

RESEARCH

Open Access



# Transmission of the human respiratory microbiome and antibiotic resistance genes in healthy populations

Lili Ren<sup>1,2†</sup>, Jing Yang<sup>3,4†</sup>, Yan Xiao<sup>1,2†</sup>, Li Guo<sup>1,5†</sup>, Jian Rao<sup>6</sup>, Chao Wu<sup>1,5</sup>, Xinming Wang<sup>1</sup>, Ying Wang<sup>1</sup>, Linfeng Zhang<sup>3</sup>, Li Zhang<sup>3</sup>, Xiaoqing Jiang<sup>3</sup>, Jiaxin Zhong<sup>3,4,6</sup>, Jingchuan Zhong<sup>1</sup>, Weizhong Yang<sup>7</sup>, Chen Wang<sup>7</sup>, Jianwei Wang<sup>1,2\*†</sup> and Mingkun Li<sup>3,4\*†</sup>

## Abstract

**Background** The human microbiome is transmissible between individuals, including pathogens and commensals with metabolic and immune-modulating effects, which could influence susceptibility, severity, and outcomes of both infection and non-infection diseases. However, limited studies of respiratory microbiome transmission within populations have been conducted. Herein, we performed species- and strain-level metagenomic analyses on oropharyngeal (OP) swabs from 1046 healthy urban dwellers across 13 districts, including 111 households with at least two cohabitants, to elucidate the transmission dynamics of the respiratory microbiome within households and communities.

**Results** We found that geographic districts accounted for the greatest variation in the OP microbiome, with unrelated individuals from the same district showing greater microbiome similarity and higher strain-sharing rates than those from different districts. Cohabitants, especially spouses and siblings, exhibited similar microbial abundances and shared more strains, with 16.7% (IQR 0.0–33.3%) of strains shared among cohabitants, compared to 0.0% (IQR 0.0–11.1%) in non-cohabiting pairs ( $p < 0.05$ ). Both respiratory commensals and opportunistic pathogens were shared among cohabitants. In contrast, no evidence of vertical transmission was detected between mother–offspring pairs. Additionally, the OP microbiome contained diverse antibiotic resistance genes (ARGs), with 15.0% linked to mobile genetic elements (MGEs) or plasmids; the flanking sequences of these ARGs were more conserved across species than those of non-MGE-associated ARGs, suggesting horizontal transfer of ARGs among respiratory microorganisms.

**Conclusions** In summary, we characterized the transmissible nature of the OP microbiome and the risk of ARG dissemination among respiratory microorganisms. These findings underscore the role of respiratory microbes and ARGs

<sup>†</sup>Lili Ren, Jing Yang, Yan Xiao and Li Guo contributed equally to this study.

<sup>†</sup>Jianwei Wang and Mingkun Li contributed equally as senior authors to this work.

\*Correspondence:

Jianwei Wang  
wangjw28@163.com  
Mingkun Li  
limk@big.ac.cn

Full list of author information is available at the end of the article



exchange in shaping the microbiome of healthy populations and emphasize their relevance to public health strategies for respiratory health management.

**Keyword** Metagenomics, Respiratory microbiome, Microbiome transmission, Antibiotic resistance genes dissemination

## Introduction

The airway mucosa is colonized by various bacteria, fungi, viruses, and archaea that comprise the respiratory microbiome, and its composition and role in respiratory health have been preliminary revealed [1, 2]. The upper respiratory microbiome serves as a gatekeeper, defending against the invasion of respiratory pathogens through direct competition and interaction with the host immune response [3]. Previous studies reported that the oropharyngeal (OP) microbiome impacts the disease progression of COVID-19 patients, and the susceptibility to influenza virus infection [4–7]. Moreover, opportunistic pathogens asymptotically colonize the upper respiratory tract, such as *Streptococcus pneumoniae* and *Haemophilus influenzae*, which could be the source of respiratory tract infection [8]. Understanding the ecological characteristics and influencing factors of the OP microbiome could help uncover mechanisms underlying respiratory diseases and inform the development of novel targeted therapies.

Recent studies reveal that the OP microbiome is primarily composed of *Streptococcus*, *Neisseria*, *Veillonella*, and *Prevotella*, and is influenced by mode of delivery, lifestyle, cigarette smoking, age, and air pollutants [9–13]. Additionally, intimacy between individuals significantly affects OP microbiome composition, which may result from the socially transmissible nature of the human microbiome [14–17]. For instance, the upper respiratory microbiomes of children resemble those of their closest siblings, which may explain the associations between asthma and infectious disease risk among siblings [18]. Adults with regular contact with children exhibit an OP microbiota enriched in pathobionts typically carried by children [19]. However, previous findings largely rely on 16S rRNA gene sequencing [18, 19], which has limited resolution and may yield inaccurate conclusions. An in-depth analysis of the microbiome at the species and strain levels is essential for making more accurate inferences about the extent and specific components involved in transmission between individuals.

The human microbiome serves as a reservoir of antibiotic resistance genes (ARGs), which can be transmitted among microorganisms, via plasmids, phages, outer membrane vesicles, and transposons, thereby facilitating the spread of antimicrobial resistance [20]. Genomic and metagenomic data have proven invaluable for profiling

ARGs in individual microorganisms and entire microbiota. While the prevalence and transmission of ARGs have been investigated in healthy populations' gut and oral cavities [21], research on airway resistome in healthy individuals remains limited.

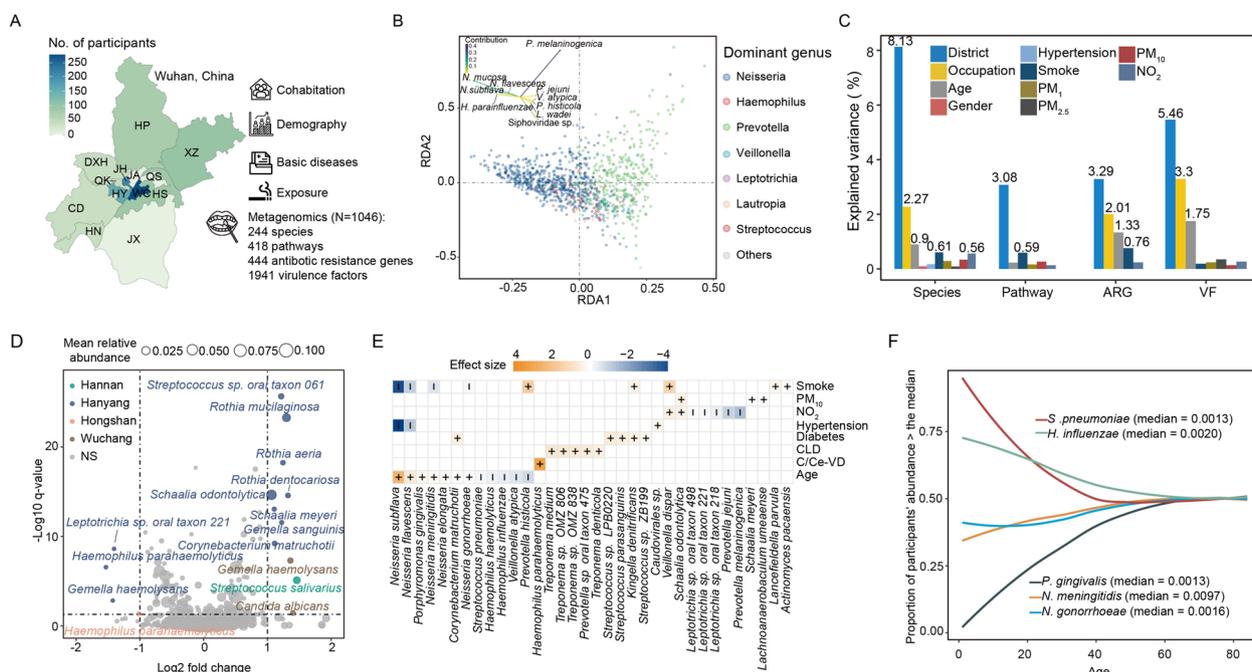
To resolve the aforementioned questions, we conducted high-resolution metagenomic sequencing on OP samples from 1046 individuals. The study aims to investigate: (1) the composition of the OP microbiota, as well as the prevalence and distribution of antibiotic resistance genes (ARGs) and virulence factors (VFs) in healthy individuals; (2) the demographic and environmental variables that correlated with the OP microbiota; (3) the possibility of microbial transmission between cohabiting members; (4) the possibility of ARGs transmission between respiratory microorganisms. In summary, our analyses revealed significant correlations between the OP microbiome and various factors, including district, age, smoking, and the concentrations of PM<sub>10</sub>, NO<sub>2</sub>. Additionally, there are higher species and strain-level similarities in the oropharyngeal microbiota among cohabiting members, particularly those with closer contact, suggesting potential microbiome transmission. Furthermore, our findings indicate that ARGs may transfer between different OP microorganisms via mobile genetic elements, underscoring the risk of disseminating antimicrobial resistance.

## Results

### Overview of the samples and sequencing data

Oropharyngeal (OP) swabs were collected from 1046 participants across 13 districts in Wuhan (with an area of 3280 m<sup>2</sup>), China (Fig. 1A). Among them, 289 participants were from 111 cohabiting households, with 2–5 individuals per household. All participants reported no respiratory symptoms in the month before sampling. Demographic, exposure, and medical data were collected (Supplementary Table S1, Fig. S1A). Environmental exposure data (concentrations of air pollutants NO<sub>2</sub>, PM<sub>1</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>) were obtained from the China High Air Pollutants datasets.

Metagenomic sequencing was performed on all samples and 99 negative controls (NCs). The median proportion of human DNA reads of OP samples was 83.3% (interquartile range (IQR) 66.1–91.5%). The median number of quality-controlled reads of OP samples was 2,144,301 (IQR 910,656–5,028,742), and the median



**Fig. 1** Overview of oropharyngeal microbiome and its associated factors. **A** Schematic diagram for data collection. Geographic heatmap showing the sample numbers from the 13 districts of Wuhan, China. HP: Huangpi, XZ: Xinzhou, DXH: Dongxihu, JH: Jianghan, JA: Jiangnan, QS: Qingshan, QK: Qiaokou, HY: Hanyang, WC: Wuchang, HS: Hongshan, CD: Caidian, HN: Hannan, JX: Jiangxia. **B** Redundancy analysis (RDA) plot of species composition of 1046 OP samples, colored by dominated genus in each sample. The upper left corner of the figure: top 10 species of explaining species compositional variation. The length and the color of each species segment represent the proportion of explained variation by that species, and the direction of the segments was determined by species coordinates. **C** Variation (adjusted  $R^2$ ) explained by each metadata to metagenomic profiles using univariate distance-based redundancy analysis (db-RDA). Only significant results are shown (FDR < 0.05). **D** District-specific species. The species were identified by comparing the relative abundance of each species in one district against that in all other districts using the Wilcoxon rank-sum test (two-sided). The y-axis of the plot represents the logarithmic q value (FDR adjusted  $p$  value) of the test, and the x-axis displays the log2 transformation of the fold change in the relative abundance of the tested microbe, comparing one district to all other districts. The points were colored by districts and the size represents the relative abundance of the species in the corresponding district. 'NS' denotes not significant. **E** Heatmap displaying the top 10 strongest associations between individual species and metadata, calculated by multivariate MaAsLin2, the effect size is represented by the z-score value of the coefficient from the model, and q values (FDR adjusted  $p$  value) < 0.05 are indicated with plus and minus signs to reflect the directionality. **F** Distribution of the relative abundance of five age-related opportunistic pathogens across age. The y-axis denotes the proportion of participants younger than the age of the x-axis, who had the relative abundance of the species above its median in the population

estimated metagenome coverage, as assessed by Nonpareil, was 74.5% (IQR 58.2–86.3%, Fig. S1B), a proportion that is sufficient for adequate assembly and detection of differentially abundant genes, as described in the Nonpareil method paper [22]. After decontamination and removing library batch effects, 244 species with a relative abundance above 1% in at least one sample were retained for downstream analysis.

**Composition of OP microbiome and its associated factors**

The predominant microbes, found in over 95% of individuals with a relative abundance greater than 0.1%, including *Neisseria subflava*, *Prevotella intermedia*, *Prevotella melaninogenica*, *Haemophilus parainfluenzae*, *Schaalia odontolytica*, *Veillonella atypica*, *Veillonella dispar*, *Veillonella parvula*, *Campylobacter concisus*, *Fusobacterium*

*nucleatum*, *Myoviridae sp.*, and *Siphoviridae sp.*, collectively comprised 43.2% of the microbial reads (IQR 36.4–50.4%). No distinct microbiota subtypes were observed (Fig. 1B, Fig. S1C). A total of 444 ARGs were identified in 603 samples, including macrolide-lincosamide-streptogramin (MLS), multi-drug, beta-lactams, tetracyclines, fluoroquinolones and aminoglycosides resistance genes (Fig. S1C). Additionally, among 1941 VFs found across 632 samples, immune modulation-related VFs were the most prevalent category, potentially aiding commensal bacteria in adapting to the host immune environment and establishing colonization in the oropharyngeal niche (Fig. S1C). Besides, we found that the number of ARGs and VFs showed a significant positive correlation, suggesting a potential co-selection mechanism (Spearman’s  $\rho = 0.47$ ,  $p < 0.001$ , Fig. S1D).

Distance-based redundancy analysis (dbRDA) revealed that geographic districts accounted for the largest proportion of variance in OP microbiome, followed by occupation, age, smoke status, and NO<sub>2</sub> concentration, each contributing to over 0.5% of the variance of at least one microbial profile (Fig. 1C). The OP microbiome composition significantly differed among each pair of districts, with Hanyang and Hannan showed the greatest deviation from other districts ( $p < 0.05$ , pairwise PERMANOVA, Fig. S2 A, Supplementary Table S2). Hanyang had the highest number of species with differential abundances, enriched with species, such as *Schaalia odontolytica*, and *Rothia mucilaginoso* (Fig. 1D). Notably, occupation and the concentration of NO<sub>2</sub>, PM<sub>1</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>, which were correlated with the district (Fig. S1A), could explain 1.02% to 9.55% of the variance elucidated by geography (Variation partitioning analysis, Fig. S2B). Moreover, the OP microbiome showed greater similarity within the same sub-district (Fig. S2C), with sub-districts explaining a substantial variance of the OP microbiota in three districts (Fig. S2D), possibly due to the aforementioned district-associated factors (Fig. S2E).

*Neisseria* species, including opportunistic pathogens *N. meningitidis* and *N. gonorrhoeae*, were positively associated with age. In contrast, *S. pneumoniae* and *H. influenzae* were negatively correlated with age (Fig. 1E). Specifically, 34 of 35 (97%) participants under 5 years old carried *S. pneumoniae* with a relative abundance greater than 0.1%, compared to 57% (571/1011) of participants older than 5 ( $p < 0.001$ , Fisher's exact test, two-sided, Fig. 1F). Additionally, four *Treponema* species were enriched in individuals with chronic lung disease (CLD), including *T. denticola*, which was reported to be more abundant in COPD patients [23]. *N. subflava* and *N. flavescens* were significantly reduced in samples from individuals with hypertension, with *N. subflava* known as a hypertension biomarker [24]. *Schaalia odontolytica* showed a positive correlation with NO<sub>2</sub> and PM<sub>10</sub> concentration, while several *Leptotrichia* species exhibited a negative correlation with NO<sub>2</sub> concentration. Smoking was associated with decreased abundance of *Neisseria* species, consistent with findings from previous studies [11, 25].

At the functional level, multiple fatty acids and lipid biosynthesis pathways were correlated with age (Fig. S3A). PWY- 5100 (pyruvate fermentation to acetate and lactate), largely contributed by *Streptococcus* and *Veillonella*, was enriched in participants with CLD and SARS-CoV- 2 antibody positivity and showed a positive correlation with NO<sub>2</sub> and PM<sub>10</sub> concentrations (Fig. S3B). Additionally, the multi-drug resistance gene *lsaC* was enriched in participants with CLD, while two tetracycline resistance genes and several capsule genes were

enriched in samples from individuals with diabetes (Fig. S3C).

#### No correlation between OP microbiome and SARS-CoV- 2 infection history

Among the participants, 165 (15.8%) tested positive for SARS-CoV- 2 antibodies, indicating prior infection during the first SARS-CoV- 2 epidemic wave in Wuhan (approximately 1–4 months before sample collection). These individuals were from 145 households, with 48 of them having at least one member who tested negative for SARS-CoV- 2 antibodies and showed no symptoms. This allowed us to explore the potential relationship between the OP microbiota and susceptibility to SARS-CoV- 2, or the OP microbiome alterations post-COVID- 19.

First, we found no significant differences in OP microbiota between antibody-positive participants and their antibody-negative cohabitants (Fig. S4A and B). Additionally, no individual microbial features were significantly correlated with SARS-CoV- 2 antibody positivity, as estimated by MaAsLin2 and logistic regression analysis, which adjusted for age, gender, smoking status, concentrations of NO<sub>2</sub> and PM<sub>10</sub>, and underlying diseases, utilizing an unpaired comparison strategy (all antibody-positive participants vs. all antibody-negative cohabitants) (FDR >0.05). Moreover, the paired comparison strategy (one antibody-positive participant matched with one antibody-negative cohabitant) also failed to identify differential features regarding SARS-CoV- 2 antibody positivity (FDR >0.05, paired Wilcoxon-sum rank test, two-sided).

#### Cohabitants have similar OP microbiome

We found that participants who live together have more similar microbiota composition, functional pathways, and ARGs compositions (Fig. 2A), with spouses exhibiting the greatest similarity among all family relationships (Fig. 2B). The relative abundance of individual microbial features, including 42 species, 59 pathways, 14 ARGs, and 4 VFs, showed a significant positive correlation among cohabitants (Fig. 2C, Supplementary Table S3), implying possible transmission of microorganisms between cohabitants. These features involved commensals (e.g., *N. subflava* and *P. melaninogenica*), opportunistic pathogens (e.g., *F. nucleatum*, *T. denticola*, and *S. pneumoniae*), pyruvate fermentation pathway (PWY- 5100), as well as ten ARGs potentially resistant to tetracyclines. In contrast, no such correlation was detected among non-cohabiting pairs, and the correlation coefficients were significantly higher among cohabitants compared to those calculated from an equivalent number of non-cohabiting pairs ( $p < 0.05$ , Permutation test, Fig. S4C, Supplementary Table S3). In addition, we observed that

the correlation between cohabitants varied across different relationships. For instance, *P. melaninogenica*, and *S. oralis* exhibited more similar abundance in siblings, whereas *F. nucleatum* and *T. denticola* showed higher correlations among spouses (Fig. 2C, Fig. S4D).

### OP microorganisms transmitted between cohabitants

The potential for transmission of microorganisms between cohabitants was further investigated at the strain level. Genotypes (sequences of the marker genes) of 39 species present in at least one cohabiting pair were profiled by StrainPhlan for each sample [26]. Hamming distances between genotypes of the same species were calculated for 141 cohabiting pairs, with an average of four species per pair (Fig. S5A). Overall, genotype distances were significantly lower between cohabitants compared to non-cohabiting pairs ( $p < 0.05$ , Kolmogorov–Smirnov test, two-sided, Fig. 3A). Specifically, 29.4% of genotype distances between cohabiting pairs were below 0.025, compared to 8.7% in non-cohabiting pairs ( $p < 0.05$ , Fisher’s exact test, two-sided). The genotype distances in 5 species were significantly lower between cohabiting pairs than between non-cohabiting pairs ( $p < 0.05$ , Wilcoxon rank-sum test, two-sided, Fig. 3B). Moreover, permutation tests revealed that the frequency of the two closest genotypes originating from cohabiting pairs was significantly higher than expected by random chance for 20 species (see Methods, Fig. S6), with *P. sp. oral. taxon 306* exhibited the highest frequency (66.7%, Fig. 3C). Additionally, gene synteny of the same species was more homologous between cohabiting pairs than non-cohabiting pairs ( $p < 0.05$ , Kolmogorov–Smirnov test, two-sided, Fig. 3D).

By applying species-specific hamming distance cutoffs to define the same strain (see details in the “Methods” section, Fig. S5B), we identified 116 putative transmission events (shared strains) within households. The median strain transmission rate (STR) among cohabiting pairs with at least five comparable species was 16.7% (IQR 0.0–33.3%), significantly higher than the 0.0% (IQR 0.0–11.1%) in non-cohabiting pairs ( $p < 0.05$ , Wilcoxon rank-sum test, two-sided). Non-cohabiting pairs from the same district (median STR 0.0%, IQR 0.0–16.7%) shared

more strains than those from different districts (median STR 0.0%, IQR 0.0–8.5%,  $p < 0.05$ , Wilcoxon rank-sum test, two-sided). Additionally, 65.9% of cohabiting pairs and 27.2% of non-cohabiting pairs, including those residing in the same or different districts, shared at least one strain ( $p < 0.05$ , Fisher’s exact test, two-sided). Among non-cohabiting pairs, those residing in the same district were more like to share strains than those in different districts (35.2% vs. 25.7%,  $p = 0.045$ , Fisher’s exact test, one-sided).

Interestingly, the transmission rate of the same species in OP (this study) was positively correlated with that reported in the oral cavity [17] (Fig. 3E). The reliability of defining the same strain using the hamming distance was further supported by gene synteny analysis, as the putative same strain exhibited a higher synteny score (Fig. S5C). Additionally, we found that the transmission of microorganisms was not correlated with their abundances, while species within the same genus tended to be transmitted together (Fig. S5D–F). This may be because species within the same genus tend to colonize proximal niches, facilitating co-transmission, or due to the limitations in species-level resolution within the same genus.

The genotype distances between cohabitants of the same species varied across different familial relationships, with siblings and spouses exhibiting the highest similarity (Fig. 3F and Fig. S5G). Spouses also showed the highest proportion of putative microorganism transmissions, followed by siblings and parent–child relationships (Fig. 3G). Notably, the genotype distances between father–child and mother–child of the same species were lower than those between unrelated non-cohabiting pairs (Fig. S5H). However, there was no significant difference in genotype distances between father–child and mother–child pairs, regardless of the child’s age (Fig. S5I), suggesting no strong evidence for vertical transmission of microorganisms.

### OP microbiome as a reservoir of antibiotic resistance genes with horizontal transfer potential

After de novo assembly, we identified 6208 ARG-carrying contigs (ACCs) in 906 samples. Of these, 4827 ACCs from 872 samples were assigned species labels,

(See figure on next page.)

**Fig. 2** Similarity of OP microbiome between cohabitants. **A** Density plots illustrating the distribution of distances between cohabitation and non-cohabitation pairs based on different microbiome features. The dotted lines represent the median values. \*\*\*:  $p < 0.001$  (Wilcoxon rank-sum test, two-sided). **B** Comparison of the distance between different types of family relationships based on different microbiome features. The central line indicates the median. The lower and upper hinges indicate the first and third quartiles. The lower and upper whiskers extend from the hinge to the smallest and largest values no further than 1.5 times the interquartile range from the hinge. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . In A and B), the distance of species and pathway compositions was calculated based on the Aitchison distance, while the distance of ARG and VF compositions was the Horn–Morisita distance. **C** Circular heatmap illustrating the 43 species with abundance correlated between cohabitation pairs. G–C denotes grandparent–child pairs, and P–C denotes parent–child pairs. \*\* represents FDR adjusted  $p < 0.05$

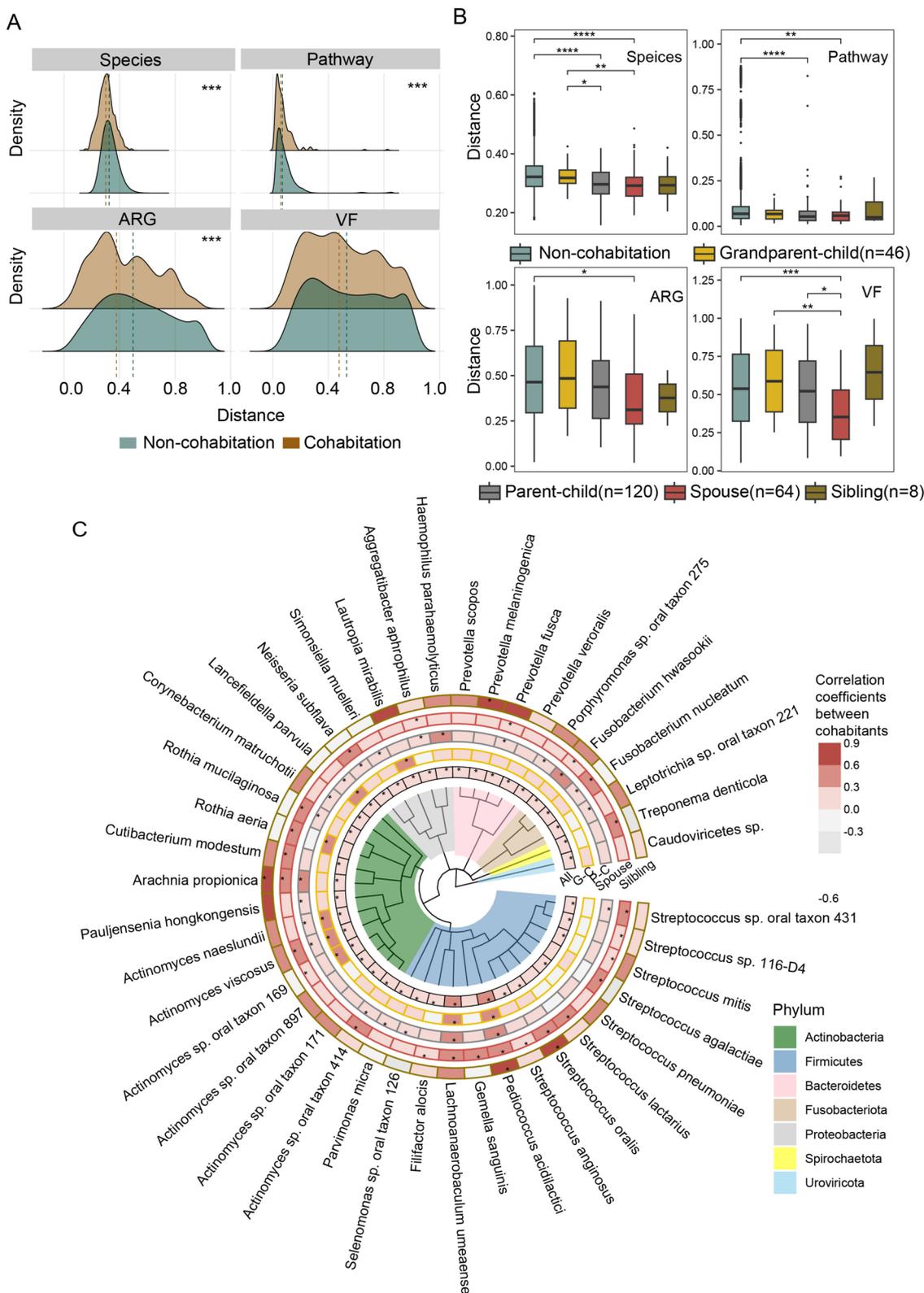
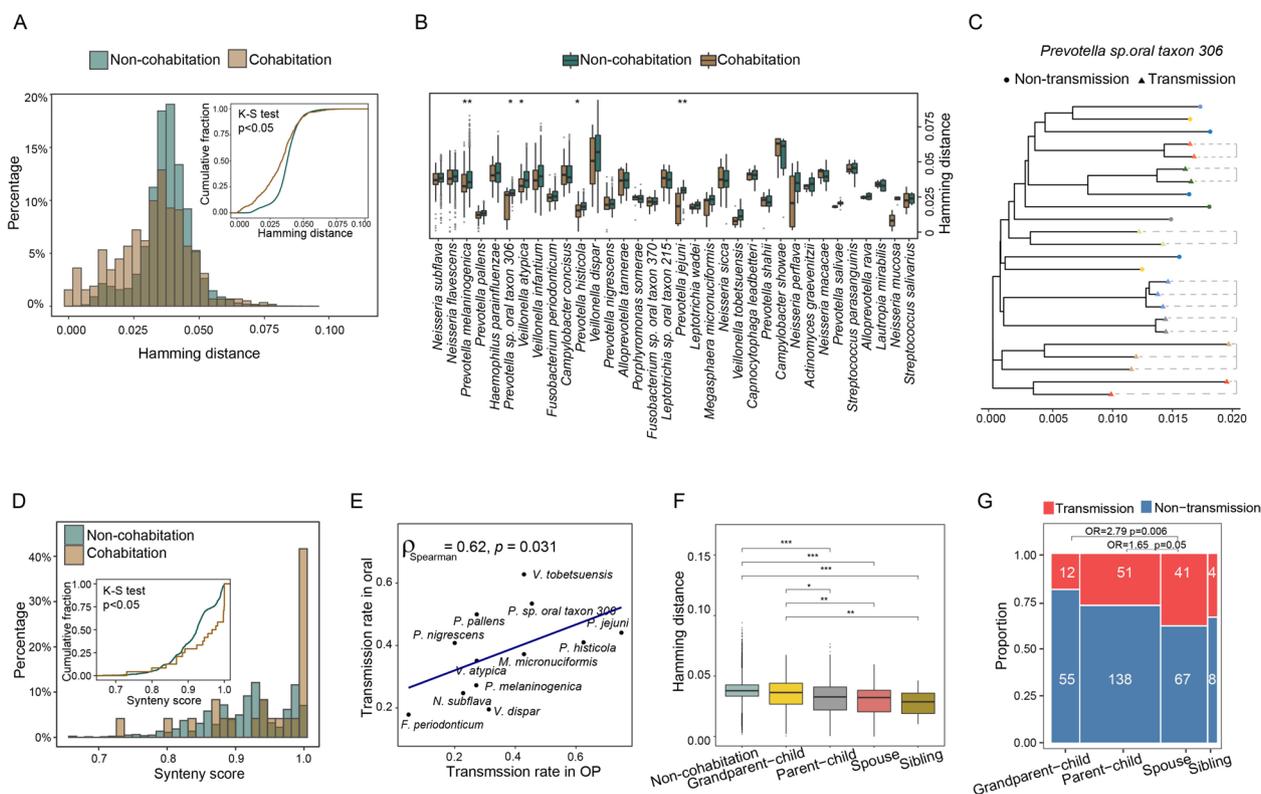


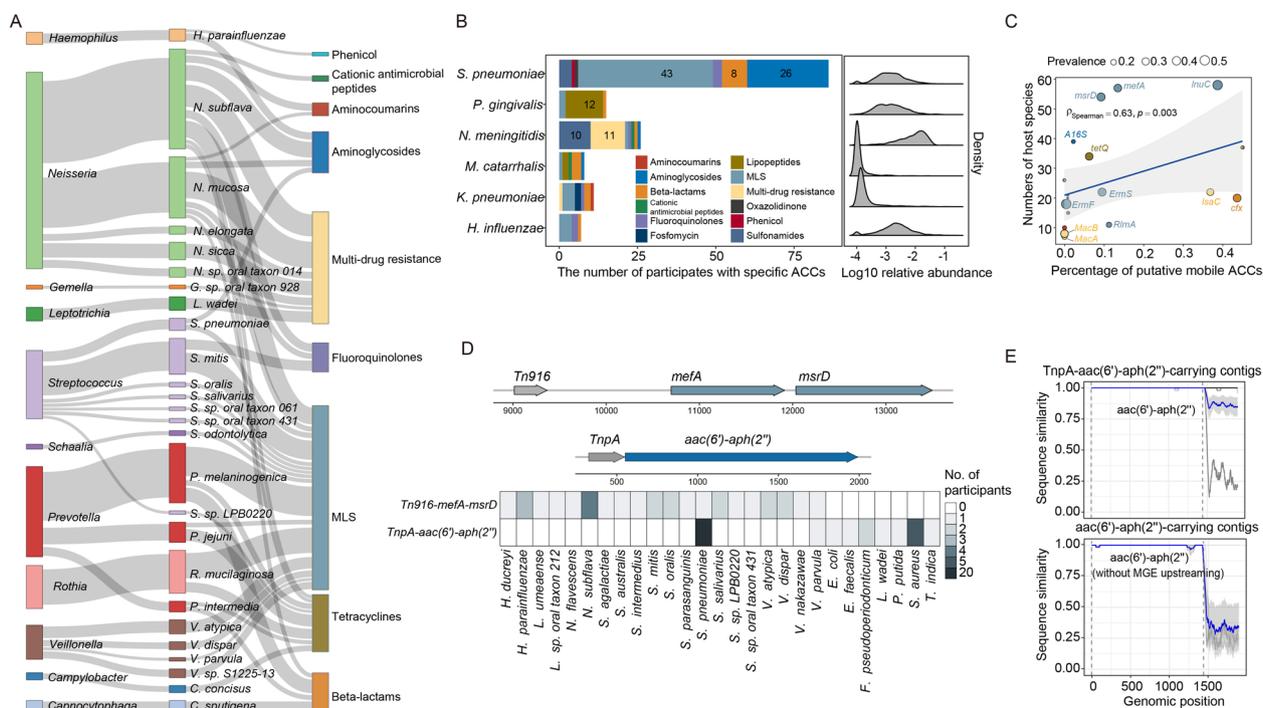
Fig. 2 (See legend on previous page.)



**Fig. 3** Possible OP microbiome transmission between cohabitants. **A** Distribution of genomic distances between genotypes in cohabitation and non-cohabitation pairs. Genomic distance was calculated using the Hamming distance, which measures the proportion of differing positions between two genotypes. The *p* value was calculated by Kolmogorov–Smirnov (K-S) test (two-sided). **B** Box plots showing the hamming distances between genotypes in cohabitation and non-cohabitation pairs across different species. *P* values were obtained using the Wilcoxon rank-sum test (two-sided). **C** Phylogenetic tree of genotypes of *Prevotella sp. oral taxon 306* from different individuals. The color of the tip points represents different families, and close pairs from the same household are marked with triangles and connected by dotted lines, all of which were defined as strain transmission events in the following analysis. **D** Distribution of genomic synteny scores between genotypes in cohabitation and non-cohabitation pairs. The *p* value was calculated by Kolmogorov–Smirnov (K-S) test (two-sided). **E** Correlation between the transmission rate of species in OP samples and oral samples. **F** Box plots displaying the distance between genotypes in different relation types. *P* values were obtained using the Wilcoxon rank-sum test (two-sided). **G** Comparing the number of putative transmission events among different relation types. Odds ratios (OR) and *p* values were calculated using Fisher’s exact test (two-sided). \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001, \*\*\*\**p* < 0.0001

with a median length of 4886 base pairs (IQR 2360–11,353 bp). First, we found that many species harbored multiple ARGs, and most ARGs were shared among different species (Fig. 4A). For instance, the most abundant ARGs, conferring MLS resistance, were detected in most respiratory commensals and opportunistic pathogens such as *S. pneumoniae*. Tetracyclines resistance genes were found in *Prevotella* and *Rothia* species, while beta-lactams resistance genes were mostly identified in *Neisseria*, *Prevotella*, and *Capnocytophaga*. Among opportunistic pathogens, *S. pneumoniae* harbored the highest number of ARGs, including those conferring resistance to MLS, aminoglycoside, and Beta-lactams antibiotics. This observation is unlikely to be attributed to a higher abundance of *S. pneumoniae*, as its abundance was comparable to other bacteria in the samples (Fig. 4B).

ARGs can be transmitted between different microorganisms by hitchhiking with mobile genetic elements (MGEs) or located on plasmids. We found that 15.0% of the ACCs contained MGEs within 5000 base pairs upstream or downstream ARGs, or the ACCs originated from plasmids (Fig. S7A). We identified *Tn916* as the most frequent MGE adjacent to ARGs on ACCs (41.1%), followed by transposase (18.6%), *qacEdelta* (12.4%), integrase (12.0%), and *tnp-ISCR* (7.8%). This finding highlights *Tn916* as the predominant element that may facilitate ARG dissemination in the oropharyngeal microbiome. ARGs with higher putative mobility (more frequently linked to MGEs) were detected across a broader range of microbial hosts (Fig. 4C), suggesting that MGEs may facilitate horizontal transfer of ARGs among microorganisms in the oropharynx. Beta-lactams resistance genes, 31.2% of which flanked by MGEs or



**Fig. 4** OP is a reservoir of antibiotic resistance genes with the potential for horizontal transfer. **A** Species annotation of ARG-carrying contigs (ACCs). Species contributing more than 30 ACCs of at least a type of ARG are shown. **B** Distribution of the ARGs within ACCs in opportunistic pathogens colonizing the upper respiratory tract. Species abundance distribution is shown on the right side of the figure. **C** Correlation between the proportion of mobile ACCs and the number of host species for different ARGs. ARGs detected in at least 100 ACCs are shown. **D** Illustrations of two ACCs that are associated with the highest numbers of microbial hosts. The upper part displays the structures of two ACCs, while the lower part displays the number of participants with these two ACCs in each species host. **E** Sequence similarity of the *aac(6)-aph(2)* gene and its downstream 500 bp between different species in contigs with MGEs (upper) and without MGEs (lower). Gray lines represent the similarity between sequences in two different species (with 50 bp sliding windows used), the blue line indicates the mean of all gray lines, and the shadow represents the 95% confidence interval

located on plasmids, exhibited the highest potential for horizontal transfer. This aligns with reports indicating that Beta-lactams resistance genes are most frequently associated with horizontal gene transfer events [27]. These MGE-associated ARGs were predominantly found in the genomes of *P. melaninogenica*, *P. intermedia*, and *C. sputigena*. Aminoglycoside resistance ARGs were the next most associated, with 25.0% linked to MGEs, primarily carried by *S. pneumoniae* (Fig. S7B).

The most widely distributed MGE-ARG pair was the MLS resistance genes *mefA* and *msrD* linked to the conjugative transposon *Tn916* (*Tn916-mefA-msrD*), identified in 20 species across 5 genera, carried by 30 unrelated individuals in 8 districts (Fig. 4D). The second most prevalent MGE-ARG pair was the aminoglycoside resistance genes *aac(6)-aph(2)*, preceded by a *TnpA* gene (*TnpA-aac(6)-aph(2)*), found in nine species across 8 genera, carried by 33 unrelated individuals in 11 districts. Notably, the *S. pneumoniae* genome in 20 samples harbored the *TnpA-aac(6)-aph(2)* gene cluster. The distribution of this gene cluster was not

correlated with age, suggesting that it is unlikely to be driven by antibiotic pressure after birth (Fig. S7C). Furthermore, we observed higher sequence similarity in regions adjacent to the *Tn916-mefA-msrD* or *TnpA-aac(6)-aph(2)* gene clusters between different species, which may be transferred along with conserved ARGs, compared to those without the MGE (Fig. 4E, Fig. S7D and E), suggesting MGE-mediated horizontal transfer of ARGs among species.

Besides, the distribution of MGEs and species together explained 82.0% of the variation in ARG composition. Specifically, MGE abundance accounted for 12.8% of the variation, while species composition explained 18.1%. The interaction between these two factors contributed to 51.1% of the variance (Variation Partitioning Analysis). These findings highlight a strong correlation among microorganisms, MGEs, and ARGs within the OP microbiome.

Additionally, we found that 154 out of 4827 (3.2%) ACCs annotated at the species level also carried VFs. However, none of these ACCs contained MGEs,

suggesting that VFs and ARGs are unlikely to be co-transferred in the OP microbiome of healthy individuals.

## Discussion

In this study, we profiled the oropharyngeal microbiome at the species level and found that geography was the predominant factor correlating with its composition, consistent with previous studies on respiratory and other human microbial niches [25, 28]. The concentration of air pollutants, especially PM<sub>10</sub> and NO<sub>2</sub>, which were correlated with geography, accounted for substantial geographic variation at both district level (tens of kilometers apart) and sub-district level (several kilometers apart) scales. Among the 13 districts investigated, Hanyang showed the most pronounced deviations. Taxa such as *Schaalia odontolytica*, *Schaalia meyeri*, and *Leptotrichia sp. oral taxon 221*, which were either enriched or depleted in Hanyang, were significantly correlated with NO<sub>2</sub> or PM<sub>10</sub> concentration. Notably, the genus *Schaalia* and *Leptotrichia* in the oropharynx have previously been linked to NO<sub>2</sub> levels [29]. Additionally, unmeasured factors like humidity and residential area greening rate may also contribute to geographic-related microbiome variations [30, 31]. Furthermore, the association between the OP microbiome and geography may reflect increased social interactions or exposure to shared environmental microbiomes within district [15].

Age explained 1.4% of the variation in the OP microbiome and correlated with the abundance of certain opportunistic pathogens. *S. pneumoniae* and *H. influenzae*, both known to cause lower respiratory infections in children [32] were more abundant in the OP microbiome of younger individuals. Conversely, *N. meningitidis*, *N. gonorrhoeae*, and *P. gingivalis*, causative agents for meningitis, gonorrhea, and periodontitis, respectively, increased with age. This age-related shift may reflect changes in contact patterns, social behavior, infection history, and receptor numbers in the oropharyngeal epithelium over time [33, 34]. Colonization of these pathogens in the upper respiratory may increase the risk of invasive infections through hematogenous spread or proliferation from the site of colonization [8].

Human microbiomes could transfer through contact, shared environments, and the exchange of body fluids [35], making cohabitation a key factor influencing microbiome composition, as seen in gut and oral microbiomes [17]. Our study suggests that OP microbiome may also be transmitted between cohabitants, particularly among those with close relationships like spouses and siblings, supported by correlations in microbial abundance and genome similarity of the same species between cohabitants. Among OP species shared among cohabitants, 16.7% were likely transmitted between cohabitants, lower

than oral (32%) but higher than gut (12%) [17]. Interestingly, the transmission rates of the same species in the OP and oral cavity were significantly correlated. Since these two sites are anatomically connected and share a large number of microorganisms, the transmission of the OP microbiome may be partially or primarily driven by the transmission of the oral microbiome through physical interactions, such as kissing and sharing utensils. However, OP microbiome transmission might also occur independently of oral microbiome transmission, as the former is more susceptible to airborne and environmental microorganisms. This raises an interesting avenue for exploring the transmission of the nasopharyngeal microbiome transmission, which is less likely to be influenced by the oral microbiome. Furthermore, we found no significant difference in genetic distance between species in mother–child and father–child pairs, even for families with children younger than 5 years old, indicating no evidence of vertical transmission of OP microorganisms, similar to oral but unlike gut microbiome transmission [17]. Additionally, although some VFs are associated with bacterial invasion and colonization [36], we found no significant correlation between VF burden—defined as the average number of VFs per contig for each species—and species transmission rates (Spearman's  $\rho = -0.47$ ,  $p = 0.07$ ). Further investigation with deeper sequencing coverage is needed to clarify the link between specific VFs and bacterial transmission.

Meanwhile, shared environments, beyond direct contact, may also influence strain sharing. A recent study on the Amboseli baboons found that baboons who never co-resided or overlapped in lifetime shared a similar proportion of gut microbiota with closely grooming partners. This was likely due to shared diets or exposure to seasonal environmental factors, such as the duration of the rainy season [37]. This emphasizes the impact of environmental factors on microbiome composition beyond social interactions. Similarly, our observation that non-cohabiting pairs in the same district exhibited higher strain-sharing rates and more similar microbial composition than those residing in different districts may be attributed to shared environmental exposures, such as air pollutants. Previous studies have identified shared microbes and ARGs between human airway samples and indoor or outdoor airborne particulate matter and dust [38, 39]. Future studies incorporating detailed environmental sampling and longitudinal monitoring of human airways could better quantify the relative contributions of direct transmission and shared environmental exposure to microbiome similarity.

The transmission of the OP microbiome involved both commensals (e.g., *N. subflava*, *P. melaninogenica*) and opportunistic pathogens (*F. nucleatum*, *T. denticola*

and *S. pneumoniae*). These findings align with previous studies on household transmission of pathogens, such as *S. pneumoniae* carriage being influenced by having siblings and periodontal pathogens transmitted between spouses [40–42]. Given that the respiratory microbiome can influence susceptibility to infections and immune homeostasis [2], the transmission of the OP microbiome among cohabitants could affect the incidence and severity of respiratory diseases, making it an important factor in assessing disease risk.

Our study revealed that the human oropharynx serves as a reservoir for ARGs, similar to the gut and oral cavity [21]. The ARG spectrum in the healthy population is similar to that observed in respiratory specimens from patients with COPD and infectious diseases, predominantly comprising ARGs for macrolides-lincosamides-streptogramins (MLS), multi-drug, and beta-lactams antibiotics [43, 44]. This contrasts with the gut microbiome, which primarily harbors ARGs for tetracyclines, MLS, and beta-lactams antibiotics [21]. The difference likely reflects variations in microbial communities across different body sites and/or differences in antibiotic pressure [45]. Among common respiratory tract pathogens, *S. pneumoniae* harbored more ARGs than other species, conferring excess resistance to MLS and aminoglycoside antibiotics. This aligns with the high resistance rate of *S. pneumoniae* to erythromycin (an MLS antibiotic), reported at 97.0% in Hubei province (Wuhan is the capital of the province) [46]. Aminoglycoside resistance, however, was not reported in governmental surveillance, highlighting a need for further monitoring and clinical attention. Notably, ARGs in the OP microbiome may transfer between resident microorganisms, with 15.0% of ARG-carrying contigs showing putative mobility (with MGEs located within 5000 base pairs upstream or downstream of ARGs). This raises concerns that resistance genes in OP commensals could transfer to invading respiratory pathogens. We noted that the proportion of MGE-associated ACCs in the OP was lower than that reported in the human gut (19.6–36.2%) [47, 48], which may reflect more frequent horizontal gene transfer and stronger selective pressure from antibiotic exposure in the human gut [49].

The study has several limitations. First, the metadata obtained from the questionnaire was limited, leading to potential missing variables that could impact the composition and functional elements of the OP microbiome. For instance, the unrecorded history of antibiotic use might influence the abundance and diversity of microorganisms and ARGs. Second, the human oropharyngeal samples contained a high number of human cells, resulting in relatively low microbial reads compared to gut microbiome studies, which may result in

the underrepresentation of low-abundance microorganisms and functional elements.

## Conclusions

In summary, our study provided a high-resolution profile of the human OP microbiome and its functional genes in a healthy population, identifying numerous demographic, environmental, and clinical factors associated with the OP microbiome. Our findings suggest that the OP microbiome, which is directly exposed to external environments, is influenced by external matters. Additionally, the OP microbiome can be transmitted between cohabitants, particularly those with close contact, and antibiotic resistance genes can transfer between different microorganisms, which may influence immune homeostasis in the respiratory tract, susceptibility to infectious or non-infectious diseases, and effectiveness of antibiotic treatments.

## Methods

### Cohort design and sample collection

The study cohort, comprising 1046 participants, was part of a population-level survey of SARS-CoV-2 antibodies in Wuhan, China on 14–15 April 2020 [50]. Inclusion criteria were those who had lived in Wuhan for at least 14 days since December 1, 2019, and without any self-report respiratory symptoms in the latest month. Households with individuals whose sera tested positive for SARS-CoV-2 antibodies were preferentially selected. The recruitment covered 69 subdistricts across 13 districts in Wuhan, China, with 289 participants having at least one cohabitant included in this study. All participants filled in a questionnaire including demographic, clinical, occupational, residential address, and smoke status. Ambient air pollutants (NO<sub>2</sub>, PM<sub>1</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>) were assessed for each individual as the average concentration of each pollutant in the residing area over the past year before the sampling date, assessed at a 1 km × 1 km spatial resolution using the China High Air Pollutants datasets [51–54]. Oropharyngeal samples were collected from each participant at community health-care center using a flocced swab (MT0301, Yocon), by swabbing both sides of the palatal arch and posterior pharynx and then immediately placed into viral transport medium and stored at –80 °C until transported to the laboratory for processing.

### DNA extractions and metagenomic sequencing

DNA was extracted using the chemagic™ 360 instrument (Revvity). Then, DNA libraries were constructed with DNA fragmentation, adapter ligation, PCR amplification, and purification by Celero™ EZ DNA-Seq Kit (TECAN). Libraries were sequenced on an Illumina NovaSeq 6000 platform using the mode of 150-bp paired-end reads. To

control for contaminations introduced during the sample processing, 10 deionized water samples and 89 unused swabs were processed following the same protocol and were used as negative controls.

#### Taxonomic and functional profiling of metagenomic data

Raw metagenomes were processed with fastp (v0.20.0) [55], which involved removing adapters and ploy-G tails, trimming positions with quality <15, removing low-quality reads (mean quality <25), and discarding reads shorter than 50 nt. Human reads were removed by mapping to a custom database that includes the human reference genome (GRCH38) and UNiVec sequences (version May 2020) by BowTie2 (v2.3.0, sensitive mode) [56]. The estimated coverage of nonhuman metagenome was assessed by Nonpareil 3 [57].

Species-level taxonomic profile was obtained by Kraken2 and Bracken (v2.1.2) [58] with default parameters and a custom database that consisted of human, virus, bacterium, archaea, and fungus genome sequences from the NCBI nucleotide database (version July 2021). To distinguish true species from potential contaminants, we assessed species contamination using the quantileTest function in the R package EnvStats (v2.7.0) [59]. Specifically, species whose 95 th percentile of abundance distribution in OP samples was significantly greater than that in NCs were retained for further analysis. Library batch bias was corrected by ConQuR (v2.0) [60].

HUMAnN2 (v2.8.1) [61] was used to profile MetaCyc functional pathways with default settings. Quality-controlled reads were also used to search for antibiotic resistance genes (ARGs), virulence factors (VFs), and mobile genetics elements (MGEs) by aligning them to different databases using BWA-MEM (v0.7.10) [62]. Specifically, MEGARes (v2.0) [63], was used to identify ARGs. Genes that confer resistance by mutation were not considered. VFDB was used to identify VFs [64]. A collective custom MGE database was used to identify MGEs [65]. The alignment bam files were processed by SAMtools (v0.1.18) [66] to obtain sequencing coverage and depth of the gene. For samples with more than 1000 mappable reads, functional genes with sequence coverage above 50% and supported by at least 10 reads were retained. The abundance of these functional genes was normalized by the number of the 16S rRNA gene reads to represent a “copy of ARG/VF/MGE per copy of 16S rRNA gene” [67]. The number of 16S rRNA gene sequences per sample was obtained by METAXA2 (v2.0) [68].

#### Strain-level profiling of the respiratory microbiome

The genotypes of species were profiled by StrainPhlan (v3.0.14) [26] using the default marker gene database,

with parameters “-min\_base\_coverage 3” for sample2 markers.py script, “-marker\_in\_n\_samples 20 -sample\_with\_n\_markers 20 -debug” for strainphlan script. The consensus sequences of all the marker genes were concatenated as a genotype (strain) of the species. To compare the genomic similarity of the same species between cohabiting pairs or non-cohabiting pairs, only genotypes with comparable regions exceeding 1000 bp were included. The pairwise genetic distance between genotypes was calculated by the dist.dna function in R packages ape (v.5.6.2)(model = “raw”, pairwise.deletion = TRUE) [69]. The hamming distance is computed as the number of mutations between pairs of genotypes divided by the length of the total comparable region. Phylogenies trees were constructed by RAxML (v8.2.12)(parameters: -p 1989 -m GTRCAT) [70] using genotypes obtained from StrainPhlan, visualizing by R package ggtree (v3.2.1) [71].

#### Identifying strain transmission events among cohabitants

To determine the transmission events between individuals, a threshold of genetic distance is needed to define the same strain. Since different species have different mutation rates, thus we applied a species-specific genetic difference threshold to define a strain. For each species, the threshold was selected as the maximum genetic distance ensured that the probability of sharing the same strain in cohabiting pairs was two times higher than that in non-cohabiting pairs, with significance in Fisher’s exact test. The threshold ranged between 0.006 to 0.038 for different species, similar to that applied in previous studies [17]. In addition, SynTracker [72] was used to compare the genomic synteny of species between cohabitation pairs and non-cohabitation pairs, and also between putative transferred strains and non-transferred ones.

To test the statistical significance of the correlation in species abundance between cohabitation pairs, we created a background distribution of species abundance correlation using unrelated individuals by randomly sampling 269 non-cohabitation pairs (the same number of cohabitation pairs) 1000 times and calculated the frequency of observing a correlation coefficient higher than that observed between cohabitants.

To test whether the genomes of shared species among cohabitants were more similar than those among non-cohabitants, we created a background distribution of the proportion of the nearest strains from cohabitants, by randomly disrupting the household labels of samples 1000 times, and calculated the frequency of observing a proportion higher than that observed between cohabitants.

### Analysis of the mobility potential of ARGs

Quality-controlled sequencing reads were assembled using megaHIT (v1.2.9) [73] by default parameters for each sample. ARG-carrying contigs (ACCs) were identified by mapping ARG reference sequences from the MEGARes database [63] to the assembled contigs using BLASTn (v2.7.1) [74], requiring identity percentage greater than 80% and coverage greater than 70%. Taxonomic classification of ACCs was assigned using Kraken2 (v2.1.2) [58]. To explore the mobility of ARGs, Mobile genetic elements (MGEs) on the ACCs were detected by aligning known MGEs reference sequences against ACCs by BLASTn (v2.7.1) [74], requiring identity percentage greater than 80% and coverage greater than 70%. The reference MGE sequences from the aforementioned custom MGE database [65] were created by retrieving coding sequences for genes annotated as *IS\**, *ISCR\**, *int11*, *int2*, *istA\**, *istB\**, *qacEdelta*, *tniA\**, *tniB\**, *tnpA\**, and *Tn916* transposon from the NCBI nucleotide database and PlasmidFinder database. Notably, these references do not contain bacteriophage sequences. ACCs harboring an MGE located within 5 kb of an ARG were recognized as putative mobile ACCs. Besides, ACCs predicted to be located on plasmids by PlasFlow [75] (using default settings) were also annotated as putative mobile ACCs.

### Statistical analysis

The proportion of variance in microbiome composition explained by each host or environmental factor was accessed using distance-based RDA with Aitchison distance on the relative abundance of species and Horn-Morisita distance on the MetaCyc functional pathway profile and normalized functional gene profiles, implemented in the R package *vegan* [76]. Variation partitioning analysis, performed using the *varpart* function in the R package *vegan*, was used to estimate the contribution of district-associated factors to geographic variations in the microbiome. The associations between metadata features and microbial compositions (species, pathway, ARG, VF) were assessed using a multivariate linear regression model implemented in *MaAsLin2* (v1.8.0) [77], with all the host and environmental factors adjusted as covariates, except for  $PM_1$ ,  $PM_{2.5}$ , district, and occupation, to avoid multicollinearity. All statistical tests were followed by multiple-testing correction using the Benjamini–Hochberg method whenever applicable.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-025-02107-9>.

Supplementary Material 1. Figure S1. Overview of metadata and microbiome composition. Figure S2. Geographic variation of OP microbiome. Figure S3. Associations between microbial features and metadata. Figure

S4. Cohabitants have similar OP microbiome. Figure S5. Definition and validation of microbial transmission among cohabitants. Figure S6. Permutation test for the nearest genotypes being from cohabitants. Figure S7. ARG-carrying contigs (ACC) analysis.

Supplementary Material 2. Table S1. Overview of study participants. Table S2. Pairwise PERMONAVA tests between the microbiomes of samples from different districts. Table S3. Correlated microbial features between cohabitation pairs

### Acknowledgements

We thank all the participants who donated their specimens to this study.

### Authors' contributions

L.R., M.L., and J.W. designed the study. J.Y., and M.L. analyzed, validated and interpreted the data and drafted the initial version of the manuscript. L.R., Y.X. and L.G. recruited participants and collected oropharyngeal swabs and clinical metadata. J.R., X.W., Y.W. and Jingchuan.Z. prepared sequencing libraries of the specimens. Chao.W., Linfeng.Z., Li.Z., X.J. and Jiaxin.Z. contributed to computational analysis. W.Y., and Chen.W. provided support through ethics and participant recruitment. M.L., L.R., and J.W. revised the manuscript. All authors reviewed the manuscript.

### Funding

This study was supported by funding from the National Key R&D Program of China (grant no. 2022YFA1304300 to M.L. and L.G.), Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS) (2020-I2M- 2-013 to L.R., 2023-I2M- 2- 001 to Chen. W. and W. Y.), Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2019PT310029 to L.R.) and Foundation for Innovative Research Groups of the National Natural Science Foundation of China (NSFC82221004 to J. W.). Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS),2020-I2M- 2- 013,2023-I2M- 2- 001,2023-I2M- 2- 001,Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences,2019PT310029,National Key R&D Program of China,2022YFA1304300 ,2022YFA1304300,Foundation for Innovative Research Groups of the National Natural Science Foundation of China,NSFC82221004

### Data availability

The sequencing data have been deposited in the GSA in the National Genomics Data Center (under accession CRAO20169, publicly accessible at <https://ngdc.cncb.ac.cn/gsa>). All analyses were implemented in RStudio and the scripts can be accessed at <https://github.com/Jing-Y-oss/Human-oropharyngeal-microbiome-analysis>.

### Declarations

#### Ethics approval and consent to participate

This study was approved by the Ethical Review Board of the Institute of Pathogen Biology, Chinese Academy of Medical Sciences (IPB- 2020-04). Sample collection was approved by the China Human Genetic Resources Management Office (No. 2021 CJ0597). Written informed consent was obtained from all participants before they were enrolled. For individuals younger than 18 years, consent was provided by parents or a legal guardian.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>National Health Commission Key Laboratory of Systems Biology of Pathogens, State Key Laboratory of Respiratory Health and Multimorbidity and Christophe Mérieux Laboratory, National Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. <sup>2</sup>Key Laboratory of Respiratory Disease Pathogenomics, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. <sup>3</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, and China

National Center for Bioinformation, Beijing, China. <sup>4</sup>University of Chinese Academy of Sciences, Beijing, China. <sup>5</sup>Key Laboratory of Pathogen Infection Prevention and Control (Ministry of Education), National Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China. <sup>6</sup>Changping Laboratory, Beijing, China. <sup>7</sup>School of Population Medicine and Public Health, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.

Received: 7 January 2025 Accepted: 8 April 2025

Published online: 06 May 2025

## References

- Marsland BJ, Yadava K, Nicod LP. The airway microbiome and disease. *Chest*. 2013;144:632–7.
- Kumpitsch C, Koskinen K, Schöpf V, Moissl-Eichinger C. The microbiome of the upper respiratory tract in health and disease. *BMC Biol*. 2019;17:87.
- Clark SE. Commensal bacteria in the upper respiratory tract regulate susceptibility to infection. *Curr Opin Immunol*. 2020;66:42–9.
- Bradley ES, Zeamer AL, Buccì V, Cincotta L, Salive MC, Dutta P, et al. Oropharyngeal microbiome profiled at admission is predictive of the need for respiratory support among COVID-19 patients. *Front Microbiol*. 2022;13: 1009440.
- Merenstein C, Liang G, Whiteside SA, Cobián-Güemes AG, Merlino MS, Taylor LJ, et al. Signatures of COVID-19 severity and immune response in the respiratory tract microbiome. *mBio*. 2021;12:e0177721.
- Ren L, Wang Y, Zhong J, Li X, Xiao Y, Li J, et al. Dynamics of the upper respiratory tract microbiota and its association with mortality in COVID-19. *Am J Respir Crit Care Med*. 2021;204:1379–90.
- Tsang TK, Lee KH, Foxman B, Balmaseda A, Gresh L, Sanchez N, et al. Association between the respiratory microbiome and susceptibility to Influenza virus infection. *Clin Infect Dis*. 2020;71:1195–203.
- Siegel SJ, Weiser JN. Mechanisms of bacterial colonization of the respiratory tract. *Annu Rev Microbiol*. 2015;69:425–44.
- Brumbaugh DE, Arruda J, Robbins K, Ir D, Santorico SA, Robertson CE, et al. Mode of delivery determines neonatal pharyngeal bacterial composition and early intestinal colonization. *J Pediatr Gastroenterol Nutr*. 2016;63:320–8.
- Guo J, Zhang X, Saiganesh A, Peacock C, Chen S, Dykes GA, et al. Linking the westernised oropharyngeal microbiome to the immune response in Chinese immigrants. *Allergy Asthma Clin Immunol*. 2020;16:67.
- Turek EM, Cox MJ, Hunter M, Hui J, James P, Willis-Owen SAG, et al. Airway microbial communities, smoking and asthma in a general population sample. *EBioMedicine*. 2021;71: 103538.
- Whelan FJ, Verschoor CP, Stearns JC, Rossi L, Luinstra K, Loeb M, et al. The loss of topography in the microbial communities of the upper respiratory tract in the elderly. *Ann Am Thorac Soc*. 2014;11:513–21.
- van Kersen W, Bossers A, de Steenhuijsen P, Pijpers WAA, de Rooij MMT, Bonten M, Fluit AC, et al. Air pollution from livestock farms and the oropharyngeal microbiome of COPD patients and controls. *Environ Int*. 2022;169: 107497.
- Beghini F, Pullman J, Alexander M, Shridhar SV, Prinster D, Singh A, et al. Gut microbiome strain-sharing within isolated village social networks. *Nature*. 2025;637:167–75.
- Sarkar A, McInroy CJA, Harty S, Raulo A, Ibatá NGO, Valles-Colomer M, et al. Microbial transmission in the social microbiome and host health and disease. *Cell*. 2024;187:17–43.
- Brito IL, Gurry T, Zhao S, Huang K, Young SK, Shea TP, et al. Transmission of human-associated microbiota along family and social networks. *Nat Microbiol*. 2019;4:964–71.
- Valles-Colomer M, Blanco-Míguez A, Manghi P, Asnicar F, Dubois L, Golzato D, et al. The person-to-person transmission landscape of the gut and oral microbiomes. *Nature*. 2023;614:125–35.
- Christensen ED, Hjelmsø MH, Thorsen J, Shah S, Redgwell T, Poulsen CE, et al. The developing airway and gut microbiota in early life is influenced by age of older siblings. *Microbiome*. 2022;10:106.
- Paulo AC, Lança J, Almeida ST, Hilty M, Sá-Leão R. The upper respiratory tract microbiota of healthy adults is affected by *Streptococcus pneumoniae* carriage, smoking habits, and contact with children. *Microbiome*. 2023;11:199.
- von Wintersdorff CJ, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front Microbiol*. 2016;7:173.
- Carr VR, Witherden EA, Lee S, Shoaie S, Mullany P, Proctor GB, et al. Abundance and diversity of resistomes differ between healthy human oral cavities and gut. *Nat Commun*. 2020;11:693.
- Rodriguez RL, Konstantinidis KT. Estimating coverage in metagenomic data sets and why it matters. *ISME J*. 2014;8:2349–51.
- Zhou X, Wang J, Liu W, Huang X, Song Y, Wang Z, et al. Periodontal status and microbiologic pathogens in patients with chronic obstructive pulmonary disease and periodontitis: a case-control study. *Int J Chron Obstruct Pulmon Dis*. 2020;15:2071–9.
- Barbadoro P, Ponzio E, Coccia E, Prospero E, Santarelli A, Rappelli GGL, et al. Association between hypertension, oral microbiome and salivary nitric oxide: a case-control study. *Nitric Oxide*. 2021;106:66–71.
- Lin L, Yi X, Liu H, Meng R, Li S, Liu X, et al. The airway microbiome mediates the interaction between environmental exposure and respiratory health in humans. *Nat Med*. 2023;29:1750–9.
- Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife*. 2021;10:e65088.
- Zhou H, Beltrán JF, Brito IL. Functions predict horizontal gene transfer and the emergence of antibiotic resistance. *Sci Adv*. 2021;7: eabj5056.
- Gupta VK, Paul S, Dutta C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front Microbiol*. 2017;8: 1162.
- Mousavi SE, Delgado-Saborit JM, Adivi A, Pauwels S, Godderis L. Air pollution and endocrine disruptors induce human microbiome imbalances: a systematic review of recent evidence and possible biological mechanisms. *Sci Total Environ*. 2022;816: 151654.
- Aggarwal N, Kitano S, Puah GRY, Kittelmann S, Hwang IY, Chang MW. Microbiome and human health: current understanding, engineering, and enabling technologies. *Chem Rev*. 2023;123:31–72.
- Zhang YD, Fan SJ, Zhang Z, Li JX, Liu XX, Hu LX, et al. Association between residential greenness and human microbiota: evidence from multiple countries. *Environ Health Perspect*. 2023;131:87010.
- Man WH, van Houten MA, Mèrelle ME, Vlieger AM, Chu M, Jansen NJG, et al. Bacterial and viral respiratory tract microbiota and host characteristics in children with lower respiratory tract infections: a matched case-control study. *Lancet Respir Med*. 2019;7:417–26.
- Christensen H, May M, Bowen L, Hickman M, Trotter CL. Meningococcal carriage by age: a systematic review and meta-analysis. *Lancet Infect Dis*. 2010;10:853–61.
- Regev-Yochay G, Raz M, Dagan R, Porat N, Shainberg B, Pincó E, et al. Nasopharyngeal carriage of *Streptococcus pneumoniae* by adults and children in community and family settings. *Clin Infect Dis*. 2004;38:632–9.
- Sessitsch A, Wakelin S, Schloter M, Maguin E, Cernava T, Champomier-Verges MC, et al. Microbiome interconnectedness throughout environments with major consequences for healthy people and a healthy planet. *Microbiol Mol Biol Rev*. 2023;87: e0021222.
- Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol*. 2008;6:288–301.
- Debray R, Dickson CC, Webb SE, Archie EA, Tung J. Shared environments complicate the use of strain-resolved metagenomics to infer microbiome transmission. *Microbiome*. 2025;13:59.
- Konecna E, Videnska P, Buresova L, Urik M, Smetanova S, Smatana S, et al. Enrichment of human nasopharyngeal bacteriome with bacteria from dust after short-term exposure to indoor environment: a pilot study. *BMC Microbiol*. 2023;23:202.
- Zhou ZC, Liu Y, Lin ZJ, Shuai XY, Zhu L, Xu L, et al. Spread of antibiotic resistance genes and microbiota in airborne particulate matter, dust, and human airways in the urban hospital. *Environ Int*. 2021;153: 106501.
- O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*. 2009;374:893–902.
- Koliou MG, Andreou K, Lamnisis D, Lavranos G, Iakovides P, Economou C, et al. Risk factors for carriage of *Streptococcus pneumoniae* in children. *BMC Pediatr*. 2018;18:144.
- Van Winkelhoff AJ, Boutaga K. Transmission of periodontal bacteria and models of infection. *J Clin Periodontol*. 2005;32 Suppl 6:16–27.

43. Mac Aogáin M, Lau KJX, Cai Z, Kumar Narayana J, Purbojati RW, Drautz-Moses DJ, et al. Metagenomics reveals a core macrolide resistome related to microbiota in chronic respiratory disease. *Am J Respir Crit Care Med*. 2020;202:433–47.
44. Zhang L, Forst CV, Gordon A, Gussin G, Geber AB, Fernandez PJ, et al. Characterization of antibiotic resistance and host-microbiome interactions in the human upper respiratory tract during influenza infection. *Microbiome*. 2020;8:39.
45. Kolár M, Urbánek K, Látal T. Antibiotic selective pressure and development of bacterial resistance. *Int J Antimicrob Agents*. 2001;17:357–63.
46. National bacterial drug resistance surveillance report of China, 2020. <https://www.carss.cn/Report>. Accessed 20 May 2024.
47. Li X, Breyndrod A, Trivedi U, Russel J, Thorsen J, Shah SA, et al. Co-localization of antibiotic resistance genes is widespread in the infant gut microbiome and associates with an immature gut microbial composition. *Microbiome*. 2024;12:87.
48. Lee K, Raguideau S, Sirén K, Asnicar F, Cumbo F, Hildebrand F, et al. Population-level impacts of antibiotic usage on the human gut microbiome. *Nat Commun*. 2023;14:1191.
49. Lopatkin AJ, Syssoeva TA, You L. Dissecting the effects of antibiotics on horizontal gene transfer: analysis suggests a critical role of selection dynamics. *BioEssays*. 2016;38:1283–92.
50. He Z, Ren L, Yang J, Guo L, Feng L, Ma C, et al. Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study. *Lancet*. 2021;397:1075–84.
51. Wei J, Liu S, Li Z, Liu C, Qin K, Liu X, et al. Ground-level NO<sub>2</sub> surveillance from space across china for high resolution using interpretable spatiotemporally weighted artificial intelligence. *Environ Sci Technol*. 2022;56:9988–98.
52. Wei J, Li Z, Lyapustin A, Sun L, Peng Y, Xue W, et al. Reconstructing 1-km-resolution high-quality PM<sub>2.5</sub> data records from 2000 to 2018 in China: spatiotemporal variations and policy implications. *Remote Sensing of Environment*. 2021;252:112136.
53. Wei J, Li Z, Guo J, Sun L, Huang W, Xue W, et al. Satellite-derived 1-km-resolution PM<sub>10</sub> concentrations from 2014 to 2018 across China. *Environ Sci Technol*. 2019;53:13265–74.
54. Wei J, Li Z, Xue W, Sun L, Fan T, Liu L, et al. The ChinaHighPM<sub>10</sub> dataset: generation, validation, and spatiotemporal variations from 2015 to 2019 across China. *Environ Int*. 2021;146: 106290.
55. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90.
56. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
57. Rodriguez RL, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems*. 2018;3:10.
58. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. *Nat Protoc*. 2022;17:2815–39.
59. Millard SP. EnvStats, an R package for environmental statistics. In: *Encyclopedia of environmetrics*. 2001.
60. Ling W, Lu J, Zhao N, Lulla A, Plantinga AM, Fu W, et al. Batch effects removal for microbiome data via conditional quantile regression. *Nat Commun*. 2022;13:5418.
61. Franzosa EA, Mclver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods*. 2018;15:962–8.
62. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997* 2013.
63. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, et al. MEG-ARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Res*. 2020;48:D561–d569.
64. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res*. 2022;50:D912–d917.
65. Pärnänen K, Karkman A, Hultman J, Lyra C, Bengtsson-Palme J, Larsson DGJ, et al. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nat Commun*. 2018;9:3891.
66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
67. Li B, Yang Y, Ma L, Ju F, Guo F, Tiedje JM, et al. Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J*. 2015;9:2490–502.
68. Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DG, et al. METAXA2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Mol Ecol Resour*. 2015;15:1403–14.
69. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20:289–90.
70. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3.
71. Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinformatics*. 2020;69: e96.
72. Enav H, Ley RE. SynTracker: a synteny based tool for tracking microbial strains. *bioRxiv* 2021:2021.2010.2006.463341.
73. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6.
74. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10: 421.
75. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res*. 2018;46: e35.
76. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14:927–30.
77. Mallick H, Rahnavard A, Mclver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol*. 2021;17: e1009442.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.