Additional information for

Diverse functional genes harbored in extracellular vesicles from environmental and human microbiota

Authors: Li-Ting Zhu[1,2#], Hai-Ning Huang[1#], Ricardo David Avellán-Llaguno[1], Yifei Qin[1,2], Xin-Li An[1], Jian-Qiang Su[1], Qiansheng Huang[1,3*], Yong-Guan Zhu[1*]

Correspondence to: Q.S. H (qshuang@iue.ac.cn), Y.G. Z (ygzhu@iue.ac.cn)

This file includes:

Supplementary Methods

Extended Data Figs. 1 to 8

Captions for supplementary table 1 to 9

Other Supplementary Materials for this manuscript include the following:

Supplementary Table 1 to 9 as separate Excel files

Supplementary Methods

Isolation of an ampicillin-resistant *E. coli* strain E4742. *E. coli* strain E4742 was isolated from wastewater following the procedure as below. The wastewater sample was diluted 1,000 times with PBS, and then 100 µl diluent was spread on LB solid medium containing 64 mg/l Ampicillin (Sigma-Aldrich) for 24 h at 37°C. Colonies were picked and purified on the ampicillin solid medium for 2 to 3 times to isolate resistant bacteria. A single resistant bacterial colony was obtained and applied for 16S rRNA gene sequencing. The primer sequences were 16S-27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 16S-1492R (5'-GGTTACCTTGTTACGACTT-3'). The reaction conditions consisted of 1 cycle of 5 min at 94°C for the initial denaturation, followed by 30 cycles of denaturation (95°C, 30 s), annealing (56°C, 30 s), and extension (72°C, 90 s), and a final extension step at 72°C for 10 min. After the PCR products were isolated by 1% agarose gel electrophoresis, Sanger sequencing was performed using ABI 3730XL DNA Analyzer (Applied Biosystems). The sequenced 16S rRNA sequence was blasted on the EzBioCloud's database website (http://www.ezbiocloud.net) and was found to show high similarity with that of *E. coli* strain E4742 [60].

Isolation of EVs from faeces by iodixanol density gradient centrifugation. Bottom-up density gradient centrifugation was performed according to a previous protocol with some modifications[61,62]. Briefly, 10%, 20% and 40% iodixanol density media (Sigma-Aldrich) were prepared by mixing with gradient buffer immediately. 50% iodixanol were prepared by mixing with a crude extract. The bottom-loaded 10%, 20%, 40%, and 50% gradients and 1 mL of PBS on the surface of the liquid were subjected to ultracentrifugation at 100,000 × g for 16 h at 4°C using a SW41 Ti rotor. Twelve individual fractions (each with 1 ml) were collected from the top of the centrifugation tube. EVs were enriched in the fractions with density from 1.083 to 1.201

g/ml, which was around 4-9 fraction from top to bottom. Combined fractions were resuspended in PBS and subjected to ultracentrifugation at 100,000 × g for 2 h at 4°C.

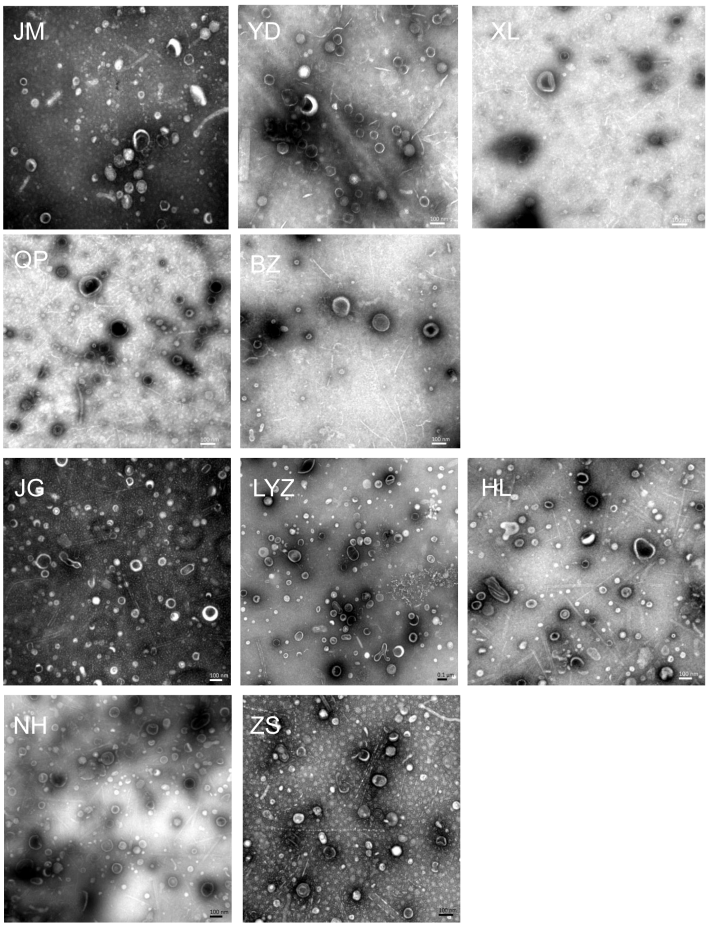
Transmission electron microscopy (TEM). 8 µl of EVs samples were placed on carbon-coated 300-mesh copper grid. After 5 min, the grid was washed with 1 mM EDTA (pH=8.0) for 10 s, stained with 2% uranyl acetate for 1 min, then allowed to dry. The image was taken using transmission electron microscope (Hitachi S4800, Japan) at an accelerating voltage of 100 kV.

Scanning Electron Microscopy (SEM). Culture media (1 ml) from *E. coli* E4742 at exponential growth stage was centrifuged (3,000 g for 10 min) to collect the bacteria. The bacteria pellet was rinsed twice with 0.1 M PBS and fixed with 2.5% glutaraldehyde overnight at 4°C. Then, the samples were dehydrated through a series of graded concentrations of ethanol (30%, 50%, 70%, 80%, 90%, 100% ethanol; 10 min each). After dehydration, the samples covered with filter paper were dried with a critical point dryer (Autosamdri 815 automatic critical point dryer apparatus, Tousimis, USA) overnight, sputter coated with gold, and then examined under an electron microscope (Hitachi S4800, Japan).

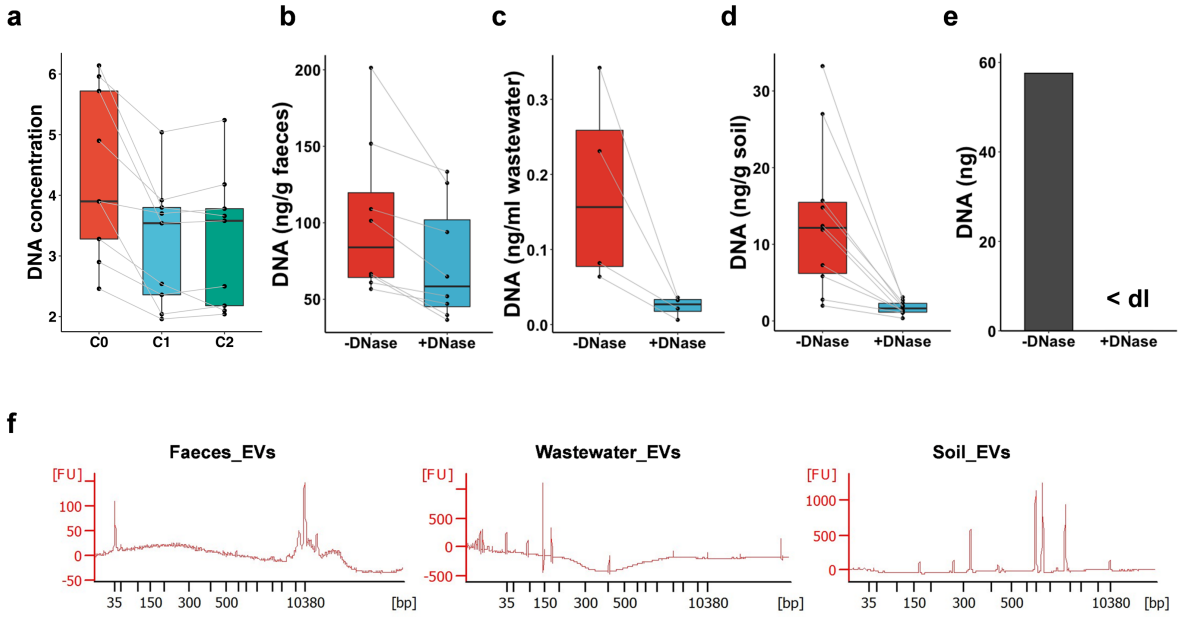
Absolute quantification of 16S rRNA gene copies. The absolute copy number of 16S rRNA gene was measured by Roche 480 using a SYBR® Green approach. A volume of 20 µl reaction solution was prepared including 10 µl 2× Light Cycler 480 SYBR® Green I Master Mix (Roche Inc., USA), 1 µM of each primer (341 F: 5′- CCTACGGGNGGCWGCAG-3′ and 806 R: 5′- GGACTACHVGGGTWTCTAAT-3′), 1 µl DNA template, and 7 µl nuclease-free water. PCR amplification was carried out with Roche 480 using the protocol of initial pre-incubation at 95°C for 5 min, 40 cycles of 95°C for 15 s, 60°C for 1 min, and 72°C for 15 s. An eight-point calibration curve was made by 10-fold dilution of a standard plasmid containing a cloned and

sequenced 16S rRNA gene fragment (1.39×10^{10} copies/l) to apply external standard calculation. The PCR reaction was run in triplicate with negative and positive controls.

Comparison of EVs-associated 16S rDNA copies between differential centrifugation and density gradient centrifugation. The absolute quantification of 16S rRNA genes was applied to estimate the concentration of bacteria in the field samples contributing DNA to the EVs. Each of three fecal samples was divided equally into two, and then applied to differential centrifugation and density gradient centrifugation, respectively. Then, the obtained EVs were treated with DNase, and the extracted DNA were applied to 16S rDNA quantification.

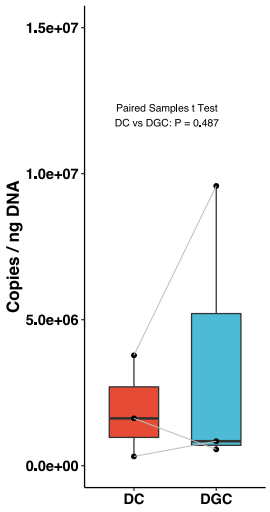


Extended Data Fig. 1| Representative transmission electron microscopy of EVs from wastewater samples (JM, YD, XL, QP, BZ) and soil samples (JG, LYZ, HL, NH, ZS). scale bar, 100 nm.

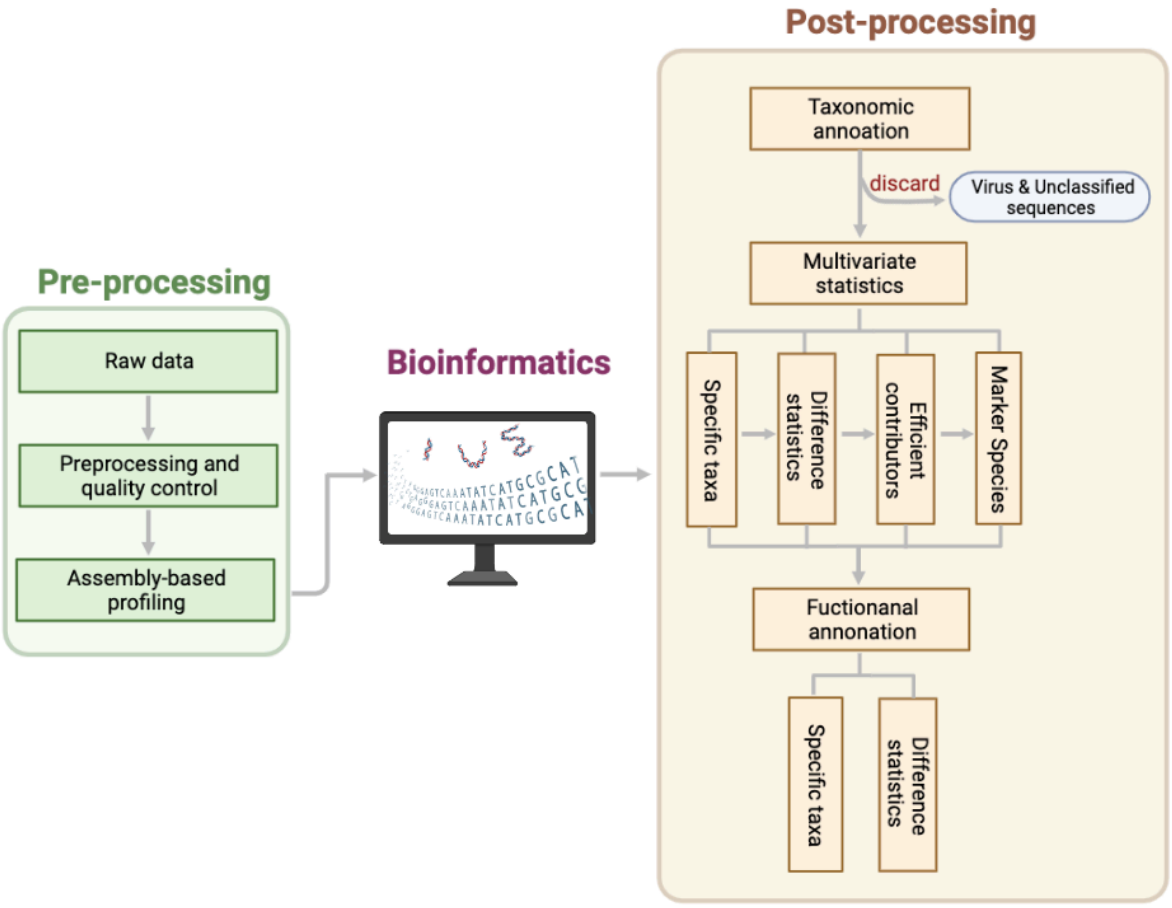


Extended Data Fig. 2| Quantitation and characterization of EVs-associated DNA. a, C0 is the initial DNA concentration of the sample, C1 is the DNA concentration after 1st DNase treatment, C2 is the DNA concentration after 2nd DNase treatment. EVs from the same faeces were linked by grey lines. EVs from **b**, faeces, **c**, wastewater, **d,** soil were treated with two rounds of 2 U Turbo DNase at 37°C for 30 min. Then, the levels of double strand DNA were assayed by Qubit 4.0 Fluorometer. Data were expressed as (middle line, median; lower hinge, first quartile; upper hinge, third quartile; lower whisker, the smallest value at most 1.5 × the interquartile range from the hinge; upper whisker, the largest value no further than 1.5 × the interquartile range from the hinge; data beyond the whiskers were outlying points). Each sample before and after DNase treatment were linked by grey lines. **e**, Total DNA from wastewater bacteria (after dilution) was used as control. '< dl' indicates that concentrations were below the assay detection limit (10 pg/µl). **f**, Bioanalyzer electropherograms of Size distribution of EVs-associated DNA fragments. Purified EVs-associated DNA from faeces, wastewater and soil, pre-

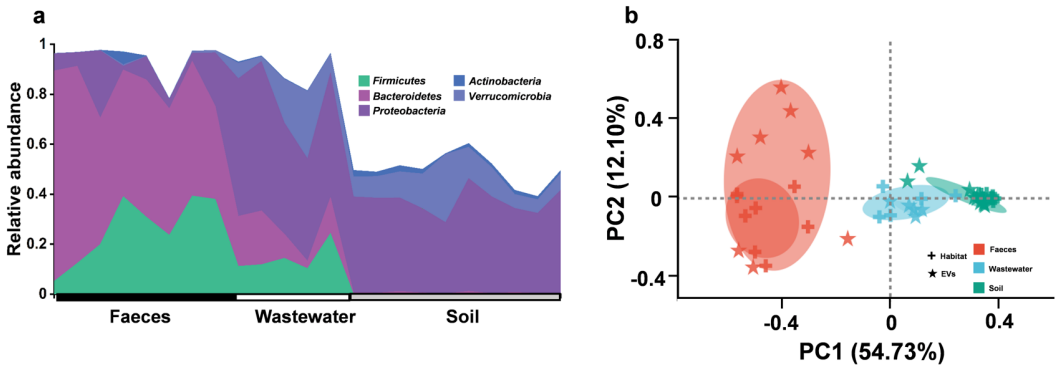
93 treated with DNase was analyzed with an Agilent Bioanalyzer High Sensitivity DNA assay. One
94 representative figure for each EVs population were shown. Peaks at 35 bp and 10380 bp were
95 from exogenous standard DNA fragments. FU, fluorescence units.
96



Extended Data Fig. 3| Quantitation of 16S rDNA copies of fecal EVs derived from differential centrifugation (DC) and density gradient centrifugation (DGC), respectively. **Half of each** fecal sample **were applied to** DC and DGC process, respectively. Then, EVs were treated with DNase treatment before DNA extraction. The red ones represent the DC group, the blue ones represent the DGC group, and the dots represent the samples. Data were expressed as (middle line, median; lower hinge, first quartile; upper hinge, third quartile; lower whisker, the smallest value at most 1.5 × the interquartile range from the hinge; upper whisker, the largest value no further than 1.5 × the interquartile range from the hinge; data beyond the whiskers were outlying points). EVs from the same faeces were linked by grey lines. Paired sample t-test was used for significance test. n=3.

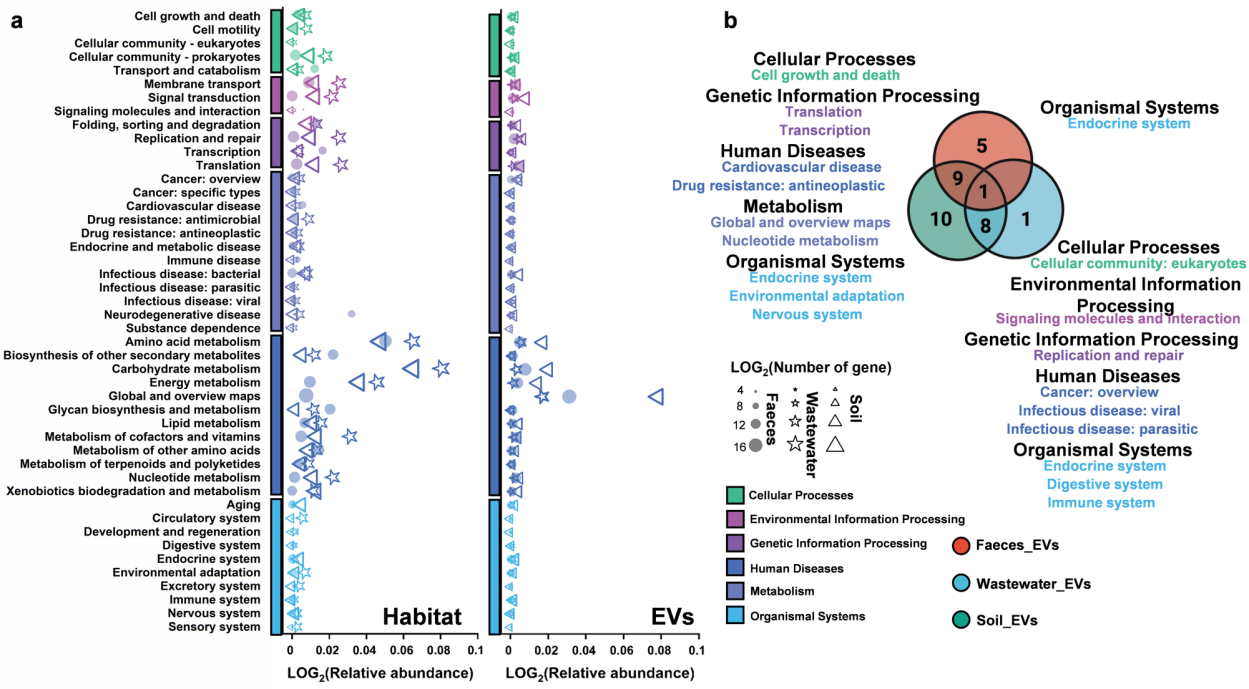


Extended Data Fig. 4| Summary of a bioinformatic workflow. This flow included pre-processing and post-processing for bioinformatic analysis.

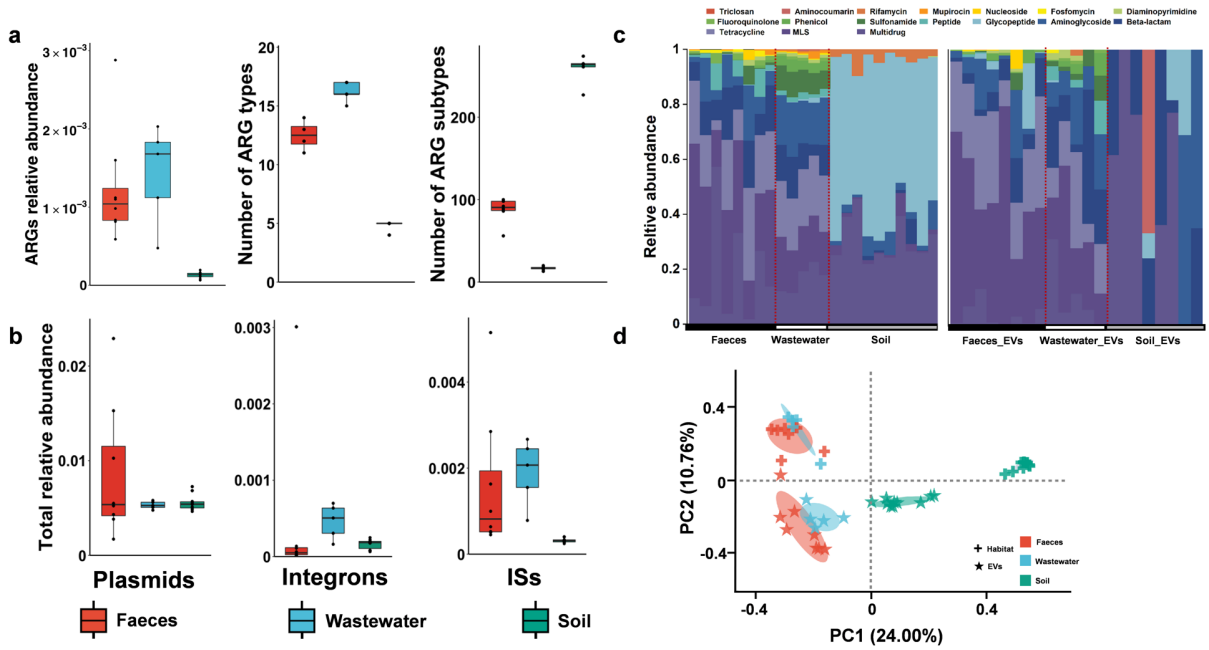


113

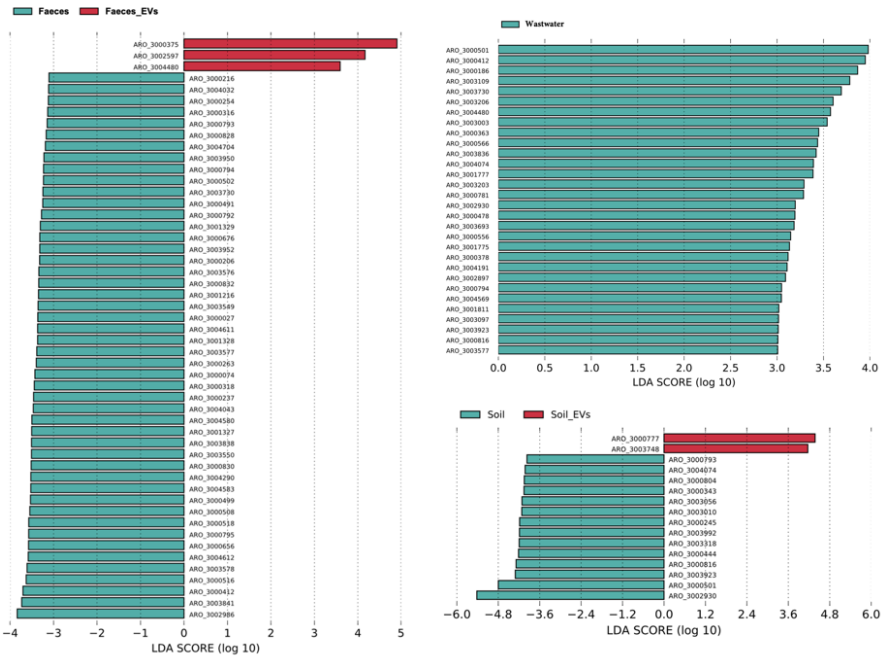
114 **Extended Data Fig. 5| Composition of microbiota and EVs in different habitats. a**, The
115 median relative abundance of microbes at the phylum level (faeces, n = 8; wastewater, n=5; soil,
116 n=10). **b**, Principal co-ordinates analysis (PCoA) plots of Bray-Curtis distance of microbial
117 composition at the phylum level. The composition was driven by different habitats. Each dot
118 represents one individual sample. Sampling types are colored, with plus and asterisk indicating
119 microbiota samples and EVs samples, respectively. Ellipses represented the 95% confidence
120 interval.



Extended Data Fig. 6| Diversity and distribution of functional categories in the habitat samples and corresponding EVs samples. a, Comparison between the habitats and EVs on level 2 of KEGG functional category. The dots represent samples of faeces and fecal vesicles, the stars indicate wastewater and EVs of wastewater, and the triangles show soil and EVs of soil. Different colors indicate different functional categories. **b**, The Venn diagram shows the number of overlapping and unique KEGG categories.



Extended Data Fig. 7| Characterization of ARG and MGE profiles in the microbiota from each habitat. a, Different levels of ARGs, including total relative abundance, antibiotic types, and antibiotic subtypes. Data were presented as box plots (middle line, median; lower hinge, first quartile; upper hinge, third quartile; lower whisker, the smallest value at most 1.5 × the interquartile range from the hinge; upper whisker, the largest value no further than 1.5 × the interquartile range from the hinge; data beyond the whiskers were outlying points). **b**, Total relative abundance of MGEs (plasmids, integrons and ISs). **c**, Characterization of ARG profiles in EVs and microbiota from corresponding habitats. The bars indicate the proportion of different ARG types as marked by colors. **d**, The interindividual variation of ARGs composition. PCoA plots of Bray-Curtis distance of the ARG subtypes in different samples. Each dot represents one sample. Sampling types are colored, with plus and asterisk indicating microbiota from habitat and EVs samples, respectively.



Extended Data Fig. 8| Characterization of ARG profiles. Differentially abundant ARGs at type level between EVs and their habitats were revealed using non-parametric factorial Kruskal-Wallis (KW) sum-rank test and the linear discriminate analysis (LDA) effect size (LEfSe) method. $p < 0.05$ and LDA score > 3.0 are considered statistically significant.

Supplementary Table 1. Sampling information and sample characteristics.

Various environmental factors were monitored for each type of habitat. For faeces samples, we recorded subject metadata including intrinsic factors (e.g. gender, age, BMI), bowel habits, medication, as well as diet (e.g. smoke, drink, staple food). For wastewater samples, water quality including temperature, pH, dissolved oxygen (DO) and conductivity (EC) were measured in situ using a portable probe (HachHQ40d, Loveland, CO, USA). For soil samples, soil characteristics including water content, pH, total organic carbon (TOC), total nitrogen (TN), total carbon (TC), total phosphorus (TP) and carbon nitrogen ratio (TC/TN) were detected.

Supplementary Table 2. Correlation between sample characteristics and particle concentrations of EVs.

We performed pairwise Spearman correlation to analyze the correlation between particle concentrations of EVs and habitat environmental factors. In the Spearman correlation analysis between the number of bacteria and the EVs particles, the data of three habitats samples were pool together for analysis. For the other factors, correlation analysis was performed individually within respective samples. Notably, *, $p < 0.05$; **, $p < 0.01$.

Supplementary Table 3. The proportion of EVs-associated DNA fragments from 200 bp to 1000 bp.

The proportion of EVs-associated DNA from faeces, wastewater and soil. Size ranged from 200bp to 1000bp.

Supplementary Table 4. Metagenomic data of each sample after quality control.

Based on the original sequencing data, fastp software was used to conduct data quality control, and low-quality and n-containing reads in the data were cut out to obtain high-quality sequences required for subsequent analysis. In addition, host contamination was removed from fecal samples. This table includes description of data production, including number of contigs, number of bp in contigs, contig N50, contig N90, the longest contig (Max), the shortest contig (Min), optimized reads number, optimized sequence length (bp), and their percentages.

Supplementary Table 5. Biological proportion and classification of samples at the phylum level.

The proportion of each sample at the level of phylum were analyzed, and the five phyla were divided into Gram-negative bacteria (GNB) and Gram-positive bacteria (GPB).

Supplementary Table 6. Detailed information of species that were enriched in EVs.

Species that were significantly enriched in EVs were selected as efficient producers (one-tailed Wilcoxon rank-sum test with FDR correction). The information includes abundance (mean and median) and *p* values for each species in both habitats and EVs samples.

Supplementary Table 7. Detailed information of the differential KEGG profiles secreted by major contributors *Firmicutes* and *Proteobacteria* between EVs and habitats.

The information includes Phylum, different levels of genetic description and Gene count of each encoding gene in the EVs samples. The number of gene count was the number of genes enriched in the EVs.

Supplementary Table 8. Detailed information of enzyme encoding genes that were enriched in EVs.

Wilcoxon unit test was used to select enzymes that were significantly enriched in EVs (one-tailed Wilcoxon rank-sum test with FDR correction). The information includes the abundance (mean and median) and *p* values of each enzyme encoding gene in the habitats and EVs samples.

Supplementary Table 9. Detailed information of the differential enzyme encoding genes secreted by major contributors *Firmicutes* and *Proteobacteria* between EVs and habitats.

The information includes Phylum, genetic family and Gene count of each encoding gene in the EVs samples. The number of gene count was the number of genes enriched in the EVs.

References

60 Yoon, S. H. *et al.* Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* **67**, 1613-1617, doi:10.1099/ijsem.0.001755 (2017).

61 Tulkens, J., De Wever, O. & Hendrix, A. Analyzing bacterial extracellular vesicles in human body fluids by orthogonal biophysical separation and biochemical characterization. *Nat Protoc* **15**, 40-67, doi:10.1038/s41596-019-0236-5 (2020).

62 Zou, Y. *et al.* 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol* **37**, 179-185, doi:10.1038/s41587-018-0008-8 (2019).