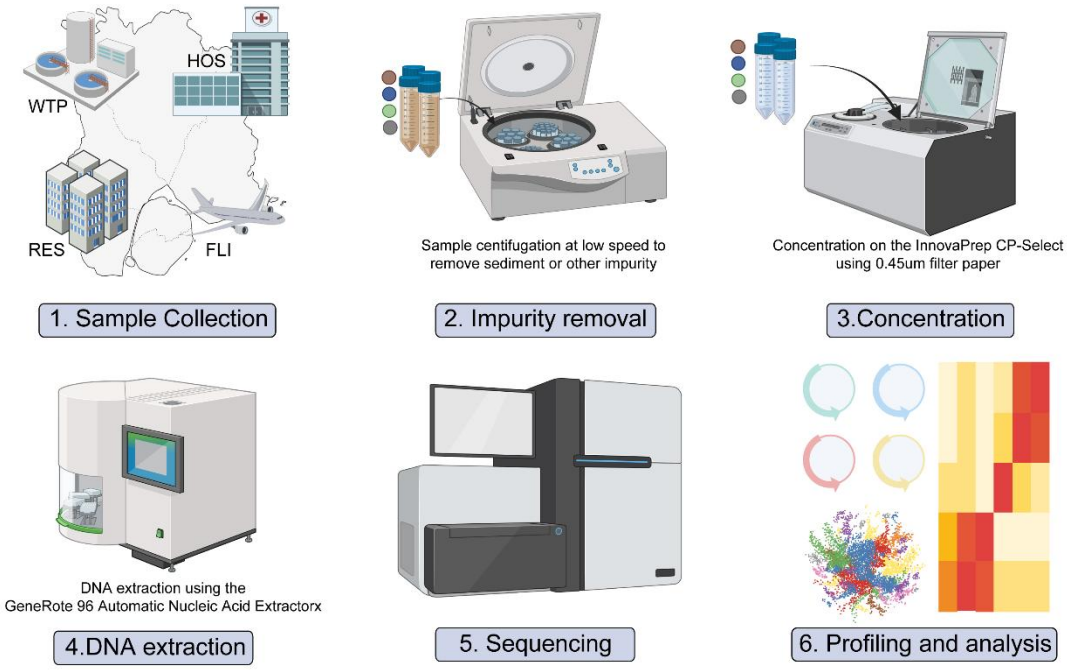
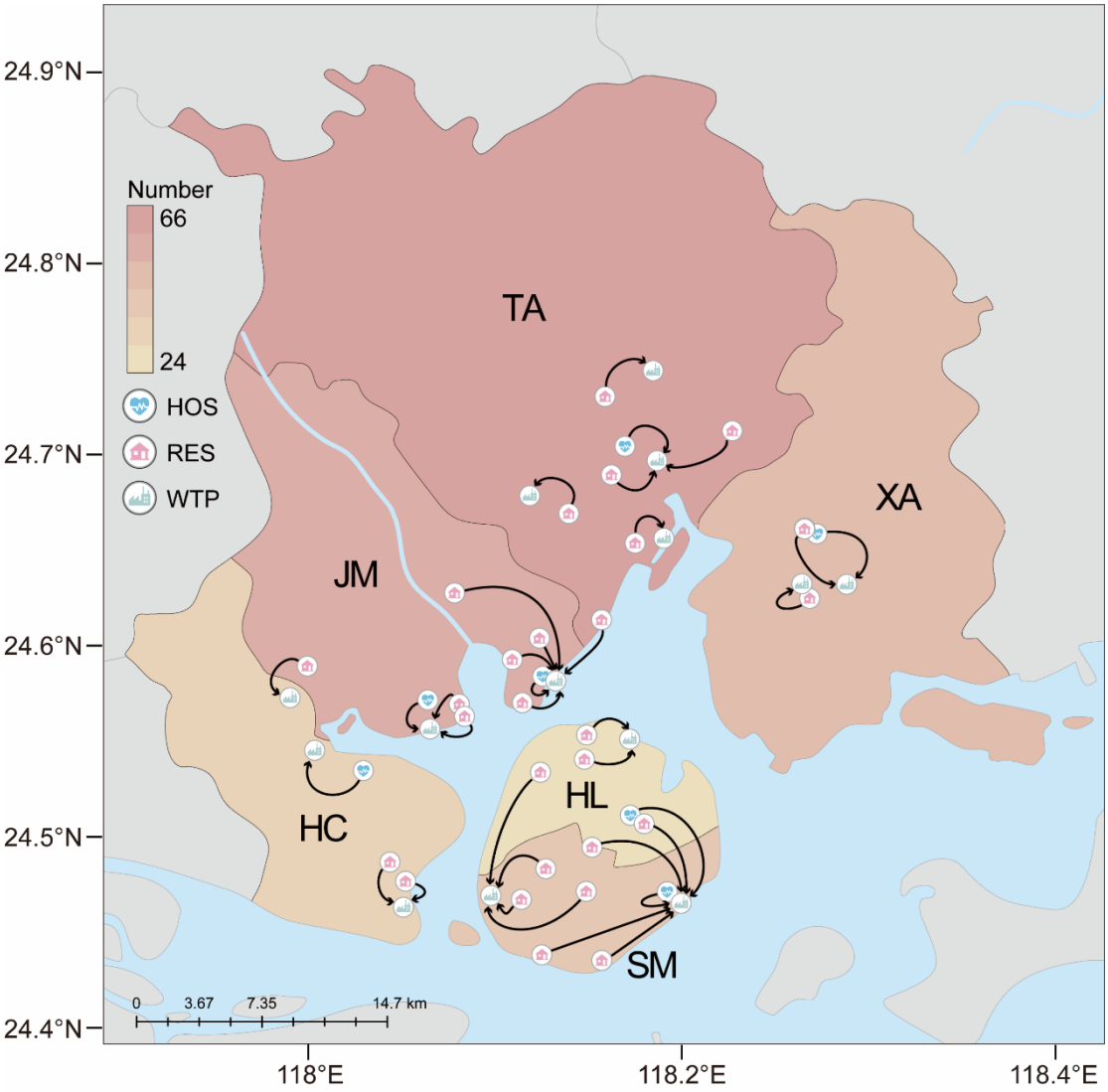
Supplementary information for "Citywide metagenomics reveals microbial community and resistome dynamics in urban wastewater".

This Supplementary Information contains Supplementary Notes, along with 19 supplementary figures and their corresponding legends.

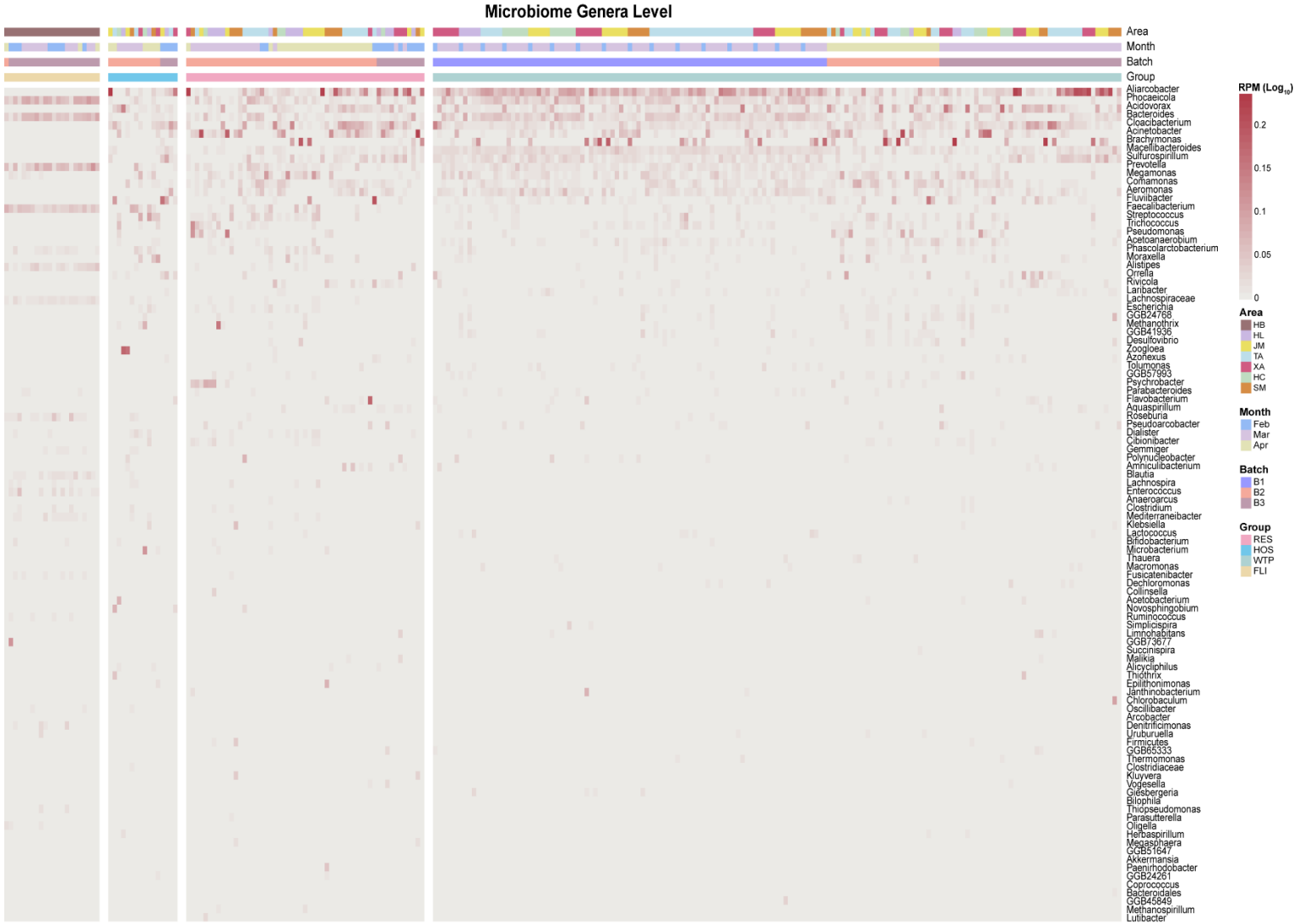
Supplementary Figures



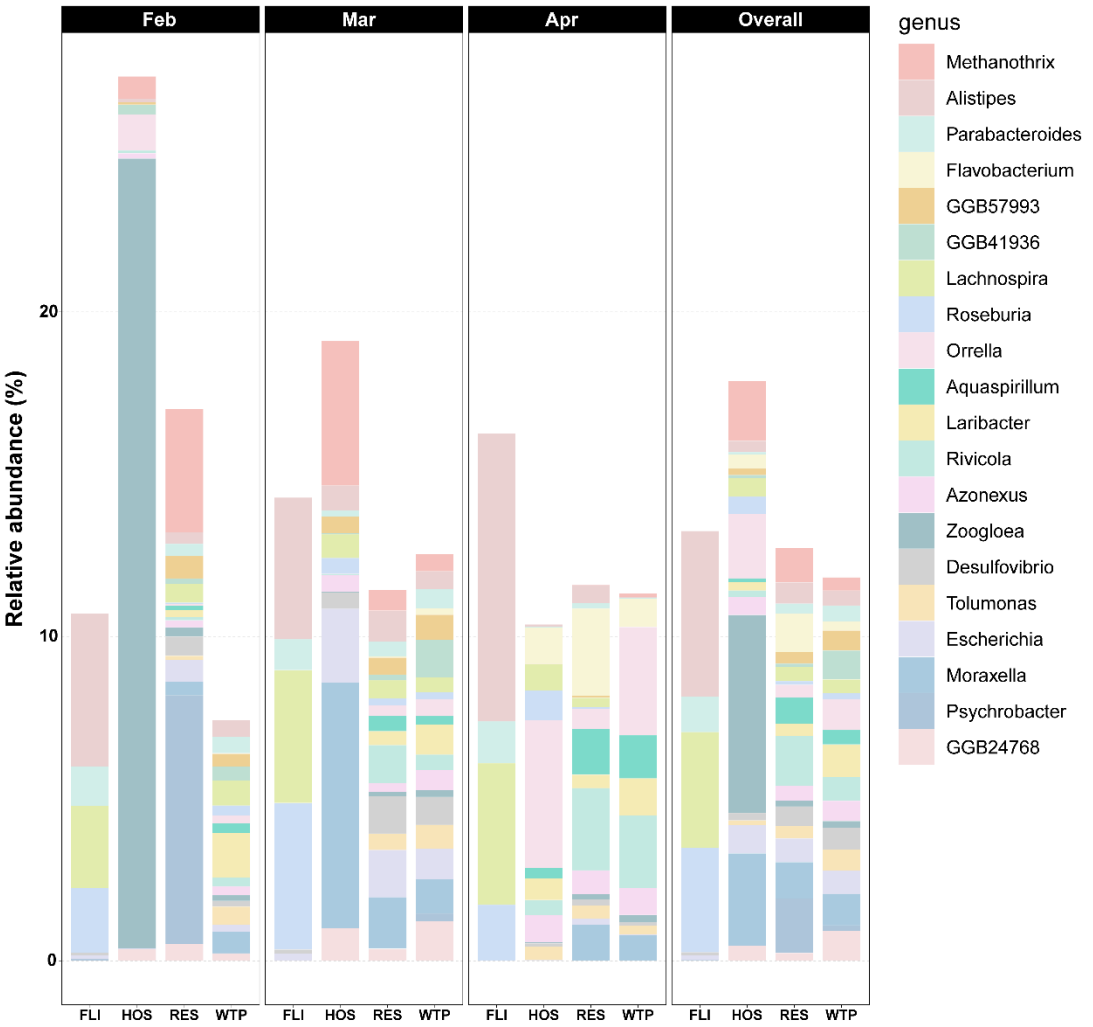
Supplementary Fig. 1. General schematic representation of the study design and experimental pipeline. The workflow included wastewater sample collection, impurity removal, concentration, DNA extraction, and the final steps of library preparation and sequencing. See the methods for detailed information. The plot was created with BioRender.com (citation: https://BioRender.com/wpsbeqx).



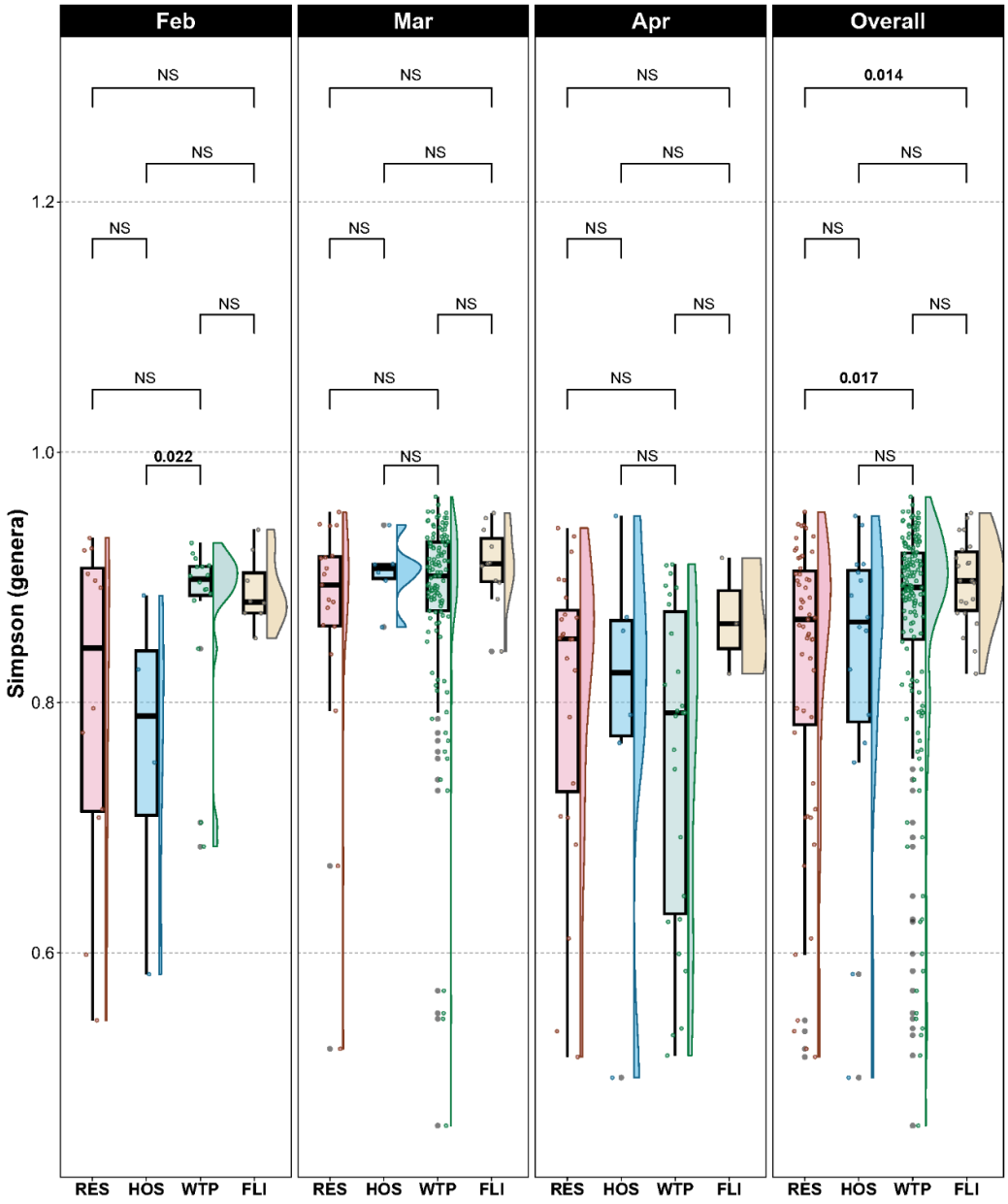
Supplementary Fig. 2. Map illustrating sampling locations with color gradients corresponding to site-specific sampling intensity. Symbolic arrows denote directional connectivity between sampling nodes, representing inferred subsurface flow pathways independent of municipal sewage infrastructure. Noted that all HOS wastewaters have undergone their respective disinfection processes before being discharged into the municipal pipeline, ultimately converging at the wastewater treatment plant. The map was generated using ArcGIS Pro 3.1.5 (Esri, Redlands, CA, USA). Sampling locations and facility types were visualized using built-in symbol libraries in ArcGIS Pro.


Microbiome Genera Level
Area
Month
Batch
Group
Aliarcobacter
Phocaeicola
Acidovorax
Bacteroides
Cloacibacterium
Acinetobacter
Brachymonas
Macellibacteroides
Sulfurospirillum
Prevotella
Megamonas
Comamonas
Aeromonas
Fluviibacter
Faecalibacterium
Streptococcus
Trichococcus
Pseudomonas
Acetoanaerobium
Phascolarctobacterium
Moraxella
Alistipes
Orrella
Rivicola
Laribacter
Lachnospiraceae
Escherichia
GGB24768
Methanothrix
GGB41936
Desulfovibrio
Zoogloea
Azonexus
Tolumonas
GGB57993
Psychrobacter
Parabacteroides
Flavobacterium
Aquaspirillum
Roseburia
Pseudoarcobacter
Dialister
Cibionibacter
Gemmiger
Polynucleobacter
Amniculibacterium
Blautia
Lachnospira
Enterococcus
Anaerorarcus
Clostridium
Mediterraneibacter
Klebsiella
Lactococcus
Bifidobacterium
Microbacterium
Thauera
Macromonas
Fusicatenibacter
Dechloromonas
Colimsella
Acetobacterium
Novosphingobium
Ruminococcus
Simplicispira
Limnohabitans
GGB73677
Succinispira
Malikia
Alicycliphilus
Thiothrix
Epilithonimonas
Janthinobacterium
Chlorobaculum
Oscillibacter
Arcobacter
Dentrificimonas
Uruburuella
Firmicutes
GGB65333
Thermomonas
Clostridiaceae
Kluyvera
Vogesella
Giesbergeria
Blophila
Thiopseudomonas
Parasutterella
Oligella
Herbaspirillum
Megasphaera
GGB51647
Akkermansia
Paenirhodobacter
GGB24261
Coprococcus
Bacteroidales
GGB45849
Methanospirillum
Lutibacter
RPM (Log10)
0.2
0.15
0.1
0.05
0
Area
HB
HL
JM
TA
XA
HC
SM
Month
Feb
Mar
Apr
Batch
B1
B2
B3
Group
RES
HOS
WTP
FLI

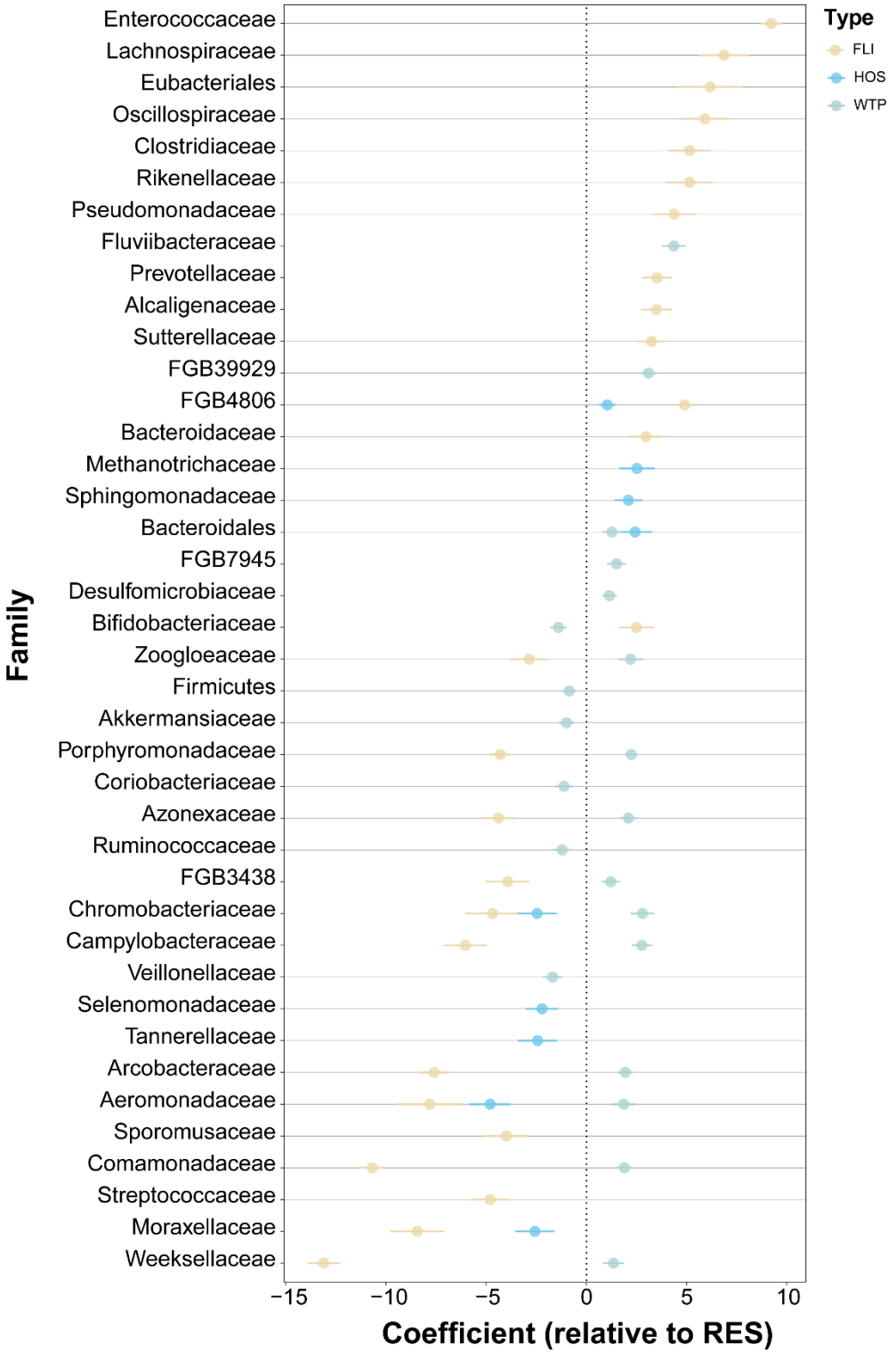
Supplementary Fig. 3. Heatmap profiling relative abundance (log10-scale RPM) of total microbiome. The samples (x-axis) are annotated at the top. The first bar annotated the collection area of samples in the top, the second bar maps the collection time of samples, the third bar indicates the sequencing batch of samples, and the fourth bar shows the type of samples. The genera of microbiome (y-axis) are directly specified on the right.



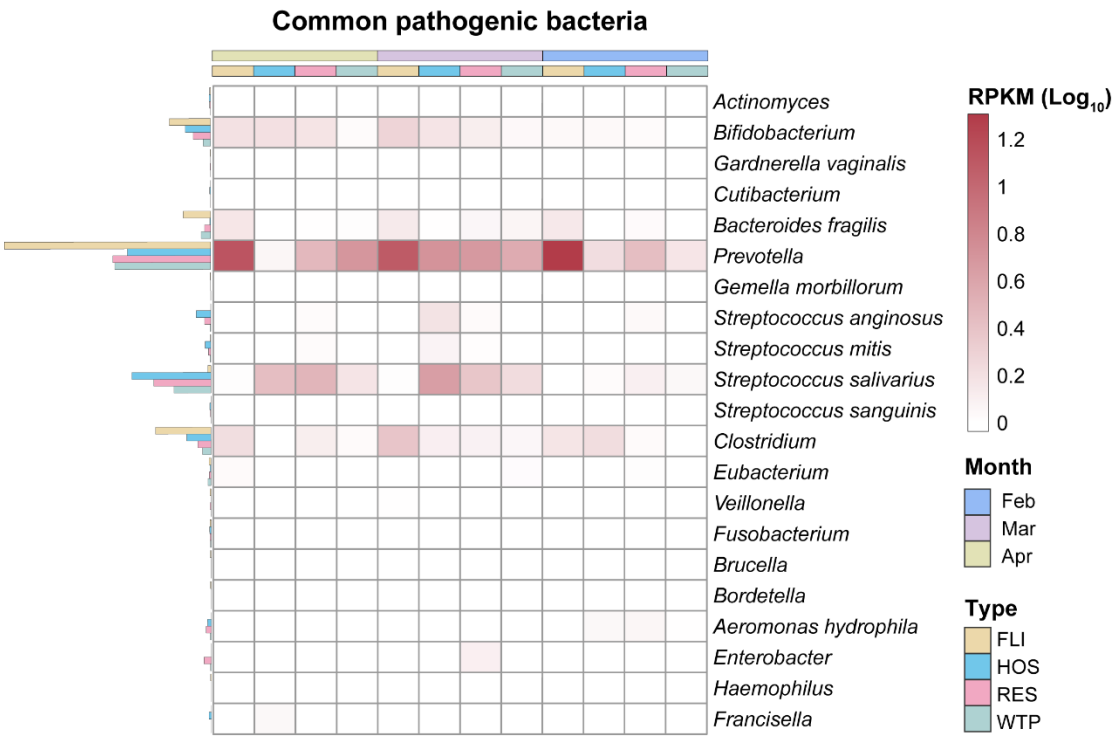
Supplementary Fig. 4. Stacked bar plot showing the relative abundance of the top 20 bacterial genera within the “Others” category in **Fig. 1D**.



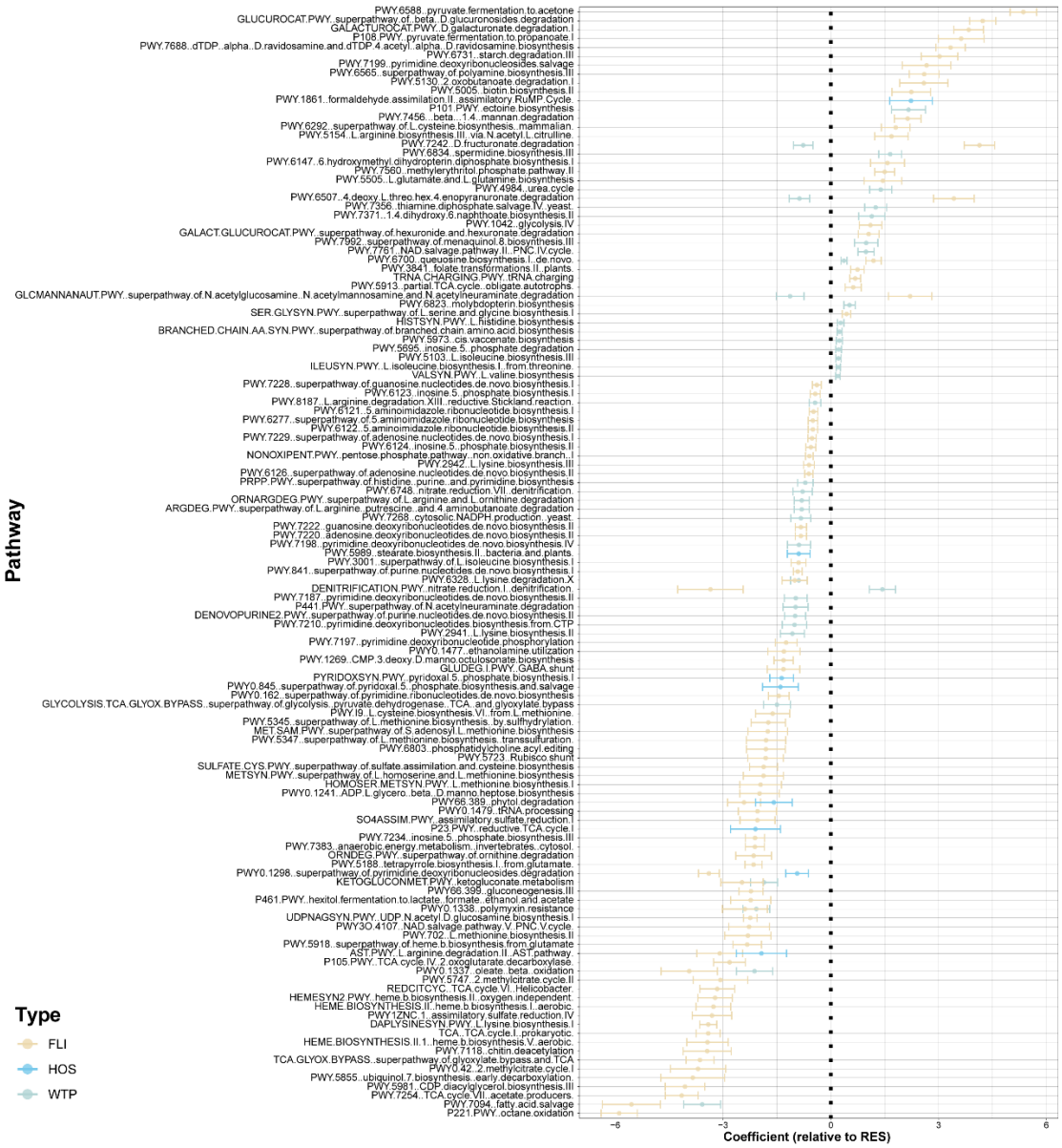
Supplementary Fig. 5. Genus-level Simpson diversity of microbiome across the four sample types over time. Individual data points correspond to distinct samples (WTP, n = 159; RES, n = 55; HOS, n = 16; FLI, n = 22). In each boxplot, the center line represents the median; the lower and upper bounds of the box indicate the 25th and 75th percentiles (first and third quartiles), respectively; the whiskers extend to the minima and maxima, excluding outliers defined as 1.5 times the interquartile range. Statistical significance was determined by two-tailed Wilcoxon rank-sum tests with Benjamini-Hochberg FDR correction ($p < 0.05$ considered significant; NS: $p > 0.05$).



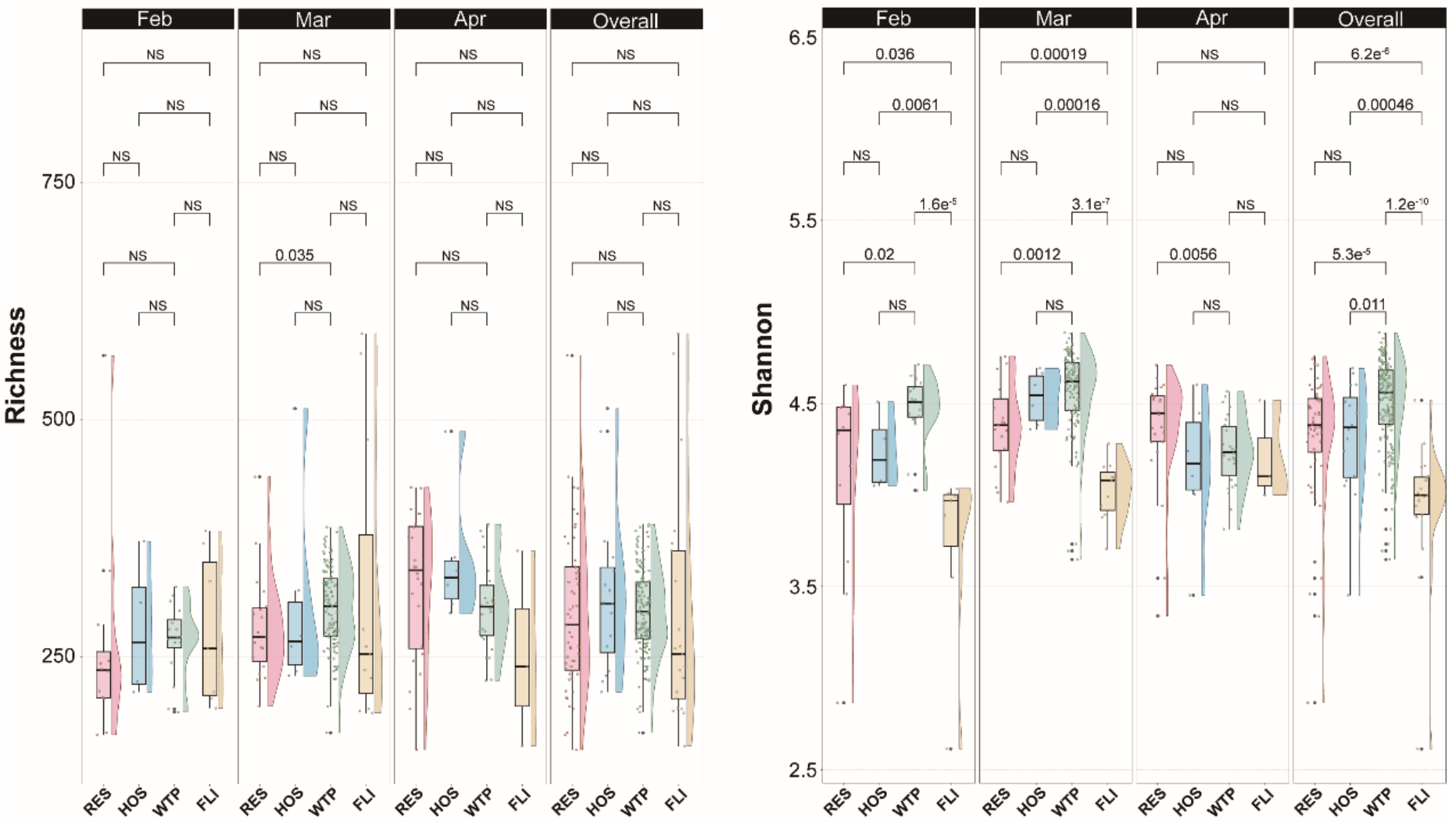
Supplementary Fig. 6. Forest plot illustrating the differential abundance of microbial community at the family level between non-RES and RES samples. Log2-transformed coefficients represent effect sizes, with symbol color corresponding to sampling types. Vertical error bars denote 95% confidence intervals for each significant pathway.



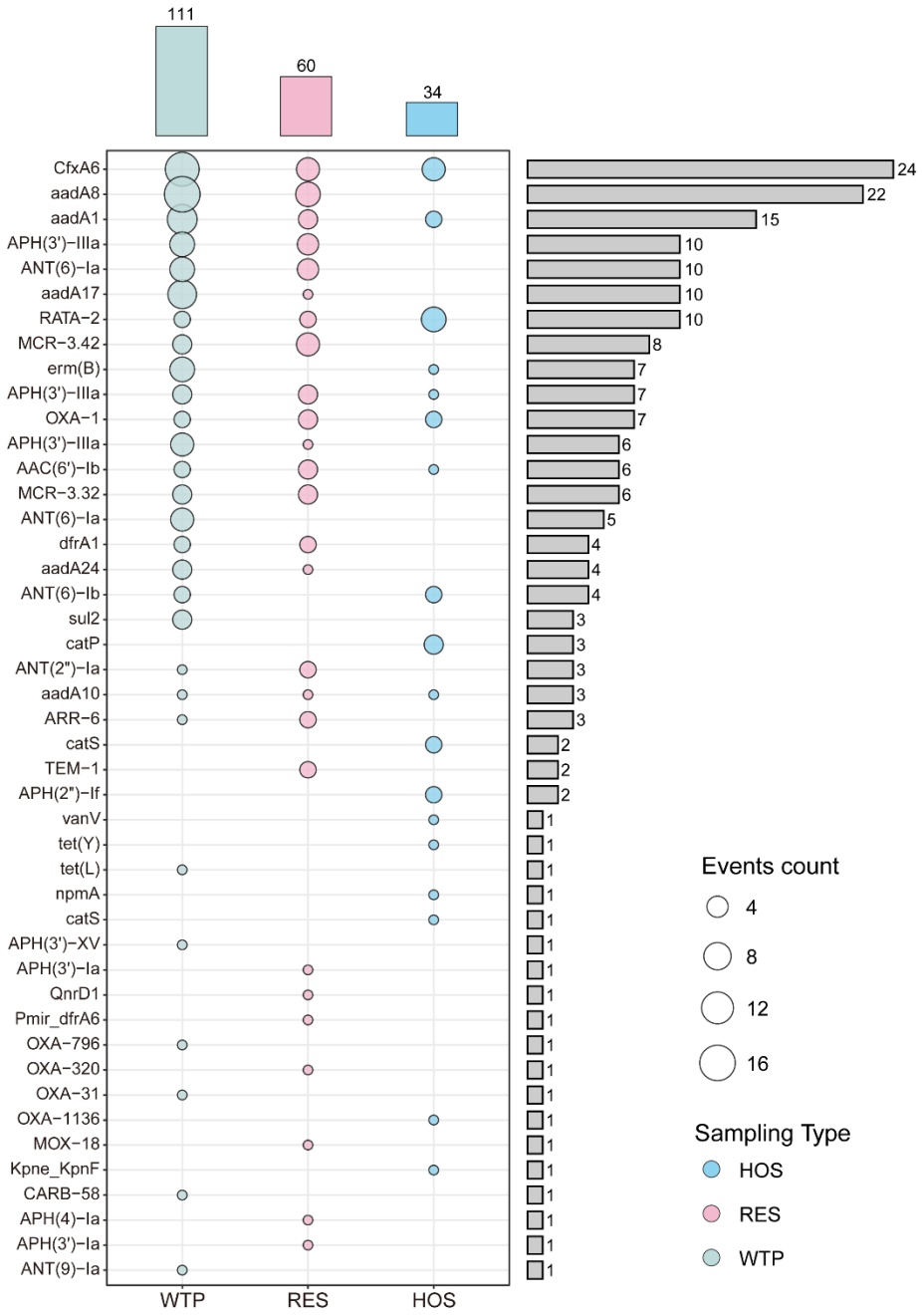
Supplementary Fig. 7. Heatmap profiling relative abundance (log10-scale RPM) of common pathogenic bacteria (https://www.msdmanuals.cn/professional/multimedia/table/classification-of-common-pathogenic-bacteria). The samples (x-axis) are annotated in the top. The first bar annotated the collection time of samples in the top, while the second bar maps the type of samples. The type of microbiome (y-axis) is directly specified on the right. See the source data for detailed information.



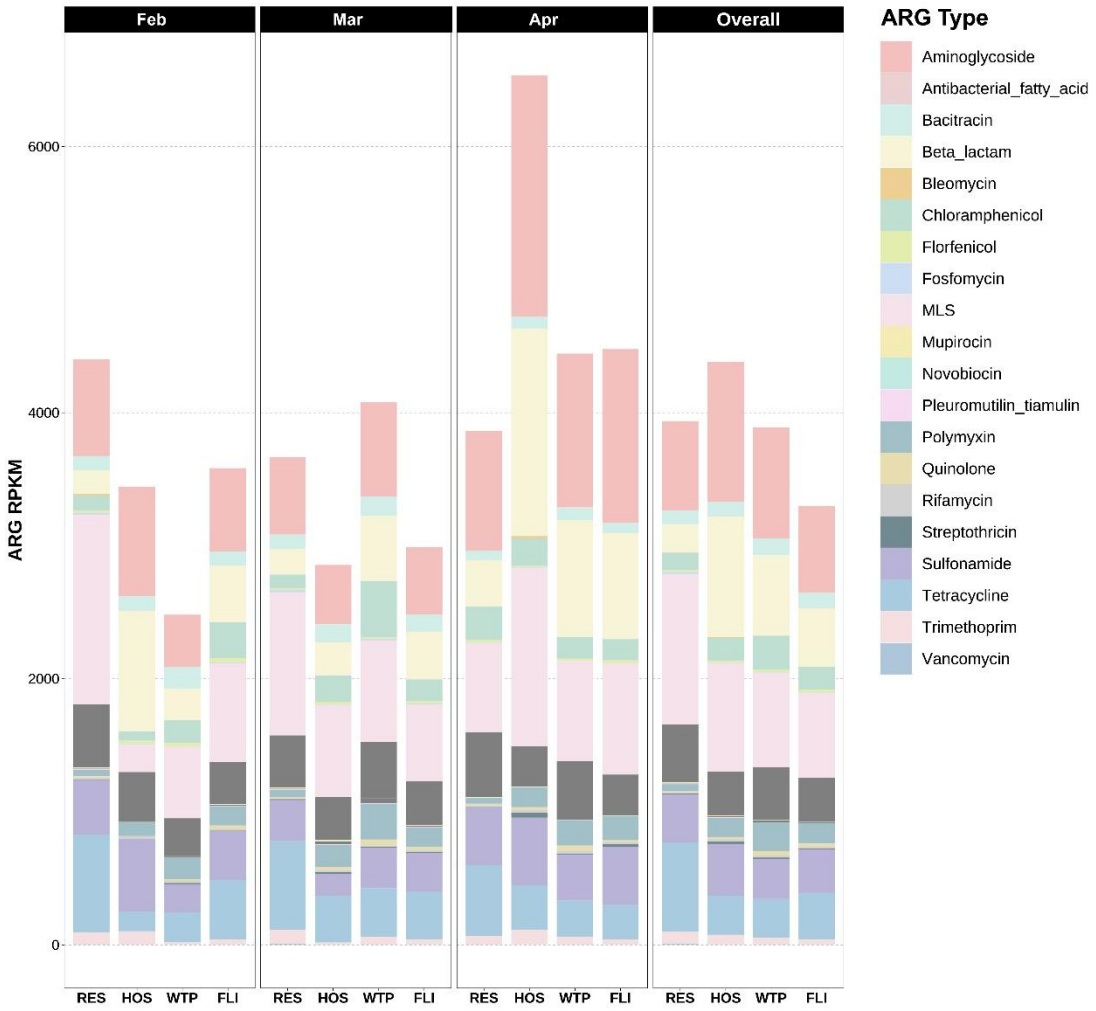
Supplementary Fig. 8. Forest plot illustrating differential abundance of microbial pathways between non-RES and RES samples. Log2-transformed coefficients represent effect sizes, with symbol color corresponding to sampling types. Vertical error bars denote 95% confidence intervals for each significant pathway.


Feb
Mar
Apr
Overall
Richness
750
500
250
NS
0.035
RES
HOS
WTP
FLI
Shannon
6.5
5.5
4.5
3.5
2.5
0.036
0.0061
1.6e-5
0.02
0.00019
0.00016
3.1e-7
0.0012
0.0056
6.2e-6
0.00046
1.2e-10
5.3e-5
0.011

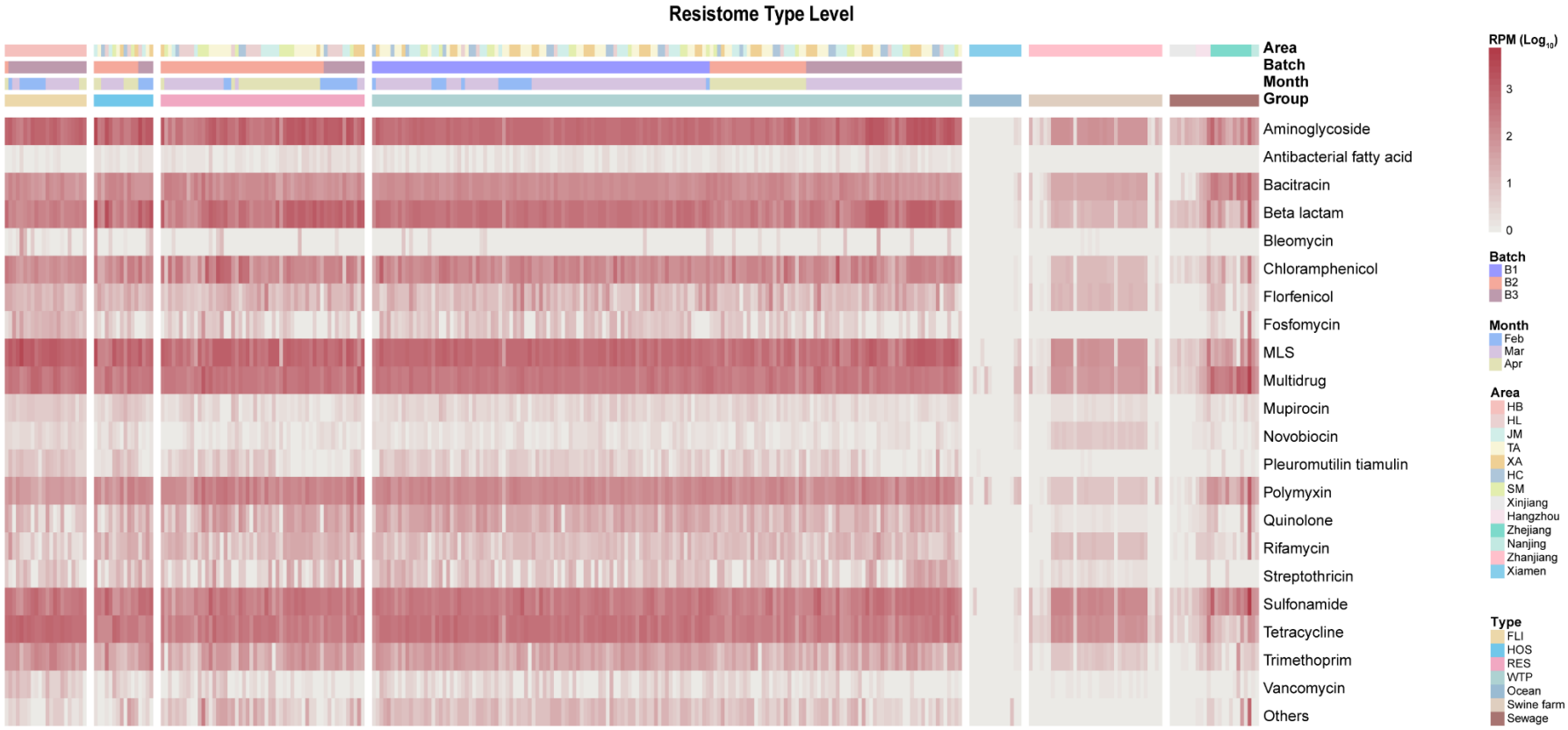
Supplementary Fig. 9. Subtype-level richness (left panel) and Shannon diversity (right panel) of ARGs across four sample types over the sampling period. In each box plot, the center line represents the median; the lower and upper bounds of the box indicate the 25th and 75th percentiles (first and third quartiles), respectively; the whiskers extend to the minima and maxima, excluding outliers defined as 1.5 times the interquartile range. Individual data points correspond to distinct samples. Statistical significance was determined by two-tailed Wilcoxon rank-sum tests with Benjamini-Hochberg FDR correction ($p < 0.05$ considered significant; NS: $p > 0.05$).



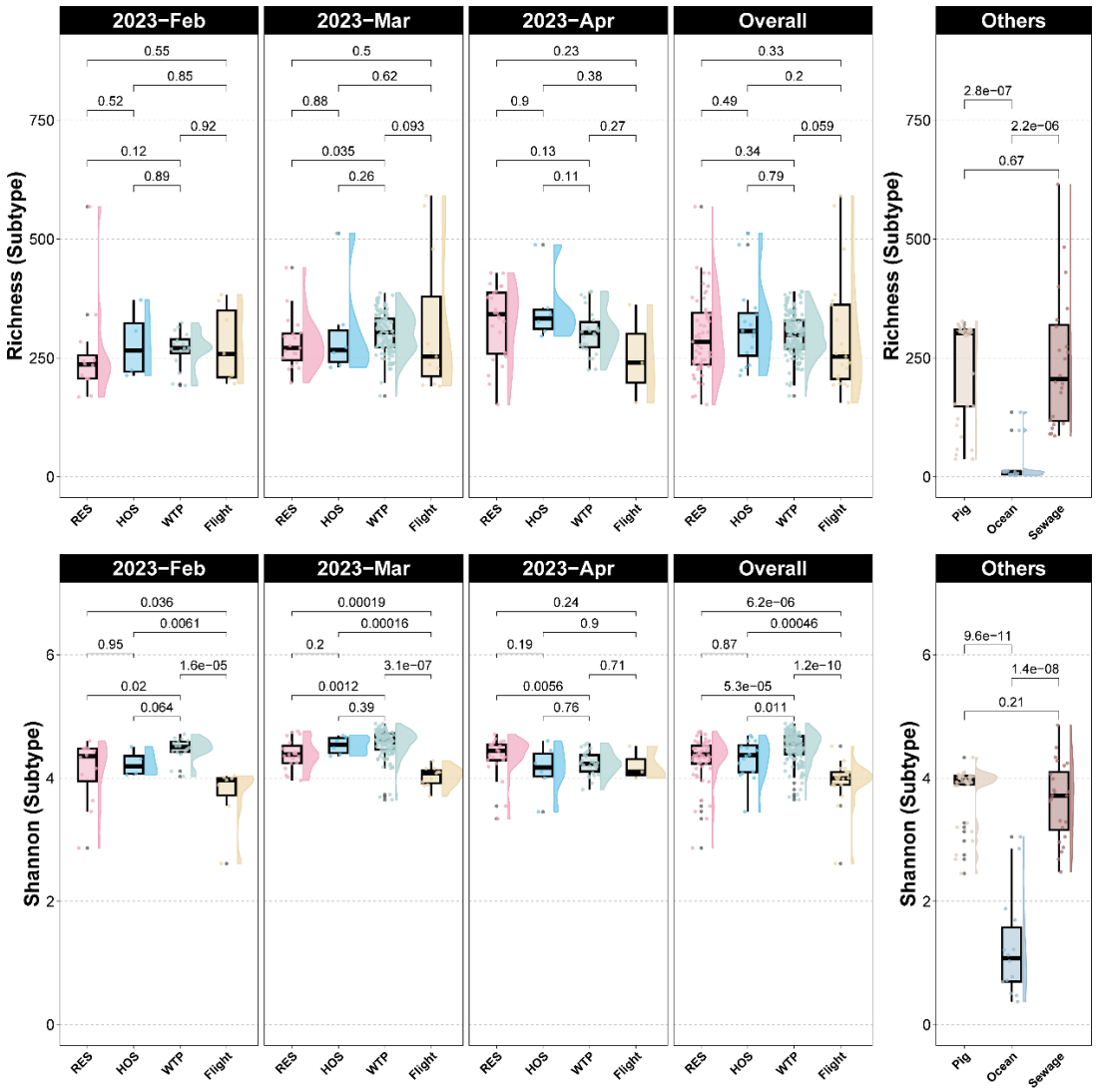
Supplementary Fig. 10. Distribution and prevalence of antibiotic resistance genes (ARGs) across different sampling environments. The bubble plot (left) illustrates the occurrence and frequency of specific ARGs detected across WTP, RES, and HOS. The size of each bubble corresponds to the event count, representing the frequency of detection of a specific gene within each sampling type. The corresponding bar chart (right) quantifies the total cumulative detection frequency for each ARG, with numerical values indicating the absolute count. Color coding distinguishes the sampling types.



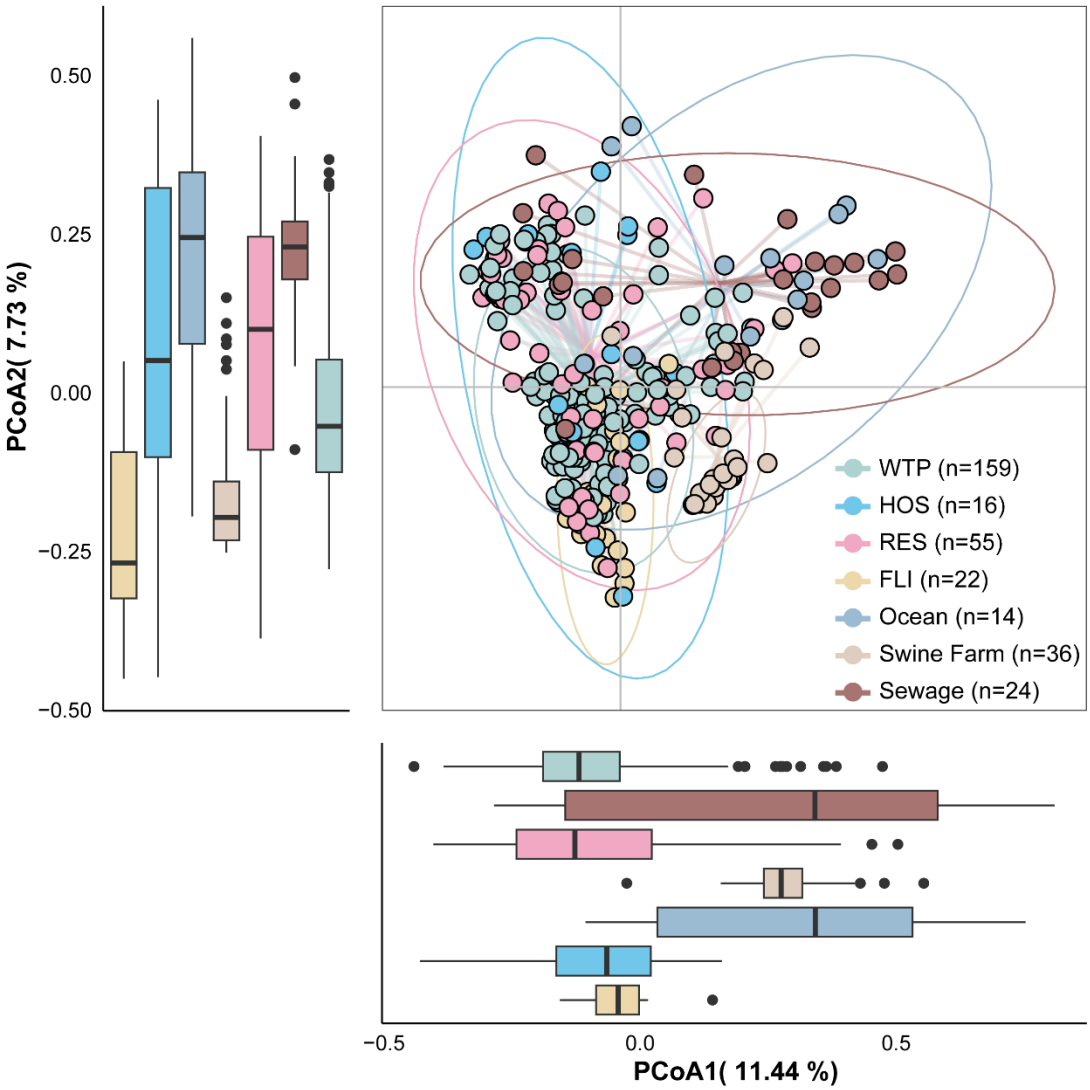
Supplementary Fig. 11. Stack plot illustrating the absolute abundance of ARG in each sampling type across months.



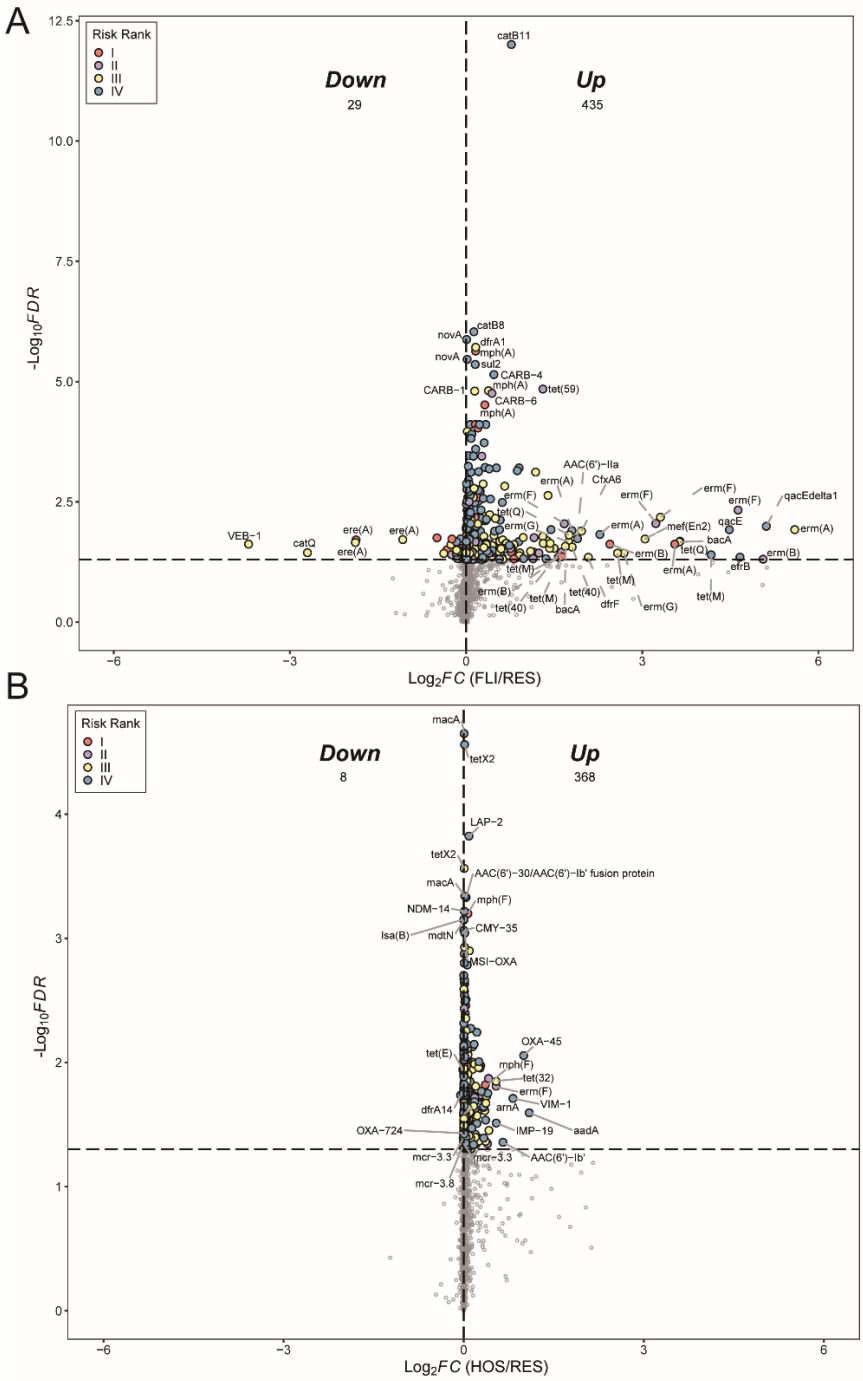
Supplementary Fig. 12. Heatmap profiling relative abundance (log10-scale RPM) of total resistome. The samples (x-axis) are annotated in the top. The first bar annotated the collection area of samples in the top, the second bar maps the collection time of samples, the third bar indicates the sequencing batch of samples, and the fourth bar shows the type of samples. The type of resistome (y-axis) is directly specified on the right. See the source data for detailed information.



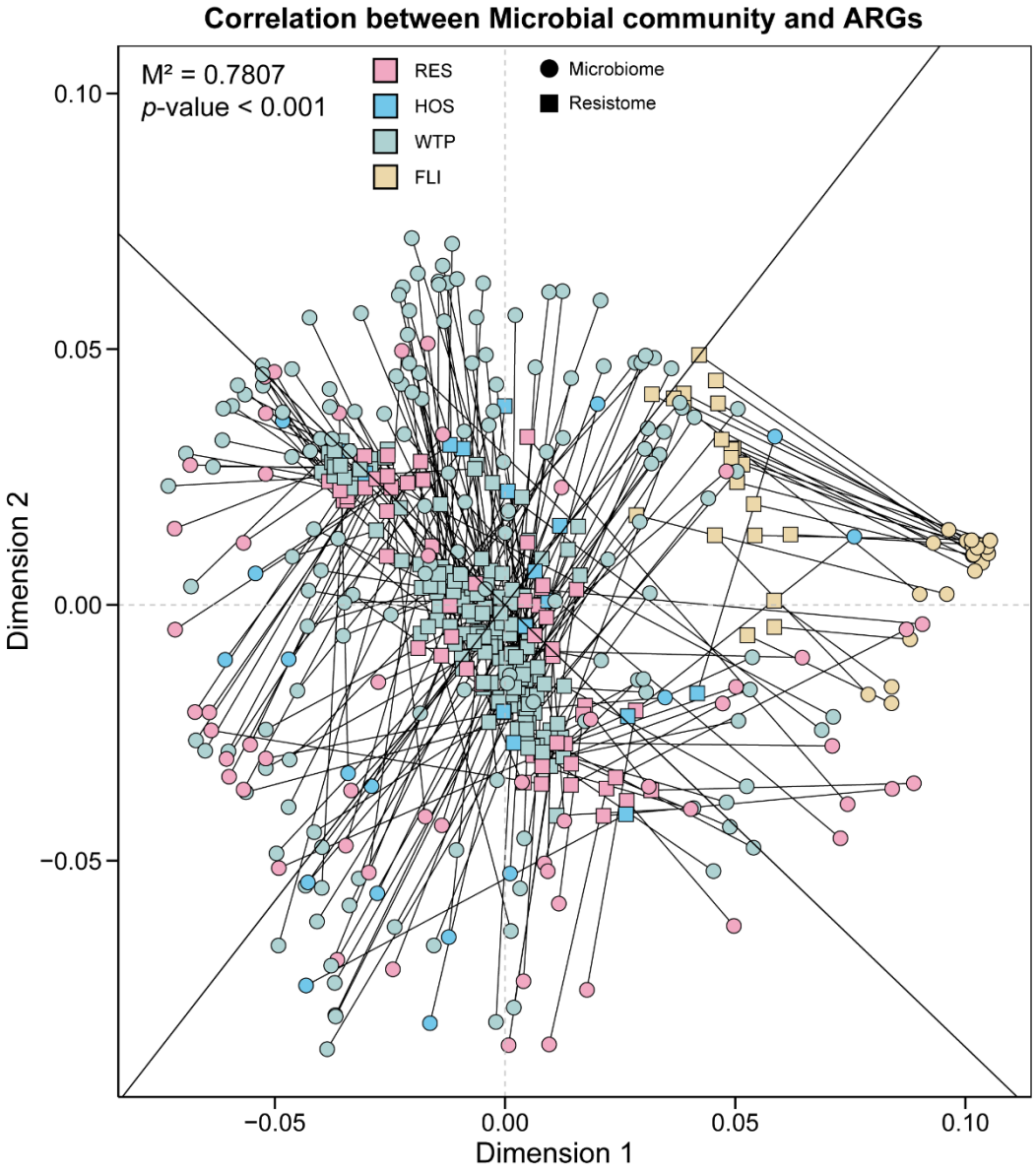
Supplementary Fig. 13. Boxplot showing the difference of Richness (Left panel) and Shannon diversity (Right panel) between this study (WTP, n = 159; RES, n = 55; HOS, n = 16; FLI, n = 22) and several studies concerning wastewater in swine farms (n = 36), the ocean near Xiamen (n = 14), and sewage of other cities in China (n = 24). In each box plot, the center line represents the median; the lower and upper bounds of the box indicate the 25th and 75th percentiles (first and third quartiles), respectively; the whiskers extend to the minima and maxima, excluding outliers defined as 1.5 times the interquartile range. Statistical significance was determined by two-tailed Wilcoxon rank-sum tests with Benjamini-Hochberg FDR correction ($p < 0.05$ considered significant; NS: $p > 0.05$).



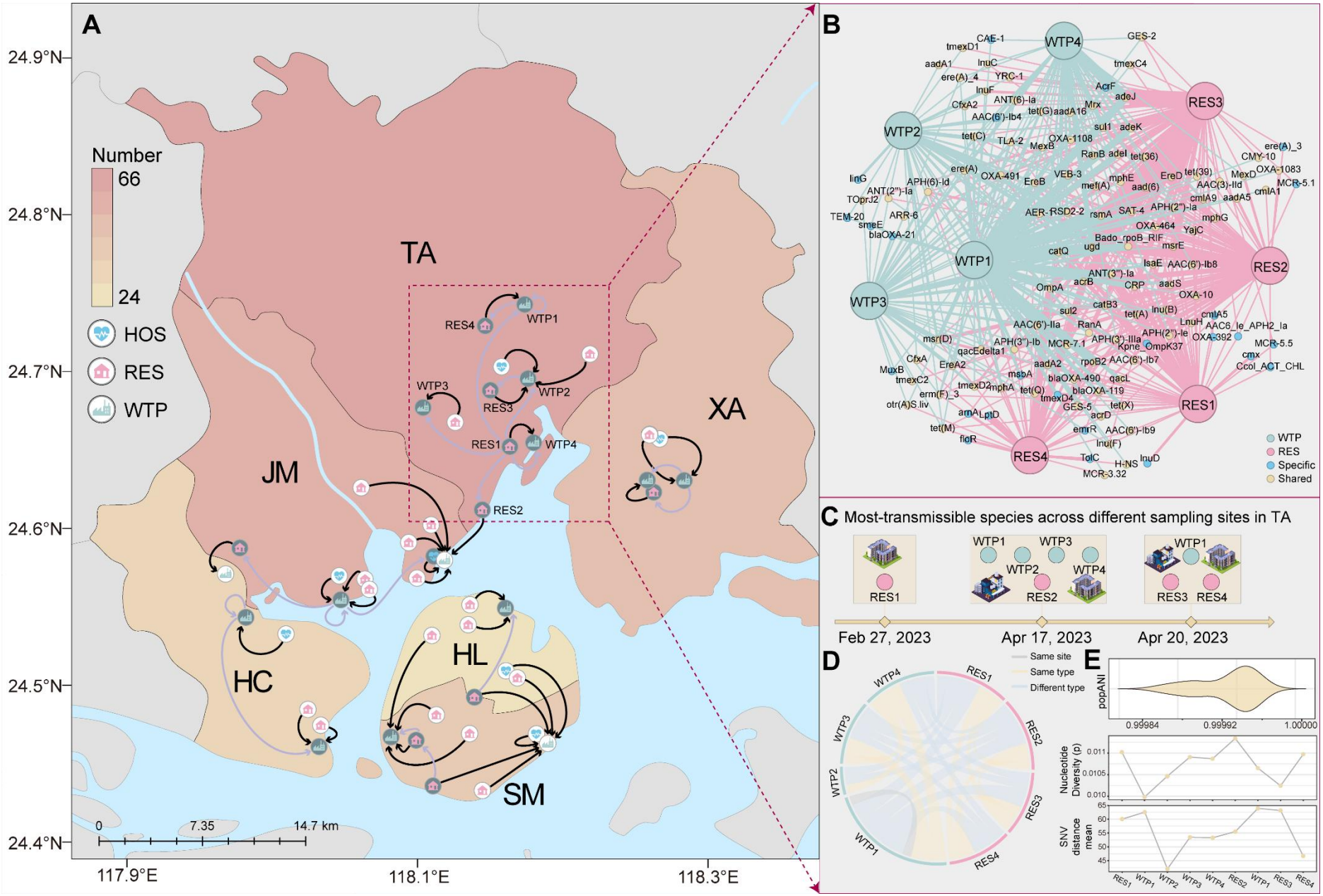
Supplementary Fig. 14. PCoA ordination based on Bray-Curtis dissimilarity matrices of ARG composition of samples in both this study and several studies concerning wastewater in swine farms, the ocean near Xiamen, and sewage of other cities in China. Boxplots represent the average Bray-Curtis distances of each sampling type. In each boxplot, the center line represents the median; the lower and upper bounds of the box indicate the 25th and 75th percentiles (first and third quartiles), respectively; the whiskers extend to the minima and maxima, excluding outliers defined as 1.5 times the interquartile range.



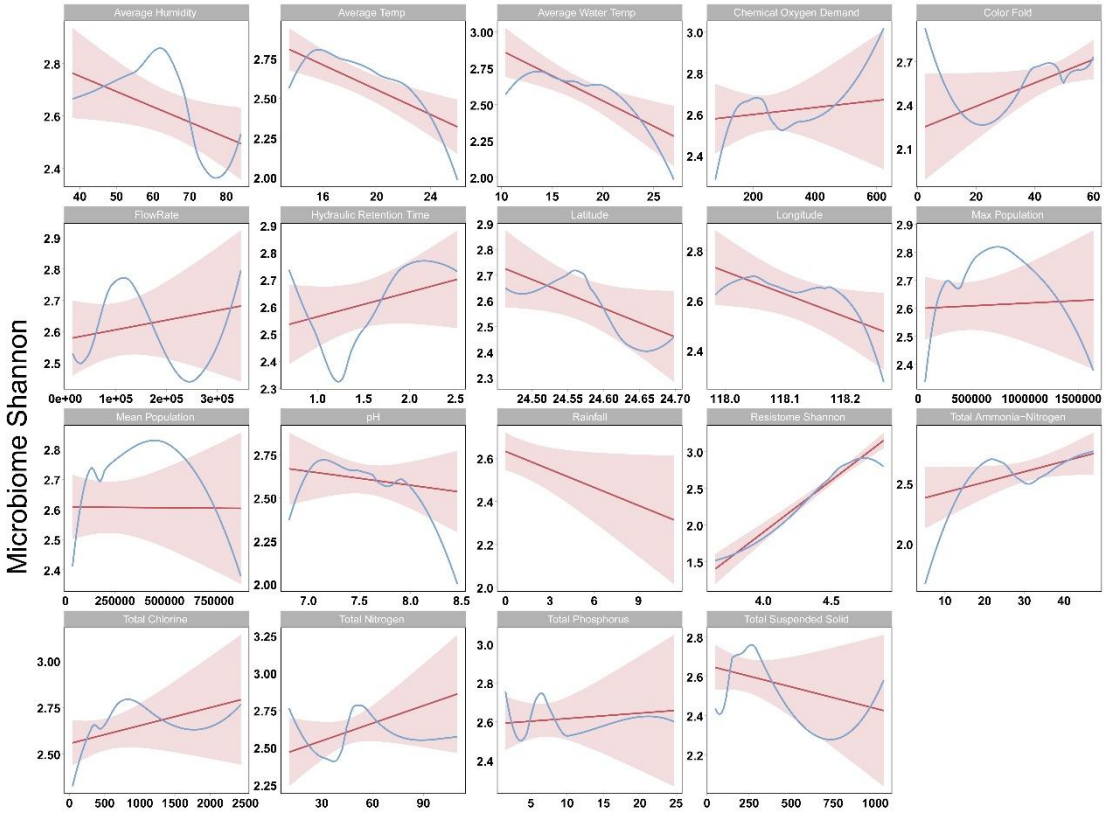
Supplementary Fig. 15. Volcano plots showing the differential enrichment of ARGs in (A) FLI and (B) HOS relative to RES. Each point represents an individual ARG, colored according to its SARG-based risk rank (I–IV). The x-axis indicates the $\log_2$fold change (FLI/RES or HOS/RES), and the y-axis denotes statistical significance ($-\log_{10}$FDR). Horizontal dashed lines mark the thresholds for significant change (FDR < 0.05).



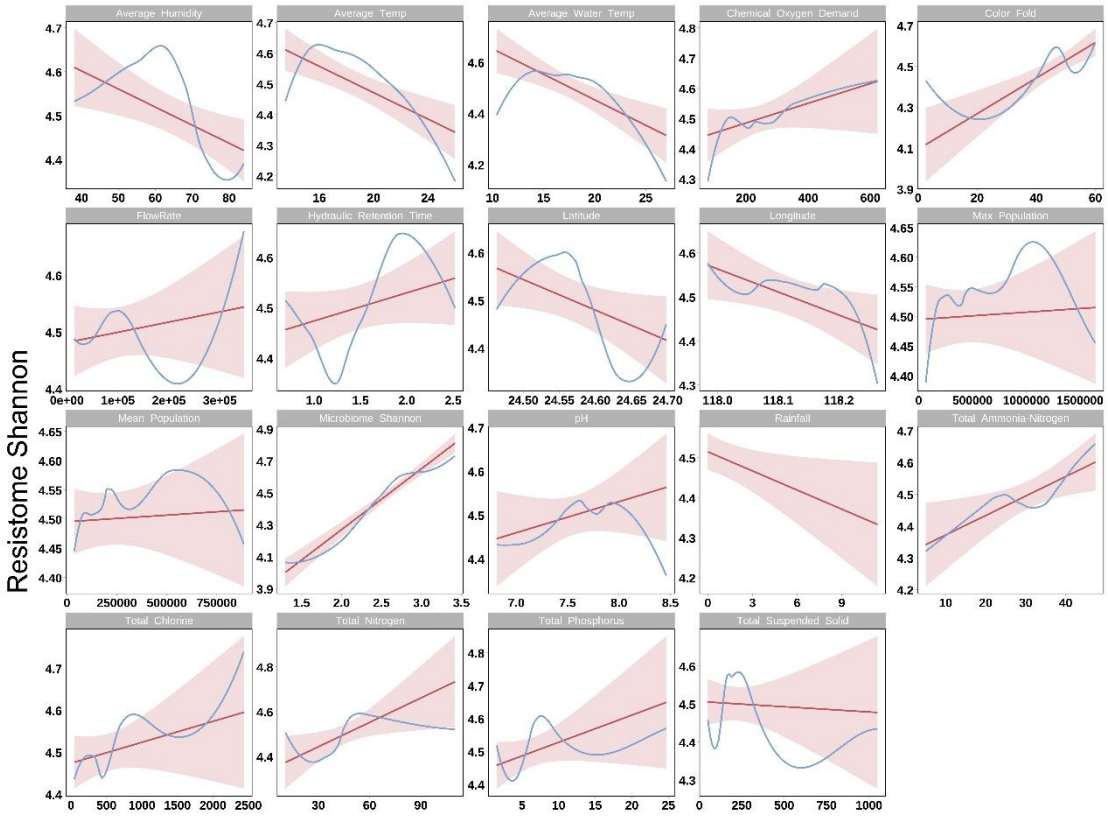
Supplementary Fig. 16. Procrustes analysis comparing microbial community structures and ARG architecture. Edges connect paired samples (n = 252), with edge length representing mismatch magnitude (M^2 = 0.7807). Significant concordance was observed (p-value < 0.001 via 999 permutations). Point colors correspond to sampling type, while shapes are differentiated based on their origination from either the microbiome or resistome.


A
24.9°N
24.8°N
24.7°N
24.6°N
24.5°N
24.4°N
117.9°E
118.1°E
118.3°E
Number
66
24
HOS
RES
WTP
TA
JM
HC
HL
SM
XA
RES4
WTP1
WTP3
RES3
WTP2
RES1
WTP4
RES2
0
7.35
14.7 km
B
WTP4
RES3
WTP2
RES2
WTP3
WTP1
RES1
RES4
WTP
RES
Specific
Shared
C Most-transmissible species across different sampling sites in TA
RES1
WTP1
WTP3
WTP2
WTP4
RES2
WTP1
RES3
RES4
Feb 27, 2023
Apr 17, 2023
Apr 20, 2023
D
Same site
Same type
Different type
WTP4
RES1
WTP3
RES2
WTP2
RES3
WTP1
RES4
E
popANI
0.99984
0.99992
1.00000
Nucleotide Diversity (p)
0.011
0.0105
0.010
SNV distance mean
65
60
55
50
45
RES1
WTP1
WTP2
WTP3
WTP4
RES2
WTP1
RES3
RES4

Supplementary Figure 17. Spatiotemporal resistome and microbiome connectivity within the TA district. **A.** Map illustrating sampling locations with color gradients corresponding to site-specific sampling intensity. All sites where *Aliarcobacter cryaerophilus* was detected are shaded in dark green. Black arrows symbolize directional connectivity between sampling nodes, while purple arrows represent the temporal sequence of transmission events of *Arcobacter cryaerophilus* within the same district. The map was generated using ArcGIS Pro 3.1.5 (Esri, Redlands, CA, USA). Sampling locations and facility types were visualized using built-in symbol libraries in ArcGIS Pro. **B.** The co-occurrence network illustrates the exchange of ARGs across different sewage matrices, with color coding distinguishing shared and type-specific ARGs. **C.** Temporal timeline identifies the chronological order of detection at all involved sites. The vector assets utilized here were designed by Magnific (https://www.magnific.com/). **D.** The chord diagram quantifies the frequency of transmission events occurring among these specific locations. **E.** The genomic characteristics of the prevalent strains and the intra-population genetic variation.



Supplementary Fig. 18. GLM diagnostic plots for environmental factors affecting microbial diversity. Each panel shows the relationship between the response variable (Microbiome Shannon Index) and a predictor variable. The blue line illustrates the original data distribution, the red line represents the fitted linear model, and the shaded area indicates the 95% confidence interval. The x-axis labels denote different environmental factors, and the y-axis represents the microbiome Shannon Index values.



Supplementary Fig. 19. GLM diagnostic plots for environmental factors affecting ARG diversity. Each panel shows the relationship between the response variable (Resistome Shannon Index) and a predictor variable. The blue line illustrates the original data distribution, the red line represents the fitted linear model, and the shaded area indicates the 95% confidence interval. The x-axis labels denote different environmental factors, and the y-axis represents the resistome Shannon Index values.

Supplementary Notes 1. Definition of environmental factor

1. Average Humidity: The mean relative humidity of the ambient air near the influent sampling site.
2. Average Water Temp: The mean temperature of the incoming wastewater.
3. Average Temp: The average ambient air temperature near the influent sampling location.
4. Chemical Oxygen Demand: Indicates the total quantity of oxygen required to oxidize both organic and inorganic compounds in the influent chemically.
5. Color Fold: A measure of the color intensity of the influent relative to a standard reference.
6. Flow Rate: The volume of raw wastewater entering the treatment facility per unit of time.
7. Hydraulic Retention Time: The average time wastewater remains in the primary treatment or equalization tank, based on influent flow and tank volume.
8. Latitude: The geographic coordinate (north–south) of the influent monitoring site.
9. Longitude: The geographic coordinate (east–west) of the influent monitoring site.
10. Max Population: The maximum number of people contributing wastewater to the treatment system during the observation period.
11. Mean Population: The average number of people contributing wastewater to the treatment system during the observation period.
12. pH: Indicates the degree of acidity or alkalinity of the incoming wastewater.
13. Rainfall: The total precipitation near the service area during the monitoring period.
14. Total Ammonia-Nitrogen: The total concentration of ammonia (NH_3) and ammonium (NH_4^+) in the influent.

15. Total Chlorine: The total concentration of chlorine compounds in the influent.
16. Total Nitrogen: The total nitrogen content in all forms (organic nitrogen, ammonia, nitrite, and nitrate) in the influent.
17. Total Phosphorus: The total phosphorus concentration (organic and inorganic) in the influent.
18. Total Suspended Solids: The total mass of solid particles suspended in the incoming wastewater.