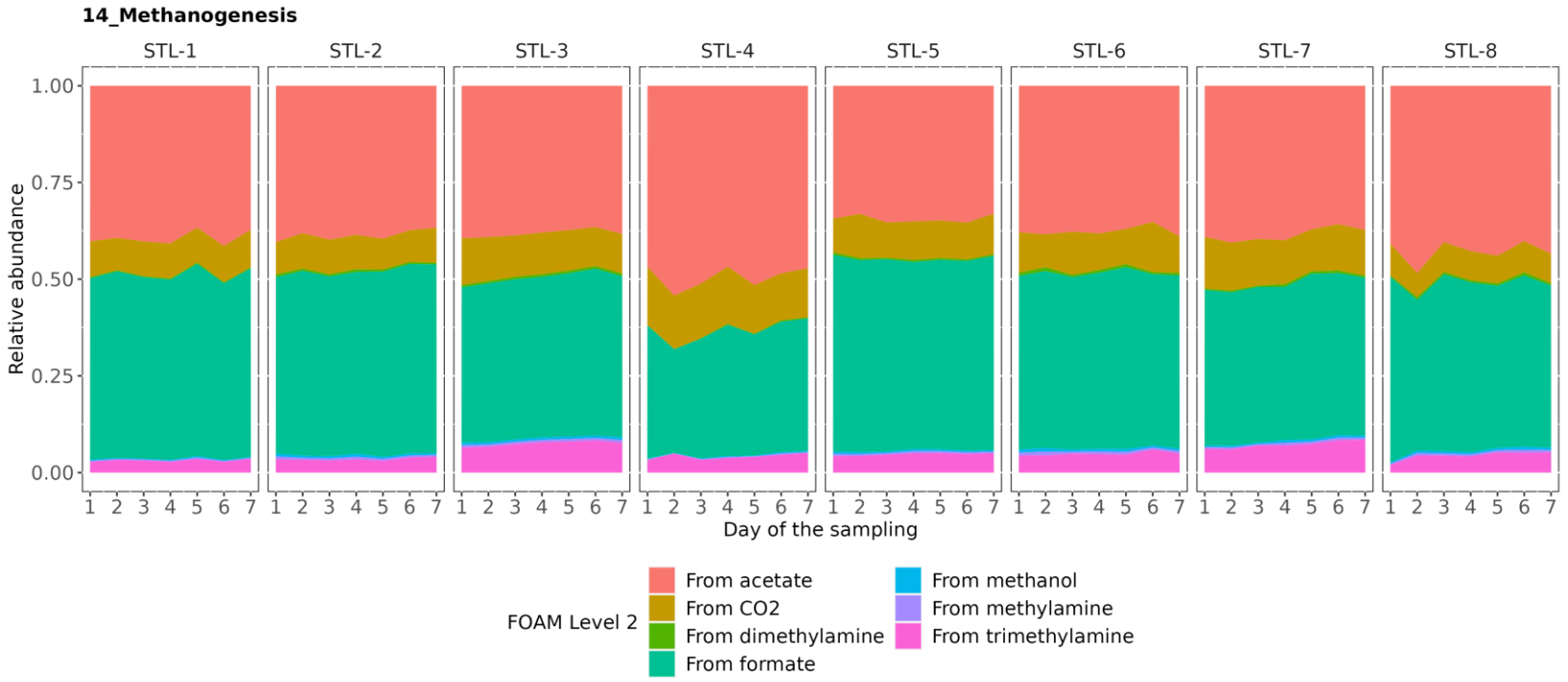
Supplementary Fig. 12: An area chart showing relative abundances of FOAM level 2 functions in nitrogen cycle. Relative abundances were computed on RPKMs.



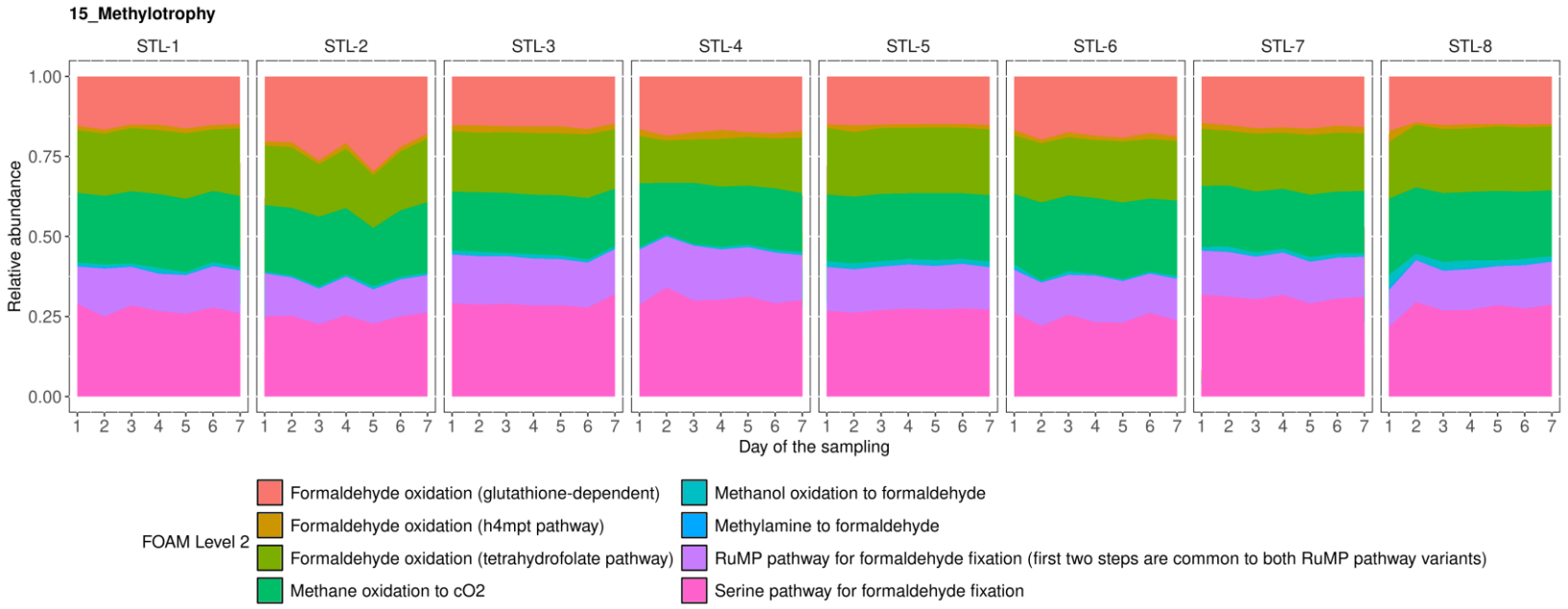
Supplementary Fig. 13: An area chart showing relative abundances of FOAM level 2 functions in transporters. Relative abundances were computed on RPKMs.



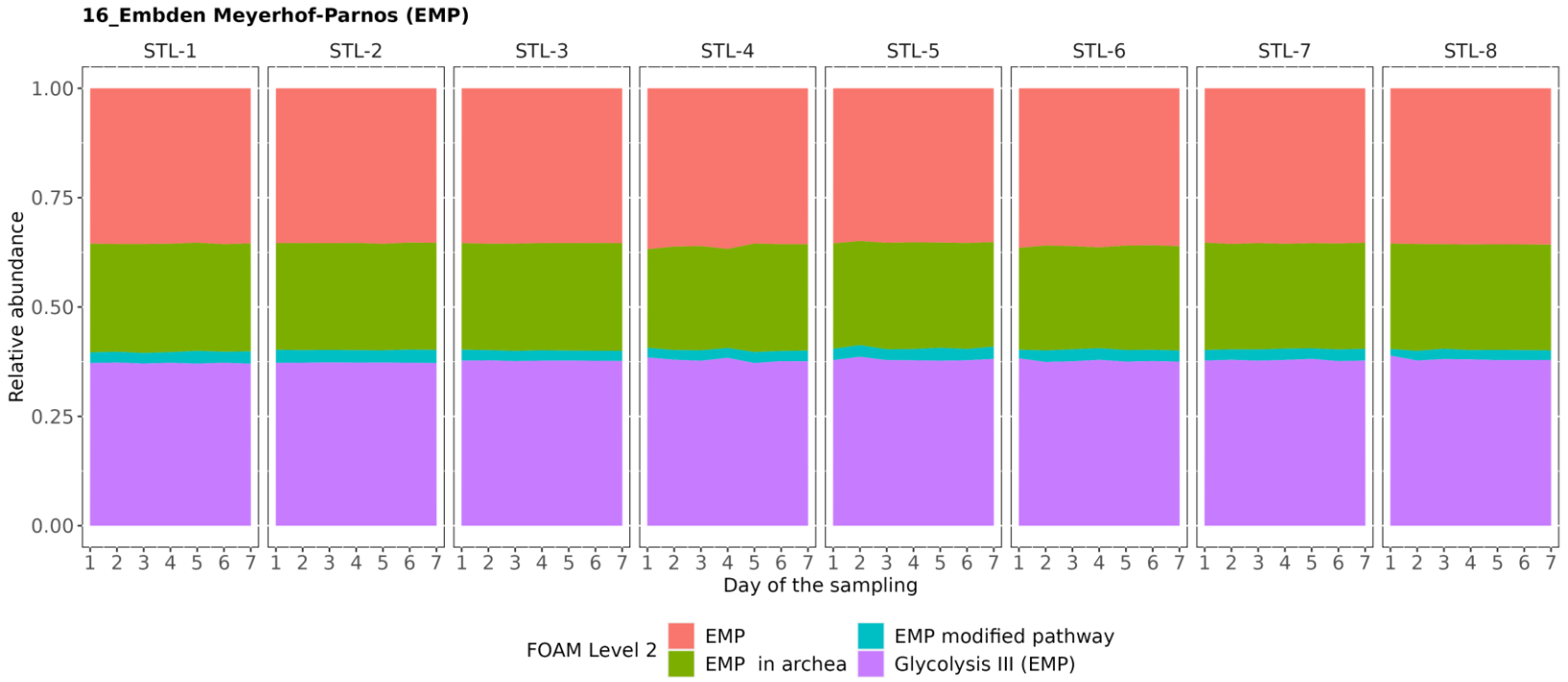
Supplementary Fig. 14: An area chart showing relative abundances of FOAM level 2 functions in hydrogen metabolism. Relative abundances were computed on RPKMs.



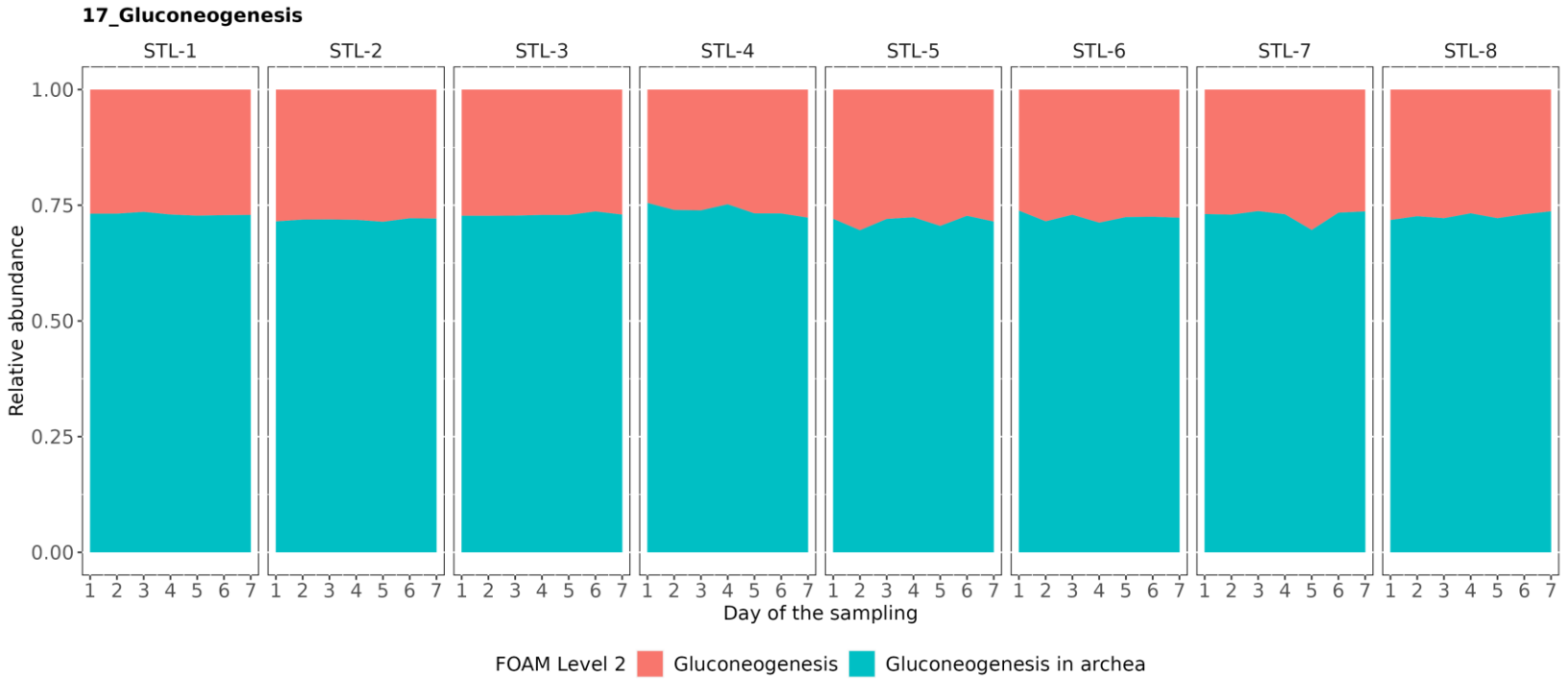
Supplementary Fig. 15: An area chart showing relative abundances of FOAM level 2 functions in methanogenesis. Relative abundances were computed on RPKMs.



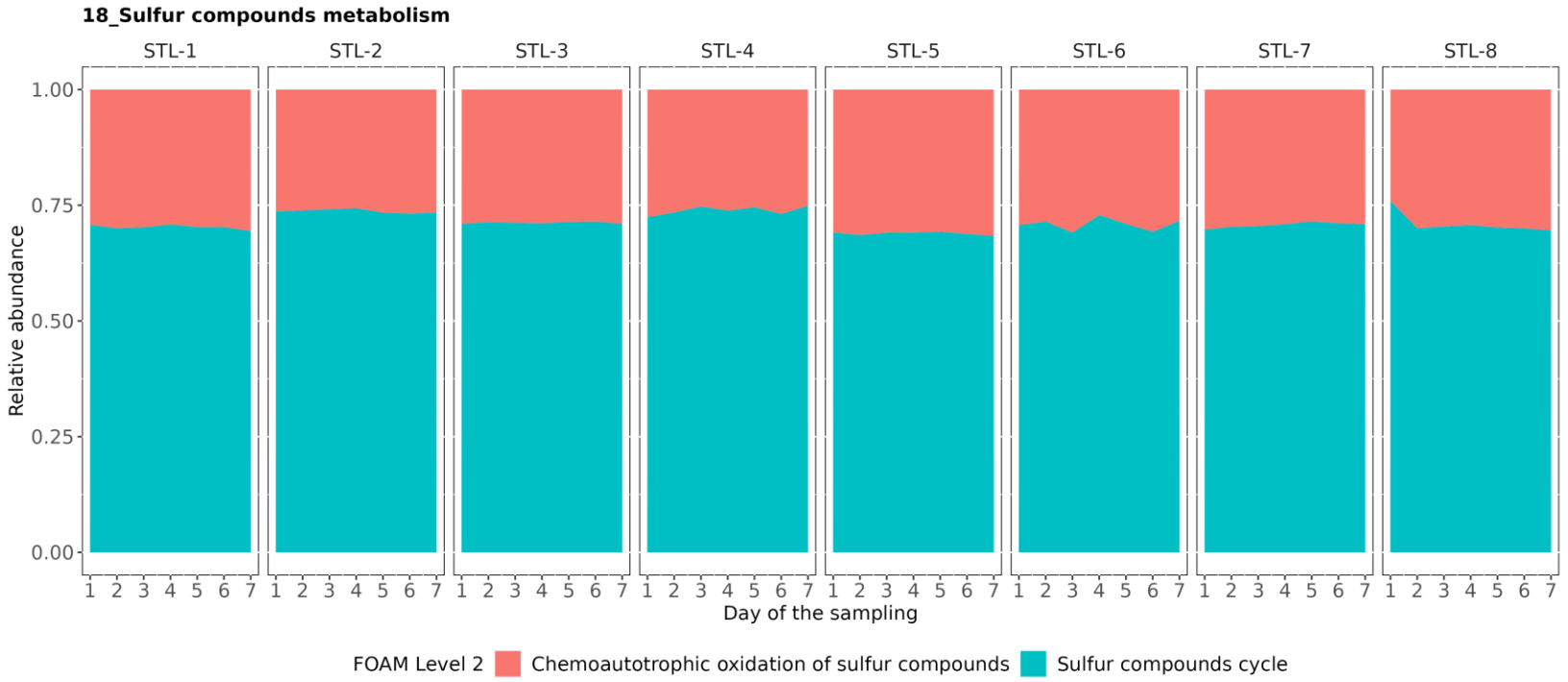
Supplementary Fig. 16: An area chart showing relative abundances of FOAM level 2 functions in methylotrophy. Relative abundances were computed on RPKMs.



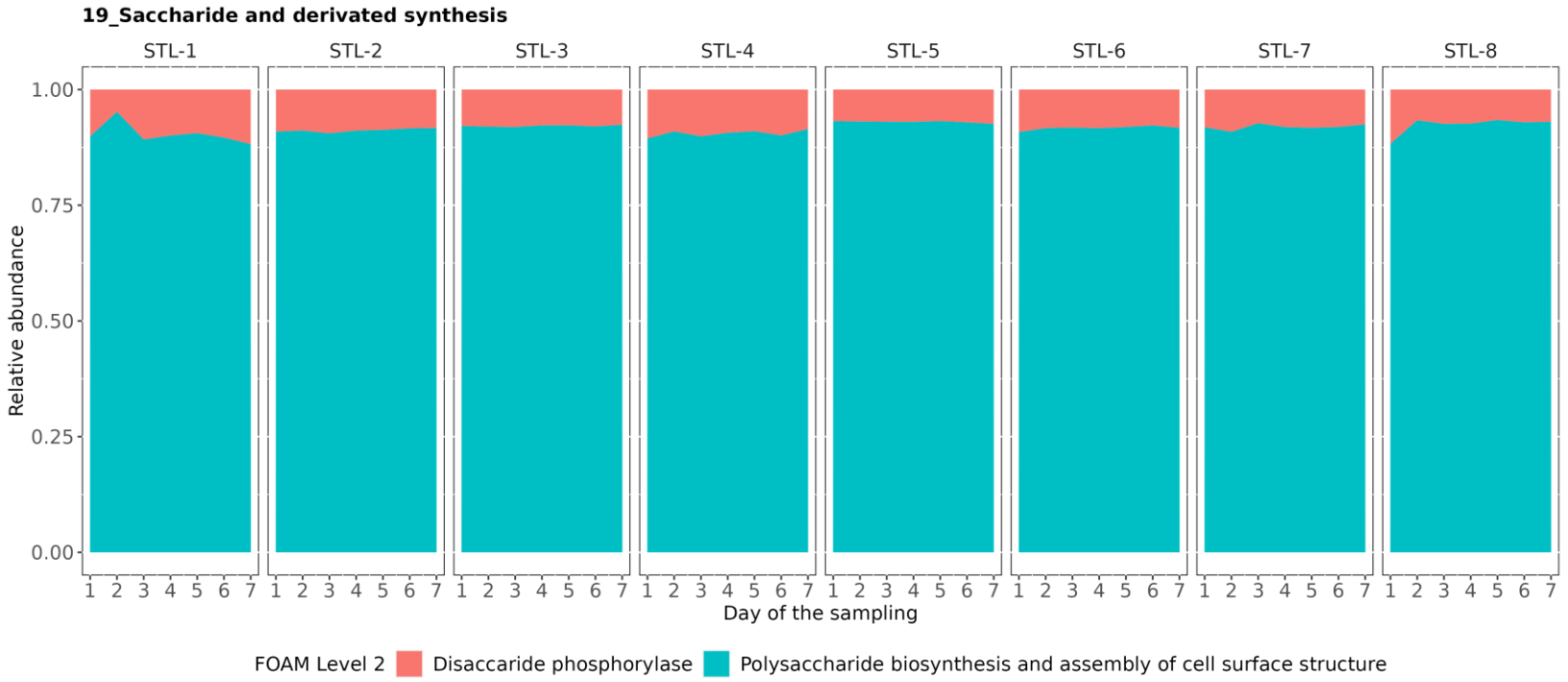
Supplementary Fig. 17: An area chart showing relative abundances of FOAM level 2 functions in Embden Meyerhof-Parnos. Relative abundances were computed on RPKMs.



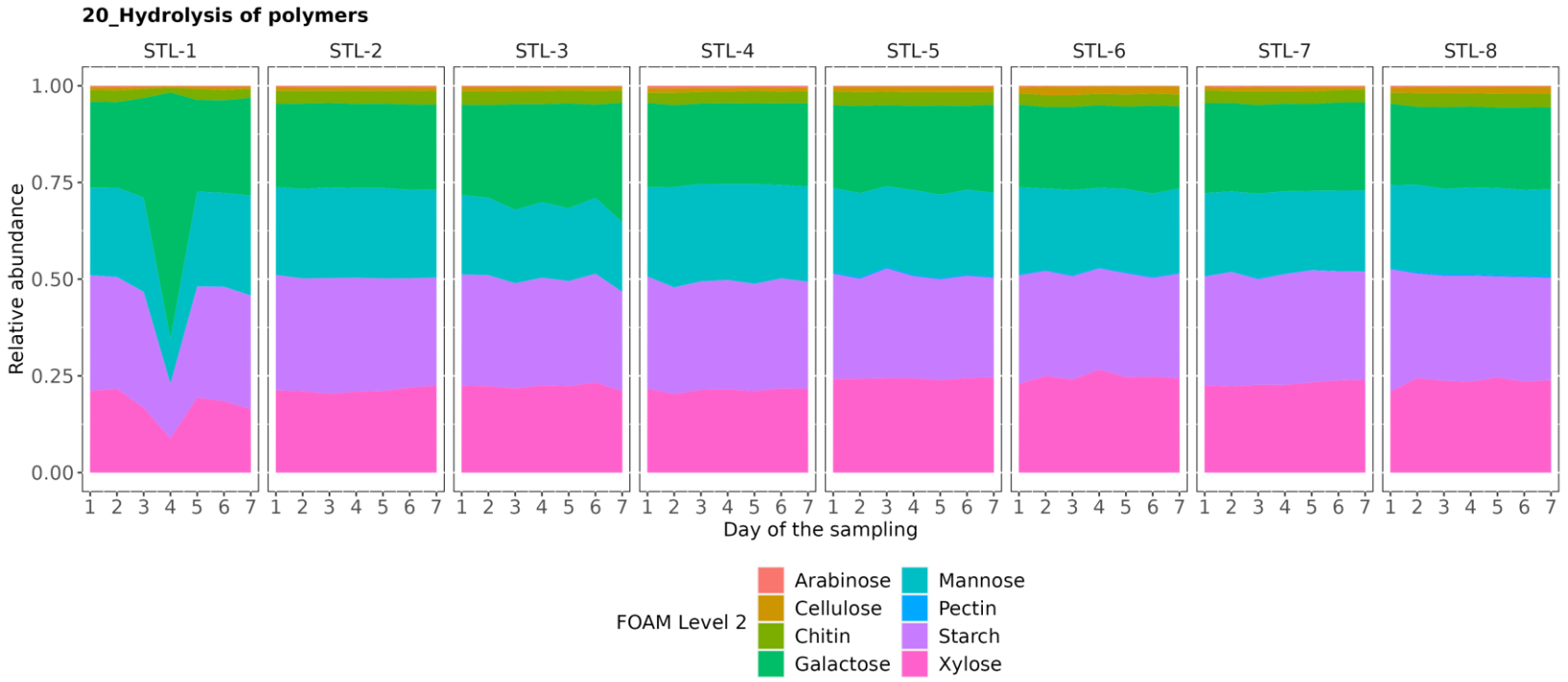
Supplementary Fig. 18: An area chart showing relative abundances of FOAM level 2 functions in gluconeogenesis. Relative abundances were computed on RPKMs.



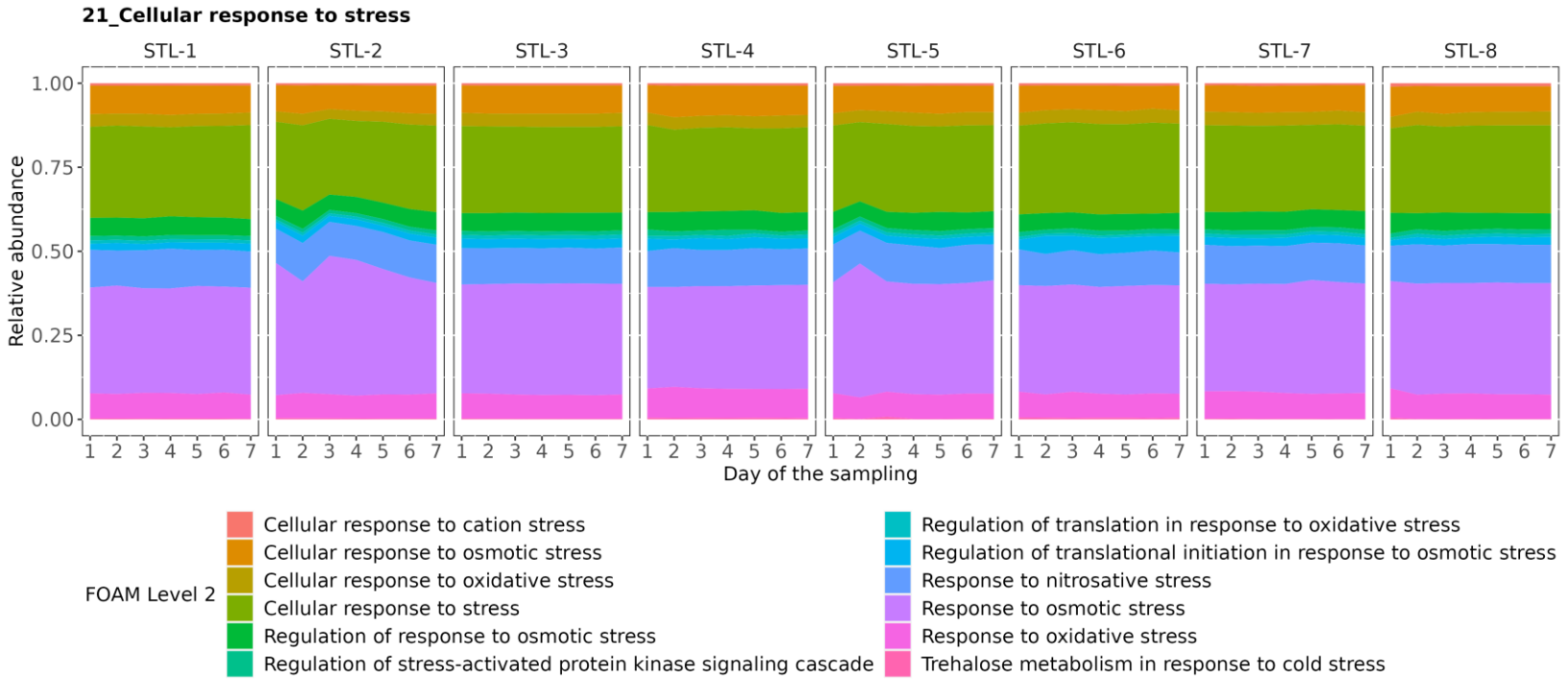
Supplementary Fig. 19: An area chart showing relative abundances of FOAM level 2 functions in sulfur compounds metabolism. Relative abundances were computed on RPKMs.



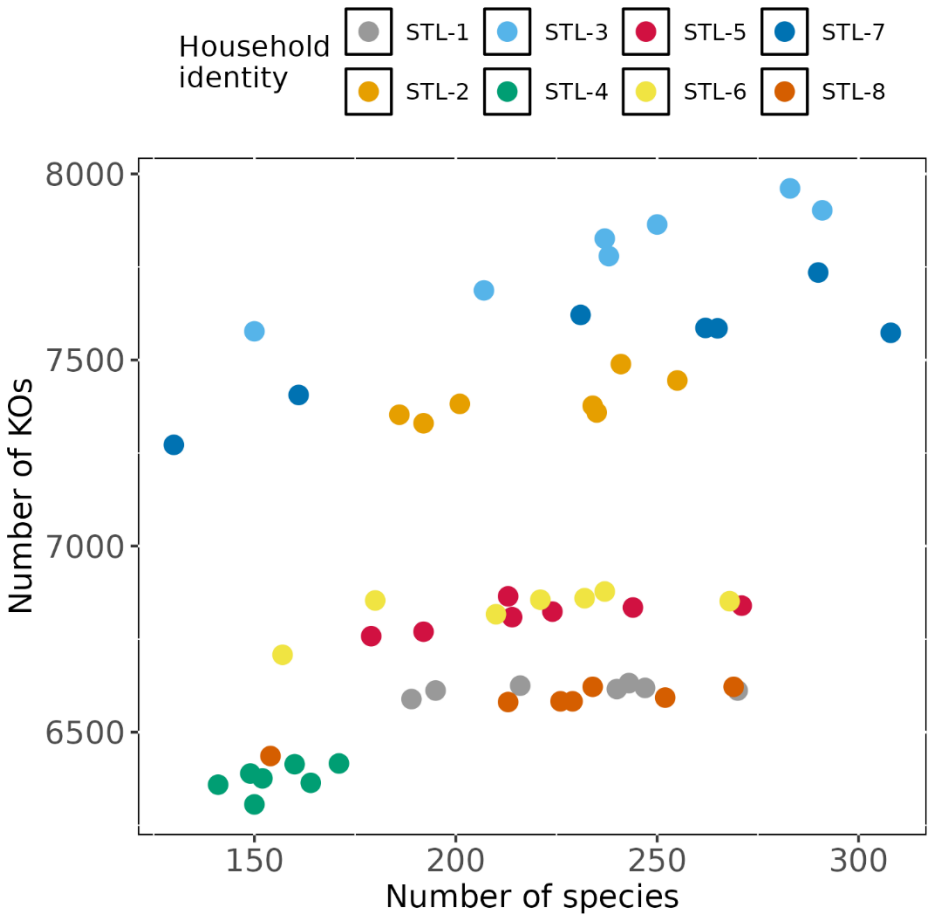
Supplementary Fig. 20: An area chart showing relative abundances of FOAM level 2 functions in saccharide and derivated synthesis. Relative abundances were computed on RPKMs.



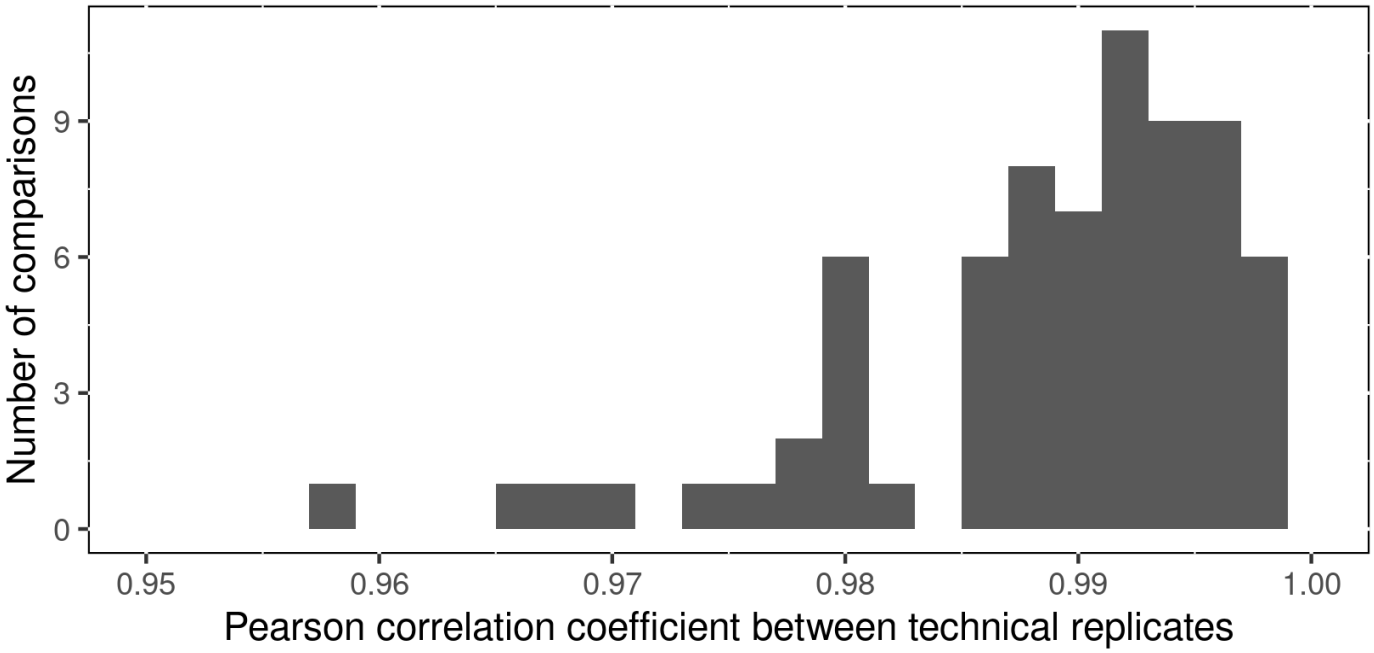
Supplementary Fig. 21: An area chart showing relative abundances of FOAM level 2 functions in hydrolysis of polymers. Relative abundances were computed on RPKMs.



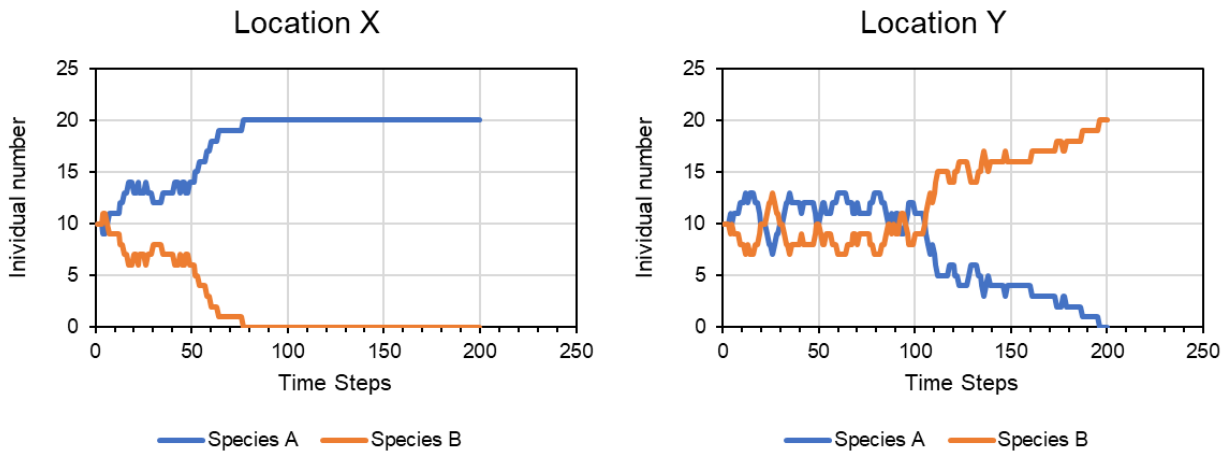
Supplementary Fig. 22: An area chart showing relative abundances of FOAM level 2 functions in cellular response to stress. Relative abundances were computed on RPKMs.



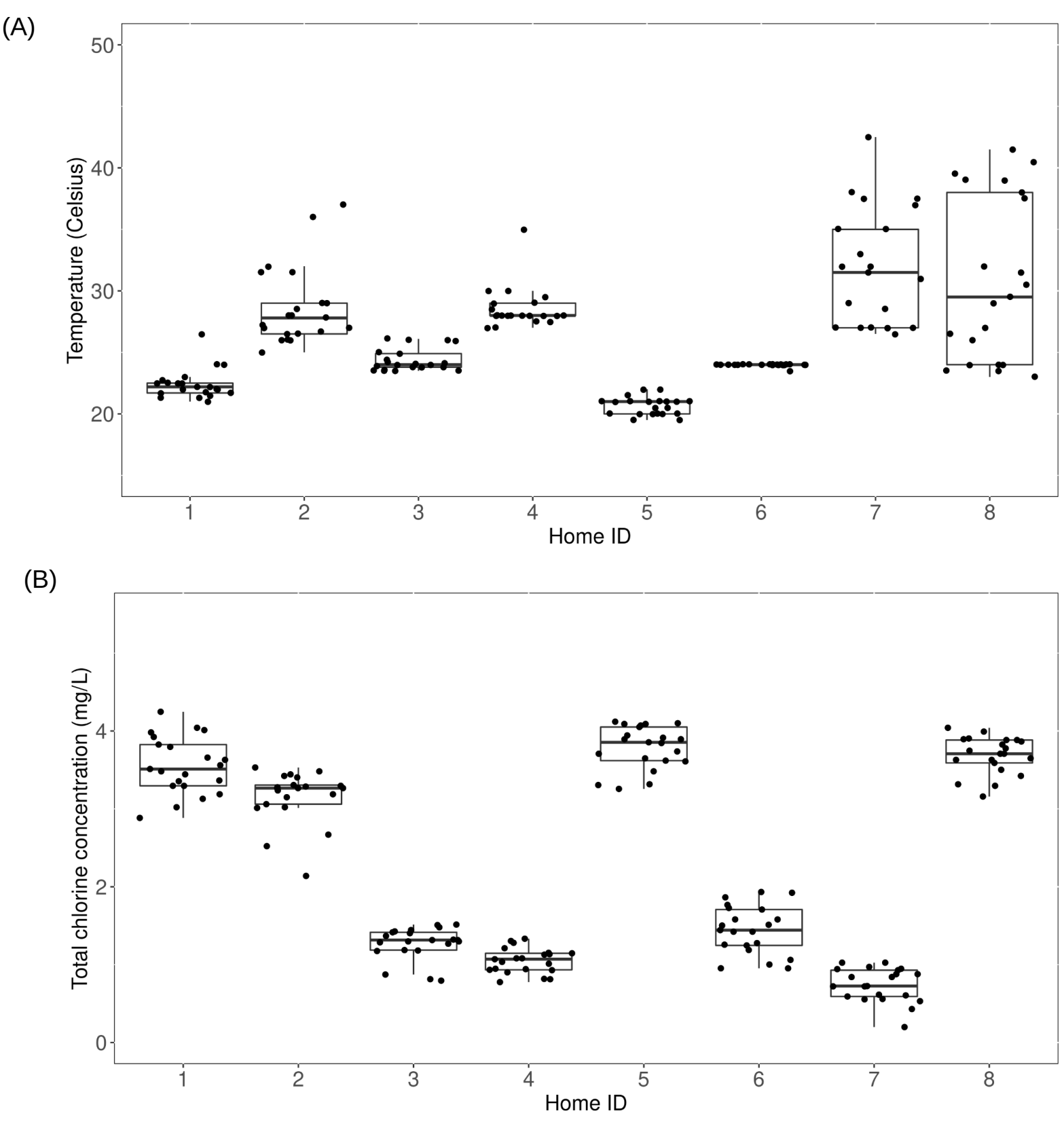
Supplementary Fig. 23: Relationships between the number of KOs and the number of species. Dots of the same color indicate bathtub faucet microbiome samples from the same household.



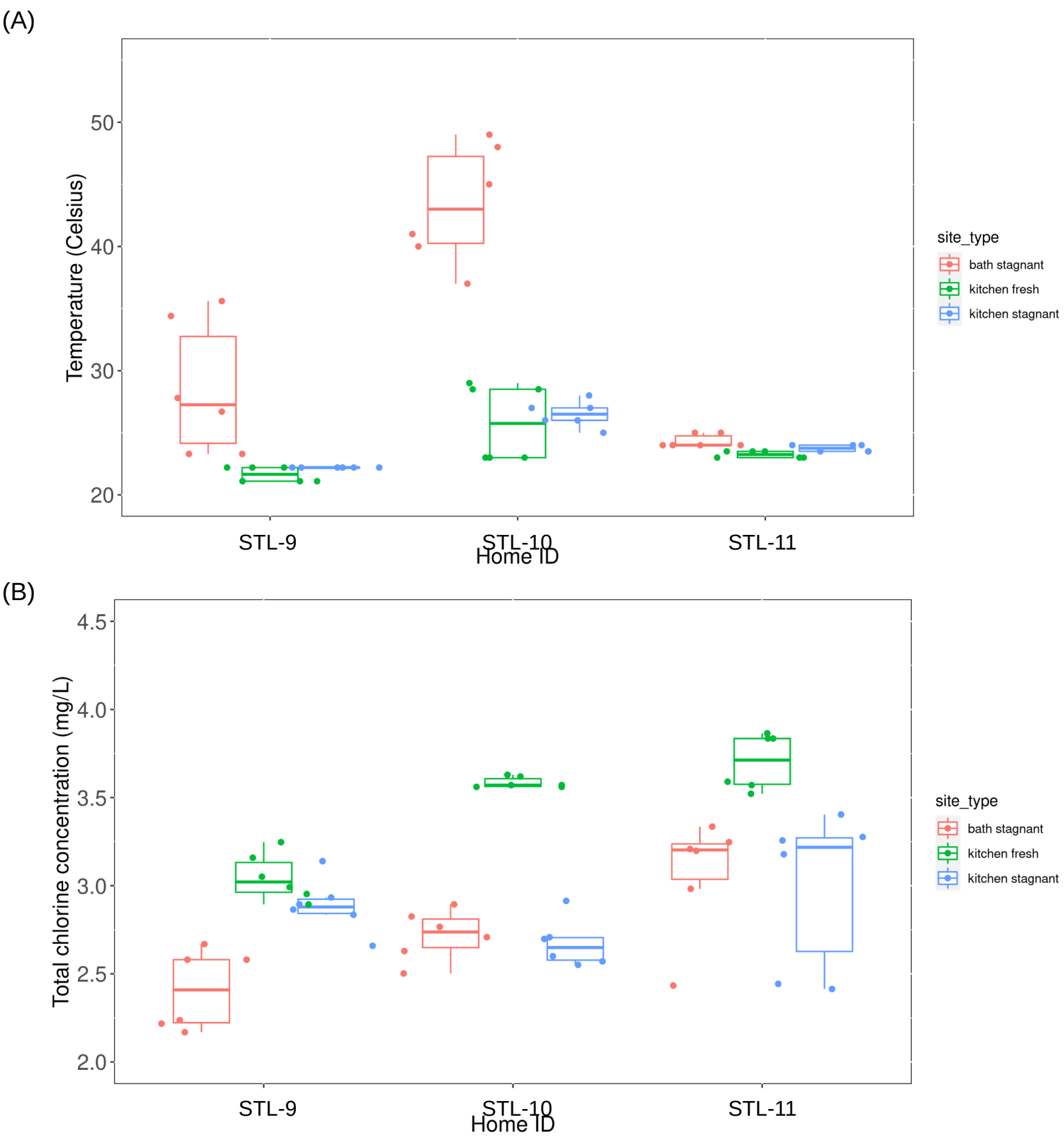
Supplementary Fig. 24: Technical replicates showed a high level of correlation (Pearson correlation coefficient >0.958).



Supplementary Fig. 25: Location-specific pattern driven by drift.

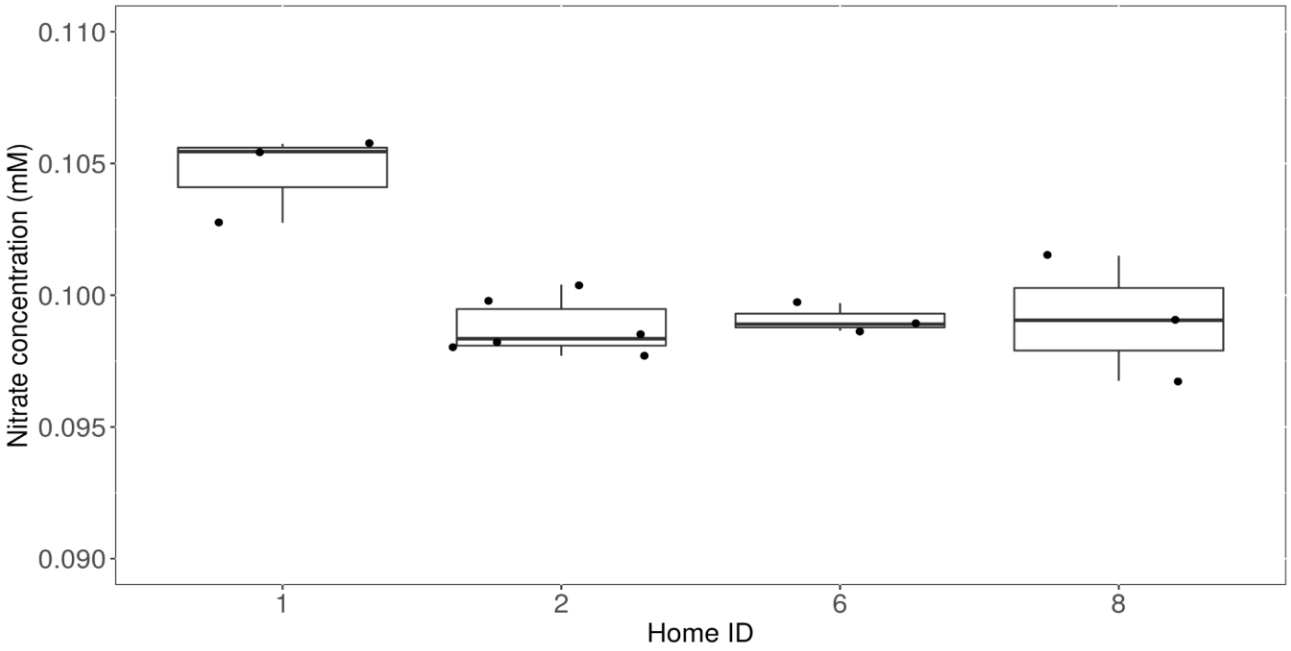


Supplementary Fig. 26: (A) Temperature and (B) total chlorine measurements for samples collected in Experiment 1. For both temperature and total chlorine, each home has 21 measurements (the first, second, and third liter of bathtub faucet water were measured separately on seven days). In the box plot, 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).

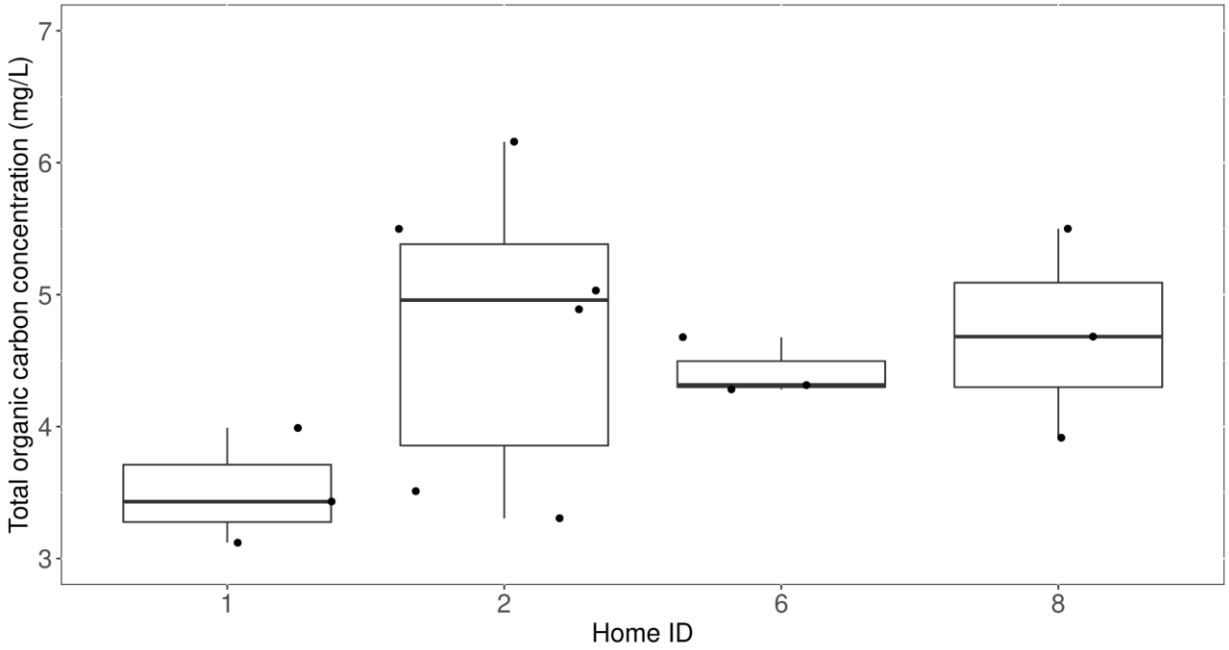


Supplementary Fig. 27: (A) Temperature and (B) total chlorine measurements for samples collected in Experiment 2. For both temperature and total chlorine, each home has 18 measurements (the first, second, and third liter were measured separately for each type of sample on two days). In the box plot, 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).

(A)



(B)



Supplementary Fig. 28: (A) Nitrate and (B) total organic carbon (TOC) concentrations of samples collected from a subset of homes in Experiment 1. For both nitrate and TOC, each home has 3 measurements (the first, second, and third liter of bathtub faucet water were measured separately on one day), except for home 2, which has 6 measurements for samples on two days. 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).