The Innovation, Volume *5*

Supplemental Information

Ecology and risks of the global plastisphere as a newly expanding microbial habitat

Changchao Li, Michael R. Gillings, Chao Zhang, Qinglin Chen, Dong Zhu, Jie Wang, Kankan Zhao, Qicheng Xu, Polly Hangmei Leung, Xiangdong Li, Jian Liu, and Ling Jin

This PDF file includes:

Other Supplemental Materials for this manuscript include the following:

Tables S1 to S29 are in a separate excel file (captions listed below).

Table S1 Metadata source.

Table S2 Sample information.

Table S3 Relative abundance of genera.

Table S4 Design of studies investigating microbial information of both the plastisphere and biofilms from other substrates.

Table S5 Genus relative abundance of the plastisphere and biofilms from other substrates.

Table S6 Taxonomic composition.

Table S7 Relative abundance of families.

Table S8 The importance of each microbial family on the accuracy of the random forest model in each ecosystem.

Table S9 Relative abundance of the plastisphere biomarkers in each ecosystem.

Table S10 Source analysis of microorganisms in the plastisphere based on the fast expectation-maximization for microbial source tracking (FEAST).

Table S11 Habitat niche breadth.

Table S12 The modified stochasticity radio (MST).

Table S13 Meta-network links.

Table S14 Meta-network nodes.

Table S15 Relative abundance of nodes in the meta-network.

Table S16 Sub-network links.

Table S17 Sub-network nodes.

Table S18 Sub-network indexes.

Table S19 Robustness of sub-networks in the plastisphere and the natural environment in each ecosystem.

Table S20 Relative abundance of ecologically relevant functions.

Table S21 Comparison of ecologically relevant functions between the plastisphere and the natural environment.

Table S22 Relative abundance of functional genes based on the metagenomic data (logTPM).

Table S23 Detected animal, plant and zoonotic pathogens in samples based on the MBPD database.

Table S24 Total abundance of animal, plant and zoonotic pathogens in each sample.

Table S25 Relative abundance of clinical pathogens detected based on the 16SPIP tool.

Table S26 Total abundance of clinical pathogens in each sample detected based on the 16SPIP tool.

MATERIALS AND METHODS

Field sampling strategy and processing of samples

We collected plastic samples and associated water samples from three freshwater bodies (the Wulong River, the Moshui River, and the Dagu River) and three seawater areas (Dingzi Bay, Southwest of Jiaozhou Bay, and Northeast of Jiaozhou Bay) in Qingdao and Yantai of Shandong province, China, during the month of September 2020. A Manta trawl (333 μm) was used to capture the plastic debris. A plastic sample was obtained in every 30 minutes of capture. Plastic debris trapped in the trawl were placed in a 50-mL centrifuge tube. Simultaneously, 2 L of surface water were collected in a sterile glass bottle. All of the samples were immediately placed on dry ice. A total of 36 plastic samples and 36 bulk water samples were obtained during the field sampling. The aim of this study was to reveal microbial ecological patterns and associated risks in the plastisphere, a huge and expanding man-made ecosystem, generated by environmental plastics. And it is generally accepted that the size of plastic debris does not significantly affect the structure of its residents as long as it is not so small as to affect the colonization of microorganisms.[1-4] Furthermore, if the plastic is too small to form a biofilm, *e.g.*, nanoplastics, it cannot form a plastisphere but instead forms an eco-corona,[5] which is beyond the aim and scope of our study. Therefore, the plastic debris included in this study include microplastics, mesoplastics, and macroplastics, and our findings and conclusions are generalizable for plastisphere research.

Each water sample was vacuum filtered successively through an 80–120 μm quantitative filter (to remove interfering substances) and a 0.22 μm membrane filter (to collect microorganisms). DNA was extracted from cells retained on the 0.22 μm filters and from plastic debris using a cetyltrimethylammonium bromide (CTAB) method.[6,7] A portion of the 16S rRNA gene was amplified with primer pairs of 515F and 806R, and subsequently sequenced to obtain 2 × 250 bp paired-end reads using an Illumina Novaseq 6000 platform.

Metadata collection and data preprocessing

To expand our view of the ecological patterns and threats associated with the plastisphere, we made an extensive effort to collect data from publications and bioprojects that used high-throughput sequencing to examine bacterial populations in the plastisphere. We first reorganized the dataset created by Wright et al. [8] in January 2020, which included 35 studies from the Web of Science Core Collection and Science Direct. Then, in October 2021 we obtained another 25 studies by searching NCBI using the search term “plastisphere”. To filter the metadata, the following criteria were applied: (i) The plastisphere had been collected or incubated in freshwater, seawater, or terrestrial ecosystems; (ii) the raw sequence data were available; and (iii) the sample information was clear, or could be obtained from the corresponding authors. Sequences satisfying all of these criteria were downloaded.

Paired-end sequences were joined, primer-cut, and quality-filtered in each project (with our own field-collected samples also treated as a project) using USEARCH [9] and VSEARCH [10]. Then, the sequences of all of the projects were combined into one file for subsequent analysis. Since the dataset was composed of thousands of samples with complex sources, it was more appropriate in this study to cluster the sequences as OTUs with a 97% similarity threshold, in order to avoid overestimations of diversity. OTUs were mapped to the RDP database to remove sequences generated from chimera, mitochondria, and chloroplasts. Then, an OTU table was generated using USEARCH. The taxonomic identity of representative sequences was annotated with the RDP classifier [11].

We first obtained 2,035 microbial samples, including plastisphere samples, the associated environmental samples, and biofilm samples on other substrates, from these plastisphere studies performed around the globe. To minimize the effect of different sequencing depths, samples with < 2,000 reads were removed, and all subsequent analyses were performed based on relative abundance. To address the effect of different sequencing regions in different studies, data at the genus level were used for all potentially affected analyses. To avoid data bias, 80 samples were randomly selected if the number of plastisphere or environmental samples in one study was greater than 80. Further, the number of seawater plastisphere samples was large compared with other subgroups, so 300 seawater plastisphere samples were randomly selected to avoid data bias. After carrying out the above data-trimming processes, we finally obtained a total of 1,192 samples from 35 bioprojects, including 143 freshwater-environment samples, 120 freshwater-plastisphere samples, 132 seawater-environment samples, 300 seawater-plastisphere samples, 148 terrestrial-environment samples, 170 terrestrial-plastisphere samples, and 179 biofilm samples from other substrates such as glass, natural seston, and plant leaves. The starting point of this study was that the plastisphere was a new microbial habitat with a vast and expanding area. This study aimed to clarify the differences in the microbial ecology between this new habitat and natural habitats, and to reveal the associated ecological threats. Therefore, we focused on the analysis of the microbial communities from the plastisphere and the natural environment (see Tables S1 and S2 for sample sources, Figure 1A for sample distributions, and Table S3 for the abundance of genera). The comparison of the plastisphere with biofilms from other substrates is presented in Result S1, and the result shows that the plastisphere is indeed a unique ecological niche that differ from other substrates significantly (Result S1, Figure S4, and Tables S4 and S5).

Microbial community structure analysis

Factors such as different sample handling, different primers, and different sequencing platforms potentially influence the microbial information of samples. To demonstrate the robustness of our findings and the fundamental difference between the plastisphere and natural environments, CCA was carried out and the relative importance of potential drivers of compositional variations in the global meta-community were quantified. These potential drivers included the ecosystem identity (*i.e.*, the freshwater ecosystem, the seawater ecosystem, and the terrestrial ecosystem), the carrier identity (*i.e.*, the plastisphere and the natural environment), the location latitude, and the study ID. Consistent with the approach applied in a Earth Microbiome Project study for revealing multi-scale microbial diversity on Earth,[12] we used the study ID as a proxy for a wide range of other potential drivers because the explanation of the variation in the meta-community composition by the study ID covered the explanation by factors like different research methods in different studies. The three categorical variables, the ecosystem identity, the carrier identity, and the study ID, needed to be converted to dummy variables for CCA and relative importance calculations. Since the carrier identity contained only two categories (the plastisphere and the environment), we replaced them directly with 0 and 1. The ecosystem identity contained three categories (the freshwater ecosystem, the seawater ecosystem, and the terrestrial ecosystem), and three different distance relationships might occur between the three groups, which we used “1,2,3”, “1,3,2”, and “2,1,3” to replace the three ecosystem IDs, respectively. For the study ID, we performed random permutations of the study IDs and then replaced the IDs with numerical values and repeated the process for 99 times. Then, with the rdacca.hp package,[13] we performed CCA analysis using the replaced variables and computed the explanation of each potential driver to the meta-community structural variation for a total of 3 × 99 = 297 times. Finally, we obtained the importance ranking of the drivers using the scores of each driver derived from the 297 calculations.

An unconstrained PCoA based on Bray-Curtis distance was carried out to analyze differences in microbial community structure between the plastisphere and the natural environment, and between different ecosystems. A PERMANOVA was used to test the statistical significance of the difference. A linear regression model between community similarity (1 – Bray-Curtis dissimilarity) and geographic distance was implemented to explore the distance-decay pattern of microbial communities in the plastisphere and the natural environment. A Wilcoxon rank sum test was used to compare the similarity in communities between the plastisphere and the natural environment. The FEAST tool [14] was used to quantify the impact of the natural environment as well as the traits of the plastisphere itself on the structure of the plastisphere microbial community. To avoid data bias due to sample size, 100 samples were randomly selected from each potential source for a FEAST analysis.

Community assembly mechanism

To reveal the community assembly mechanisms underlying microbial ecological patterns, including community structure and diversity, we computed the ecological niche breadth and the modified stochasticity ratio (MST). Habitat niche breadth is a key feature that influences species sorting and dispersal limitation in community assembly processes.[15] Microbiota with wider niches are usually more metabolically flexible at the community level, implying less influence from environmental filtering.[16] Using Levins' niche breadth index,[17] we estimated the habitat niche breadth of each genus in a metacommunity and then evaluated the community-level niche breadth by calculating the average habitat niche breadth of all taxa present in the community. The MST based on a null model is usually used to quantify the relative importance of stochasticity and determinism in the community assembly process. The MST model reflects the community assembly process by relative difference, rather than by the significance of the difference between the observed situation and the null expectation, and therefore provides a better quantitative measure of the stochasticity in community assembly.[18,19] The values of MST range from 0 to 1, with MST = 0 representing completely deterministic assembly and MST = 1 representing completely stochastic assembly, with 0.5 as the boundary defining deterministic (MST < 0.5) or stochastic (MST > 0.5) dominated assembly processes.

Ecological network construction and analysis

A meta-network was constructed to explore co-occurrence patterns of the global microbial meta-community. Genera with a relative abundance of > 0.001% and occurring in more than 60 samples were selected for a correlation calculation. Spearman's rank correlations were computed using the Benjamini-Hochberg method for multiple-testing-correction. Links with Spearman's $\rho \leq 0.4$ or P-value ≥ 0.05 were discarded. Further, we constructed ecological sub-networks of the plastisphere and the natural environment in each ecosystem to reveal the difference in the co-occurrence pattern between the plastisphere and the natural environment. To avoid data bias caused by sample size, 100 samples in the plastisphere or the natural environment subgroup in each ecosystem were randomly selected, and genera occurring in more than 10 samples were selected to construct the sub-networks based on the Spearman's correlation with the Benjamini-Hochberg correction method. Since the number of samples in the sub-datasets was much smaller than that in the meta-dataset, more stringent criteria were used when selecting the links used to build the sub-networks. Only links with Spearman's $\rho \geq 0.6$ and P-value ≤ 0.05 were chosen for the further construction of sub-networks. Node properties, module partition, and topological characteristics were analyzed using the igraph package. The small-world property of the network was tested using the power-law model with a good fit representing a scale-free and non-random network. To compare the robustness of the ecological networks in the plastisphere and the natural

environment, we further calculated the average degree and the natural connectivity after nodes were randomly removed to simulate species extinction.

Ecologically functional signatures

The FAPROTAX platform v.1.2.3 [20] was used to extrapolate the functional potential of the plastisphere. FAPROTAX is a tool that maps prokaryotic taxa to their corresponding metabolically or ecologically relevant functions based on current literature on cultured strains.[20] Unconstrained PCoA with PERMANOVA was carried out to test the difference in the overall functional signatures between the plastisphere and the natural environments. We extracted ecologically important functions involved in nitrogen metabolism and organic compound metabolism and examined the difference in potentials of these functions between the plastisphere and natural environments using the Wilcoxon rank sum test with false discovery rate (FDR) correction. Furthermore, using the usearch_global command in VSEARCH, we mapped our sequences to PlasticDB [21] and mibPOPdb [22] databases to evaluate the functional potential of plastic biodegradation and POP biodegradation in the plastisphere and the natural environment.

Pathogenic risks

The MBPD database is a newly established, specialized, large, and curated database for the monitoring of animal, plant, and zoonotic pathogens in biological and environmental samples under the “One Health” vision.[23] We annotated potential animal, plant, and zoonotic pathogens in our samples by aligning our sequences to the MBPD database with the usearch_global command. The 16SPIP pipeline is a comprehensive tool for rapid pathogen detection in clinical samples and also widely applied in environmental samples.[24-27] Using the 16SPIP pipeline, we further explored potential human pathogens in the plastisphere and in the natural environment. Moreover, we mapped our sequences to the Fish Pathogen Database [28] to specifically identify potential fish pathogens in the samples and reveal the threat from the plastisphere to fish health. The abundance of identified potential pathogens in the plastisphere and the natural environment was compared using the Wilcoxon rank sum test with FDR correction.

Metagenomic sample collection and analysis

To validate the robustness of the functional potential evaluations based on global bacterial communities, we conducted metagenomic-based analyses on our 38 paired-, field-collected plastisphere and natural environmental samples. These samples were also obtained from freshwater, seawater, and terrestrial ecosystems. Seawater plastic debris and the bulk water samples were pair-collected from two sites (120.315° E, 36.255° N and 120.3° E, 36.071° N) in coastal areas across Qingdao, China, in August 2021. At each site, no less than three sample pairs were collected. The sampling method was consistent with that used to collected samples for amplicon sequencing as described before, the Manta-trawl method. A total of seven pairs of plastic and water samples were obtained in the seawater ecosystem. Each water sample was vacuum filtered successively through an 80–120 μm quantitative filter (to remove interfering substances) and a 0.22 μm membrane filter (to collect microorganisms). Total genomic DNA from the collected plastic debris and filters was extracted using the QIAamp DNA Mini Kit (QIAGEN, Germany) according to the manufacturer’s protocol. The extrated DNA from plastic and water samples were shotgun-metagenomic sequenced on the MGISEQ-2000 platform using a pair-end (2 × 150 bp) sequencing strategy. Freshwater plastic debris and the bulk water samples were pair-collected from nine sampling sites along the Huangpu River in Shanghai, China, in October 2021. At each site, the plastic debris for metagenomic sequencing was collected by passing 5 L

of water through a 50 μm mesh sieve, and additional 5 L of water was collected for the detection of the bulk water genome. The detailed methods for sample treatment, DNA extraction and sequencing can be found in our previous publication (to separate samples collected from the freshwater ecosystem and from the seawater ecosystem, one pair of samples collected from the estuary of the Huangpu River included in the previous paper was excluded from this study).[29] For the metagenomic-based investigation of the plastisphere in the terrestrial ecosystem, we employed an *in-situ* incubation strategy in Harbin, China. The microplastics were purchased from Youngling Electromechanical Technology Co. (Shanghai, China). Before incubation, these microplastics were soaked in 1% sodium hypochlorite for 30 min and then washed with sterile water five times to remove the microorganisms inherent in the microplastics. The microplastics were transferred into a nylon mesh bag and then buried in the soil. After eight weeks of incubation, the microplastic samples and the surrounding soil samples were collected. The detailed methods for sample treatment, DNA extraction and sequencing can be found in our previous publication.[30] In our previous study, we performed metagenomic sequencing for ten microplastic samples and three soil samples.[30] Since the plastic and environment samples for metagenomic sequencing in freshwater and seawater ecosystems were obtained using a paired-sampling strategy, three samples were randomly selected from the ten terrestrial plastic samples to balance the sample sizes of the two groups (the plastisphere and the environment). Finally, we obtained a total of 38 metagenomic samples (including nine freshwater-plastisphere samples, nine freshwater-environment samples, seven seawater-plastisphere samples, seven seawater-environment samples, three terrestrial-plastisphere samples, and three terrestrial-environment samples) for the characterization of functional genomes in the plastisphere and the natural environment to support our global sample-based findings on the ecological risks posed by the plastisphere.

Metagenomic raw sequences of each sample were quality-filtered to remove adapters and low-quality sequences with fastp v0.23.2 [31] with default parameters. The filtered sequences were assembled using MEGAHIT v1.2.9 [32]. Assembled contigs with length >500 bp were selected for further analysis. Using Prodigal v2.6.3,[33] open reading frames (ORFs) were predicted from the assembled contigs. All the predicted ORFs were further clustered to generate a non-redundant gene set by employing CD-HIT v4.8.1 [34] at 95% sequence identity with >90% coverage. The filtered reads were mapped to the non-redundant gene set to quantify the relative abundance (transcripts per million, TPM) of each gene in each sample with Salmon v1.10.1 [35]. Specialized functional gene databases including NCycDB,[36] CAZy,[37] PlasticDB,[21] mibPOPdb,[22] and VFDB [38] were employed to identify and quantify the genes encoding for nitrogen cycle-related functions, carbohydrate-active enzymes (CAZymes), plastic biodegradation functions, POP biodegradation functions and bacterial virulence factors, respectively. Non-redundant genes were translated into protein sequences with Seqkit v2.4.0 [39], and then the protein sequences were aligned to the above target functional gene datasets using DIAMOND v 2.1.6 (For the CAZymes annotation, the recommended *e*-value threshold of 1e–102 was adopted, and for other databases, the *e*-value threshold was set as 1e–5).[40]

SUPPLEMENTAL RESULTS

S1 Plastisphere microbial community distinct from other biofilms

Although the aim of this study was to reveal the microbial ecology in a new microbial habitat with a huge and expanding area – the plastisphere, its difference with the natural habitats, and the accompanying ecological threats of the plastisphere, we still tested the compositional difference between the plastisphere microbial community and other natural or unnatural biofilms to further illustrate the distinctiveness of the plastisphere as a microbial habitat. We screened 16 studies [1,30,41-54] from the metadata set that investigated microbial information of both the plastisphere and other biofilms, and obtained 289 plastisphere samples and 179 biofilms samples from other substrates including glass, natural seston, plant leaves, plant litters, tile, aluminium, cardboard, cellulose, and rock (see Tables S4 and S5 for the sample design and compositional information). The unconstrained principal coordinate analysis (PCoA) with the permutational multivariate analysis of variance (PERMANOVA) showed that significant differences existed in the microbial community composition between the plastisphere and other biofilms, both overall ($P < 0.001$) and specifically in each ecosystem ($P < 0.001$; Figure S4), indicating that the plastisphere was indeed a unique ecological niche for microorganisms. The underlying mechanism is that plastics are a persistent, inert, hydrophobic, buoyant, organic, and long-distance transportable substrate, which is distinguished from other natural or unnatural substrates. In addition, the whole area of plastics is huge and expanding with an unabated momentum in the near future, but the size of individual plastics can be small enough to enter into plants, animals, and even humans, which is the starting point of this study to decipher the microbial ecology of the plastisphere.

S2 Plastisphere biomarkers in each ecosystem

To identify a set of microbial features, which could best distinguish the plastisphere from the natural environment in each ecosystem, among numerous microbial taxa with significant difference between the plastisphere and the natural environment, we carried out a random-forest machine-learning model.[55-57] The model was established based on relative abundances of microbial families in the plastisphere and the natural environment in each ecosystem (Table S7) using the randomForest package [58] in R.

In each ecosystem, the model explained >97% of the variation in microbial communities between the plastisphere and the natural environment, showing the reliability of the models and the fundamental difference between the plastisphere and the natural habitats. Ten-fold cross-validation with five repeats was carried out in each ecosystem to evaluate the importance of each microbial feature. The error-rate curves stabilized before the 20 most relevant microbial features were used by the model, so we uniformly selected the top 20 microbial features that were most important for the accuracy of the models to discriminate between the plastisphere and the natural environment as biomarkers of the plastisphere in each ecosystem (Figure S11 and Table S8).

The plastisphere biomarker taxa in the freshwater ecosystem were from 5 phyla (Figure S12A), of which 9 taxa were enriched (namely, *Enterobacteriaceae*, *Rhizobiaceae*, *Burkholderiales incertae sedis*, *Erythrobacteraceae*, *Bacillaceae-1*, *Sphingomonadaceae*, *Bacillales Incertae Sedis XII*, *Halomonadaceae*, and *Xanthomonadaceae*) while 11 taxa were depleted (namely, *Streptomycetaceae*, *Cryomorphaceae*, *Microbacteriaceae*, *Burkholderiaceae*, *Demequinaceae*, *Flammeovirgaceae*, *Sutterellaceae*, *Puniceicoccaceae*, *Chitinophagaceae*, *Flavobacteriaceae*, and *Cyclobacteriaceae*) in the plastisphere (Figure S12B, C and Table S9).

The plastisphere biomarker taxa in the seawater ecosystem were from 4 phyla (Figure S13A), of which 9 taxa were enriched (namely, *Erythrobacteraceae*, *Saprospiraceae*, *Arenicellaceae*, *Rhizobiaceae*, *Hyphomonadaceae*, *Alteromonadaceae*, *Burkholderiaceae*, *Phyllobacteriaceae*, and *Hyphomicrobiaceae*) while 11 taxa were depleted (namely, *SAR11*, *Methylophilaceae*, *Euzebyaceae*, *Cryomorphaceae*, *Pseudomonadaceae*, *Verrucomicrobiaceae*, *Puniceicoccaceae*, *Rhodospirillaceae*, *Flavobacteriaceae*, *Chitinophagaceae*, and *Oceanospirillales incertae sedis*) in the plastisphere (Figure S13B, C and Table S9).

The plastisphere biomarker taxa in the terrestrial ecosystem were from 6 phyla (Figure S14A), of which 10 taxa were enriched (namely, *Pseudomonadaceae*, *Nocardiaceae*, *Burkholderiaceae*, *Moraxellaceae*, *Caulobacteraceae*, *Chromatiaceae*, *Peptococcaceae-2*, *Phyllobacteriaceae*, *Enterobacteriaceae*, and *Nocardioidaceae*) while 10 taxa were depleted (namely, *Gaiellaceae*, *Actinomycetaceae*, *Conexibacteraceae*, *Thermomonosporaceae*, *Hyphomicrobiaceae*, *Solirubrobacteraceae*, *Ktedonobacteraceae*, *Rhodocyclaceae*, *Planctomycetaceae*, and *Gemmatimonadaceae*) in the plastisphere (Figure S14B, C and Table S9).

S3 Ecosystem identity controls the microbial coexistence pattern

Based on the Spearman's rank correlations corrected by the Benjamini-Hochberg method, we constructed a global ecological meta-network to explore the dominate factor of the global microbial co-occurrence pattern (Figure S16 and Tables S13 and S14). The degree of the meta-network followed a power-law distribution (R^2 = 0.858; Figure S17), displaying non-random and scale-free features. The meta-network contained 660 nodes that formed 11,752 significant associations (Figure S16A and Tables S13 and S14). The top three large modules formed in the meta-network encompassed more than 96% of the nodes (Figure S16A, B and Table S15). By analyzing the relative abundance of the nodes in each sample, we found that each of the three modules reflected a corresponding ecosystem. Module 1 consisted mainly of members of the *Alphaproteobacteria*, *Gamaproteobacteria*, and *Bacteroidetes*, and was prevalent mainly in seawater ecosystems (Figure S16B, C and Table S15). Module 2 was comprised mainly of members of the *Firmicutes*, *Actinobacteria*, and *Alphaproteobacteria*, and reflected terrestrial ecosystems (Figure S16B, C and Table S15). Module 3 was mainly formed by members of the *Betaproteobacteria*, *Bacteroidetes*, and *Gammaproteobacteria*, and represented freshwater ecosystems (Figure S16B, C and Table S15). These phenomena indicated that the ecosystem identity was a more important driver of the co-occurrence pattern of the global microbiome than differences between the plastisphere and the natural environments.

S4 Increased risk from clinical pathogens in the plastisphere

We explored the potential for clinical pathogens to be present in the plastisphere based on the 16SPIP (16S Pathogenic Identification Process),[27] a comprehensive tool for rapid pathogen detection in clinical samples and also widely applied in environmental samples.[24-26] A total of 40 pathogenic species were observed in our dataset after matching with >99% similarity in sequence (Table S25). Overall, the plastisphere exhibited a significantly higher pathogenic potential compared to the natural environment (Figure S26 and Table S26). In the freshwater ecosystem, pathogens accounted for 10.4% of the plastisphere community, which was more than four times the proportion in the natural environment. In the terrestrial ecosystem, pathogens accounted for 9.3% of the plastisphere community, 5.7 times that of the community in the natural environment (Table S26).

Notably, in each ecosystem, all pathogens detected in the natural environment also occurred in the plastisphere, but the plastisphere harbored additional pathogens that were not detected in the

corresponding natural environment (Figure S26). By comparing the plastisphere and the natural environment in terms of the abundance of each pathogen, we found that a significant proportion of pathogenic species showed higher abundance in the plastisphere in all studied ecosystems (Figure S26 and Table S27). This suggests that the plastisphere promotes the growth of diverse pathogens. For example, 25 of 40 pathogens were enriched in the plastisphere in the freshwater ecosystem (Figure S26 and Table S27). In particular, the relative abundance of four pathogenic species (*Erysipelothrix rhusiopathiae*, *Proteus vulgaris*, *Citrobacter freundii*, and *Morganella morganii*) in the freshwater plastisphere was two to three orders of magnitude higher than that in the natural environment (Table S28). Similarly, a total of 17 out of 39 pathogens were plastisphere-enriched in the terrestrial ecosystem (Figure S26 and Table S27). Of these, the relative abundance of *Escherichia coli*, *Acinetobacter lwoffii*, *Citrobacter freundii*, *Acinetobacter baumannii*, and *Nocardia asteroides* in the plastisphere was again two to three orders of magnitude higher than that in the natural environment (Table S28). Pathogens unique to the plastisphere and plastisphere-enriched pathogens were different between ecosystems (Figures S26 and S27), showing that the plastisphere could pose different health threats in these different ecosystems.

S5 The plastisphere shelters its residents from external disturbances

Using our own field-collected samples, we explored the driving effect of the physicochemical properties of the surrounding medium on the plastisphere microbiome. The measurement methods for environmental physicochemical parameters were as described in our previous study.[57] Procrustes analysis and Mantel test showed that significant correlation existed between variation in the physicochemical properties of the surrounding medium and variation in the structure of the plastisphere community (Procrustes: $r = 0.563$, $P < 0.001$; Mantel: $r = 0.252$, $P < 0.001$; Figure S30A). Among the measured physicochemical factors, oxidation-reduction potential, concentrations of nutrients (dissolved organic carbon, NO_3^-, and NH_4^+), and salinity, explained more of the variation in the plastisphere microbial community and may be significant environmental drivers of the plastisphere microbial community (Figure S30B). Compared to the ambient microbial community (Procrustes: $r = 0.582$, $P < 0.001$; Mantel: $r = 0.366$, $P < 0.001$), the microbial community in the plastisphere were less driven by environmental physicochemical factors (Figure S30 A and C), demonstrating the sheltering effect of the plastisphere on its residents. In addition, while there was a significant association between the ambient microbial community and the plastisphere microbial community, changes in the ambient microbial community explained only a small fraction of the changes in the plastisphere microbial community (Procrustes: $r = 0.428$, $P < 0.01$; Mantel: $r = 0.119$, $P < 0.05$; Figure S30D), suggesting the selective assembly of the plastisphere with its preferred microorganisms, the sheltering effect of the plastisphere on its residents, and the potential of the plastisphere to raft its residents for long-distance transport.

SUPPLEMENTAL FIGURES

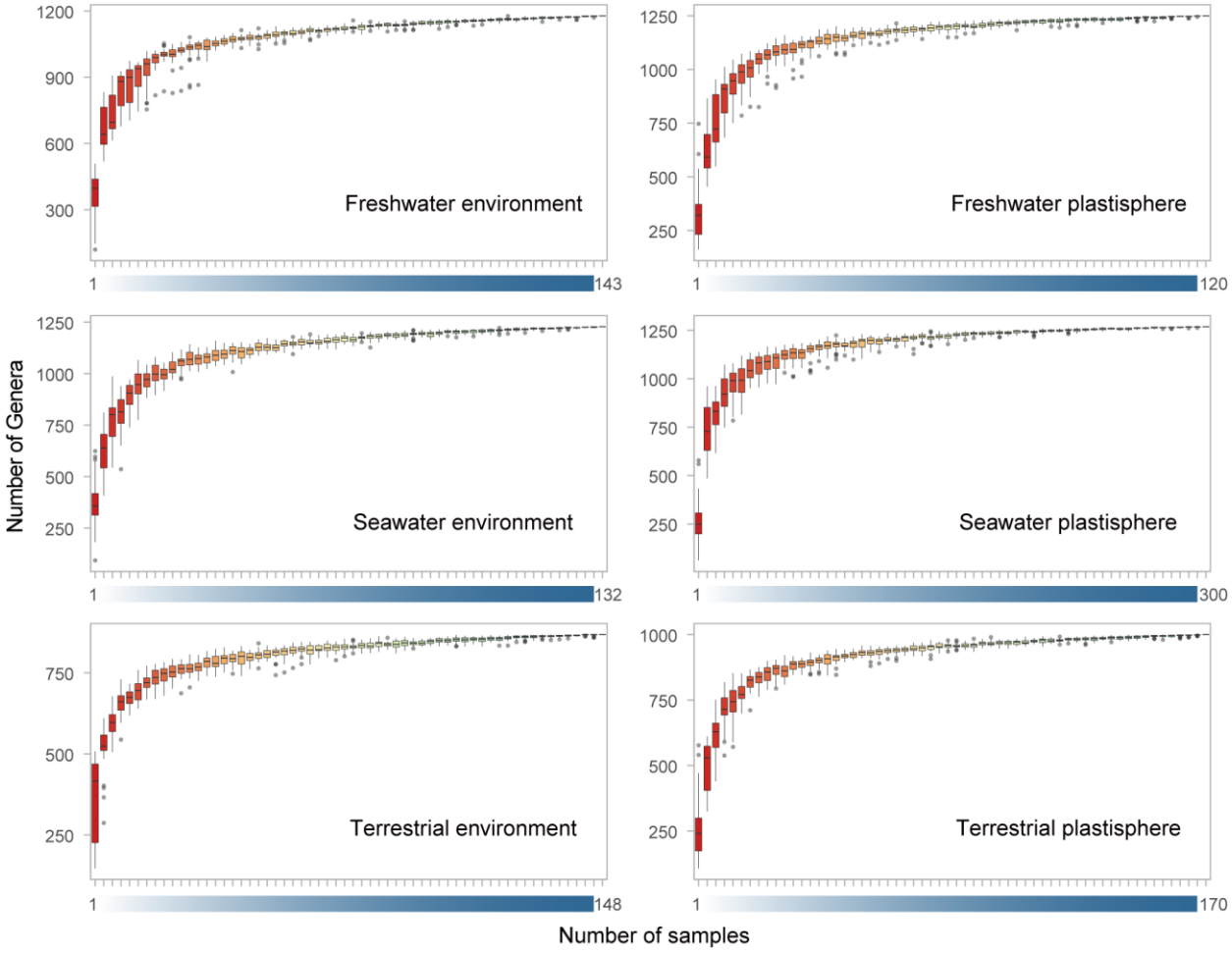

Figure S1 Rarefaction curves.

Rarefaction curves of the number of genera in the plastisphere and the natural environment in freshwater, seawater, and terrestrial ecosystems reach the saturation stage with increasing numbers of samples, indicating that the number of samples in our study is sufficient to capture most microorganisms from the plastisphere and the natural environment in each ecosystem.

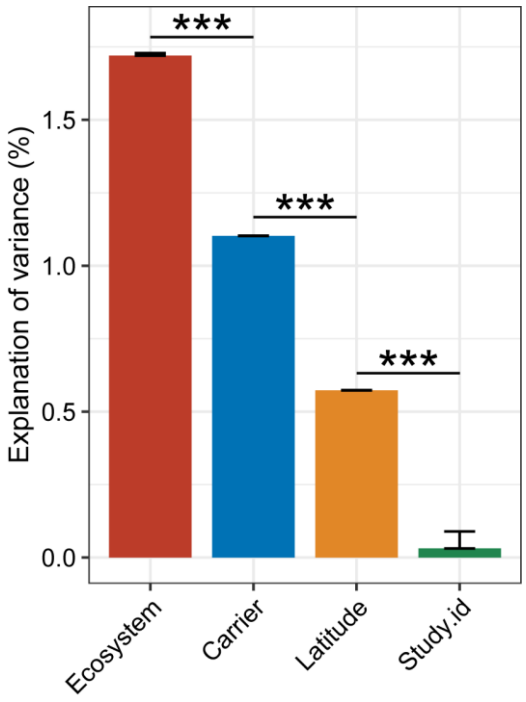

Figure S2 Explanations of the meta-community variation by different potential drivers.

The result was obtained based on the canonical correspondence analysis, and shows that, except for the ecosystem identity, the carrier identity, *i.e.*, the difference between the plastisphere and the natural environment, is the most important factor driving the variation in the meta-community structure.

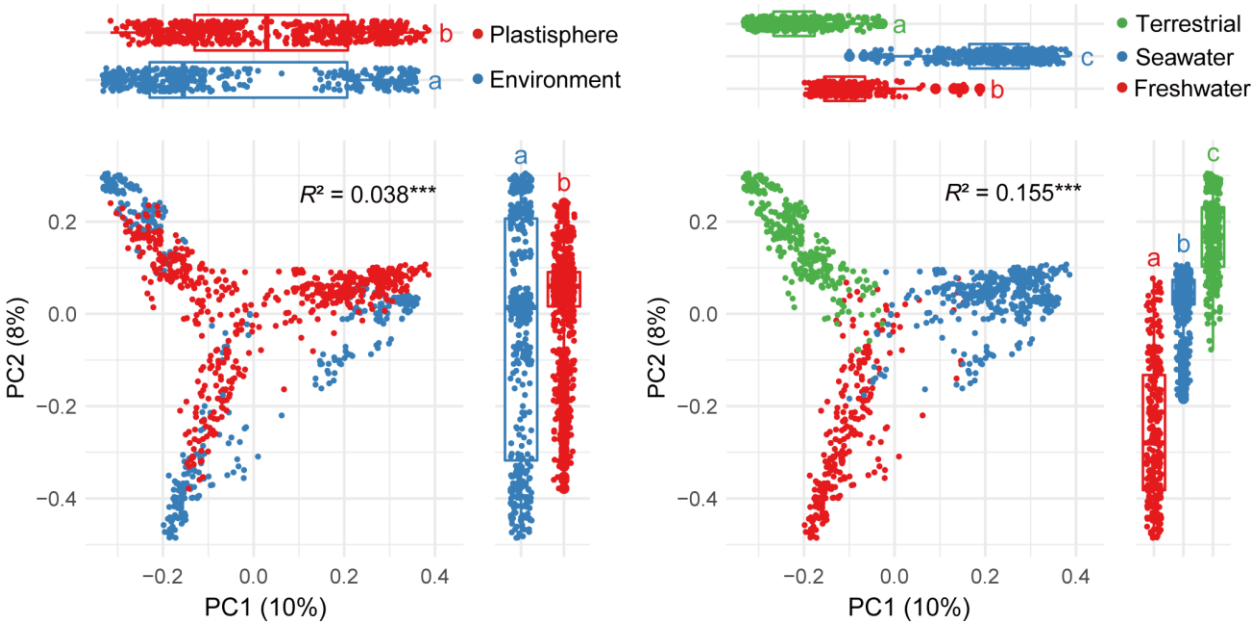

Figure S3 Differences in microbial community structure between the plastisphere and the natural environment and among different ecosystems.

Unconstrained principal coordinate analysis (PCoA) with permutational multivariate analysis of variance (PERMANOVA) showing that the plastisphere has a distinct microbial community from that of the natural environment (R^2 = 0.038, ***P < 0.001), but that the structure of the community is more dependent on the ecosystem (R^2 = 0.155, ***P < 0.001).

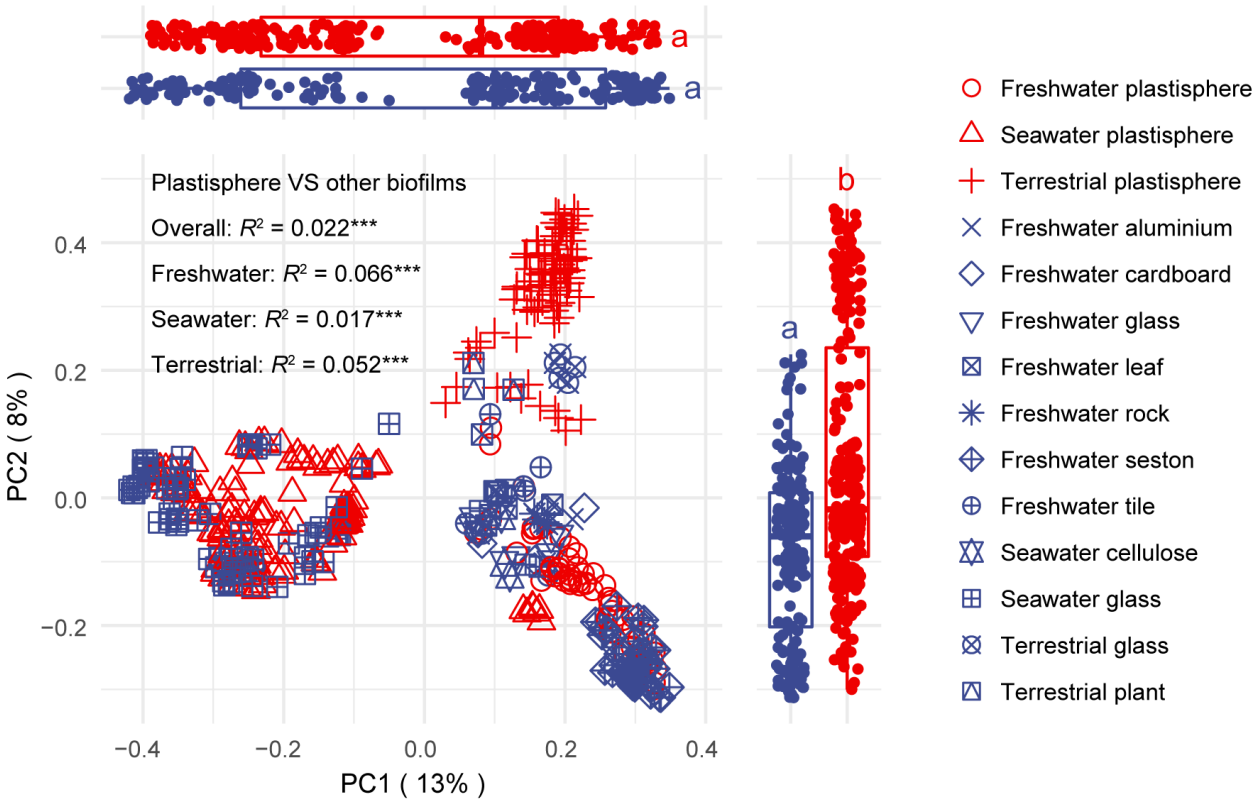

Figure S4 Significant differences between plastisphere microbial communities and other biofilms.

Unconstrained principal coordinate analysis (PCoA) with permutational multivariate analysis of variance (PERMANOVA) showing that the plastisphere has a distinct microbial community from other biofilms, both overall (R^2 = 0.022, ***P < 0.001) and in each ecosystem specifically (in the freshwater ecosystem: R^2 = 0.066, ***P < 0.001; in the seawater ecosystem: R^2 = 0.017, ***P < 0.001; in the terrestrial ecosystem: R^2 = 0.052, ***P < 0.001).

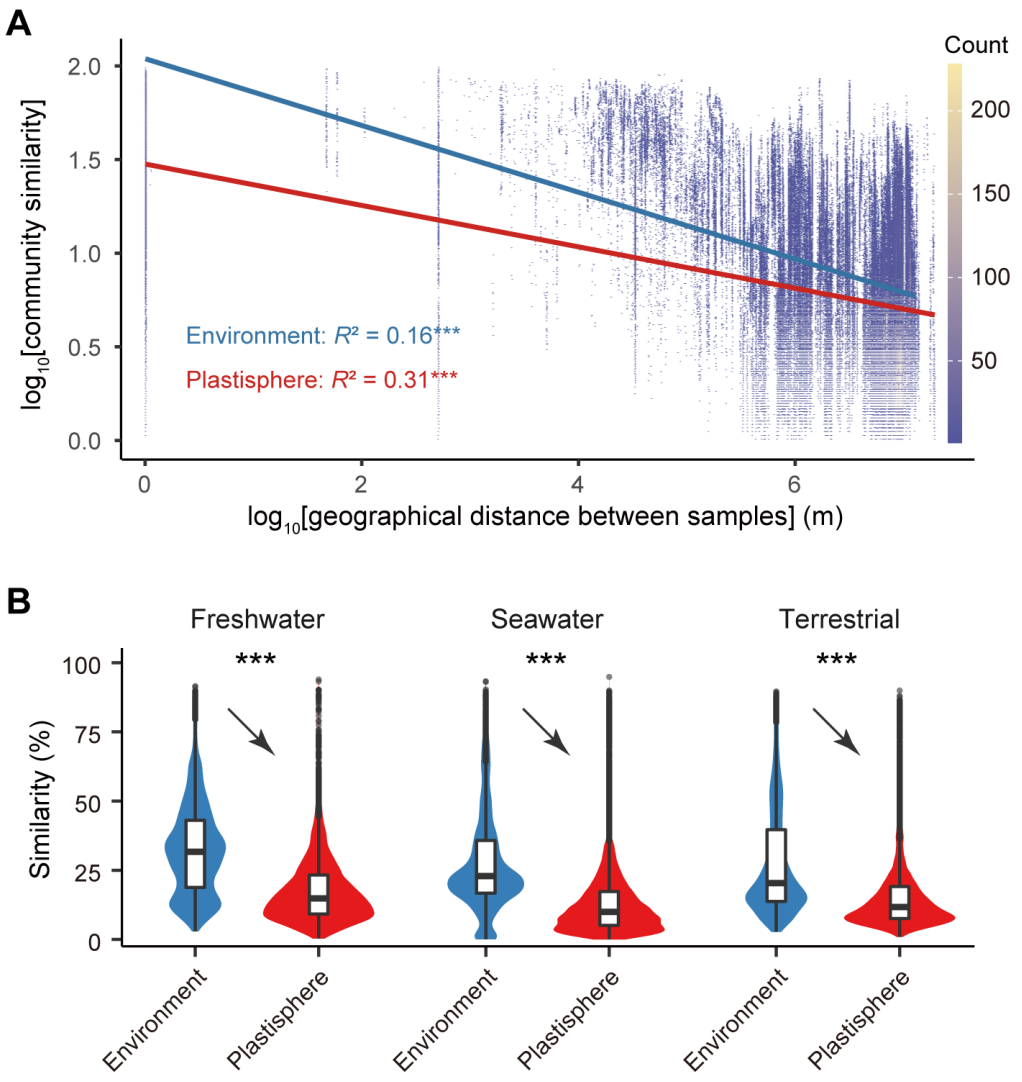

Figure S5 Between-sample compositional similarity.

(**A**) Significant distance-decay patterns in the plastisphere and the natural environment (***P < 0.001; linear regressions). (**B**) Comparisons of compositional similarity between the community in the plastisphere and that of the natural environment (***P < 0.001; Wilcoxon rank sum test), and the numbers of replicated samples are as follows: freshwater plastisphere (n = 120), freshwater environment (n = 143), seawater plastisphere (n = 300), seawater environment (n = 132), terrestrial plastisphere (n = 170), terrestrial environment (n = 148).

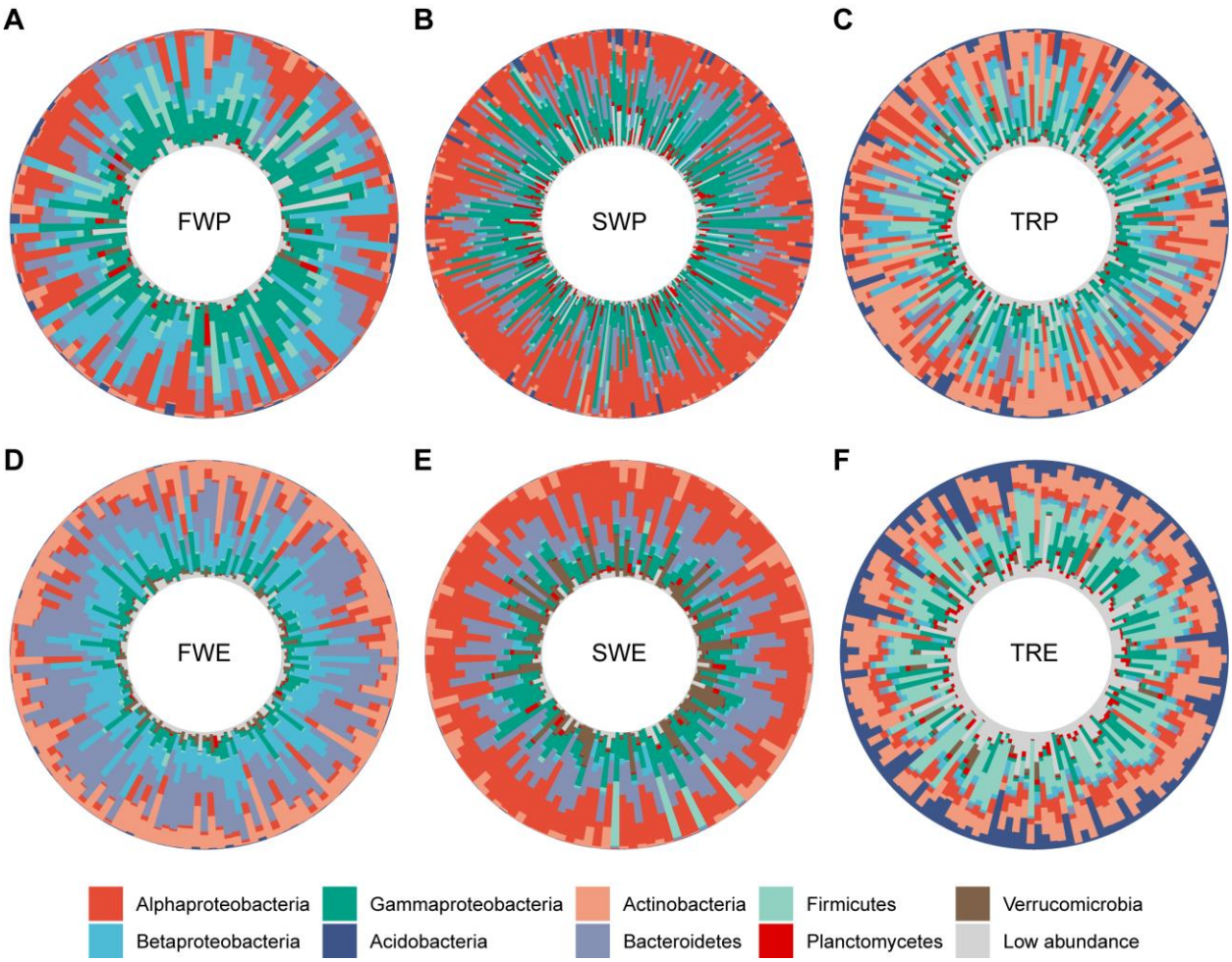

Figure S6 Taxonomic composition of microbial communities in the plastisphere and the natural environment.

(**A**-**C**) Phylum-level (with Proteobacteria being shown at the class level) composition of microbial communities in the plastisphere in freshwater (**A**), seawater (**B**) and terrestrial (**C**) ecosystems. (**D**-**F**) Phylum-level (with *Proteobacteria* being shown at the class level) composition of microbial communities in the natural environment in freshwater (**D**), seawater (**E**) and terrestrial (**F**) ecosystems. FWP = freshwater plastisphere (n = 120); FWE = freshwater environment (n = 143); SWP = seawater plastisphere (n = 300); SWE = seawater environment (n = 132); TRP = terrestrial plastisphere (n = 170); TRE = terrestrial environment (n = 148).

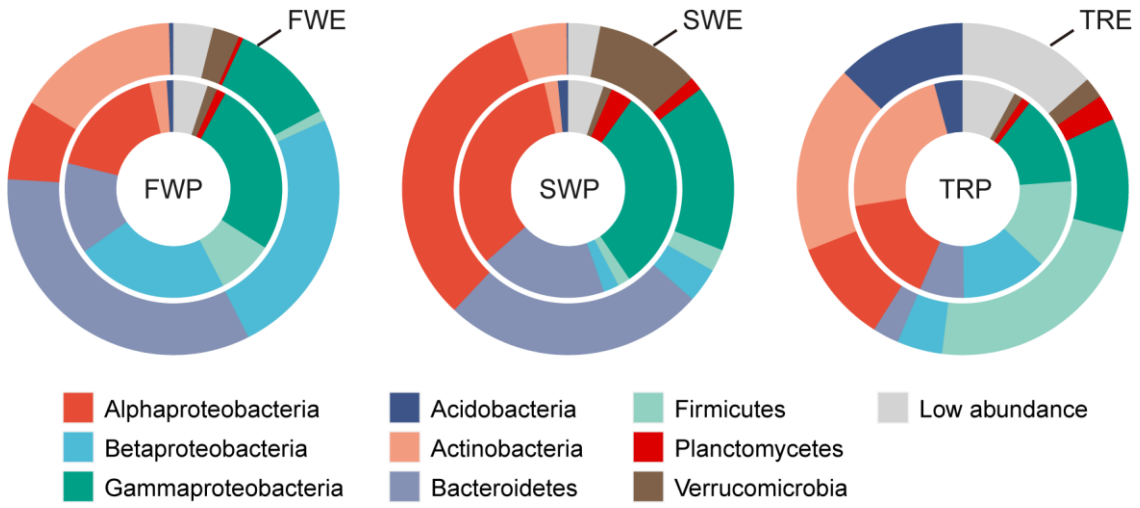

Figure S7 Taxonomic composition of microbial communities in the plastisphere (inner circles) and the natural environment (outer circles).

The numbers of replicated samples are as follows: freshwater plastisphere (*n* = 120), freshwater environment (*n* = 143), seawater plastisphere (*n* = 300), seawater environment (*n* = 132), terrestrial plastisphere (*n* = 170), terrestrial environment (*n* = 148).

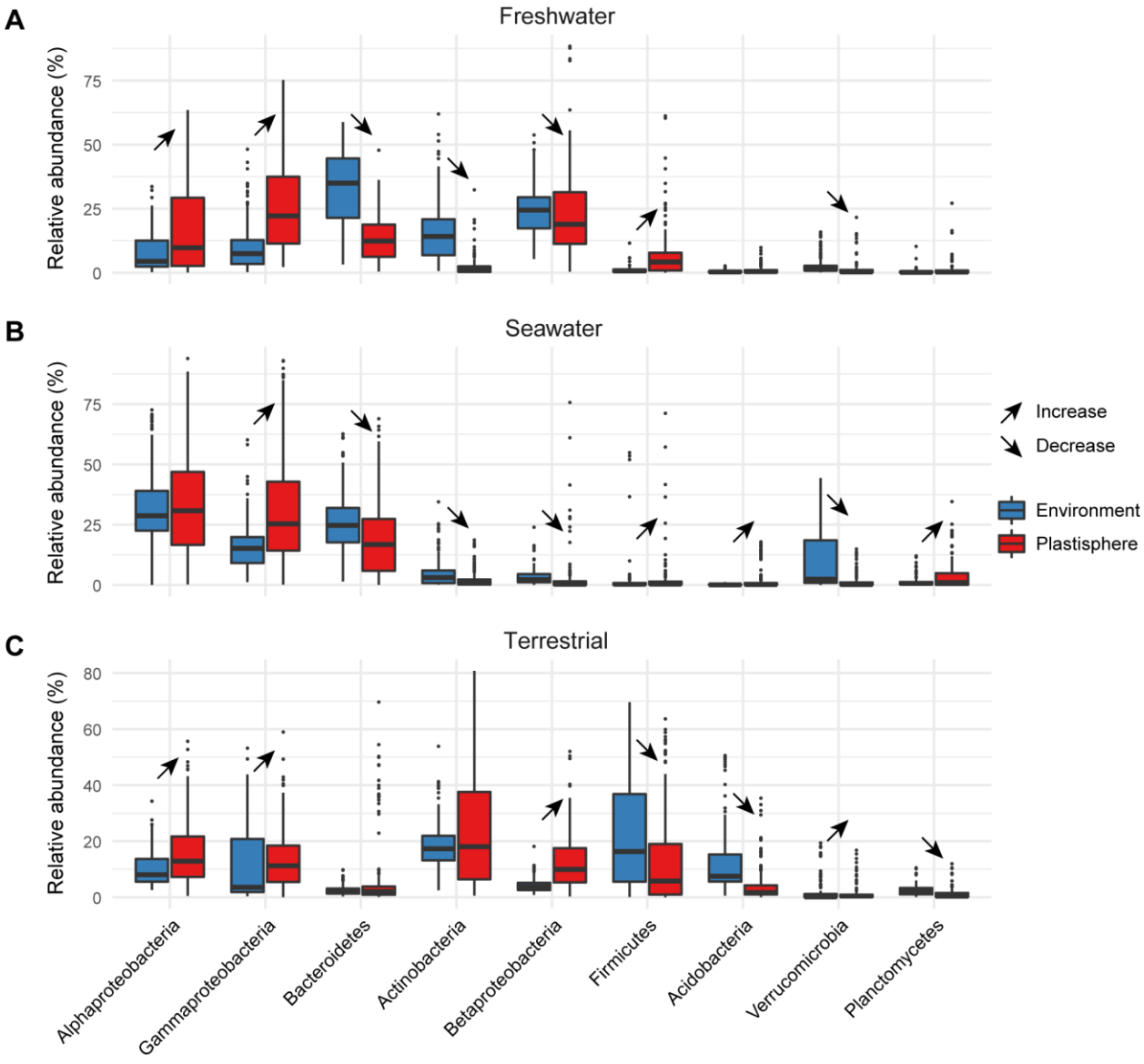

Figure S8 Differences in the relative abundance of microbial taxa between the plastisphere and the natural environment.

Wilcoxon rank sum tests for the relative abundance of the top nine most abundant microbial taxa between the plastisphere and the natural environment in freshwater (A), seawater (B), and terrestrial (C) ecosystems showing that most microbial taxa are significantly altered in the plastisphere ($P < 0.05$). An upward arrow represents that the relative abundance of the microbial taxon is significantly higher in the plastisphere than in the natural environment, while a downward arrow represents that the relative abundance of the taxon is significantly lower in the plastisphere than in the natural environment ($P < 0.05$; Wilcoxon rank sum test). The numbers of replicated samples are as follows: freshwater plastisphere ($n = 120$), freshwater environment ($n = 143$), seawater plastisphere ($n = 300$), seawater environment ($n = 132$), terrestrial plastisphere ($n = 170$), terrestrial environment ($n = 148$).

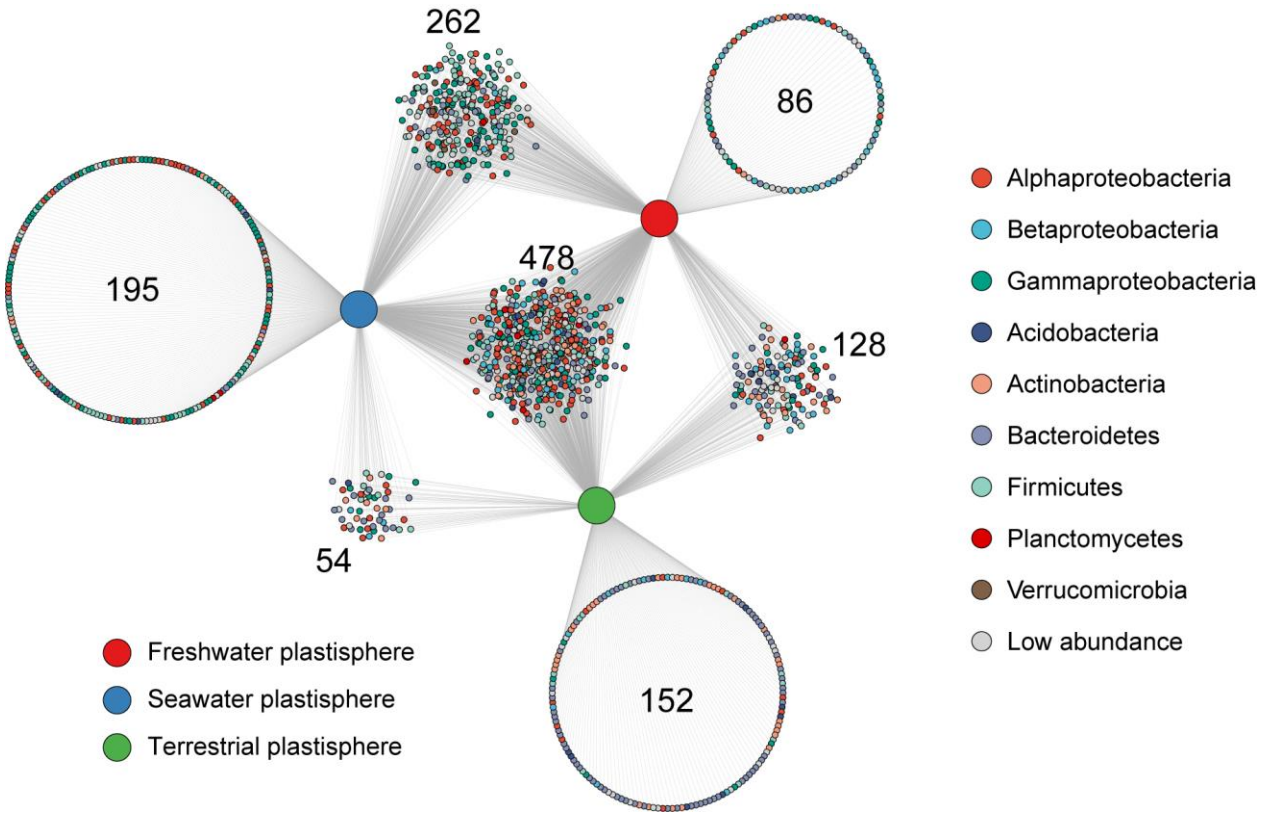

Figure S9 Shared and unique taxa between the plastisphere in freshwater, seawater, and terrestrial ecosystems.

Each small dot represents a microbial genus, and its color represents the taxonomic information. Each large dot represents a group (the freshwater plastisphere, the seawater plastisphere, and the terrestrial plastisphere). A line between a small dot and a large dot represents the presence of this taxon in the corresponding plastisphere.

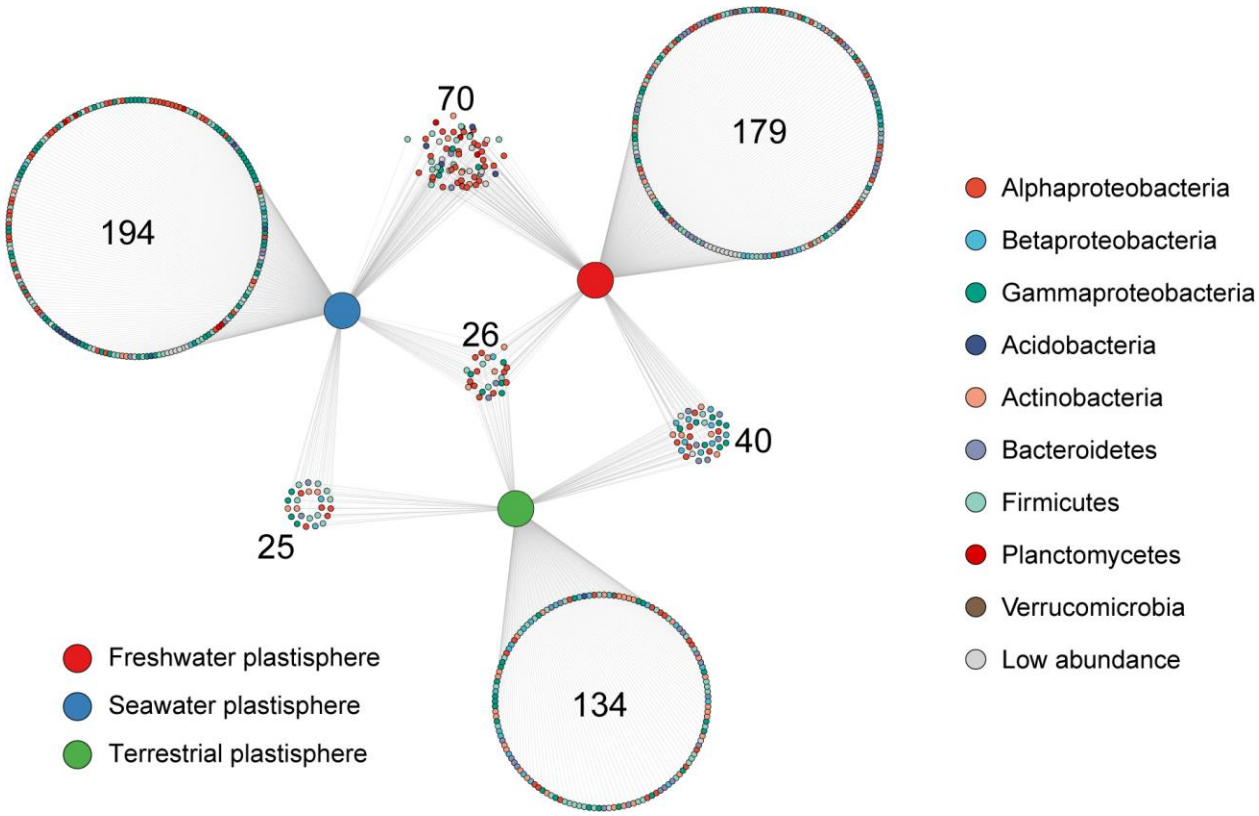

Figure S10 Commonly and uniquely enriched taxa between the plastisphere in freshwater, seawater, and terrestrial ecosystems.

Each small dot represents a microbial genus, and its color represents the taxonomic information. Each large dot represents a group (the freshwater plastisphere, the seawater plastisphere, and the terrestrial plastisphere). A line between a small dot and a large dot represents the presence of this taxon in the corresponding plastisphere.

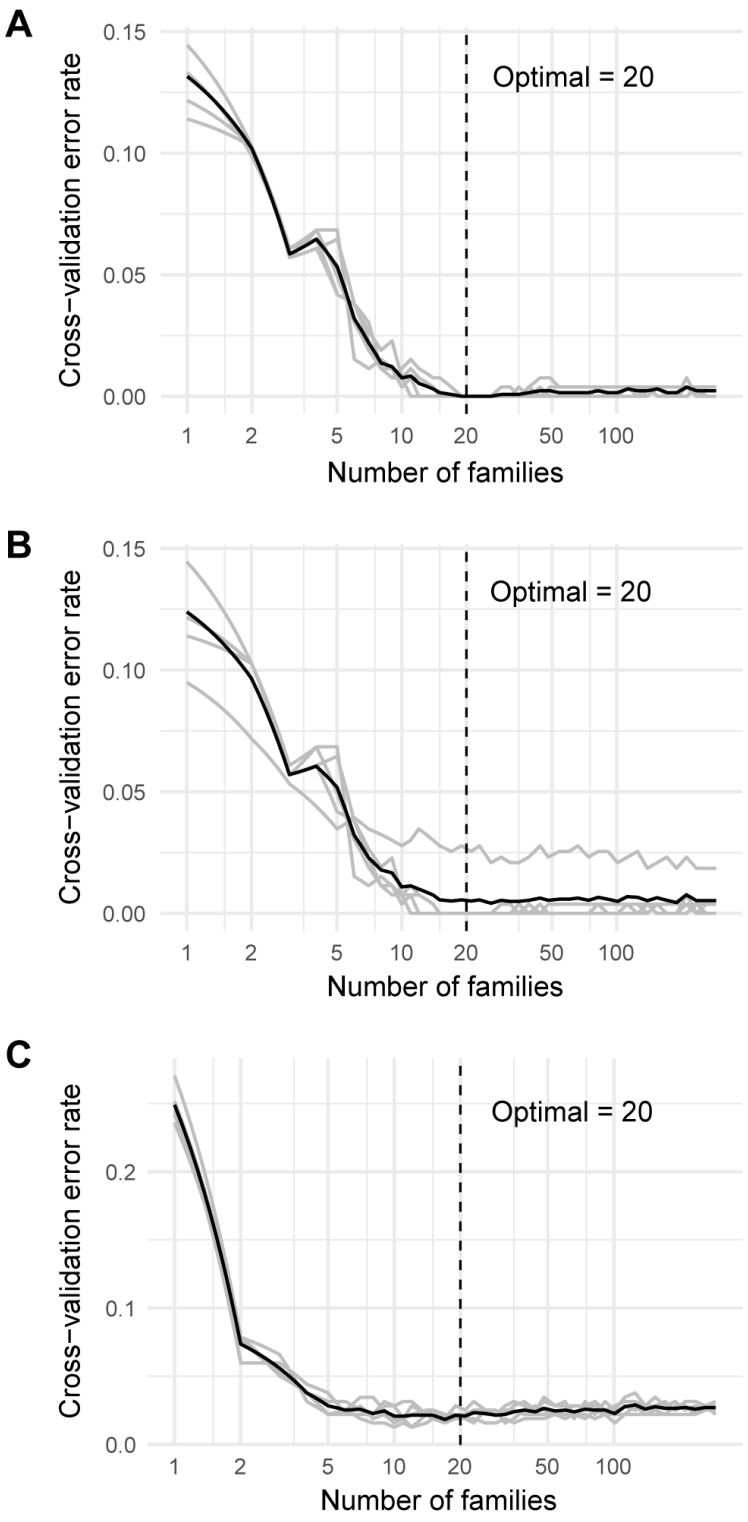

Figure S11 Identification of the number of plastisphere biomarkers.

Ten-fold cross-validation with five repeats revealing that cross-validation error curves have stabilized when 20 microbial families are included with error rates having reduced to a low level in freshwater (**A**), seawater (**B**), and terrestrial (**C**) ecosystems.

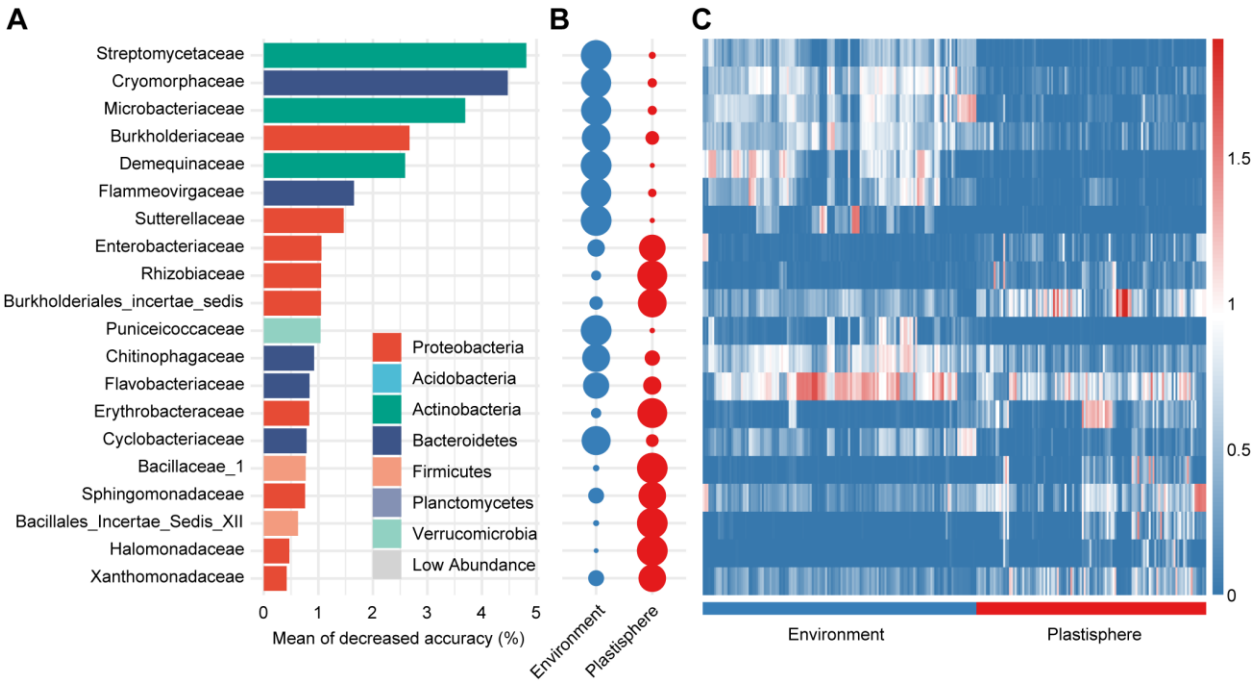

Figure S12 Plastisphere biomarkers in the freshwater ecosystem identified based on a random-forest model.

(**A**) The top 20 microbial families most important to the accuracy of the random-forest classification model for distinguishing the plastisphere from the natural environment were identified as plastisphere biomarkers in the freshwater ecosystem. The biomarker taxa are listed in descending order of importance to the accuracy of the model. (**B**) Relative proportions of mean abundance of the biomarker taxa in the plastisphere and the natural environment. (**C**) Relative abundance profiles for the biomarker taxa in each sample of the plastisphere and the natural environment. Relative abundances are log-transformed for a clear presentation in the heatmap. The numbers of replicated samples used in the model are as follows: the plastisphere (n = 120), the natural environment (n = 143).

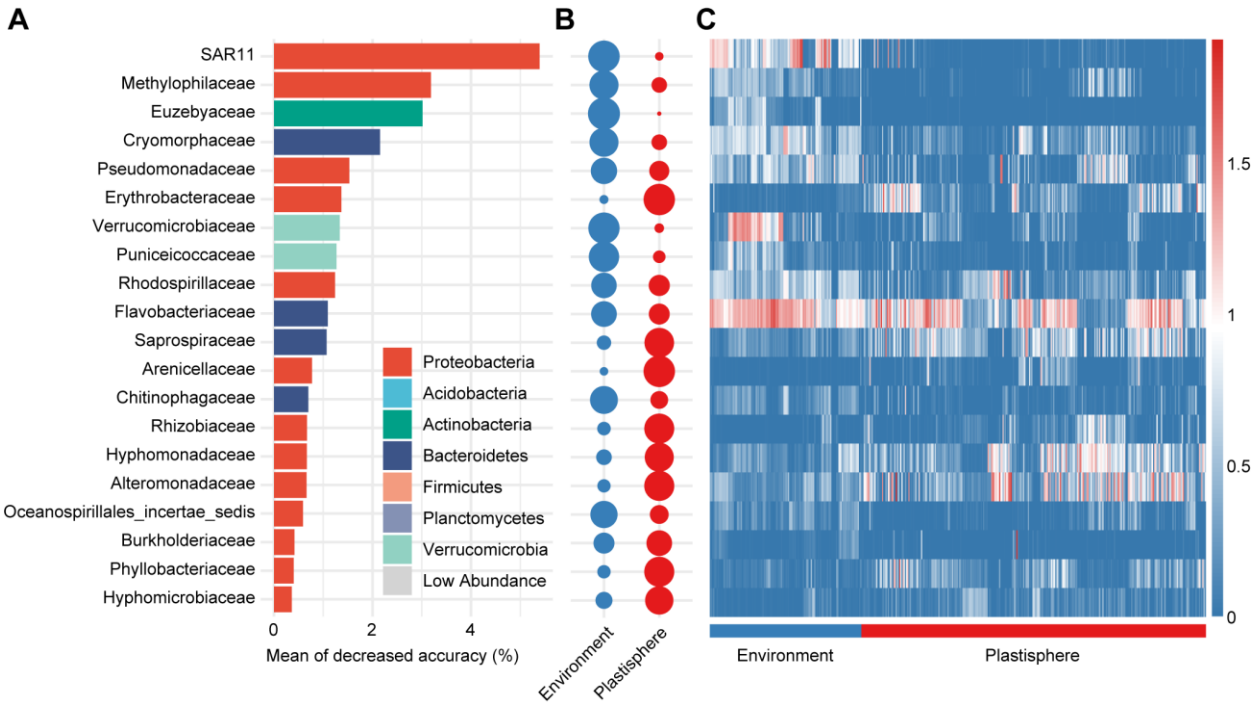

Figure S13 Plastisphere biomarkers in the seawater ecosystem identified based on a random-forest model.

(**A**) The top 20 microbial families most important to the accuracy of the random-forest classification model for distinguishing the plastisphere from the natural environment were identified as plastisphere biomarkers in the seawater ecosystem. The biomarker taxa are listed in descending order of importance to the accuracy of the model. (**B**) Relative proportions of mean abundance of the biomarker taxa in the plastisphere and the natural environment. (**C**) Relative abundance profiles for the biomarker taxa in each sample of the plastisphere and the natural environment. Relative abundances are log-transformed for a clear presentation in the heatmap. The numbers of replicated samples used in the model are as follows: the plastisphere (n = 300), the natural environment (n = 132).

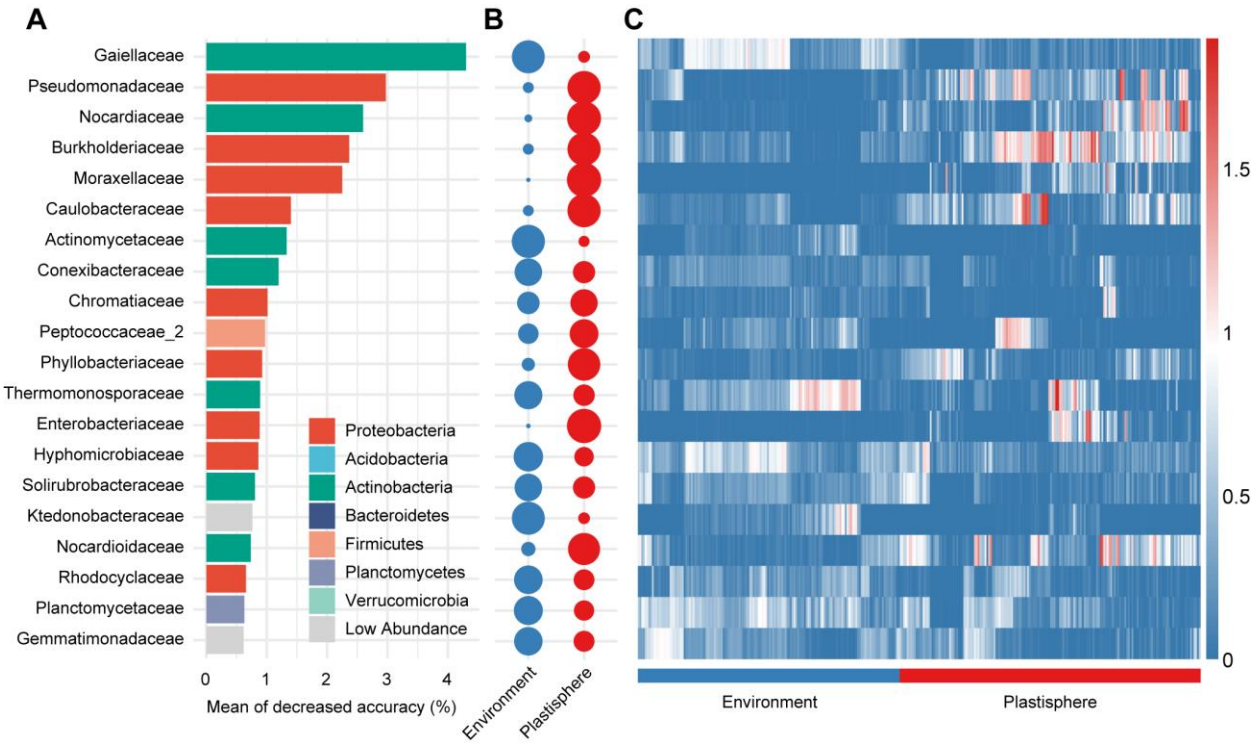

Figure S14 Plastisphere biomarkers in the terrestrial ecosystem identified based on a random-forest model.

(**A**) The top 20 microbial families most important to the accuracy of the random-forest classification model for distinguishing the plastisphere from the natural environment were identified as plastisphere biomarkers in the terrestrial ecosystem. The biomarker taxa are listed in descending order of importance to the accuracy of the model. (**B**) Relative proportions of mean abundance of the biomarker taxa in the plastisphere and the natural environment. (**C**) Relative abundance profiles for the biomarker taxa in each sample of the plastisphere and the natural environment. Relative abundances are log-transformed for a clear presentation in the heatmap. The numbers of replicated samples used in the model are as follows: the plastisphere (n = 170), the natural environment (n = 148).

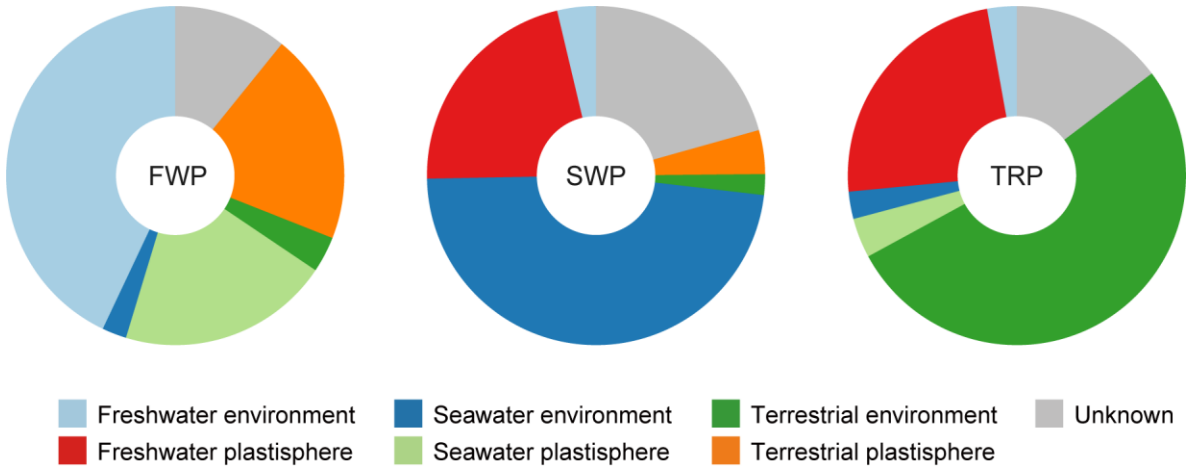

Figure S15 Source analysis.

The fast expectation-maximization for microbial source tracking (FEAST) analysis showing that the corresponding natural environment contributes the largest part, but only a subset, of the sources of microorganisms in the plastisphere.

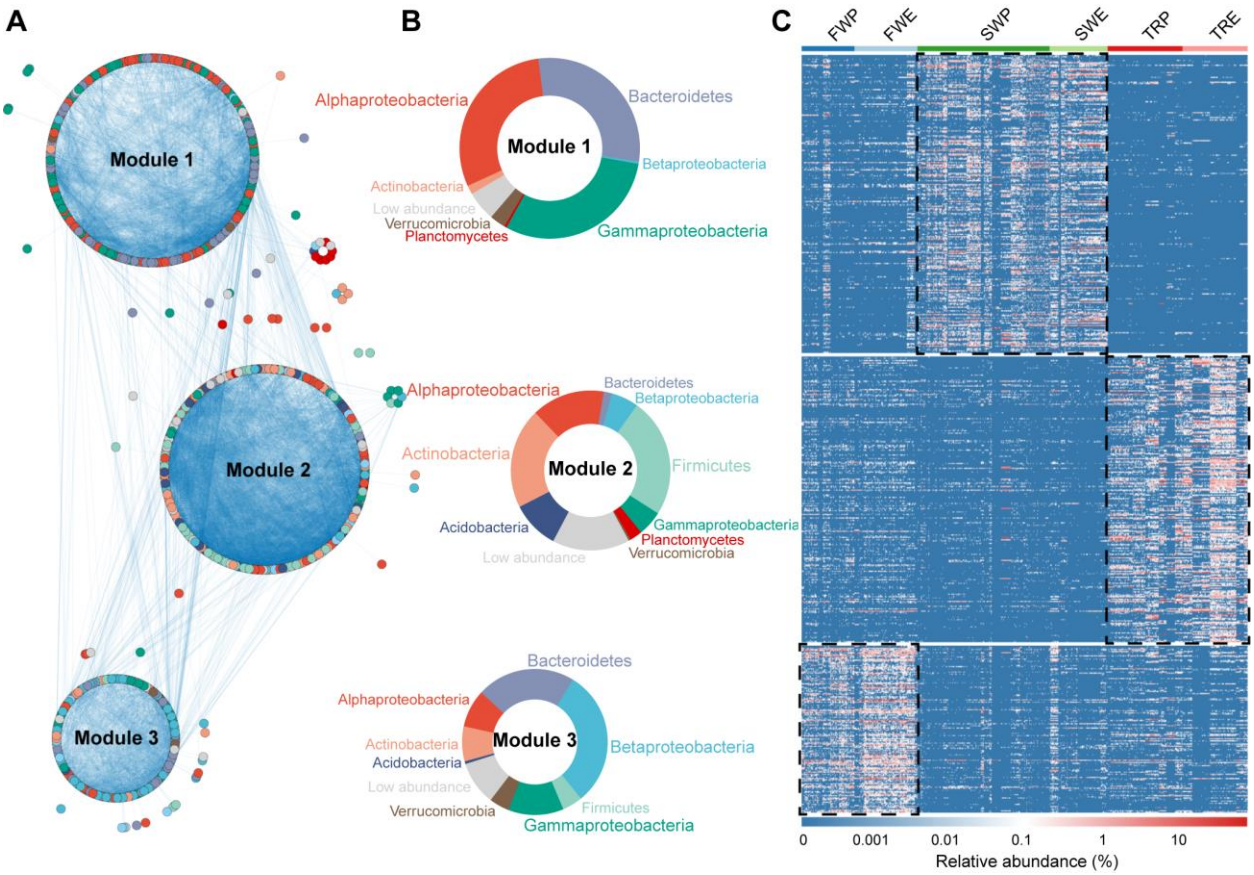

Figure S16 Global ecological meta-network.

(**A**) An overview of the meta-network. Each node represents a unique genus. Each connection between the two nodes represents a significant co-occurrence relationship (Spearman's $\rho > 0.4$ and $P < 0.05$). The size of each module indicates the number of nodes that it contains. The colors of the nodes indicate taxonomic identity. (**B**) Taxonomic composition of the top three largest modules, containing more than 96% of the total nodes in the meta-network. (**C**) Patterns of relative abundance of the nodes in different ecosystems. FWP = freshwater plastisphere; FWE = freshwater environment; SWP = seawater plastisphere; SWE = seawater environment; TRP = terrestrial plastisphere; TRE = terrestrial environment.

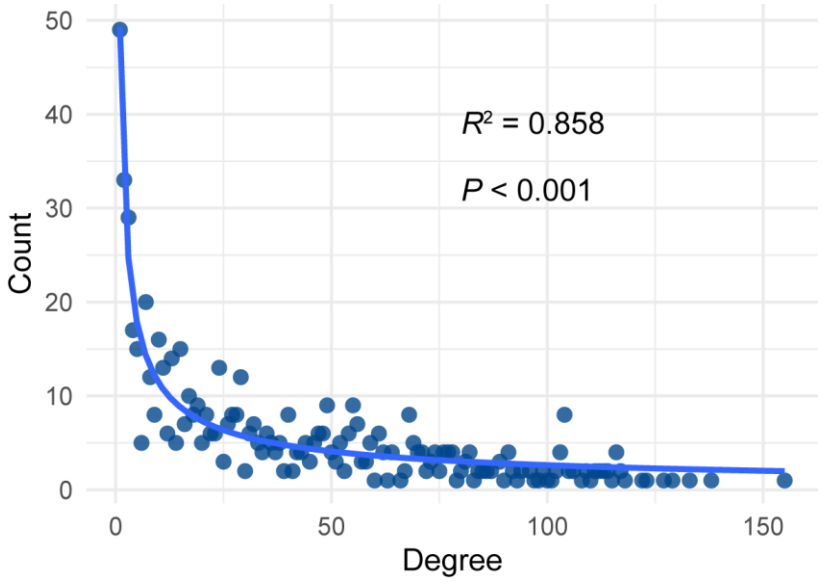

Figure S17 Degree distributions of the microbial ecological meta-network.

R^2 represents the goodness of fit of a power-law model.

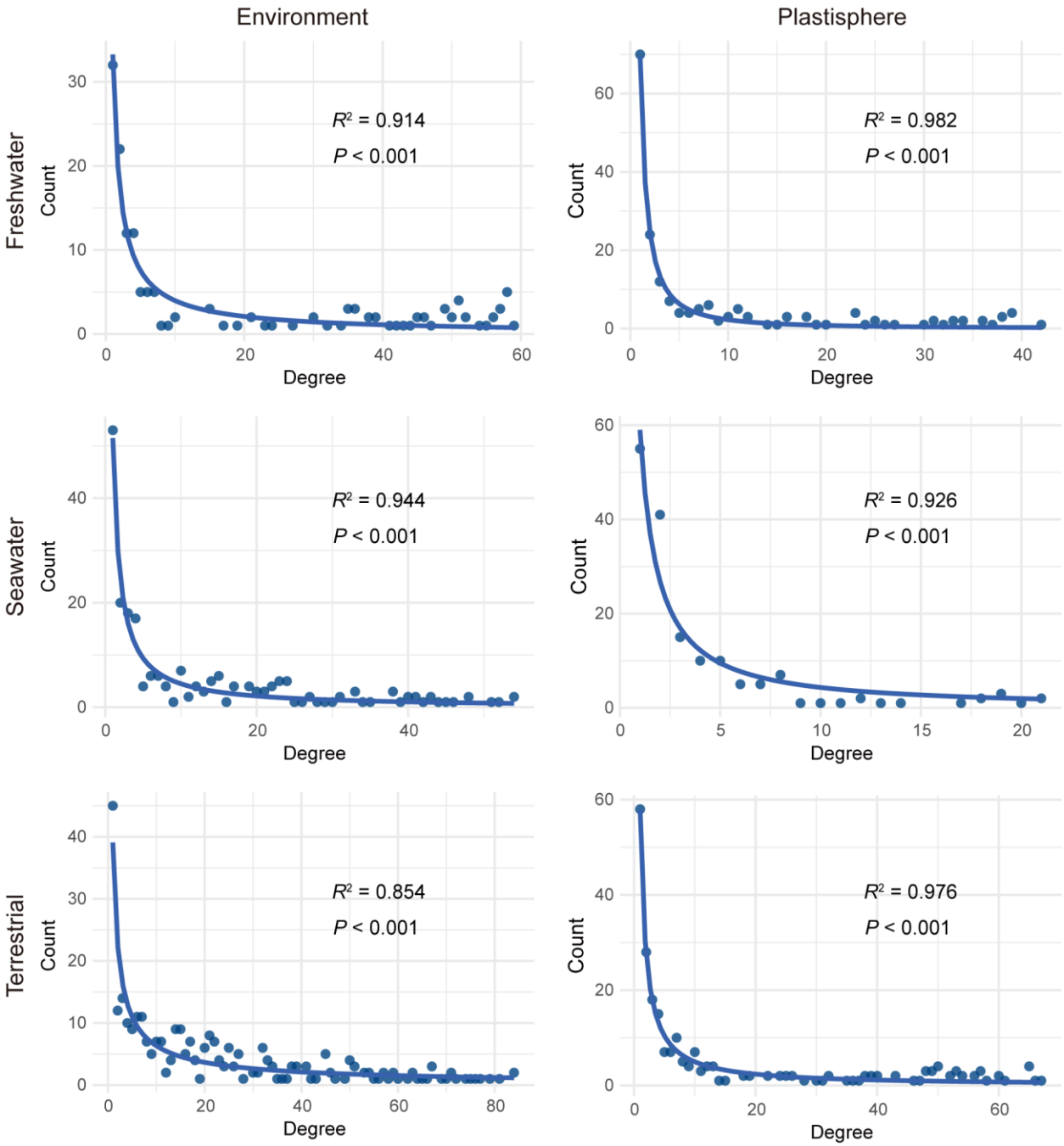

Figure S18 Degree distributions of the microbial ecological sub-networks.

R^2 represents the goodness of fit of a power-law model.

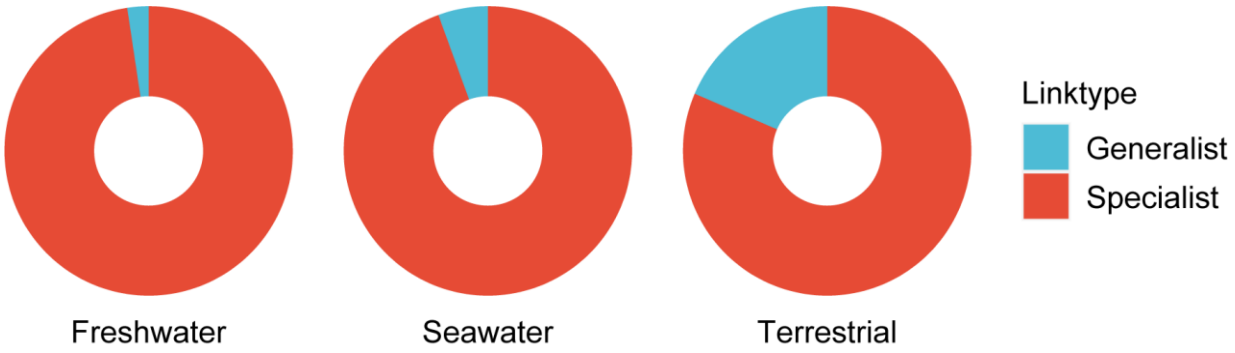

Figure S19 Proportions of specialist links in the plastisphere sub-networks.

The specialist link means that the microbial association occurs only in the plastisphere and not in the corresponding natural environment of that ecosystem. The generalist link means that the microbial association occurs in both the plastisphere and the natural environment.

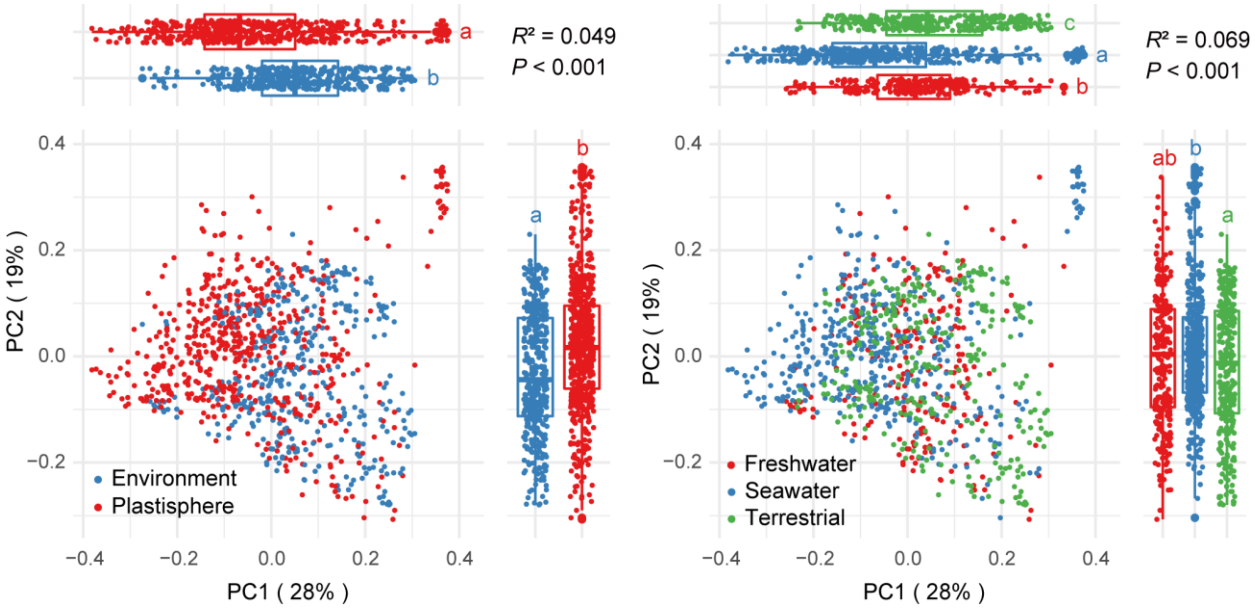

Figure S20 Differences in functional composition between the plastisphere and the natural environment among different ecosystems.

Unconstrained principal coordinates analysis (PCoA) showing the difference in ecologically functional composition between the plastisphere and the natural environment and among different ecosystems, and permutational multivariate analysis of variance (PERMANOVA) showing the statistical significance of the differences ($P < 0.001$).

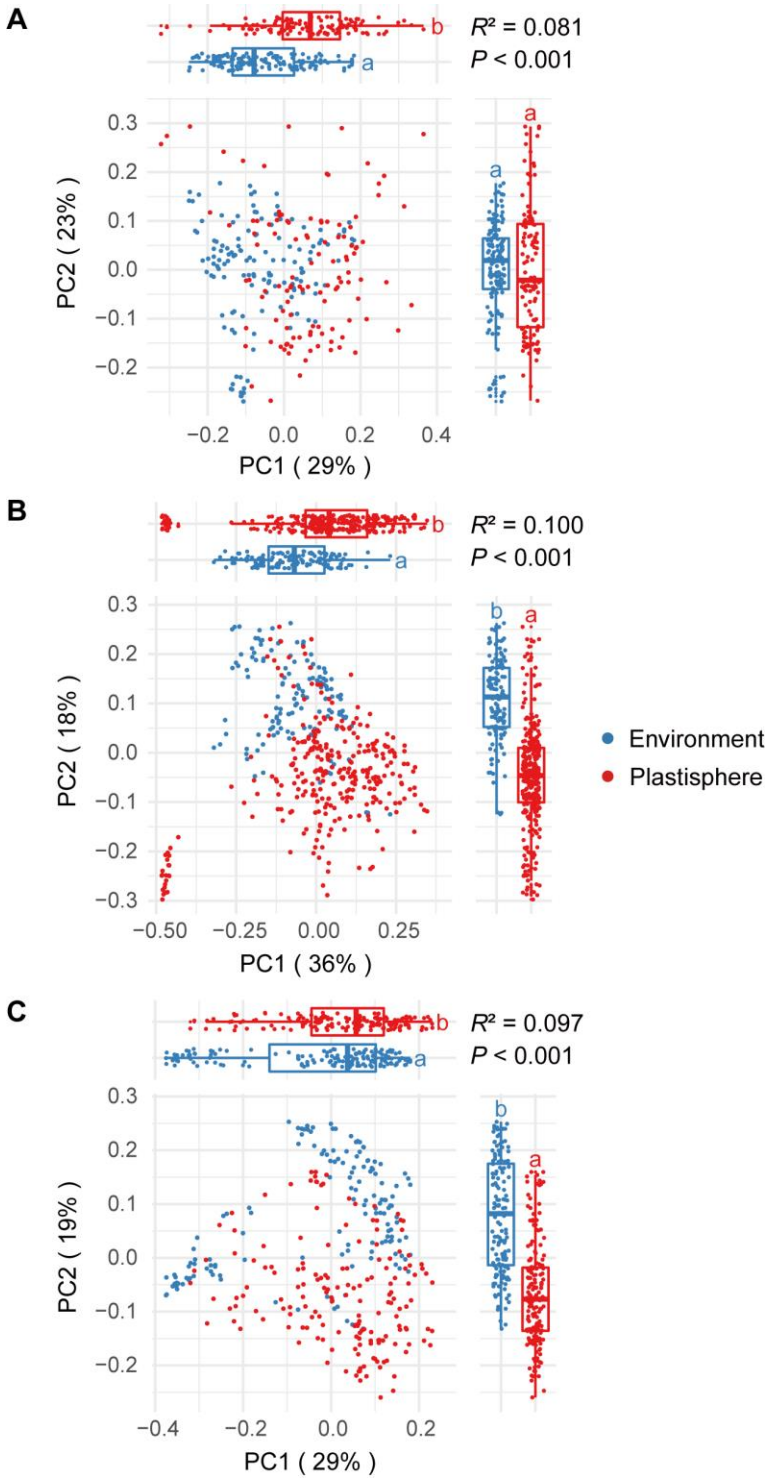

Figure S21 Differences in functional composition between the plastisphere and the natural environment in each ecosystem.

Unconstrained principal coordinates analysis (PCoA) showing the difference in ecologically functional composition between the plastisphere and the natural environment in freshwater (**A**), seawater (**B**), and terrestrial (**C**) ecosystems, and permutational multivariate analysis of variance (PERMANOVA) showing the statistical significance of the differences ($P < 0.001$).

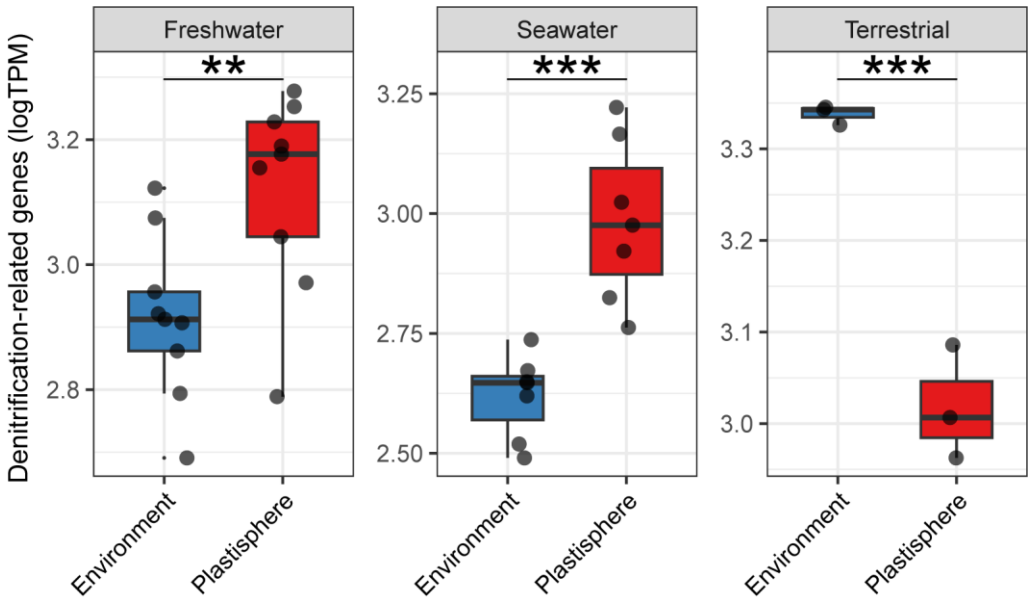

Figure S22 Comparison of the abundance of genes encoding for the denitrification function between the plastisphere and the natural environment.

TPM = transcripts per million. ***P* < 0.01, ****P* < 0.001; *t*-test. The numbers of replicated samples are as follows: freshwater plastisphere (*n* = 9), freshwater environment (*n* = 9), seawater plastisphere (*n* = 7), seawater environment (*n* = 7), terrestrial plastisphere (*n* = 3), terrestrial environment (*n* = 3).

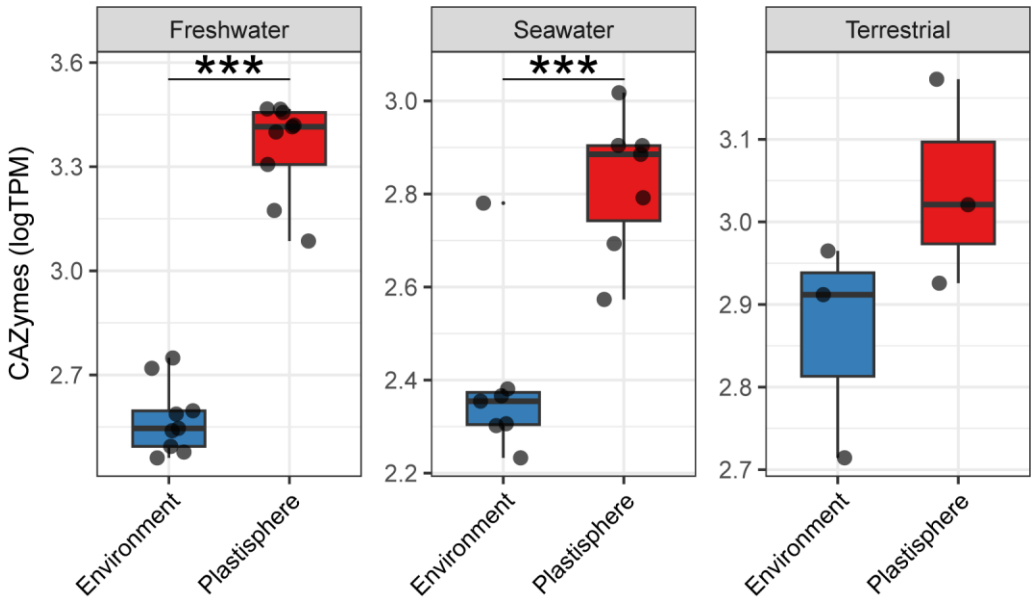

Figure S23 Comparison of the abundance of genes encoding for carbohydrate-active enzymes (CAZymes) between the plastisphere and the natural environment.

TPM = transcripts per million. ***P < 0.001; *t*-test. The numbers of replicated samples are as follows: freshwater plastisphere (n = 9), freshwater environment (n = 9), seawater plastisphere (n = 7), seawater environment (n = 7), terrestrial plastisphere (n = 3), terrestrial environment (n = 3).

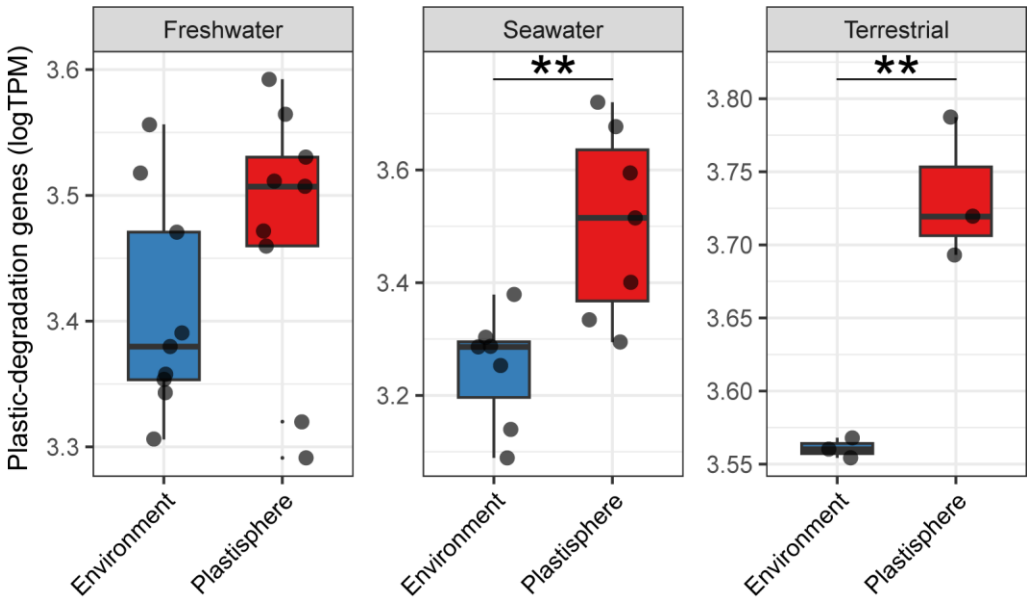

Figure S24 Comparison of the abundance of genes encoding for plastic degradation between the plastisphere and the natural environment.

TPM = transcripts per million. $^{**}P < 0.01$, *t*-test. The numbers of replicated samples are as follows: freshwater plastisphere (n = 9), freshwater environment (n = 9), seawater plastisphere (n = 7), seawater environment (n = 7), terrestrial plastisphere (n = 3), terrestrial environment (n = 3).

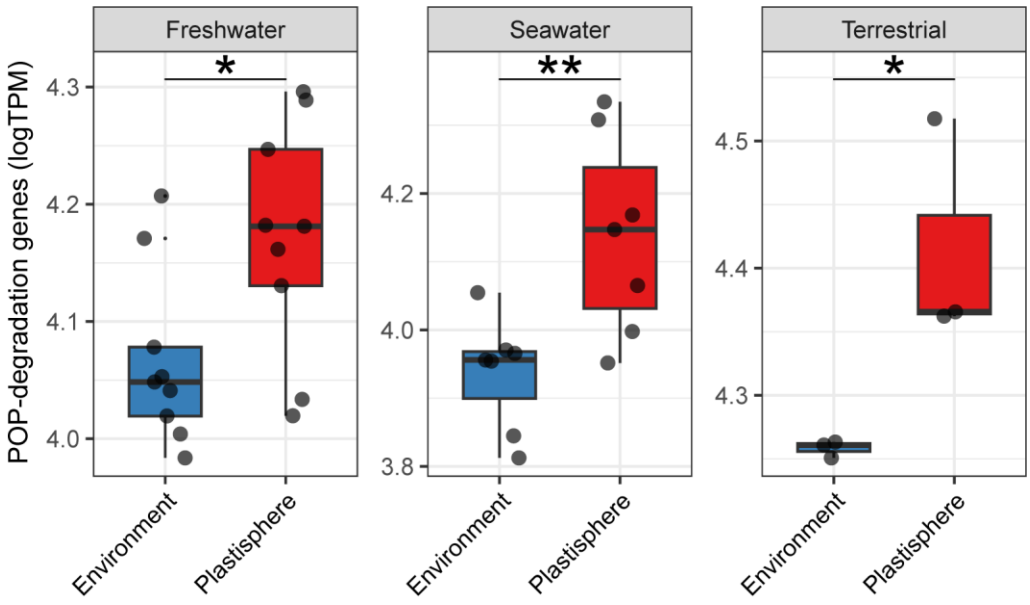

Figure S25 Comparison of the abundance of genes encoding for persistent organic pollutant (POP) degradation between the plastisphere and the natural environment.

TPM = transcripts per million. *P < 0.05, **P < 0.01; t-test. The numbers of replicated samples are as follows: freshwater plastisphere (n = 9), freshwater environment (n = 9), seawater plastisphere (n = 7), seawater environment (n = 7), terrestrial plastisphere (n = 3), terrestrial environment (n = 3).

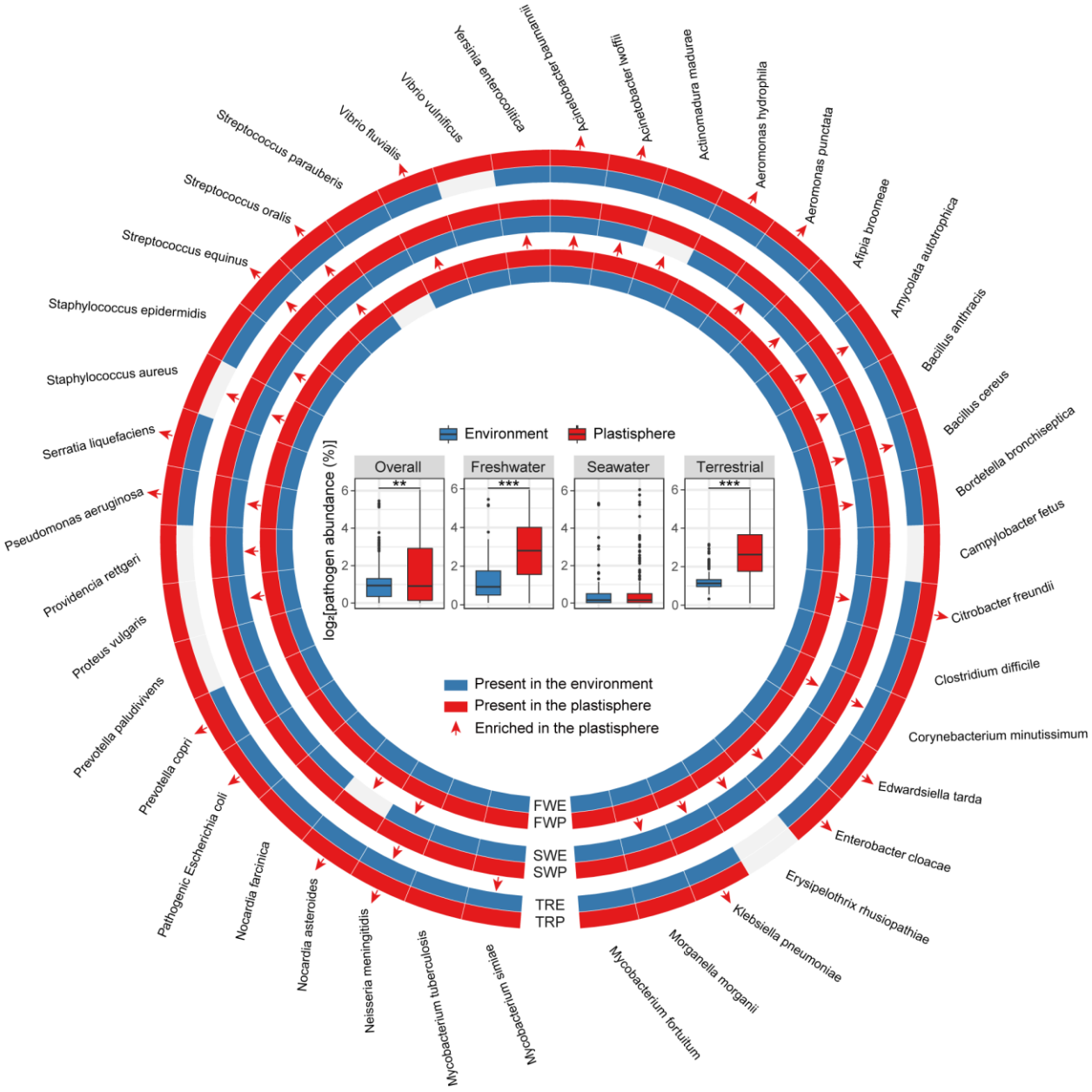

Figure S26 Clinically pathogenic threat of the plastisphere.

Potentially clinical pathogens are identified based on the 16S Pathogenic Identification Process (16SPIP). Box plots showing the difference in the total abundance of the identified pathogens between the plastisphere and the natural environment in each ecosystem (**P < 0.01, ***P < 0.001; Wilcoxon rank sum test). Circle diagram showing the species and number of pathogens that are present in the plastisphere and the natural environment in each ecosystem, and that are enriched in the plastisphere in each ecosystem (P < 0.05; ns = non-significant; Wilcoxon rank sum test). FWP = freshwater plastisphere (n = 120); FWE = freshwater environment (n = 143); SWP = seawater plastisphere (n = 300); SWE = seawater environment (n = 132); TRP = terrestrial plastisphere (n = 170); TRE = terrestrial environment (n = 148).

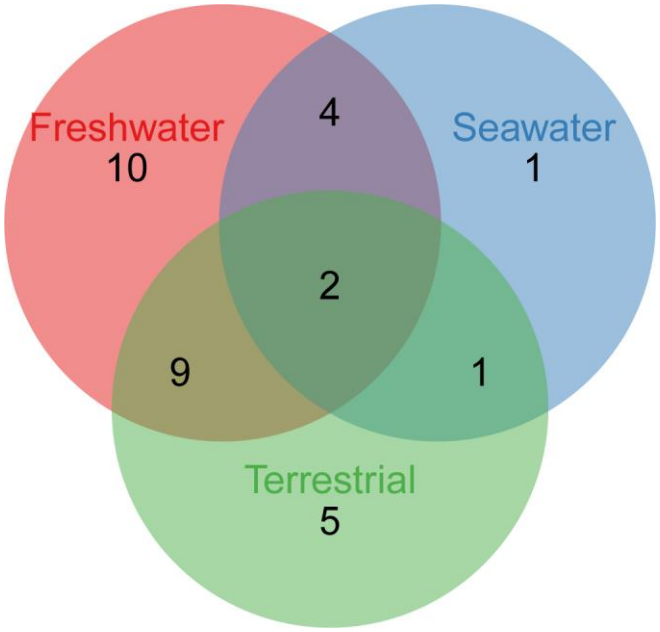

Figure S27 Venn diagram showing that the plastisphere-enriched clinically pathogenic species vary greatly among different ecosystems.

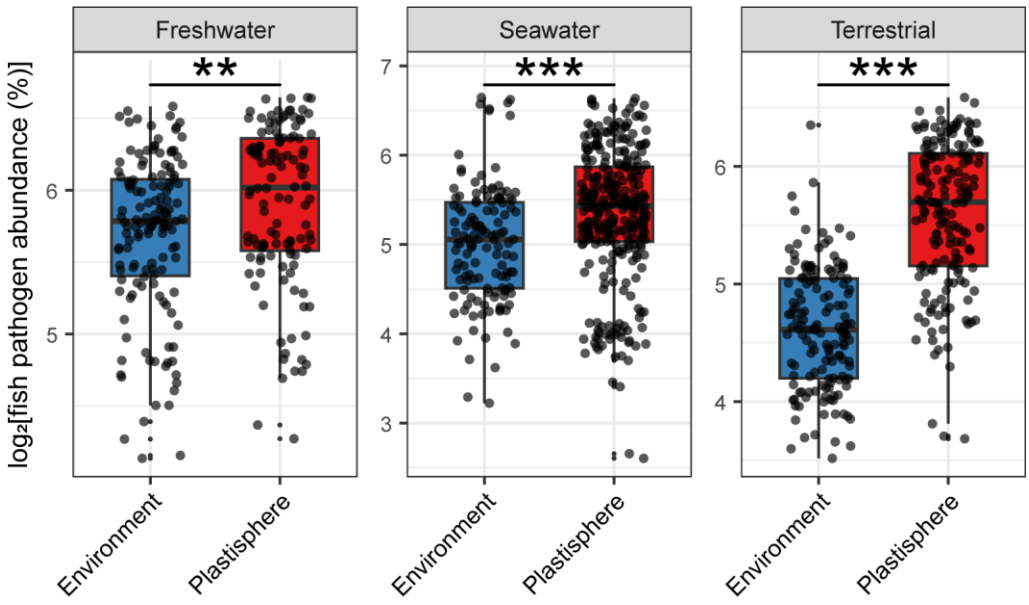

Figure S28 Comparison of the abundance of fish pathogens between the plastisphere and the natural environment.

***P* < 0.01, ****P* < 0.001; Wilcoxon rank sum test. The numbers of replicated samples are as follows: freshwater plastisphere (*n* = 120), freshwater environment (*n* = 143), seawater plastisphere (*n* = 300), seawater environment (*n* = 132), terrestrial plastisphere (*n* = 170), terrestrial environment (*n* = 148).

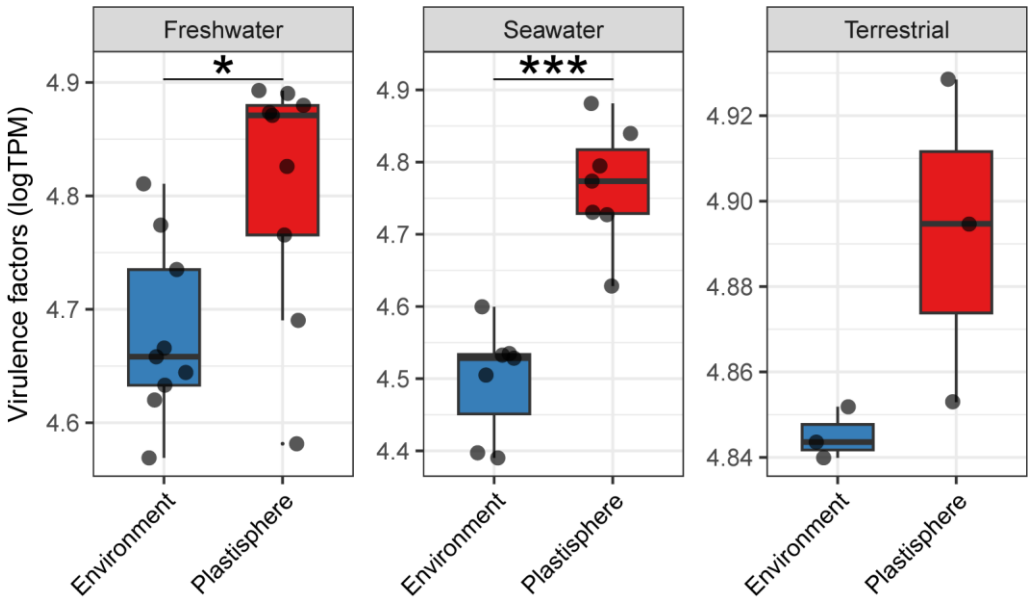

Figure S29 Comparison of the abundance of genes encoding for virulence factors between the plastisphere and the natural environment.

TPM = transcripts per million. *$P < 0.05$, ***$P < 0.001$; *t*-test. The numbers of replicated samples are as follows: freshwater plastisphere (n = 9), freshwater environment (n = 9), seawater plastisphere (n = 7), seawater environment (n = 7), terrestrial plastisphere (n = 3), terrestrial environment (n = 3).

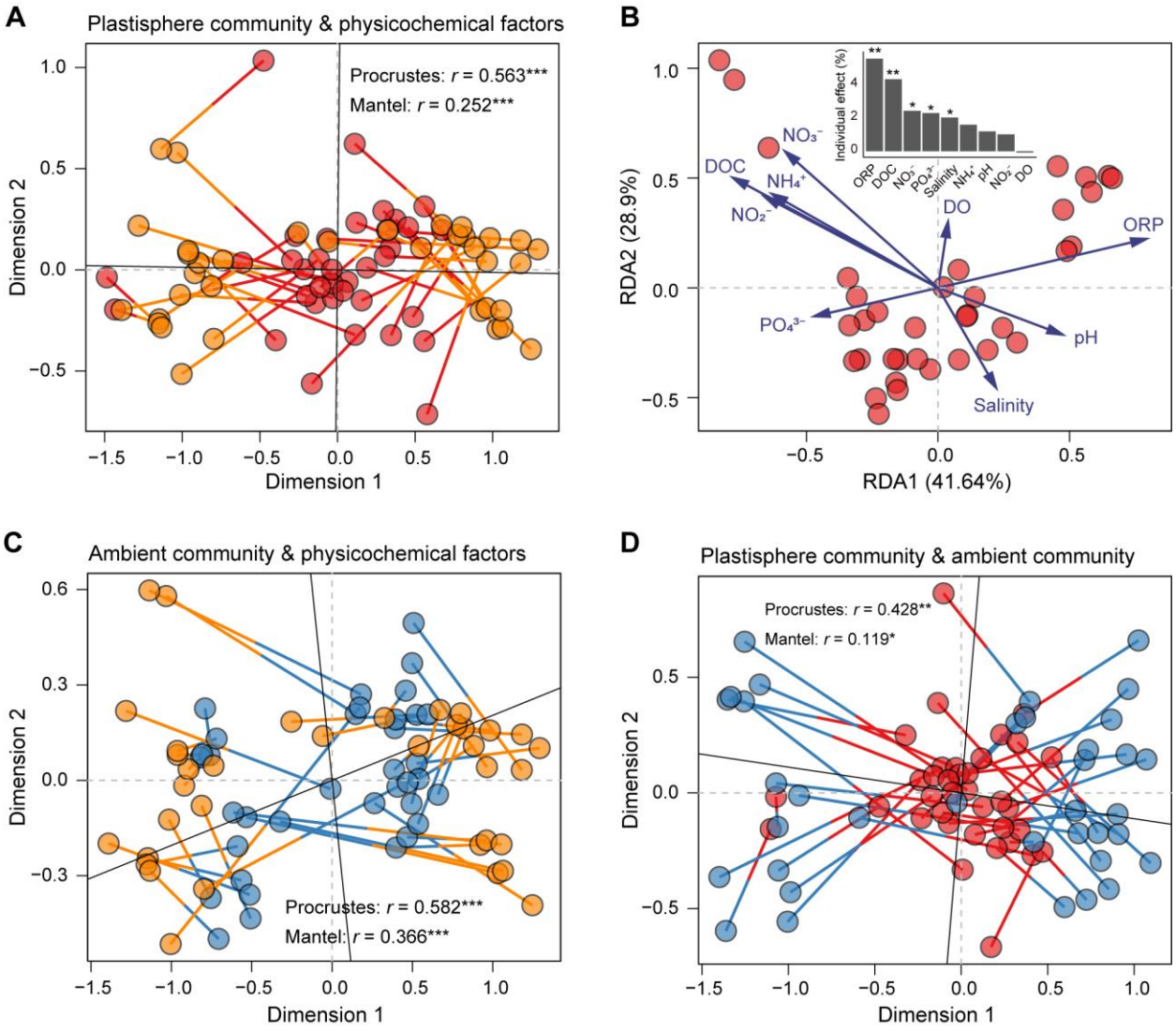

Figure S30 Driving factors of the plastisphere microbiome.

(**A**) Correlations between the plastisphere community and environmental physicochemical properties revealed by Procrustes analysis and Mantel test. (**B**) Potential environmental drivers of the plastisphere microbiome revealed by distance-based redundancy analysis (db-RDA). Correlations between the ambient community and environmental physicochemical properties (**C**), and between the plastisphere community and the ambient community (**D**) revealed by Procrustes analysis and Mantel test. ORP = oxidation-reduction potential, DOC = dissolved organic carbon, DO = dissolved oxygen. *P < 0.05, **P < 0.01, ***P < 0.001.

SUPPLEMENTAL REFERENCES

1. Cheng, J., Jacquin, J., Conan, P., et al. (2021). Relative influence of plastic debris size and shape, chemical composition and phytoplankton-bacteria interactions in driving seawater plastisphere abundance, diversity and activity. Front. Microbiol. ***11***, 610231.

2. Delacuvellerie, A., Ballerini, T., Frère, L., et al. (2022). From rivers to marine environments: A constantly evolving microbial community within the plastisphere. Mar. Pollut. Bull. ***179***, 113660.

3. Du, Y., Liu, X., Dong, X., and Yin, Z. (2022). A review on marine plastisphere: biodiversity, formation, and role in degradation. Comput. Struct. Biotechnol. J. ***20***, 975-988.

4. Frère, L., Maignien, L., Chalopin, M., et al. (2018). Microplastic bacterial communities in the Bay of Brest: Influence of polymer type and size. Environ. Pollut. ***242***, 614-625.

5. Shi, X., Chen, Z., Wei, W., et al. (2023). Toxicity of micro/nanoplastics in the environment: Roles of plastisphere and eco-corona. Soil & Environmental Health ***1***, 100002.

6. Trojánek, Z., Kovařík, A., Španová, A., et al. (2018). Application of magnetic polymethacrylate-based microspheres for the isolation of DNA from raw vegetables and processed foods of plant origin. J. Food Process. Pres. ***42***, e13384.

7. Raimundo, J., Reis, C.M.G., and Ribeiro, M.M. (2018). Rapid, simple and potentially universal method for DNA extraction from *Opuntia* spp. fresh cladode tissues suitable for PCR amplification. Mol. Biol. Rep. ***45***, 1405-1412.

8. Wright, R.J., Langille, M.G.I., and Walker, T.R. (2021). Food or just a free ride? A meta-analysis reveals the global diversity of the Plastisphere. ISME J. ***15***, 789-806.

9. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics ***26***, 2460-2461.

10. Rognes, T., Flouri, T., Nichols, B., et al. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ ***4***, e2584.

11. Wang, Q., Garrity George, M., Tiedje James, M., and Cole James, R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microb. ***73***, 5261-5267.

12. Thompson, L.R., Sanders, J.G., McDonald, D., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature ***551***, 457-463.

13. Lai, J., Zou, Y., Zhang, J., and Peres-Neto, P.R. (2022). Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. Methods in Ecology and Evolution ***13***, 782-788.

14. Shenhav, L., Thompson, M., Joseph, T.A., et al. (2019). FEAST: fast expectation-maximization for microbial source tracking. Nat. Methods ***16***, 627-632.

15. Jiao, S., Yang, Y., Xu, Y., et al. (2020). Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. ISME J. ***14***, 202-216.

16. Wu, W., Lu, H.-P., Sastri, A., et al. (2018). Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. ISME J. ***12***, 485-494.

17. Levins, R. (1968). Evolution in changing environments (Princeton University Press).

18. Xu, Q., Vandenkoornhuyse, P., Li, L., et al. (2021). Microbial generalists and specialists differently contribute to the community diversity in farmland soils. J. Adv. Res. ***40***, 17-27.

19. Ning, D., Deng, Y., Tiedje James, M., and Zhou, J. (2019). A general framework for quantitatively assessing ecological stochasticity. Proc. Natl. Acad. Sci. USA ***116***, 16892-16898.

20. Louca, S., Parfrey Laura, W., and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. Science ***353***, 1272-1277.

21. Gambarini, V., Pantos, O., Kingsbury, J.M., et al. (2022). PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. Database ***2022***, baac008.

22. Ngara, T.R., Zeng, P., and Zhang, H. (2022). mibPOPdb: An online database for microbial biodegradation of persistent organic pollutants. iMeta ***1***, e45.

23. Yang, X., Jiang, G., Zhang, Y., et al. (2023). MBPD: A multiple bacterial pathogen detection pipeline for One Health practices. iMeta ***2***, e82.

24. Zhao, J., Jin, L., Wu, D., et al. (2022). Global airborne bacterial community—interactions with Earth's microbiomes and anthropogenic activities. Proc. Natl. Acad. Sci. USA ***119***, e2204465119.

25. Li, D., Van De Werfhorst, L.C., Dunne, T., et al. (2020). Surf zone microbiological water quality following emergency beach nourishment using sediments from a catastrophic debris flow. Water Res. ***176***, 115733.

26. Li, D., Van De Werfhorst, L.C., Rugh, M.B., et al. (2021). Limited bacterial removal in full-scale stormwater biofilters as evidenced by community sequencing analysis. Environ. Sci. Technol. ***55***, 9199-9208.

27. Miao, J., Han, N., Qiang, Y., et al. (2017). 16SPIP: a comprehensive analysis pipeline for rapid pathogen detection in clinical samples based on 16S metagenomic sequencing. BMC Bioinformatics ***18***, 568.

28. Drønen, K., Roalkvam, I., Tungland, K., et al. (2023). How to define fish pathogen relatives from a 16S rRNA sequence library and Pearson correlation analysis between defined OTUs from the library: Supplementary data to the research article "Presence and habitats of bacterial fish pathogen relatives in a marine salmon post-smolt RAS". Data Brief ***46***, 108846.

29. Xu, C., Lu, J., Shen, C., et al. (2022). Deciphering the mechanisms shaping the plastisphere antibiotic resistome on riverine microplastics. Water Res. ***225***, 119192.

30. Zhu, D., Ma, J., Li, G., et al. (2022). Soil plastispheres as hotspots of antibiotic resistance genes and potential pathogens. ISME J. ***16***, 521-532.

31. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics ***34***, i884-i890.

32. Li, D., Liu, C.-M., Luo, R., et al. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. Bioinformatics ***31***, 1674-1676.

33. Hyatt, D., Chen, G.-L., LoCascio, P.F., et al. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics ***11***, 119.

34. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics ***22***, 1658-1659.

35. Patro, R., Duggal, G., Love, M.I., et al. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods ***14***, 417-419.

36. Tu, Q., Lin, L., Cheng, L., et al. (2019). NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. Bioinformatics ***35***, 1040-1048.

37. Lombard, V., Golaconda Ramulu, H., Drula, E., et al. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. ***42***, D490-D495.

38. Liu, B., Zheng, D., Zhou, S., et al. (2022). VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. ***50***, D912-D917.

39. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE ***11***, e0163962.

40. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods ***12***, 59-60.

41. Coons, A.K., Busch, K., Lenz, M., et al. (2021). Biogeography rather than substrate type determines bacterial colonization dynamics of marine plastics. PeerJ ***9***, e12135.

42. Erni-Cassola, G., Wright, R.J., Gibson, M.I., and Christie-Oleza, J.A. (2020). Early colonization of weathered polyethylene by distinct bacteria in marine coastal seawater. Microb. Ecol. ***79***, 517-526.

43. Hoellein, T., Rojas, M., Pink, A., et al. (2014). Anthropogenic litter in urban freshwater ecosystems: distribution and microbial interactions. PLoS ONE ***9***, e98485.

44. Hoellein, T.J., McCormick, A.R., Hittie, J., et al. (2017). Longitudinal patterns of microplastic concentration and bacterial assemblages in surface and benthic habitats of an urban river. Freshwater Science ***36***, 491-507.

45. Kirstein, I.V., Wichels, A., Krohne, G., and Gerdts, G. (2018). Mature biofilm communities on synthetic polymers in seawater - Specific or general? Mar. Environ. Res. ***142***, 147-154.

46. Kirstein, I.V., Wichels, A., Gullans, E., et al. (2019). The Plastisphere – Uncovering tightly attached plastic "specific" microorganisms. PLoS ONE ***14***, e0215859.

47. McCormick, A., Hoellein, T.J., Mason, S.A., et al. (2014). Microplastic is an abundant and distinct microbial habitat in an urban river. Environ. Sci. Technol. ***48***, 11863-11871.

48. McCormick, A.R., Hoellein, T.J., London, M.G., et al. (2016). Microplastic in surface waters of urban rivers: concentration, sources, and associated bacterial assemblages. Ecosphere ***7***, e01556.

49. Oberbeckmann, S., Osborn, A.M., and Duhaime, M.B. (2016). Microbes on a bottle: substrate, season and geography influence community composition of microbes colonizing marine plastic debris. PLoS ONE ***11***, e0159289.

50. Ogonowski, M., Motiei, A., Ininbergs, K., et al. (2018). Evidence for selective bacterial community structuring on microplastics. Environ. Microbiol. ***20***, 2796-2808.

51. Parrish, K., and Fahrenfeld, N.L. (2019). Microplastic biofilm in fresh- and wastewater as a function of microparticle type and size class. Environmental Science: Water Research & Technology ***5***, 495-505.

52. Pinto, M., Langer, T.M., Hüffer, T., et al. (2019). The composition of bacterial communities associated with plastic biofilms differs between different polymers and stages of biofilm succession. PLoS ONE ***14***, e0217165.

53. Wu, X., Pan, J., Li, M., et al. (2019). Selective enrichment of bacterial pathogens by microplastic biofilm. Water Res. ***165***, 114979.

54. Zhang, M., Zhao, Y., Qin, X., et al. (2019). Microplastics from mulching film is a distinct habitat for bacteria in farmland soil. Sci. Total Environ. ***688***, 470-478.

55. Zhang, J., Liu, Y.-X., Zhang, N., et al. (2019). *NRT1.1B* is associated with root microbiota composition and nitrogen use in field-grown rice. Nat. Biotechnol. ***37***, 676-684.

56. Yuan, J., Wen, T., Zhang, H., et al. (2020). Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. ISME J. ***14***, 2936-2950.

57. Li, C., Wang, L., Ji, S., et al. (2021). The ecology of the plastisphere: Microbial composition, function, assembly, and network in the freshwater and seawater ecosystems. Water Res. ***202***, 117428.

58. Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News ***2***, 18-22.