Microbiome

RESEARCH

Open Access

Disruption and adaptation: infant gut microbiota's dynamic response to SARS-CoV-2 infection

Li-Ting Zhu[1,2], Lei Zhao[3], Yue Zhu[1], Xue-Li Xu[1,2], Jing-Jing Lin[1], Yi-Fang Duan[1], Lu Long[1,2], Yang-Yu Wu[1], Wen-Juan Xu[1,2], Jing-Yu Chen[1], Yu-Han Yin[1], Alex Ujong Obeten[1,2] and Qiansheng Huang[1,4*]

Abstract

Background The responses of the infant gut microbiota to infection significantly disrupt the natural intrahost evolutionary processes of the microbiome. Here, we collected a 16-month longitudinal cohort of infant gut microbiomes affected by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Then, we developed a multicriteria approach to identify core interaction network driving community dynamics under environmental disturbances, which we termed the Conserved Variated Interaction Group (CVIgroup).

Results The CVIgroup showed significant advantages on pinpointing a sparse set associated with the disturbances, as validated both our own and publicly available datasets. Leveraging the Oxford Nanopore Technology, we found this group facilitates the ecosystem's adaptation to environmental disruptions by enhancing the mobility of mobile genetic elements, including the reinforcement of the twin-arginine translocation pathway in response to increased virulence factors. Furthermore, the CVIgroup serves as an effective indicator of ecosystem health. The timescale for the gut microbiota's adaptation extends beyond 10 months. Members of the CVIgroup, such as *Bacteroides thetaiotaomicron* and *Faecalibacterium*, exhibit varying degrees of genomic structural variants, which contribute to guiding the community toward a new stable state rather than returning to its original configuration.

Conclusions Collectively, the CVIgroup offers a snapshot of the gut microbiota's adaptive response to environmental disturbances. The disruption and subsequent adaptation of the gut microbiota in infants after COVID-19 infection underscores the necessity of re-evaluating reference standards in the context of the post-pandemic era.

Keywords SARS-CoV-2, Infant, Gut microbiota, Interaction network, Environmental disturbances

In summary, CVIgroup offers a framework to identify key interactive network and characterize the dynamics of the infant gut ecosystem in response to infections. Given that early-life intestinal disturbance significantly disrupts natural evolutionary trajectories, our findings call for future research to redefine reference systems with the advent of the post-pandemic era.

*Correspondence:
Qiansheng Huang
qshuang@iue.ac.cn
[1] Xiamen Key Laboratory of Indoor Air and Health, State Key Laboratory for Ecological Security of Regions and Cities, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China
[2] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Beijing Key Laboratory of Biodiversity and Organic Farming, College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China
[4] National Basic Science Data Center, Beijing 100190, China

Background

The stepwise assembly of human gut microbiome is a complex process that begins at birth [1–3]. While considerable research has focused on the predictable assembly of gut microbiota in healthy infants [4–11], environmental factors, including maternal microbiome [12, 13], lifestyle [4, 14] and pathogenic infections [15–17], profoundly shape the gut microbiome. Among these factors, the ongoing implications of COVID-19 have become particularly concerning, especially regarding long COVID, which affects 10% to 30% of individuals in China even five years after the pandemic began [18]. Despite increasing attention, the role of gut microbiota dysfunction as a mechanism behind long COVID remains poorly understood [19, 20]. Thus, our study seeks to investigate the response of infant gut microbiota to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). A key challenge is identifying biologically relevant features within the gut ecosystem that can effectively indicate its response to the virus.

Interaction networks, which incorporate extensive ecological data, offer more precise ecosystem evaluations compared to mere species abundance metrics [21, 22]. Prior research has established robust methodologies for their construction, such as NetMoss [23]. However, pinpointing microbial groups that exhibit specific sensitivities to distinct environmental factors remains a challenge [22]. Networks based on cross-sectional data often struggle to represent the consistency of microbial interactions over time or across different locations [23–25]. Furthermore, the techniques employed to handle high-dimensional data can profoundly influence outcomes [26–28], highlighting the necessity for improved methods to accurately depict microbial community responses to environmental changes.

In this study, we developed a multicriteria approach aimed at identifying a core set of interaction networks that drive community dynamics across temporal scales under environmental disturbances. We refer to this environmentally sensitive cluster as the Conserved Variated Interaction Group (CVIgroup). This framework builds on criteria established in previous studies [29–31], including the network's sensitivity to environmental responses, conservativeness in time series, and structural sparsity. We applied the CVIgroup framework to a longitudinal cohort of infants with severe acute respiratory SARS-CoV-2 infection. By integrating data from Oxford Nanopore Technology (ONT) and Illumina sequencing, we assembled metagenome-assembled genomes (MAGs) for the members of the CVIgroup, with the goal of clarifying how the CVIgroup responds to viral infections.

Methods

Cohort recruitment

After obtaining informed consent from the guardians, we tracked a longitudinal cohort of 58 infants on a monthly basis, collecting and analyzing infant feces from May 2022 to September 2023 in Xiamen, Fujian Province, China. Concurrently, comprehensive cohort information was compiled. In total, 397 fecal samples were included in this study (Table S1). The ethical review of this study has been formally registered in the Chinese National Medical Research Registration and Information System (https://www.medicalresearch.org.cn/) under registration number HSR-24–000233.

Sample collection and DNA extractions

Parents collected infant stool samples at home and stored them in an incubator with an ice box, maintaining a temperature of approximately 4 °C. Delivery personnel were promptly contacted to ensure the samples were transported to the laboratory within 30 min to 3 h after collection. Upon receipt at the laboratory, we aliquoted and preserved the samples in an ultra-low temperature medical freezer set at −80 °C, maintaining optimal conditions until DNA extraction. The DNA was extracted using the Fast DNA Stool Mini Kit (Qiagen, Germany) and stored in the ultra-low temperature medical freezer at −80 °C for future analysis.

16S rRNA library construction and sequencing

DNA samples were quantified using the Invitrogen Qubit 4.0 Spectrophotometer (Thermo Fisher Scientific, USA) and normalized to a concentration of 10 ng/μl. Subsequently, the V3-V4 hypervariable region of the bacterial 16S rRNA gene was sequenced on the Illumina NovaSeq instrument. The obtained paired-end reads (2 × 250 bp) were refined using the DADA2 [32] pipeline, resulting in the generation of amplicon sequence variants (ASVs). Taxonomic assignments of ASV representative sequences were performed with a confidence threshold of 0.8 by a pre-trained Naive Bayes classifier which was trained on the Greengenes2 [33]. We filtered to retain only those ASVs exhibiting a relative abundance of at least 0.1% in a minimum of two samples, resulting in a final dataset of 1,149 ASVs.

Calculation of microbiome stability index

According to a previous study, the Average Variation Degree (AVD) serves as a robust metric for assessing the stability of microbial communities [34]. Typically, a lower AVD value signifies greater community stability. First, we calculate the variation degree a_i for each of the 1,149 ASVs as follows:

$$|a_i| = \frac{|x_i - \overline{x_i}|}{\delta_i} \tag{1}$$

In this formula, x_i represents the relative abundance of ASV_i in each sample i, while $\overline{x_i}$ and δ_i denote the mean and standard deviation of the abundance of ASV_i across all samples, respectively.

Subsequently, the AVD is derived from the variation degree:

$$AVD = \frac{\sum_{i=1}^{n} |a_i|}{k \times n} = \frac{\sum_{i=1}^{n} \frac{|x_i - \overline{x_i}|}{\delta_i}}{k \times n} \tag{2}$$

Here, k refers to the number of samples and n to the number of ASVs.

Calculation of Gut Microbiome Health Index (GMHI)

To accommodate the specific characteristics of this study cohort, the GMHI was adapted from the framework proposed by Gupta et al. [35, 36]. The core concept of GMHI involves identifying an optimal set of microbial features (M_H and M_D) that are enriched in HC or depleted in HC. Initially, we employed Wilcoxon differential testing to distinguish between the HC and IC. For each sample i, the "collective abundance" ψ of species M_H and M_D is defined as follows:

$$\psi_{M_{H,i}} = \frac{R_{M_H,i}}{|M_H|} \sum\nolimits_{j \in I_{M_H}} \left| n_{j,i} \ln(n_{i,j}) \right| \tag{3}$$

$$\psi_{M_{D,i}} = \frac{R_{M_D,i}}{|M_D|} \sum\nolimits_{j \in I_{M_D}} \left| n_{j,i} \ln(n_{i,j}) \right| \tag{4}$$

Here, $R_{M_H,i}$ and $R_{M_D,i}$ represent the abundance of M_H and M_D in sample i; $|M_H|$ and $|M_D|$ denote the size of the set M_H and M_D; I_{M_H} and I_{M_D} are the index sets; and $n_{j,i}$ is the relative abundance of j in sample i.

Subsequently, the collective abundance $\psi_{M_{H,i}}$ and $\psi_{M_{D,i}}$ are compared using the ratio of their respective abundances:

$$GHMI_i = \log_{10}\left(\frac{\psi_{M_{H,i}}}{\psi_{M_{D,i}}}\right) = \log_{10}\left(\frac{\frac{R_{M_{H,i}}}{M_H} \sum_{j \in I_{M_H}} \left| n_{j,i} \ln(n_{i,j}) \right|}{\frac{R_{M_{D,i}}}{M_D} \sum_{j \in I_{M_D}} \left| n_{j,i} \ln(n_{i,j}) \right|}\right) \tag{5}$$

$GHMI_i$ reflects the relative collective abundance of M_H in sample i compared to that of M_D. Specifically, a positive $GHMI_i$ indicates a higher prevalence of M_H microorganisms in the sample, while a negative $GHMI_i$ signifies a predominance of M_D microorganisms. A ratio of zero implies an equilibrium between these two microbial groups.

Pipeline of the identification of CVIgroup

In neuroimaging research, Structural Covariance Networks (SCNs) are frequently employed to elucidate the effects of diseases on brain structure [37]. Additionally, we found that covariance can be more informative through repeated drills with simulated data (details in the supplementary methods). We then calculated taxon-taxon covariance matrices for both the HC group and the IC group, each group consisted of 1,149 ASVs. These matrices, denoted as $Cov(i,j)_{HC}$ and $Cov(i,j)_{IC}$ respectively, were calculated on a monthly basis.

For each pair of taxa i and j, we calculated the covariance $Cov(i,j)_{HC}$ using the relative abundance data in HC.

The covariance $Cov(i,j)_{HC}$ is given by:

$$Cov(i,j)_{HC} = \frac{1}{n-1} \sum\nolimits_{k=1}^{n} \left(\mathrm{X}_{ik} - \overline{\mathrm{X}}_i\right)\left(\mathrm{X}_{jk} - \overline{\mathrm{X}}_j\right) \tag{6}$$

where X_{ik} is the relative abundance of taxon i in sample k of group A; $\overline{\mathrm{X}}_i$ is the mean relative abundance of taxon i in HC; and n is the sample size of HC. Similarly, for each pair of bacteria i and j, calculate the covariance $Cov(i,j)_{IC}$ using the relative abundance data in IC.

By examining the differences between these covariance matrices, $\Delta Cov(i,j)$, we constructed 10 Differential Structural Covariance Networks (DSCNs).

The difference in covariance matrices was calculated as:

$$\Delta Cov(i,j) = Cov(i,j)_{IC} - Cov(i,j)_{HC} \tag{7}$$

Subsequently, hierarchical clustering was performed to reorder the indices based on the covariance difference in DSCNs. This reordering aimed to position similar microbial groups closer together, facilitating the identification of patterns. The covariance differences $\Delta Cov(i,j)$ were then normalized relative to the maximum monthly $\Delta Cov(i,j)$ value:

$$\Delta Cov(i,j)_{nor} = \frac{\Delta Cov(i,j)}{\Delta Cov(i,j)_{max}} \tag{8}$$

With a threshold of 10 from the reordered DSCNs [31], the DSCNs were then decomposed into eigenvalues and eigenvectors, where the eigenvector weights (v_t) indicated the importance of each microbial group within the corresponding feature dimension. We focused on the first principal component (PC1) due to its significant contribution to the overall variance (Fig. 1c).

To measure the balance, fine-grained information, smoothness, and continuity of the data, we set the window size to 3 to capture rapid dynamic changes,

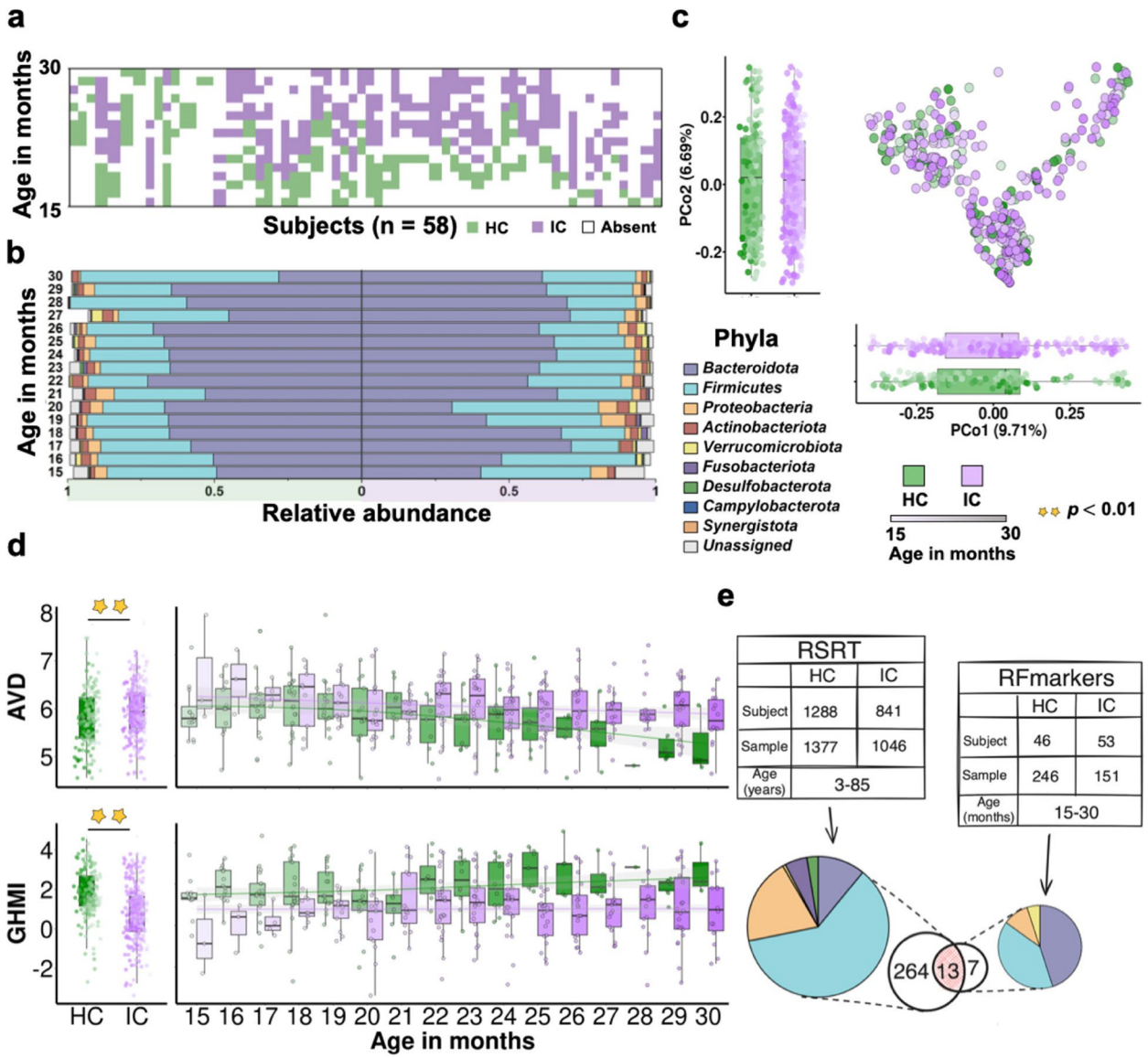

Fig. 1 Disruption of microbial ecosystems by viral infections. **a** Schematic illustration of the study cohort. **b** Taxa distribution at phylum level. Stacked bar charts compare the distribution of species at the phylum level between two distinct cohorts. **c** PCoA plots at the ASV level are presented, accompanied by box plots showing sample values on the PCo1 and PCo2 axes. Statistical analysis indicates no significant differences between the two groups on either axis (two-sided Wilcox test, $p > 0.05$). **d** Comparative distribution of AVD and GHMI. This section displays the comparative distribution of AVD and GHMI across various cohorts. **e** Data source information of RSRT and RFmarkers. Detailed information on RSRT and RFmarkers is provided, supported by pie plots illustrating the taxa composition within each dataset

employing a fixed-length sliding window (n = 8). For each window, we calculated the average eigenvector weights ($\overline{v}_t$) for the top 5% of taxa (second-sparsification). This methodology smoothed temporal fluctuations, thereby highlighting microbial groups conservatively deemed important over the window. By ranking these average eigenvector weights, we identified the most critical microbial groups within each window. These groups, constituting 20 members, were classified under the CVIgroup.

Illumina library preparation and sequencing

According to the DNA extraction method described above, metagenomic DNA was extracted from 18 faecal samples (sample information in Table S9). DNA samples were quantified using the Invitrogen Qubit 4.0 Spectrophotometer (Thermo Fisher Scientific, USA). Qualified DNA samples were sheared to 400 bp with a Covaris M220 device, and then concentrated using NEXTFLEXTM Rapid DNA-Seq Kit (Bioo Scientific, USA) following the manufacturer's instructions. Nanopore

sequencing libraries were prepared using the One-pot Native Barcoding kit (EXP-NBD104, Oxford Nanopore Technologies) and the Ligation Sequencing kit (SQK-LSK109, Oxford Nanopore Technologies).

The fast5 data from nanopore reads were base-called with Guppy v.3.03 using the recommended parameter settings. Next, Fitlong v.0.2.1 was applied to filter pass reads shorter than 500 bp, producing clean data. The reads were then assembled with metaFlye v.2.9 [38], and high-quality long reads were mapped to contigs using minimap2. After two rounds of polishing with Racon v.1.5.0, Pilon v.1.24 was used to generate the final genome assemblies.

Metagenomic binning

Metabat v.2.12.1 [39], CONCOCT v.0.5.0 [40], and Maxbin2 v.2.2.5 [41] were employed to generate the initial bins for each sample. Subsequently, DASTool v.1.1.0 [42] was utilized to merge the bins produced by multiple software tools, facilitating regeneration and refinement of the bins through RefineM v.0.0.24. The quality of the bins was assessed using CheckM v.1.0.12 [43]. Initially, marker genes within the bin sequences were identified using Hidden Markov Models (HMM) and compared against universal single-copy genes via HMMER v.3.1.2. The bins were incorporated into the reference phylogenetic tree using pplacer, which enabled the identification of a specific set of pedigree markers for quality assessment of the bins. Following this, dRep v.3.4.2 was applied to perform redundant clustering, while Mash [44] facilitated the comparison of these high-quality Metagenome-Assembled Genomes (MAGs). Two rapid clustering methods based on average nucleotide identity (ANI) were implemented, with a default threshold established at 99%. Ultimately, deredundant MAGs were obtained, followed by annotation using GTDBTK v.2.3.0.

Identification of mobile genetic elements (MGEs) of CVIgroup

Based on the taxonomic annotation results, we selected the MAGs associated with CVIgroup members at the genus level, retaining only those MAGs present in both the HC and IC cohorts. This process will yield 8 CVIgroup members and a total of 16 MAGs. We then subsequently annotate these MAGs for MGEs utilizing geNomad [45].

Functional annotation of the genome

We conducted KEGG functional annotation on the genes encoded by the MAG genomes and MGEs to quantify the number of annotated genes in each pathway. To minimize potential bias in the annotation results, we selected results with a gene count difference exceeding three between the HC and IC groups in the subsequent analysis. Besides, we looked at MAGs involved in virulence.

Results

Viral infections disrupt the gut ecosystem of infants

From June 2022 to September 2023, we conducted a comprehensive longitudinal study to analyze the gut microbiomes of infants in relation to first SARS-CoV-2 infection. We collected monthly stool samples from 58 infants aged 15 to 30 months, resulting in a total of 397 samples spanning the period before and after infection with Omicron variants (Fig. 1a). Of these, 246 samples were from 46 infected infants, referred to as the infected cohort (IC). The remaining 151 samples, obtained from

(See figure on next page.)

Fig. 2 Framework for CVIgroup identification and alterations in the microbial interaction network. **a** Dataset description. The system identification process employs training data for model development, while test data is reserved for performance validation. **b** Construction process of DSCNs. Covariance matrices are generated for both temporal cohort groups. Disparities for each DSCN unit are computed using a specific formula detailed in the Methods section. **c** Eigen decomposition. After initial sparsification, eigen decomposition is applied to the DSCNs. The eigenvector of the principal component (PC1) is extracted, and then, 8 average feature values are derived via a sliding window technique. **d** Phylogenetic tree. The phylogenetic tree represents 90 taxa post-sparsification, with phylum classifications denoted by distinct colors. The outer gray bars depict the importance fraction of the 20 CVIgroup members. **e** Predictive ability of different methods. ROC curves illustrate the classification performance using three distinct 20-marker sets: CVIgroup, NetMoss markers, and RFmarkers, within this study. **f** Interaction networks associated with CVIgroup post first-sparsification. This panel illustrates the intricate relationships within the HC and IC cohorts, stratified by age groups. Each node, depicted as either a solid or hollow circle, represents a member of the CVIgroup or other taxa, respectively, with colors indicating their taxonomic classifications. The edges between nodes signify significant taxon-taxon correlations ($P < 0.05$), with an absolute SparCC's rho threshold of 0.3 used for visualization. **g** Variability in networks structure across cohorts: Box plots illustrating the variability indices in networks structure across different age cohorts ($n = 11$), and using a two-sided Wilcoxon test for pane. **h** PCoA mapping of compositional distribution. PCoA charts the compositional distribution of CVIgroup members across various subjects, providing insights into the spatial arrangement and clustering of these members within different cohorts. i. Abundance and interaction of CVIgroup members. A scatter plot displays the differences in abundance among CVIgroup members, visualizing these abundances along with their respective contributions to networks connectivity

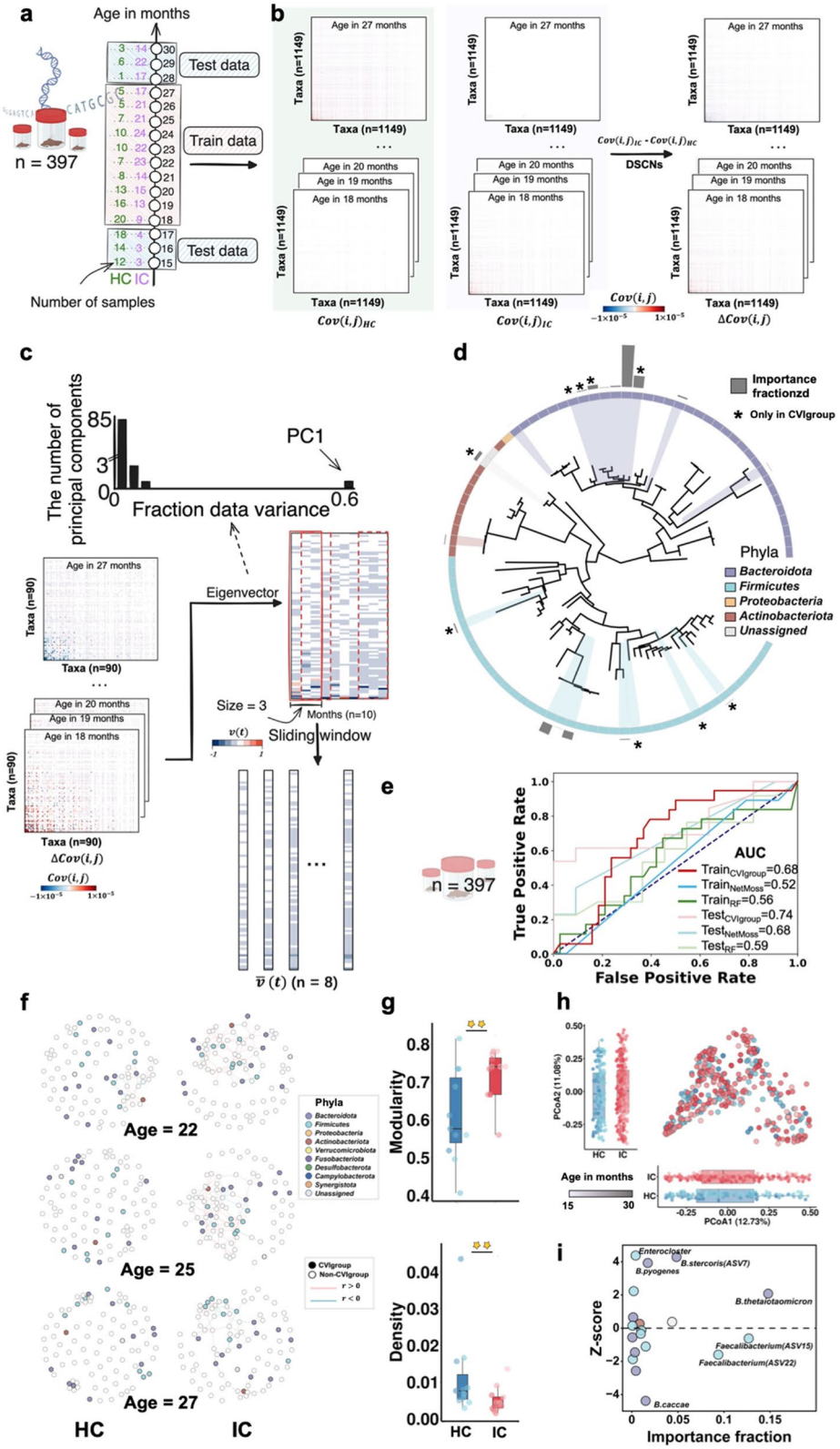

Fig. 2 (See legend on previous page.)

53 infants who had not yet been infected, constituted the healthy cohort (HC), as detailed in Table S1. Using 16S rRNA sequencing, we analyzed the dynamic shifts in microbial composition. At the phylum level, we observed significant decreases in the relative abundances of *Actinobacteriota* and *Verrucomicrobiota* in the IC compared to the HC ($p < 0.01$) among the top five categories (Fig. 1b, Fig. S1a). However, Principal Coordinates Analysis (PCoA) at the ASV level did not exhibit significant differentiation between groups (Fig. 1c).

Our findings revealed a significant increase in the Average Variation Degree (AVD) ($p < 0.01$) following viral infection, indicating a pronounced reduction in microbial community stability (Fig. 1d). Further age-based analysis revealed that AVD in HC gradually decreased with age, suggesting improved microbial community stability over time. However, this stabilization trend was attenuated in IC (Fig. 1d). To elucidate this pattern, we used a mixed linear model to compare the effects of age and viral infection on the microbial ecosystem. Both factors significantly impacted AVD, with age having a more pronounced effect than viral infection (Fig. S1b). Thus, while microbial community stability generally increases with age, this process is significantly hindered in the presence of viral infection. Furthermore, the Gut Microbiome Health Index (GMHI) was developed using 20 taxa identified through random forests, referred to as RFmarkers (Table S3).

Notably, 13 of these RFmarkers are part of the Reported SARS-COV-2 Relevant Taxa (RSRT) set (Fig. 1e, Table S4). Our analysis showed a significant decrease in GMHI during the disease state ($p < 0.01$), indicating impaired gut health. In contrast, in the HC, GMHI gradually increased with age, while in the IC, it remained relatively constant. Mixed linear model analysis demonstrated that the effect of viral infection on GMHI was significantly greater than that of age (Fig. S1b). This finding explains why GMHI in IC remains constant across age groups, suggesting that the influence of viral infection outweighs the age-related improvements in GMHI.

Taxon-taxon interaction network markedly expanded after infection

We then developed a comprehensive computational framework grounded in the principles of sensitivity, conservatism, and sparsity optimization. Initially, we partitioned the data (Fig. 2a, Table S5) and used the training dataset to construct differential covariance networks (DSCNs) (Fig. 2b). In these networks, each cell $\Delta Cov(i, j)$ reflects the change in covariance between taxon i and j in the IC relative to the HC. 90 taxa were identified after initial sparsification, and then eigen decomposition was performed on their DSCNs. Principal Component 1 (PC1) was selected for its significance (Fig. 2c). To detect taxa most responsive to viral infection over time, we analyzed consecutive time windows, ultimately pinpointing 20 taxa with the highest average weights (Fig. 2c). These taxa, designated as the CVIgroup, consist of a diverse array of microbial taxa: 10 from the phylum *Bacteroidota*, 8 from *Firmicutes*, 1 from *Proteobacteria*, and 1 unclassified member (Fig. 2d, Table S7).

Subsequently, we assessed the efficacy of the CVIgroup in distinguishing between HC and IC, comparing its performance with NetMoss markers and RFmarkers. The CVIgroup achieved a higher accuracy, evidenced by an area under the curve (AUC) of 0.68 (Fig. 2e). To ensure its external applicability, we tested the CVIgroup on a separate dataset, where it maintained superior performance with an AUC of 0.74 (Fig. 2e). Additionally, to further assess the reproducibility of the CVIgroup's computational process, we compiled an industrial cohort (INDC) and a non-industrial cohort (NINDC) from previous studies [4]. This dataset was meticulously partitioned, with 625 samples used for model training, while the remaining samples served as the test set (Table S6). Finally, CVIgroup showed a robust advantage in this analysis (Fig. S2).

Next, we characterized the connectivity among CVIgroup members across different age groups in two cohorts (Fig. 2f, Fig. S3). Compared to the HC, the IC exhibited significantly increased modularity and markedly decreased network density (Fig. 2g). This trend was generally robust in analyses with different cutoffs (Fig. 2f, Fig. S3). Interestingly, the infection did not significantly alter the overall composition of the CVIgroup community (Fig. 2h), suggesting it was unlikely a driving factor in the interactome differences. To further elucidate this, we calculated z-scores to normalize the effect size of differential abundance, the results showed no significant correlation between network structure drivers and abundance shifts among CVIgroup members (Fig. 2i, Table S8).

(See figure on next page.)

Fig. 3 Mobile genetic elements displayed high variation due to virus infection. **a** Comparisons of MGEs between the microbiomes of individuals from HC and IC. MGEs count of genes identified by geNomad. Black lines represent data of individual species. Statistical analysis indicates significant differences between the two groups (paired Wilcox test, $p < 0.01$). **b** Schematic of the pathway. Functional genes associated with the plasmid-encoded TatABC transporter are present in the mobile genome of IC but not HC. **c** Phylogenetic tree of the retrieved virulence factor-carrying MAGs in gut. **d** Correlation matrix using Spearman rank correlation between TatA, TatB and TatC transport proteins and virulence factors

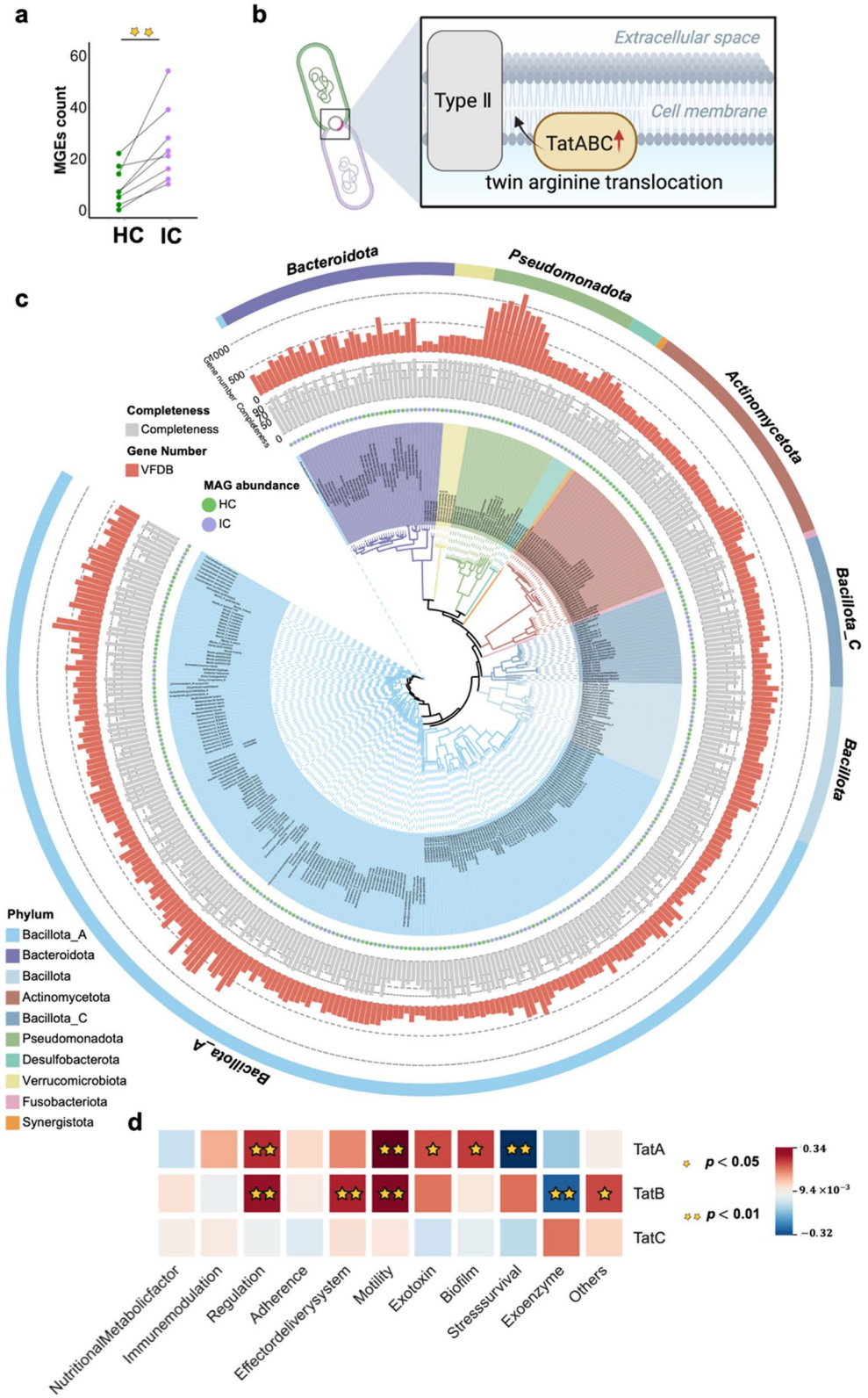

Fig. 3 (See legend on previous page.)

Mobile genome of CVIgroup response to a perturbed ecosystem

The transfer of genetic information among organisms may shed light on response mechanisms within ecosystems. We then incorporated both Oxford Nanopore Technology (ONT) and Illumina reads to compile 16 representative genomes of CVIgroup members (Table S10) and utilized geNomad [45] to identify their mobile genetic elements (MGEs). Our findings indicate a significant increase in the number of MGEs in these CVIgroup members within the IC group (Fig. 3a, $p < 0.01$). The functional analysis of these mobile genes encompasses 7 KEGG pathways (Table S11). In comparison to the HC group, following viral infection, there was an increase in classifications related to pathways involving metabolism, genetic information processing, and human diseases, including pathways associated with amino acid metabolism, replication and repair, and neurodegenerative disorders (Fig. S4, Table S11). For example, after infection, plasmid-encoded genes related to the twin-arginine translocation (Tat) pathway increased, including TatA, TatB and TatC transport proteins (Fig. 3b).

Tat pathway is an important virulence factor in different bacterial pathogens [46–48]. A worrisome trend is the increasing number of genes encoding various virulence factors within the microbial community after viral infection, including those associated with immune modulation (Fig. 3c, Fig. S5). Specifically, we quantified the spearman correlation between TatA, TatB and TatC transport proteins and virulence factors in ecosystem (Fig. 3d). 5 categories of virulence factors exhibit a significant association with TatA and TatB, particularly in the regulation and motility categories, which show a strong positive correlation with these TatA and TatB. In contrast, TatC shows no significant correlation with the virulence factor genes number.

Ecological drivers of the dynamic of disturbed ecosystems

To further investigate the use of the CVIgroup in assessing the recovery of gut ecosystem, we first applied iterative Principal Component Analysis (iPCA) to determine when stabilization occurs in the CVIgroup system [31]. This critical point was found at 20 months of age (Fig. 4a), consistent with previous studies [31]. To minimize age-related confounding effects, we performed subsequent analyses on data collected from 20 months onwards (n = 286). Concurrent AVD and GHMI results indicated that although ecosystem fluctuations gradually stabilized over time, they still exhibited a large variance compared to the state before the infection (Fig. 4b). Next, we performed eigenvector decomposition of the CVIgroup covariance matrices across different infection stages. The first three principal components (PC1, PC2, and PC3) accounted for 86.48% of the total variance, and we used these components to generate spatial coordinates for each infection period (Fig. 4c). Our findings indicated that ecosystem variability peaked during Period 3, approaching a state of recovery by Period 10 (Fig. 4c). The combined data suggest that after environmental disturbances, the ecosystem requires a considerable recovery period and may stabilize at a new equilibrium rather than fully reverting to its original healthy state.

We further elucidated the key microbial taxa driving ecosystem restoration (Fig. 4d), identifying *Bacteroides stercoris* (*B. stercoris*), *Bacteroides thetaiotaomicron* (*B. thetaiotaomicron*), *Bacteroides caccae* (*B. caccae*), *Faecalibacterium* (ASV15), and *Faecalibacterium* (ASV22). Collinearity analysis of the genomes of *B. thetaiotaomicron* and *Faecalibacterium* revealed that structural variations occurred in the genomes of these species during infection (Fig. 4e-g, Fig. S6). By aligning the genome of *B. thetaiotaomicron* from IC with that from HC, we identified five types of structural variants (SVs): deletions, insertions, duplications, inversions, and translocations. Additionally, we identified 162 SVs longer than 50 bp, 116 SVs longer than 1 Kb, with the longest fragment measuring 590,771 bp (Fig. 4f). Gene variations were mainly related to functions such as carbohydrate metabolism, replication and repair, and translation (Fig. 4g).

(See figure on next page.)

Fig. 4 Resilience of Microbial Communities. **a** iPCA test. Using 30 months as the reference time point, we performed the iPCA test. The eigenvalue represents the degree of difference at each time point relative to the reference, defined mathematically through the PC1 obtained via iPCA. The inset illustrates the degree of variation in the data relative to the preceding point. b This scatter plot presents the unique microbial community profiles (AVD and GHMI) of individuals at each infection stage. **c** This PCA plot shows the centroids of microbial communities across different infection stages. Arrows depict the temporal progression of these stages. The inset indicates the weighted Euclidean distance of each spatial point from the reference point, "Health". **d** Decomposition of each taxon's contribution to the three principal components. Arrows highlight taxa with significant contributions, elucidating their roles in the microbial community structure. This dataset was categorized into distinct infection periods: Health (Period 0), 1 month post-infection (Period 1), 2 months post-infection (Period 2), and continuing monthly up to 10 months post-infection (Period 10), resulting in 11 distinct periods. **e** Collinearity analysis between the IC genome and the HC genome of *B. thetaiotaomicron*. **f** The number of SVs identified for *B. thetaiotaomicron*. **g** The differentially pathway between the two genomes, where the red font indicates a higher number of genes in the IC group than in the HC, and vice versa in blue

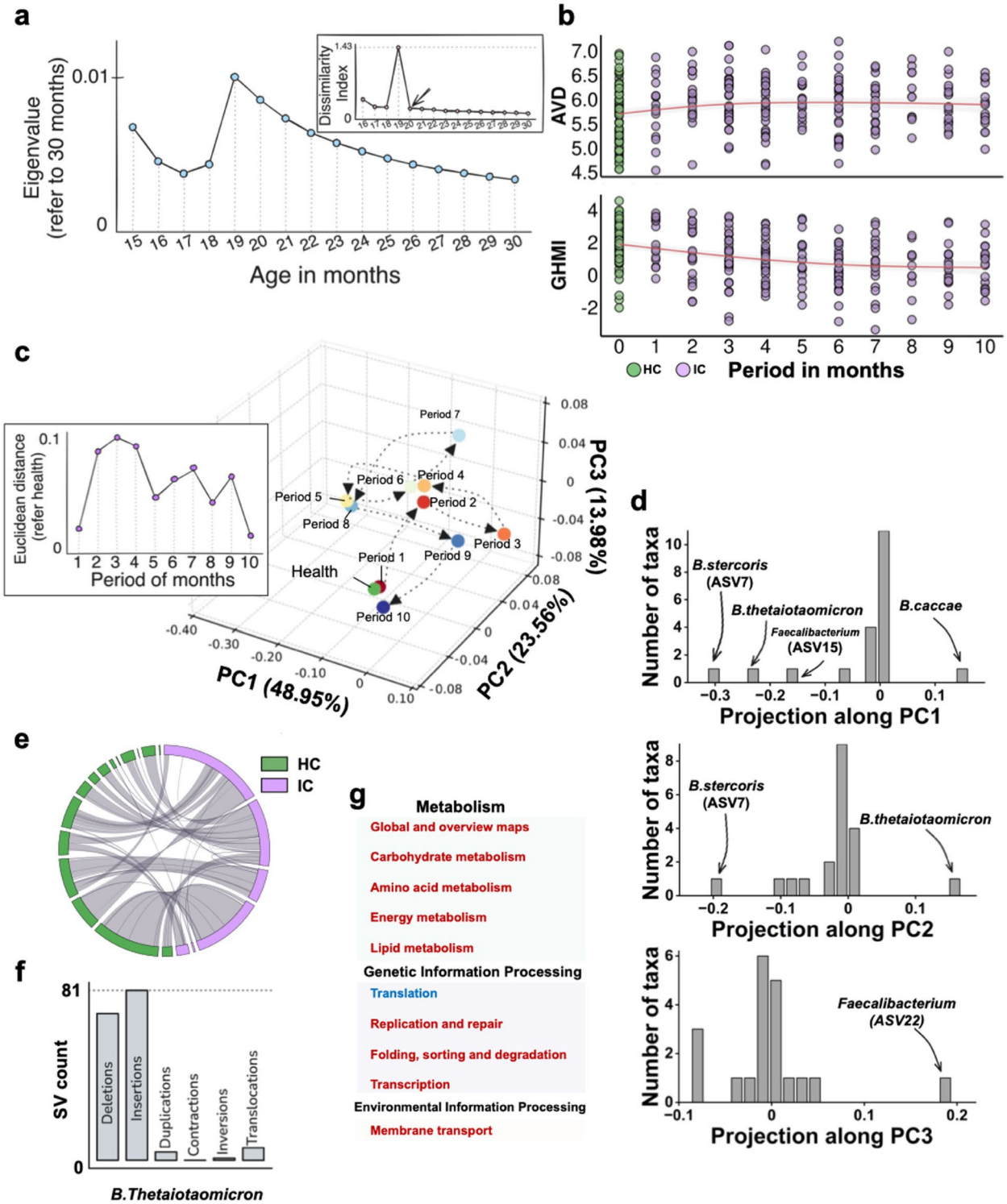

Fig. 4 (See legend on previous page.)

Discussion

Here, our extensive 16-month observation of infants firstly exposed to SARS-CoV-2 offers profound insights into their disturbed ecosystems. We observed significant changes in the microbial community AVD and GHMI, suggesting that the ecosystem has been disturbed. However, the analysis of community composition based on microbial relative abundance did not reveal significant changes, which contrasts with previous research findings [49]. Furthermore, relying solely on abundance differences to characterize the disrupted ecosystem proved overly simplistic, resulting in a variety of marker types that complicate future selection and ultimately fail to meet the original objective [23]. To address these challenges, we developed a scalable CVIgroup framework boasts several pivotal features, including the sensitivity to environmental responses, conservativeness in time series, and structural sparsity. This framework has consistently outperformed traditional metrics across various datasets by capturing higher-dimensional information and providing more nuanced insights.

We found that the interaction network of the CVIgroup may adapt to environmental disturbances by expanding the interaction network itself, rather than altering taxonomic abundance. To investigate this further, we combined data from ONT and Illumina sequencing to assemble MAGs for the members of the CVIgroup. These taxa may enhance specific community functions by sharing the MGEs [50], such as the Tat pathway. The transfer of these related genes may enhance the efficiency of delivering folded proteins into the periplasm within the gut after infection [51]. Previous studies indicate that the Tat pathway, when expressed under certain environmental conditions, is linked to the virulence factors of various bacterial pathogen [47, 48]. Our findings support that virulence factors within the microbial community increase following infection. Notably, 5 categories of virulence factors show significant associations with TatA and TatB, particularly in the regulation and motility categories, which demonstrate a strong positive correlation with these components.

Furthermore, 3 months post-infection, the variability of the microbial community peaked, this process is dynamically consistent with the host's immune response [52]. Ecosystems require prolonged periods to recover from environmental disturbances. During this dynamic process, CVIgroup members, such as *B. thetaiotaomicron* and *Faecalibacterium* display varying degrees of genomic SVs, suggesting that SVs can have a considerable impact in adaptive divergence between populations [53, 54]. Most mutations that confer resistance to adverse conditions are largely irreversible [55, 56], which means that recovery often does not restore the ecosystem to its original state. Instead, it frequently results in the establishment of a new equilibrium. This underscores the complex interplay between microbial communities, ecological resilience, and adaptive mechanisms.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-025-02029-6.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

We sincerely thank the volunteer families for their active participation and strong support. We also thank the team members and Yi-bi Wu for their valuable help during the research process.

Authors' contributions

Q.S.H. designed this cohort study, L.T.Z. was responsible for extracting DNA from human fecal samples, generating the sequencing dataset, and designing the computational pipeline for implementation. Y.Z., X.L.X., and L.L. helped with the interaction analysis and metagenomic sequencing. Z.L., Y.Y.W. and Y.F.D. provided suggestions for computational adjustments. J.J.L., W.J.X., J.Y.C., Y.H.Y. and A.U.O. helped with generating the sequencing dataset. L.T.Z. and Q.S.H. wrote the manuscript.

Funding

This work was supported by National Natural Science Foundation of China (32161143016, 42177362) and National Basic Science Data Center "Environment Health DataBase" (NO. NBSDC-DB-21).

Data availability

The raw sequencing reads have been deposited in the Genome Sequence Archive (GSA): https://bigd.big.ac.cn/gsa/browse/CRA020628.

Declarations

Ethics approval and consent to participate

After obtaining informed consent from the guardians, we tracked a longitudinal cohort of 58 infants on a monthly basis, collecting and analyzing infant feces from May 2022 to September 2023 in Xiamen, Fujian Province, China. The ethical review of this study has been formally registered in the Chinese National Medical Research Registration and Information System (https://www.medicalresearch.org.cn/) under registration number HSR-24–000233.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 7 November 2024 Accepted: 4 January 2025
Published online: 11 March 2025

References

1. Barreto HC, Gordo I. Intrahost evolution of the gut microbiota. Nat Rev Microbiol. 2023. https://doi.org/10.1038/s41579-023-00890-6.
2. Liang G, Zhao C, Zhang H, Mattei L, Sherrill-Mix S, Bittinger K, et al. The stepwise assembly of the neonatal virome is modulated by

breastfeeding. Nature. 2020;581(7809):470–4. https://doi.org/10.1038/s41586-020-2192-1.
3. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature. 2018;562(7728):583–8. https://doi.org/10.1038/s41586-018-0617-x.
4. Olm MR, Dahan D, Carter MM, Merrill BD, Yu FB, Jain S, et al. Robust variation in infant gut microbiome assembly across a spectrum of lifestyles. Science. 2022;376(6598):1220–3. https://doi.org/10.1126/science.abj2972.
5. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604(7907):732–9. https://doi.org/10.1038/s41586-022-04567-7.
6. Donald K, Finlay BB. Early-life interactions between the microbiota and immune system: impact on immune system development and atopic disease. Nat Rev Immunol. 2023;23(11):735–48. https://doi.org/10.1038/s41577-023-00874-w.
7. Peppas I, Ford AM, Furness CL, Greaves MF. Gut microbiome immaturity and childhood acute lymphoblastic leukaemia. Nat Rev Cancer. 2023;23(8):565–76. https://doi.org/10.1038/s41568-023-00584-4.
8. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. The Human Microbiome and Child Growth - First 1000 Days and Beyond. Trends Microbiol. 2019;27(2):131–47. https://doi.org/10.1016/j.tim.2018.09.008.
9. Vandenplas Y, Carnielli VP, Ksiazyk J, Luna MS, Migacheva N, Mosselmans JM, et al. Factors affecting early-life intestinal microbiota development. Nutrition. 2020;78:110812. https://doi.org/10.1016/j.nut.2020.110812.
10. Hsu CL, Schnabl B. The gut-liver axis and gut microbiota in health and liver disease. Nat Rev Microbiol. 2023;21(11):719–33. https://doi.org/10.1038/s41579-023-00904-3.
11. Gehrig JL, Venkatesh S, Chang HW, Hibberd MC, Kung VL, Cheng J, et al. Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. Science. 2019;365(6449); https://doi.org/10.1126/science.aau4732.
12. Vatanen T, Jabbar KS, Ruohtula T, Honkanen J, Avila-Pacheco J, Siljander H, et al. Mobile genetic elements from the maternal microbiome shape infant gut microbial assembly and metabolism. Cell. 2022;185(26):4921–36 e15; https://doi.org/10.1016/j.cell.2022.11.023.
13. Gomez de Agüero M, Ganal-Vonarburg SC, Fuhrer T, Rupp S, Uchimura Y, Li H, et al. The maternal microbiota drives early postnatal innate immune development. Science. 2016;351(6279):1296–302; https://doi.org/10.1126/science.aad2571.
14. You X, Rani A, Özcan E, Lyu Y, Sela DA. Bifidobacterium longum subsp. infantis utilizes human milk urea to recycle nitrogen within the infant gut microbiome. Gut Microbes. 2023;15(1):2192546; https://doi.org/10.1080/19490976.2023.2192546.
15. Fasano A, Chassaing B, Haller D, Flores Ventura E, Carmen-Collado M, Pastor N, et al. Microbiota during pregnancy and early life: role in maternal-neonatal outcomes based on human evidence. Gut Microbes. 2024;16(1):2392009. https://doi.org/10.1080/19490976.2024.2392009.
16. Ronan V, Yeasin R, Claud EC. Childhood Development and the Microbiome-The Intestinal Microbiota in Maintenance of Health and Development of Disease During Childhood Development. Gastroenterology. 2021;160(2):495–506. https://doi.org/10.1053/j.gastro.2020.08.065.
17. Lim ES, Zhou Y, Zhao G, Bauer IK, Droit L, Ndao IM, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. Nat Med. 2015;21(10):1228–34. https://doi.org/10.1038/nm.3950.
18. Qin S, Zhang Y, Li Y, Huang L, Yang T, Si J, et al. Long COVID facts and findings: a large-scale online survey in 74,075 Chinese participants. The Lancet Regional Health - Western Pacific. 2024;52; https://doi.org/10.1016/j.lanwpc.2024.101218.
19. Zhang F, Lau RI, Liu Q, Su Q, Chan FKL, Ng SC. Gut microbiota in COVID-19: key microbial changes, potential mechanisms and clinical applications. Nat Rev Gastroenterol Hepatol. 2023;20(5):323–37. https://doi.org/10.1038/s41575-022-00698-4.
20. Nagata N, Takeuchi T, Masuoka H, Aoki R, Ishikane M, Iwamoto N, et al. Human Gut Microbiota and Its Metabolites Impact Immune Responses in COVID-19 and Its Complications. Gastroenterology. 2023;164(2):272–88. https://doi.org/10.1053/j.gastro.2022.09.024.
21. Moreno-Mateos D, Alberdi A, Morrien E, van der Putten WH, Rodriguez-Una A, Montoya D. The long-term restoration of ecosystem complexity. Nat Ecol Evol. 2020;4(5):676–85. https://doi.org/10.1038/s41559-020-1154-1.
22. Faust K. Open challenges for microbial network construction and analysis. ISME J. 2021;15(11):3111–8. https://doi.org/10.1038/s41396-021-01027-4.
23. Xiao L, Zhang F, Zhao F. Large-scale microbiome data integration enables robust biomarker identification. Nat Comput Sci. 2022;2(5):307–16. https://doi.org/10.1038/s43588-022-00247-8.
24. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol. 2015;11(5):e1004226. https://doi.org/10.1371/journal.pcbi.1004226.
25. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687. https://doi.org/10.1371/journal.pcbi.1002687.
26. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017;5(1):27. https://doi.org/10.1186/s40168-017-0237-y.
27. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. Isme j. 2016;10(7):1669–81. https://doi.org/10.1038/ismej.2015.235.
28. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature. 2017;551(7681):507–11. https://doi.org/10.1038/nature24460.
29. Moqri M, Herzog C, Poganik JR, Ying K, Justice JN, Belsky DW, et al. Validation of biomarkers of aging. Nat Med. 2024;30(2):360–72. https://doi.org/10.1038/s41591-023-02784-9.
30. Moqri M, Herzog C, Poganik JR, Justice J, Belsky DW, Higgins-Chen A, et al. Biomarkers of aging for the identification and evaluation of longevity interventions. Cell. 2023;186(18):3758–75. https://doi.org/10.1016/j.cell.2023.08.003.
31. Raman AS, Gehrig JL, Venkatesh S, Chang H-W, Hibberd MC, Subramanian S, et al. A sparse covarying unit that describes healthy and impaired human gut microbiota development. Science. 2019;365(6449); https://doi.org/10.1126/science.aau4735.
32. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–3. https://doi.org/10.1038/nmeth.3869.
33. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, et al. Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol. 2024;42(5):715–8. https://doi.org/10.1038/s41587-023-01845-1.
34. Xun W, Liu Y, Li W, Ren Y, Xiong W, Xu Z, et al. Specialized metabolic functions of keystone taxa sustain soil microbiome stability. Microbiome. 2021;9(1):35. https://doi.org/10.1186/s40168-020-00985-9.
35. Gupta VK, Kim M, Bakshi U, Cunningham KY, Davis JM 3rd, Lazaridis KN, et al. A predictive index for health status using species-level gut microbiome profiling. Nat Commun. 2020;11(1):4635. https://doi.org/10.1038/s41467-020-18476-8.
36. 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(7):1750–9. https://doi.org/10.1038/s41591-023-02424-2.
37. Liu Z, Palaniyappan L, Wu X, Zhang K, Du J, Zhao Q, et al. Resolving heterogeneity in schizophrenia through a novel systems approach to brain structure: individualized structural covariance network analysis. Mol Psychiatry. 2021;26(12):7719–31. https://doi.org/10.1038/s41380-021-01229-4.
38. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17(11):1103–10. https://doi.org/10.1038/s41592-020-00971-x.
39. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165. https://doi.org/10.7717/peerj.1165.
40. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014;11(11):1144–6. https://doi.org/10.1038/nmeth.3103.
41. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets.

Bioinformatics. 2016;32(4):605–7; https://doi.org/10.1093/bioinformatics/btv638.
42. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3(7):836–43. https://doi.org/10.1038/s41564-018-0171-1.
43. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55. https://doi.org/10.1101/gr.186072.114.
44. Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour. 2014;14(6):1097–102. https://doi.org/10.1111/1755-0998.12324.
45. Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2024;42(8):1303–12. https://doi.org/10.1038/s41587-023-01953-y.
46. Lee PA, Tullman-Ercek D, Georgiou G. The bacterial twin-arginine translocation pathway. Annu Rev Microbiol. 2006;60:373–95. https://doi.org/10.1146/annurev.micro.60.080805.142212.
47. De Buck E, Lammertyn E, Anné J. The importance of the twin-arginine translocation pathway for bacterial virulence. Trends Microbiol. 2008;16(9):442–53. https://doi.org/10.1016/j.tim.2008.06.004.
48. Palmer T, Berks BC. The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol. 2012;10(7):483–96. https://doi.org/10.1038/nrmicro2814.
49. Wang J, Li Y, Mu Y, Huang K, Li D, Lan C, et al. Missing microbes in infants and children in the COVID-19 pandemic: a study of 1,126 participants in Beijing. China Sci China Life Sci. 2024. https://doi.org/10.1007/s11427-023-2488-0.
50. Forster SC, Liu J, Kumar N, Gulliver EL, Gould JA, Escobar-Zepeda A, et al. Strain-level characterization of broad host range mobile genetic elements transferring antibiotic resistance from the human microbiome. Nat Commun. 2022;13(1):1445. https://doi.org/10.1038/s41467-022-29096-9.
51. Robinson C, Bolhuis A. Protein targeting by the twin-arginine translocation pathway. Nat Rev Mol Cell Biol. 2001;2(5):350–6. https://doi.org/10.1038/35073038.
52. Wimmers F, Burrell AR, Feng Y, Zheng H, Arunachalam PS, Hu M, et al. Multi-omics analysis of mucosal and systemic immunity to SARS-CoV-2 after birth. Cell. 2023;186(21):4632-51.e23. https://doi.org/10.1016/j.cell.2023.08.044.
53. Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, dePamphilis CW, Tiffin P. Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. Proc Natl Acad Sci U S A. 2021;118(35); https://doi.org/10.1073/pnas.2102914118.
54. Lowry DB, Willis JH. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 2010;8(9); https://doi.org/10.1371/journal.pbio.1000500.
55. Dan L, Li Y, Chen S, Liu J, Wang Y, Li F, et al. A rapidly reversible mutation generates subclonal genetic diversity and unstable drug resistance. Proc Natl Acad Sci U S A. 2021;118(43); https://doi.org/10.1073/pnas.2019060118.
56. Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. Cell. 2007;128(4):655–68. https://doi.org/10.1016/j.cell.2007.01.023.

Publisher's Note