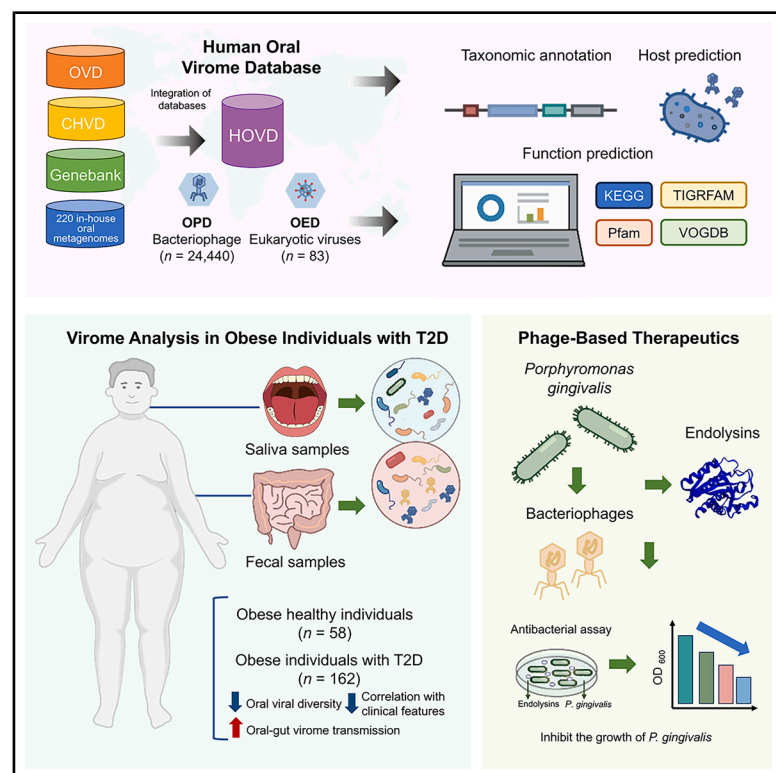


Oral virome metagenomic catalog links *Porphyromonas gingivalis* phages to obesity and type 2 diabetes

Graphical abstract



Authors

Tiansong Xu, Xian Yue Jiao, George Liu, ..., Feng Deng, Feng Chen, Guanxiang Liang

Correspondence

chenfeng2011@hsc.pku.edu.cn (F.C.),
guanxiangliang@tsinghua.edu.cn (G.L.)

In brief

Xu et al. establish the Human Oral Virome Database and reveal virome alterations and oral-gut viral transmission in obesity with T2D. They identify *Porphyromonas gingivalis* phages and validate phage-derived endolysins as potential therapeutics for periodontitis in individuals with T2D.

Highlights

- HOVD catalogs 24,440 phage OTUs and 83 eukaryotic viruses in an online resource
- Oral viral diversity decreases in obese T2D individuals, with weaker clinical associations
- Oral-gut viral transmission was enhanced in obesity/T2D
- Phage-derived endolysins inhibit *P. gingivalis*, offering T2D-periodontitis therapy

Article

Oral virome metagenomic catalog links *Porphyromonas gingivalis* phages to obesity and type 2 diabetes

Tiansong Xu,^{1,2} Xian Yue Jiao,^{3,4} George Liu,⁵ Xi Chen,¹ Qingqing Luo,^{3,4} Gaopu Zhang,^{3,4} Bowen Li,^{3,4} Yu Zhang,⁶ Xue Li,² Yoo Cheung,⁷ Xiaofen Chai,¹ Yueqing Huang,⁸ Huihui Wu,⁹ Feng Deng,⁹ Feng Chen,^{1,10,*} and Guanxiang Liang^{3,4,10,11,*}

¹Central Laboratory, Peking University School and Hospital of Stomatology & National Center for Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices, Beijing 100081, China

²Fifth Clinical Division, Peking University School and Hospital of Stomatology & National Center for Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices, Beijing 100081, China

³Center for Infection Biology, School of Basic Medical Sciences, Tsinghua University, Beijing 100084, China

⁴Tsinghua-Peking Joint Center for Life Sciences, Beijing 100084, China

⁵State Key Laboratory of Membrane Biology, Beijing Frontier Research Center of Biological Structure, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China

⁶Department of Implantology, Peking University School and Hospital of Stomatology & National Center for Stomatology & National Clinical Research Center for Oral Diseases & National Engineering Research Center of Oral Biomaterials and Digital Medical Devices, Beijing 100081, China

⁷Lingchuan Stomatology Inc., Beijing 100176, China

⁸Department of General Medicine, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Nanjing Medical University, Suzhou 215002, China

⁹GENEWIZ Inc., Suzhou 215000, China

¹⁰These authors contributed equally

¹¹Lead contact

*Correspondence: chenfeng2011@hsc.pku.edu.cn (F.C.), guanxiangliang@tsinghua.edu.cn (G.L.)

<https://doi.org/10.1016/j.xcrm.2025.102325>

SUMMARY

The human microbiota has a critical role in maintaining human microbiome homeostasis and health, yet the viral component of the oral microbiome remains largely unidentified. We establish the Human Oral Virome Database (HOVD) catalog, a freely accessible online resource cataloging 24,440 bacteriophage viral operational taxonomic units and 83 eukaryotic viral genomes. Utilizing HOVD, we investigate oral virome variation and its correlation with oral bacteria and gut virome in 220 obese individuals with or without type 2 diabetes mellitus (T2D). Obese individuals with T2D exhibit reduced oral viral diversity, lower correlations with clinical features, disrupted viral-bacterial correlations, and enhanced oral-gut virome transmission. Furthermore, we computationally identify bacteriophages that infect *Porphyromonas gingivalis* and screen six putative endolysins. Experimental validation reveals that a mixture of three endolysins significantly inhibits *Porphyromonas gingivalis* growth. These findings highlight the potential of phage-derived endolysins for periodontitis with T2D, offering a path toward oral and systemic disease intervention.

INTRODUCTION

The important role of the human microbiome is widely recognized.^{1–4} The human microbiome in the oral cavity, the second largest microbial habitat in the body, is remarkably diverse and exhibits a complex ecology. It plays crucial roles in maintaining human health and supporting biological development.^{5–9} Oral virome was a significant component of the human oral microbiome, consisting mostly of bacteriophages (phages) and eukaryotic viruses.¹⁰ Although phages were not considered as important in the oral microbiome as in other microbial systems, particularly in previous studies using static culture techniques of mature biofilms, their role may have been underestimated.¹¹ The diversity and abundance of viruses might exert significant

selective pressure and have an important role in the oral ecosystem, according to recent findings enabled by expanded sequence databases, enhanced sequencing capabilities, and advanced bioinformatics tools.^{5,12–15}

Our knowledge of viral genomes in the human oral microbiome is limited, with only a few studies focusing on the genomic resources of oral viruses. Tisza et al. created the Cenote Human Virome Database (CHVD) including 13,565 oral viruses from over 1,000 oral metagenomic samples.¹⁶ Li et al. integrated 2,792 publicly available human oral metagenomic samples to form Oral Virus Database (OVD) including 48,425 nonredundant viruses.¹⁷ These databases have provided initial insights into various unknown viral genomes, expanding the diversity and heterogeneity of the human oral virome, and now serve as

valuable annotation sources for future studies.⁵ However, several aspects of the human oral virome are still not described clearly, such as the function of the oral virome and the connection between phages and their corresponding hosts, which are essential for a comprehensive understanding of the complex and mutualistic relationships within the human oral microbiome and their effects on human development.^{10,18}

In addition, oral viruses have been identified as being associated with human diseases, such as oral cancers; periodontal diseases; hand, foot, and mouth disease; and rheumatoid arthritis, suggesting that oral viruses might influence the progression of these diseases and indicating potential approaches for prognosis and diagnosis based on microbiome composition.^{19–24} Among chronic metabolic diseases, type 2 diabetes mellitus (T2D) is a major contributor to global morbidity and mortality, leading to enormous economic and societal consequences.^{25,26} The population of adults with diabetes, nearly 90% of whom were T2D, was estimated to grow from 108 million in 1980 to 537 million in 2020–2021.^{27–29} However, little is known about the alteration of human virus related to T2D in the oral cavity and whether it has a latent role in treating and predicting disorder development, which might have a microbial connection within the context of disease pathogenesis.^{30–32}

To expand the existing research and provide a complementary perspective of the human oral virome, we performed a large-scale integration of oral viral genomes from three human oral virome databases previously published and viral genomes identified from 220 oral metagenomes obtained from obese patients with or without T2D. We used these data to establish the Human Oral Virome Database (HOVD) catalog, which contained 24,440 phage viral operational taxonomic units (vOTUs) made up of the Human Oral Phage Database (OPD) and 83 eukaryotic viral contigs composed of the Human Oral Eukaryotic Virus Database (OED), estimated to be over 50% complete. According to the developed database, we characterized the changes in the oral virome and their correlation with oral bacteria and the gut virome in obese subjects with T2D and explored potential therapeutic strategies for periodontitis individuals with T2D. This study would advance our knowledge of viruses identified in the oral microbiome and host-virus interactions and provide clues for the diagnosis and treatment of these human diseases.

RESULTS

A genomic catalog of the human oral virome

We applied integrative virome pipeline (iVirP) pipeline to 220 oral metagenomes obtained from obese patients with ($n = 158$) or without type 2 diabetes ($n = 62$), identifying 2,155 high-quality viral genomes.³³ A total of 706 contigs (≥ 1 kbp in length) that could not be annotated as viral by iVirP but could be aligned to sequences in the Nucleotide Sequence Database (NT database) were retained for further analysis. We then clustered oral viral genomes from our samples and genomes publicly available from human oral viral genome databases, including OVD,¹⁷ CHVD,¹⁶ and oral virome selected from GenBank (Figure 1A; Table S1). After assessing genome completeness with CheckV and distinguishing viral genomes annotated as phage or eukaryotic virus using previous databases, 29,033 oral phage contigs with

>50% completeness and 83 eukaryotic viral contigs were preserved (Figures 1B and S1; Table S2).³⁴

All the phage contigs were clustered to generate a Human OPD, comprising 24,440 vOTUs at 95% average nucleotide identity over 85% alignment fraction of the shorter sequence.³⁵ OPD and OED were integrated to establish HOVD. HOVD is now publicly available at <https://hovd.org/>, where the online website offers more detailed information about our virome dataset. Our phage database sets spanned over 3,000 samples from nine countries across four continents, with genomes from North America (52.6%) and Asia (36.7%) accounting for the largest proportion (Figures 1C and 1D; Tables S3 and S4). Viral genomes were largely derived from various sources, including supragingival plaque (32.7%), saliva (28.7%), tongue dorsum (25.1%), buccal mucosa (4.6%), and subgingival plaque (3.4%) (Figure 1D). The length of vOTUs ranged from 5,013 to 680,541 bp, with an average length of 43,553 bp and an N50 length of 35,067 bp. We then evaluated the quality of oral phage vOTUs in OPD, including 15.7% complete, 36.5% high-quality, and 47.8% medium-quality genomes, with over 95% of vOTUs having low contamination (Figure S1). Detailed information on the 83 eukaryotic viral contigs in the Human OED is provided in Figure S2.

Higher quality viral genomes in OPD

We then compared our database with two recently published human-associated metagenome-assembled oral virus catalogs, OVD and CHVD.^{16,17} We first assessed the completeness and quality of viral genomes of three databases. Compared to the other two databases, such as OVD (where over 56% of contigs were estimated to be low-quality genome fragments), OPD was populated with a higher proportion and number of complete, high-, and medium-quality virus genomes (Figure S3). Our database contributed 576 complete viral genomes (13.8%), 557 high-quality viral genomes (5.2%), and 1,023 medium-quality viral genomes (7.2%) to all the genomes of OPD, demonstrating the high quality and efficiency of the iVirP pipeline (Table S3).

To recognize higher ranking viral clades, we clustered 24,440 species-level vOTUs into genus- and family-level groups approximately and revealed 1,240 genus-level vOTUs and 342 family-level vOTUs (Figure 2A). Notably, 6.4% (1,574 of 24,440) species-level vOTUs, 16.1% (200 of 1,240) genus-level vOTUs, and 17.3% (59 of 342) family-level vOTUs in the OPD catalog were not identified in previously published databases, highlighting the contributions of our study to the understanding of the human oral virome.

The taxonomic annotation and host prediction of oral phages

To explore the taxonomic annotation of 24,440 vOTUs in OPD, we annotated the database based on the viral proteins from the MMSeqs2 virus protein database (Table S3). A total of 24,291 vOTUs were annotated covering 9 viral classes, with Caudoviricetes accounting for the largest proportion ($n = 23,523$, 96.2%). The remaining classes were sparsely represented, such as Faserviricetes ($n = 553$), Malgrandaviricetes ($n = 194$), and Howeltoviricetes ($n = 6$) (Figure 2B). Additionally, six vOTUs belonged to the order Cressdnaviricota,

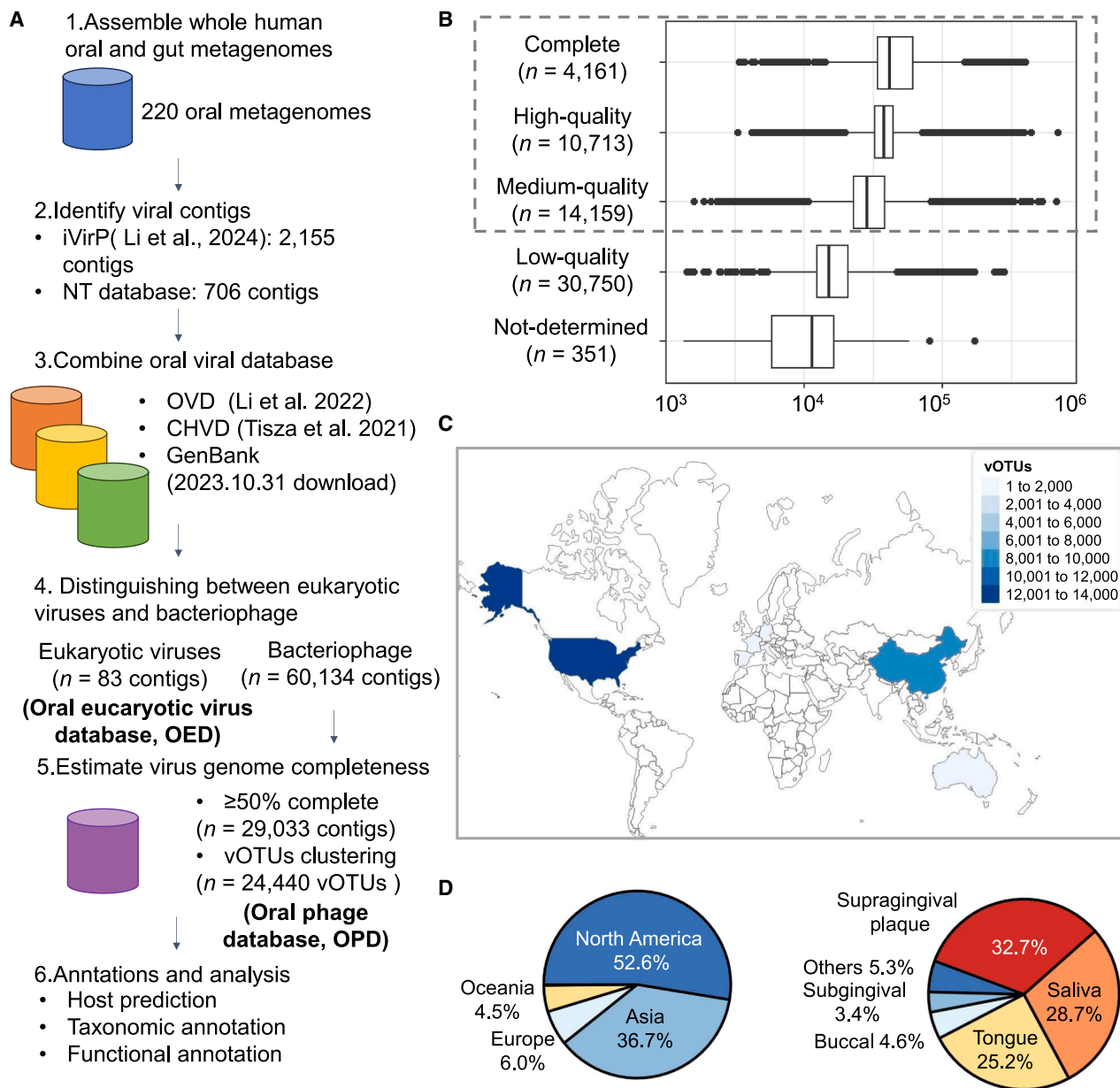


Figure 1. Overview of viral genomes recovered from human oral metagenomes

(A) Overview of oral viral discovery effort and formation of the HOVD catalog, including the OED and OPD catalog.

(B) Distribution of genome completeness and classification of OPD (complete, $n = 4,161$; high-quality [$>90\%$ complete], $n = 10,713$; medium-quality [$50\%–90\%$ complete], $n = 14,159$). Low-quality genomes were excluded from OPD (low-quality [$<50\%$ complete], $n = 30,750$; completeness not determined, $n = 351$). In the boxplots, the middle line indicates the median, the box represents the interquartile range (IQR), and the length of the whiskers indicates $1.5 \times$ the IQR.

(C) Map of the world showing the number of vOTU samples of OPD from four continents.

(D) Metadata for 24,440 vOTUs with over 50% completeness, including distribution of continents and sampling sites.

which were not classified at the class level. Also, most vOTUs (23,618 vOTUs, 96.6%) were not identified at the family level, and 24,429 vOTUs (nearly 100%) were not recognized at the genus level, suggesting that there was still a significant knowledge gap in the current taxonomy of human oral phage that needed to be explored. Besides, based on the taxonomic annotation of the previous virus catalog, we identified 6 viral

families in OED, including *Coronaviridae*, *Orthoherpesviridae*, and *Herpesviridae* (Table S4).

To understand phage predation and design innovative phage therapies, we applied host-virus interactions to predict the potential cellular hosts of viruses with VirHostMatcher.³⁶ We leveraged a large-scale oral prokaryotic genome collection represented by 3,569 species-level genome bins as the potential

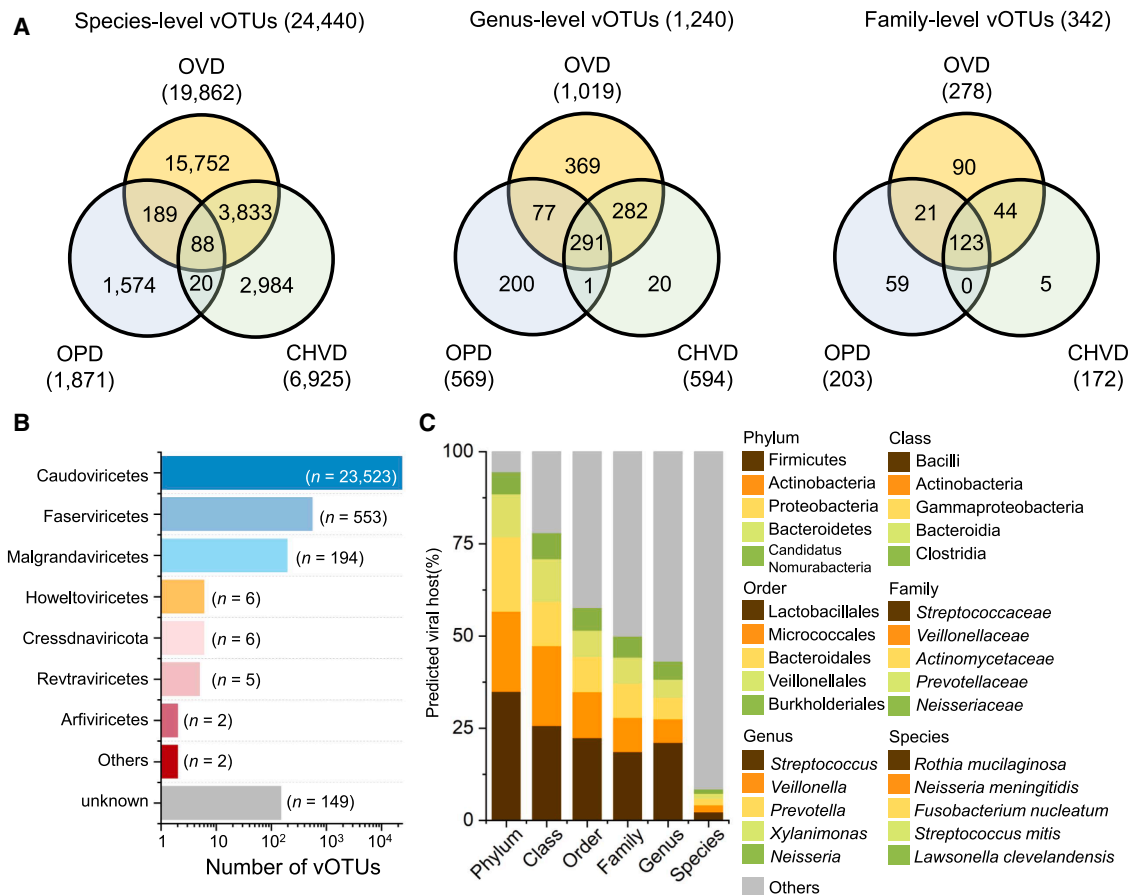


Figure 2. Taxonomic annotation, host prediction, and cluster analysis of 24,440 vOTUs in OPD

(A) 29,033 oral phage genomes in OPD were clustered into 24,440 species-level, 1,240 genus-level, and 342 family-level vOTUs approximately. (B) Taxonomic annotation of 24,440 species-level vOTUs in OPD. The number in parentheses shows the vOTU number. (C) The proportion of the top five predicted hosts at different taxonomic ranks, including phylum, class, order, family, genus, and species.

hosts of oral phages.³⁷ We identified bacterial hosts for 28.4% of 24,440 vOTUs ($n = 6,931$) in species level, with the most common phylum host being Firmicutes ($n = 8,541$), followed by Actinobacteriota ($n = 5,306$), Proteobacteria ($n = 4,914$), and Bacteroidota ($n = 2,868$) (Figure 2C). The most common bacterial family host was involved in *Streptococcaceae* ($n = 4,544$), *Veillonellaceae* ($n = 2,269$), *Actinomycetaceae* ($n = 2,242$), *Prevotellaceae* ($n = 1,748$), and *Neisseriaceae* ($n = 1,402$). At the genus level, *Streptococcus* hosted 21.10% ($n = 5,149$) of the vOTUs, significantly more than any other bacterial genus (Table S5).

Functions of the human oral virome

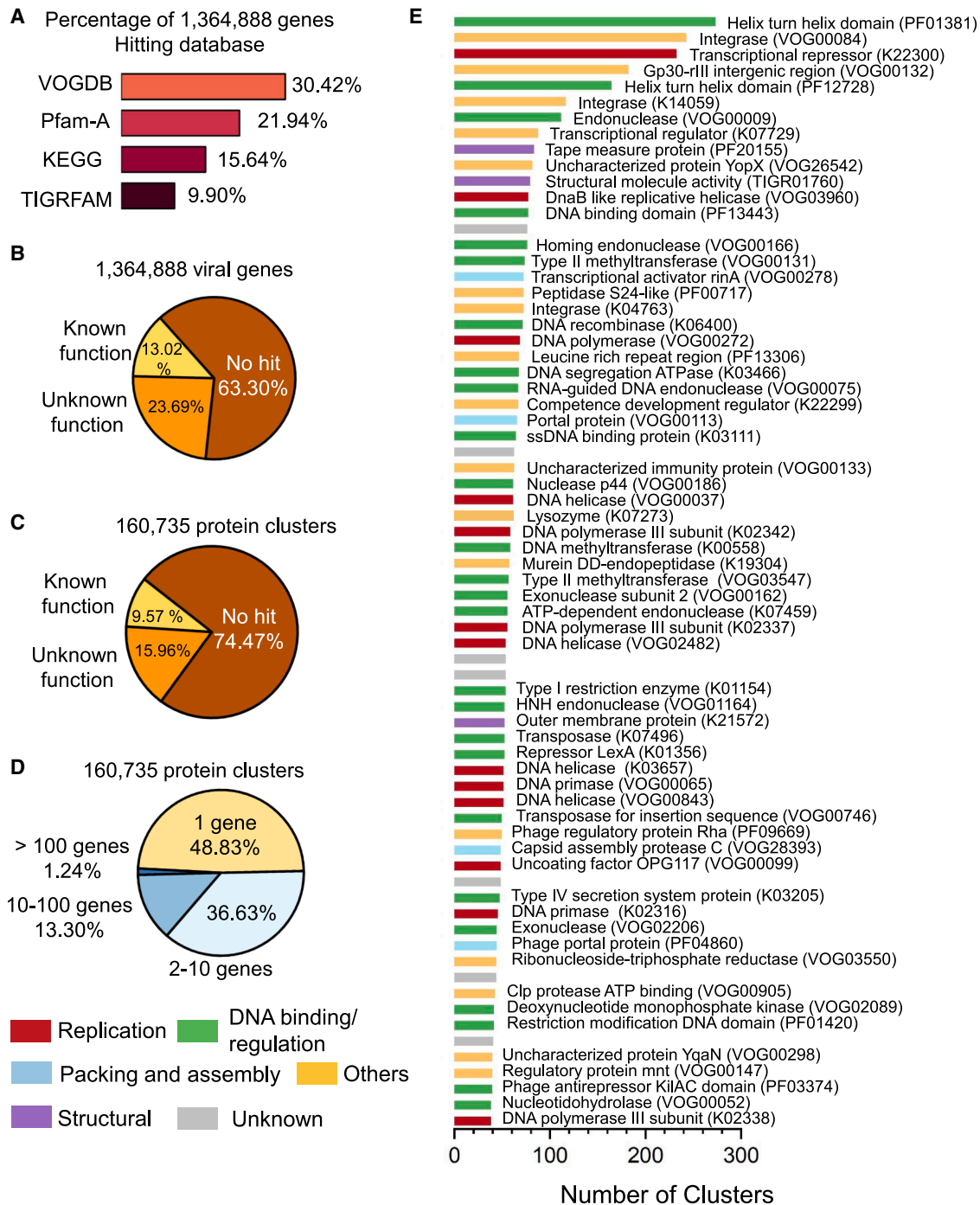
Although the function of human oral bacteria was extensively studied and predicted,^{5,38,39} our understanding of the human oral virome remained limited. To explore the potential function of the human oral virome, we identified 1,364,888 protein-coding genes from all the contigs of HOVD and aligned them to protein-coding genes in Hidden Markov Model (HMM) databases, including Kyoto encyclopedia of genes and genomes (KEGG),⁴⁰ TIGRFAM,⁴¹ Pfam,⁴² and VOGDB (<http://vogdb.org>) (Figure 3A). On the whole, only 13.02% of viral genes were compared with

known functional genes, 23.69% were not assigned any biological function, and 63.30% had no significant matches in databases (Figure 3B), indicating a vast unknown domain regarding the functional potential of human oral viruses.

Since lots of unknown functions were identified according to the number of genes (Figure S4), it was challenging to discern the most prevalent functions in the oral virome. We clustered the 1.3 million viral genes at 30% average amino acid identity (AAI) into 160,735 protein clusters using MMseqs2 (Figure 3C), where over 50% of clusters contained more than 2 genes (Figure 3D). The largest protein clusters were annotated with viral replication, DNA binding or regulation, packaging and assembly, and structural proteins. Additionally, some larger clusters lacked predictive annotation, underscoring significant research potential (Figure 3E; Table S6).

Factors influencing the oral virome in obese participants with T2D

Although various studies have explored the associations between the virome and human diseases,^{16,43,44} there was still a lack of evidence characterizing the alteration of the oral virome



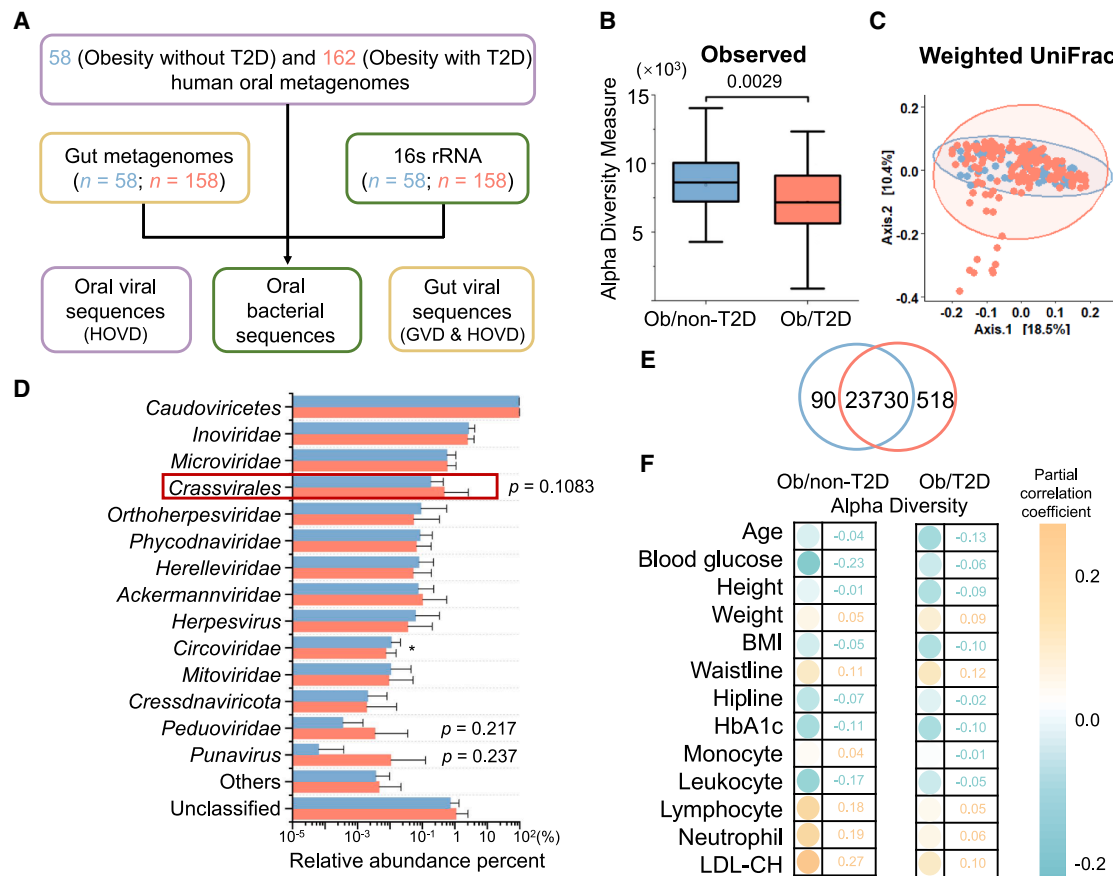


Figure 4. Comparison of the human oral virome in obese patients with or without T2D

(A) The pipeline comparing the human oral virome in obese patients with (red) or without (blue) T2D with oral and gut metagenomes. (B) Alpha diversity (observed vOTUs) of the oral virome in the Ob/non-T2D and Ob/T2D groups was compared using Wilcoxon rank-sum test ($p = 0.0029$). The boxplots display the interquartile range (IQR), with the middle line representing the median and the whiskers extending to 1.5 times the IQR. (C) Principal coordinate analysis (PCoA) ordination of the oral virome in obese participants without and with T2D measured by weighted UniFrac distance using permutational multivariate analysis of variance (PERMANOVA) test ($p = 0.285$). (D) The bar graph showing relative abundance percent of the oral virome in the Ob/non-T2D and Ob/T2D groups. Data are presented as mean $\pm$ SD; t test was used for statistical analysis. (E) Venn diagram indicating the shared vOTUs between obese participants without and with T2D. (F) Relationship between the observed vOTU diversities and clinical parameters after adjusting for confounding variables using partial correlation analysis in the Ob/T2D and the Ob/non-T2D group. Partial correlation coefficients are shown in the corresponding square (all p values > 0.05). Clinical parameters contained age, blood glucose, height, weight, body mass index (BMI), waistline, hipline, glycated hemoglobin (HbA1c), monocyte, leukocyte, lymphocyte, neutrophil, and low-density lipoprotein-cholesterol (LDL-Ch).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

in obese participants with or without T2D. We collected saliva and fecal samples from 220 participants, including 58 obese healthy individuals (Ob/non-T2D; females [$n = 31$] and males [$n = 27$]), and 162 obese participants with T2D (Ob/T2D; females [$n = 99$] and males [$n = 63$]). Detailed demographics and clinical parameters of these participants were presented in Table S7. The composition of metagenomic sequencing data used in our study is illustrated in Figure 4A. The gut sequencing presented in our study was published previously.⁴⁵ Based on HOVD, we obtained scaled vOTUs matrices from human oral metagenome sequences. Oral bacteriome feature tables were calculated from 16S rRNA gene sequences from these participants.

To analyze the variation of the oral virome between obese participants with and without T2D samples, alpha diversity and beta diversity were measured using observed vOTU diversities and Bray-Curtis dissimilarities, respectively (Figures 4B, 4C, and S5A). Observed vOTUs in the Ob/T2D group were lower than those in the Ob/non-T2D group ($p < 0.05$). Alpha diversity of the oral bacteriome in the Ob/T2D group was similar to that in the Ob/non-T2D group ($p = 0.64$) (Figure S5B). It might indicate that oral phage communities might be different from each healthy individual while representing a trend toward homogenization in a diseased state.⁴⁶ The discordance of phage and bacterial community diversities might be explained by the existence of cross-infective phages.⁴⁶ In the principal coordinate

analysis plot based on weighted UniFrac distance, the virome profiles between the Ob/T2D and Ob/non-T2D groups exhibited substantial overlap ($p = 0.246$), indicating no significant oral viral compositional separation (Figure 4C). Several dispersed data points were observed, but the underlying reasons for these outliers remain unclear despite our attempts to associate them with clinical parameters.

After annotating by oral virome taxonomy, we compared the relative abundance percent between the two groups (Figure 4D). The proportion of Crassvirales increased in the Ob/T2D group, which was also a predominant phage in the human gut virome.⁴⁷ Although taxa such as Crassvirales exhibited visually distinct abundance patterns between groups, these differences did not reach statistical significance, likely due to high inter-individual variability that is commonly observed in human virome studies. Also, most types of vOTUs were shared in obese participants with and without T2D, while more types of vOTUs were detected in ones with T2D ($n = 518$) compared with ones without T2D ($n = 90$) (Figure 4E).

To determine whether viral diversity was associated with clinical parameters of obesity and T2D, we analyzed partial correlation coefficients (Figure 4F) and Spearman's correlation coefficients (Figure S5C) between observed vOTU diversities and indices in the Ob/non-T2D and Ob/T2D group. Most partial correlation coefficients in the Ob/non-T2D group were higher than that in the Ob/T2D group, suggesting that correlated tendency had remained decreased in a diseased state (Figure 4F). However, several pairs of oral virome diversity, relative abundance percentages of viral families, and clinical parameters showed statistically significant Spearman's correlation coefficients, which might be due to the multicollinearity of clinical parameters (Figures S5C and S5D).

Interactions between the oral virome and bacteriome associated with T2D

Previous studies have highlighted the influence of the oral virome on the oral bacteriome.^{18–20,46} To further explore this, we examined the correlation between the oral virome and bacteriome associated with T2D. We first examined the relationship between alpha diversities of the oral virome and bacteriome, showing a higher significant correlation in the Ob/non-T2D group ($R = 0.61$, $p = 6e-07$) compared to the Ob/T2D group ($R = 0.41$, $p = 1.4e-7$), which aligned with the expectation of lower correlation in disease states (Figure 5A). Then, we analyzed whether the oral virome and bacteriome might contribute to the disordered state and have the capacity to provide a powerful diagnostic performance for T2D using the R package XGboost. As the number of predictors in the model increased, the area under the curve value became higher generally, suggesting that both the oral virome and bacteriome might contribute to the development of T2D (Figures 5B and S6).

We then analyzed the interactive networks between the oral virome and bacteriome grouped by the following clinical parameters, including T2D history (Ob/non-T2D and Ob/T2D), gender (female and male), and exercise level (level 1, 2, and 3, whose intensity of exercise gradually increased). Consistent with the correlation coefficient estimated earlier (Figure 5C), the microbial networks in the non-T2D group exhibited greater connectivity

than those in the T2D group, as reflected by a higher average degree (3.299 vs. 2.75) and graph density (0.024 vs. 0.019), suggesting a stronger coordination between the oral virome and bacteriome in the non-T2D group compared to the T2D group. No obvious difference was shown between female and male groups. Regarding exercise, although the network in level 2 exhibited greater structural complexity than in level 3 based on visual inspection, both showed higher overall complexity than level 1 (average degree: 2.588 in level 1 vs. 3.894 in level 2 and 3.021 in level 3; graph density: 0.019 vs. 0.030 and 0.021, respectively), suggesting that increased physical activity may enhance microbial co-occurrence patterns (Table S8).

Transmission of viruses between the oral cavity and gut is linked with obese participants with T2D

Based on the potential causal link between oral-gut microbial transmission and human diseases,^{43,48,49} it was necessary to explore the latent relationship between oral-gut microbial communities and their dysregulation in disease progression. The gut virome was annotated by the Gut Virome Database (GVD) and HOVD,⁵⁰ respectively, to determine whether oral viral genomes could be detected in the gut microbiota. More types of vOTUs were shared in the Ob/T2D group ($n = 20,710$) compared to the Ob/non-T2D group ($n = 13,812$) (Figure 6A). This suggested that the oral and intestinal barriers were disrupted, allowing more oral microorganisms to enter the intestine in diseased states, consistent with previous studies.^{31,51,52} Correspondingly, more gut viral genomes at species level compared with HOVD were annotated in the Ob/T2D group ($n = 7,470$) than in the Ob/non-T2D group ($n = 2,986$) (Figure S7A). Also, we estimated the correlation between alpha diversities of the oral and gut virome, revealing that the correlation coefficients became lower in the Ob/T2D group ($R = 0.49$, $p = 4.8e-011$), similar to the results observed between the oral virome and bacteriome (Figure 6B).

We also analyzed correlations among the oral, gut virome, and T2D-related clinical indices. No obvious differences were observed in the gut viral families and T2D-related clinical indices across two groups and the entire sample (Figures 6C and S7B). More significant differences were displayed in the correlations between the oral and gut virome (Figure 6D) and in the correlations between the gut virome annotated by GVD and HOVD in the Ob/T2D group (Figure S7C), consistent with the higher number of shared gut viral genomes in the Ob/T2D group.

Human oral phages hosted by *Porphyromonas gingivalis*

Periodontitis was considered the “sixth complication” and the most critical oral complication of diabetes mellitus.⁵³ In patients with diabetes mellitus, periodontitis could negatively impact glycemic control and contribute to the development of complications. Periodontal therapy appeared to lead to a modest improvement in glycemic control.⁵⁴ It was possible to search for phages regulating bacterial dysbiosis to prevent the prevalence of oral diseases.¹⁸ Therefore, we focused on the human oral vOTUs hosted by *Porphyromonas gingivalis*, which was considered to have an important role in disrupting the oral immune homeostasis and promoting periodontal disease.⁵⁵ Similarly, *Porphyromonas* showed a slightly higher mean relative abundance in

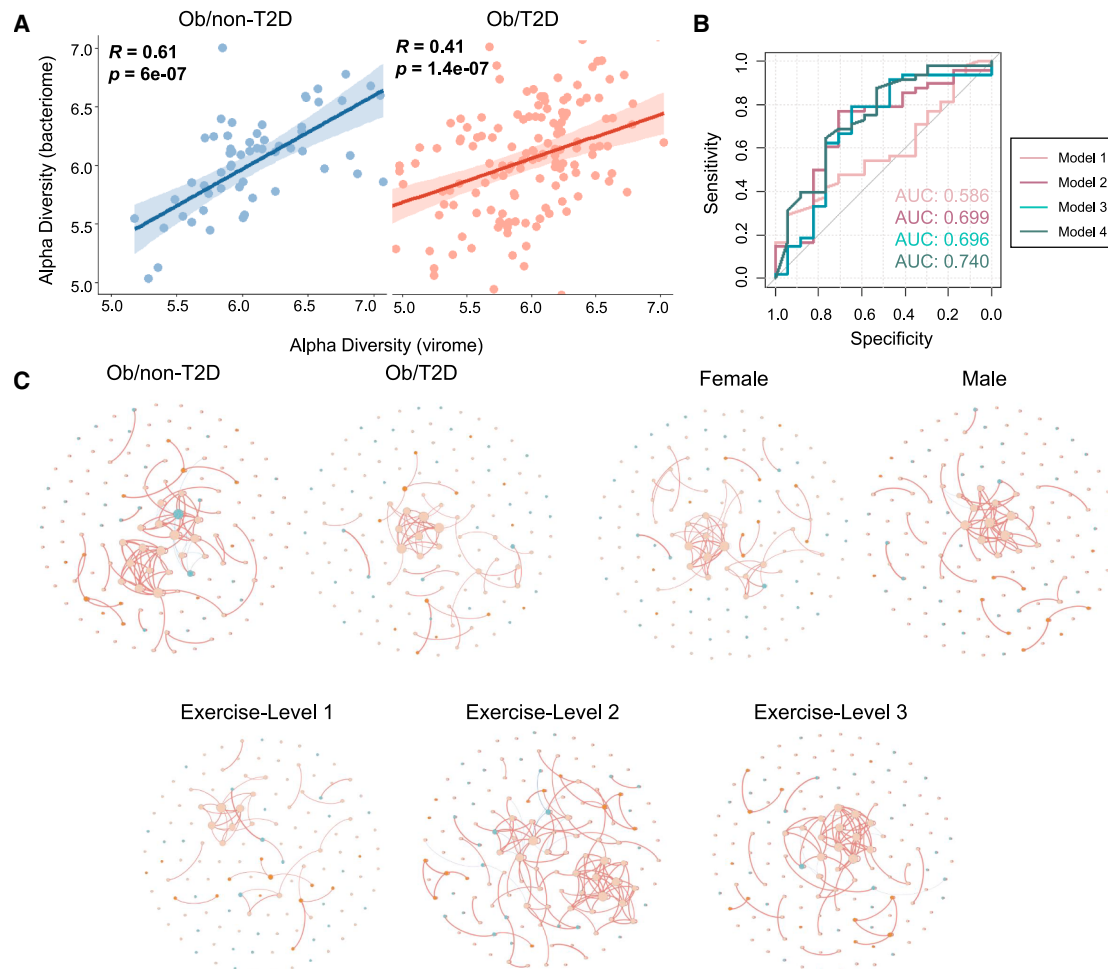


Figure 5. Relevance between the oral virome and bacteriome related to T2D

(A) Correlation between alpha diversity of the oral virome and bacteriome in obese individuals with and without T2D. The p values were measured with linear regression, and the shaded area shows the estimated 95% confidence interval.

(B) Prediction of the contribution of the oral virome and bacteriome to the disorder state. Four models were established as described in STAR Methods. Area under the curve (AUC) values, defined as the area under the receiver operating characteristic (ROC) curve, are shown in the lower left corner. AUC values of each model are colored by model number.

(C) Network analysis of the relevance of the oral virome and bacteriome, group by the following clinical parameters, including T2D history (Ob/non-T2D and Ob/T2D), gender (female and male), and exercise level (exercise level 1, 2, and 3, whose intensity of exercise gradually increased).

the Ob/T2D group (6.0%) compared to the Ob/non-T2D group (5.3%), while the difference did not reach statistical significance ($p > 0.05$; Figure 7A). We first identified phages that were potentially hosted by *P. gingivalis*. A total of 150 *Porphyromonas* genomes were retrieved from the Human Oral Genome Database and RefSeq database. All phage vOTUs were queried against these bacterial genomes using VirHostMatcher to search for their hosts (Figure S8A; Tables S9 and S10). We selected $d2^*$ value to reflect oligonucleotide frequency dissimilarity between viral and host genomes. A smaller $d2^*$ value suggests greater genomic similarity and a higher likelihood of host-virus association. As a result, we identified 25 vOTUs with $d2^*$ dissimilarity <0.25 and 5 vOTUs with $d2^*$ dissimilarity <0.2 . The sum relative abundance of 25 vOTUs and 5 vOTUs in the Ob/T2D group was higher than that in the Ob/non-T2D group (Figure S8B). Correlations between

these vOTUs and blood sugar and HbA1c decreased in the Ob/T2D group compared to the Ob/non-T2D group, matching the associations observed with all the vOTUs and clinical parameters (Figure S8C).

Interestingly, although not all of the 25 vOTUs with $d2^*$ dissimilarity <0.25 showed higher mean relative abundance in the Ob/T2D group compared to the Ob/non-T2D group, all 5 vOTUs with $d2^*$ dissimilarity <0.2 exhibited increased mean relative abundance in the Ob/T2D group (Figures 7B and S8D). In addition, the correlation between the relative abundance of these 5 vOTUs and the genus *Porphyromonas* was stronger in the Ob/T2D group (Figure 7C). Subsequent characterization of the five vOTUs revealed that all corresponding contigs exceeded 50 kb in length and were taxonomically assigned to the class Caudoviricetes, indicating high sequence

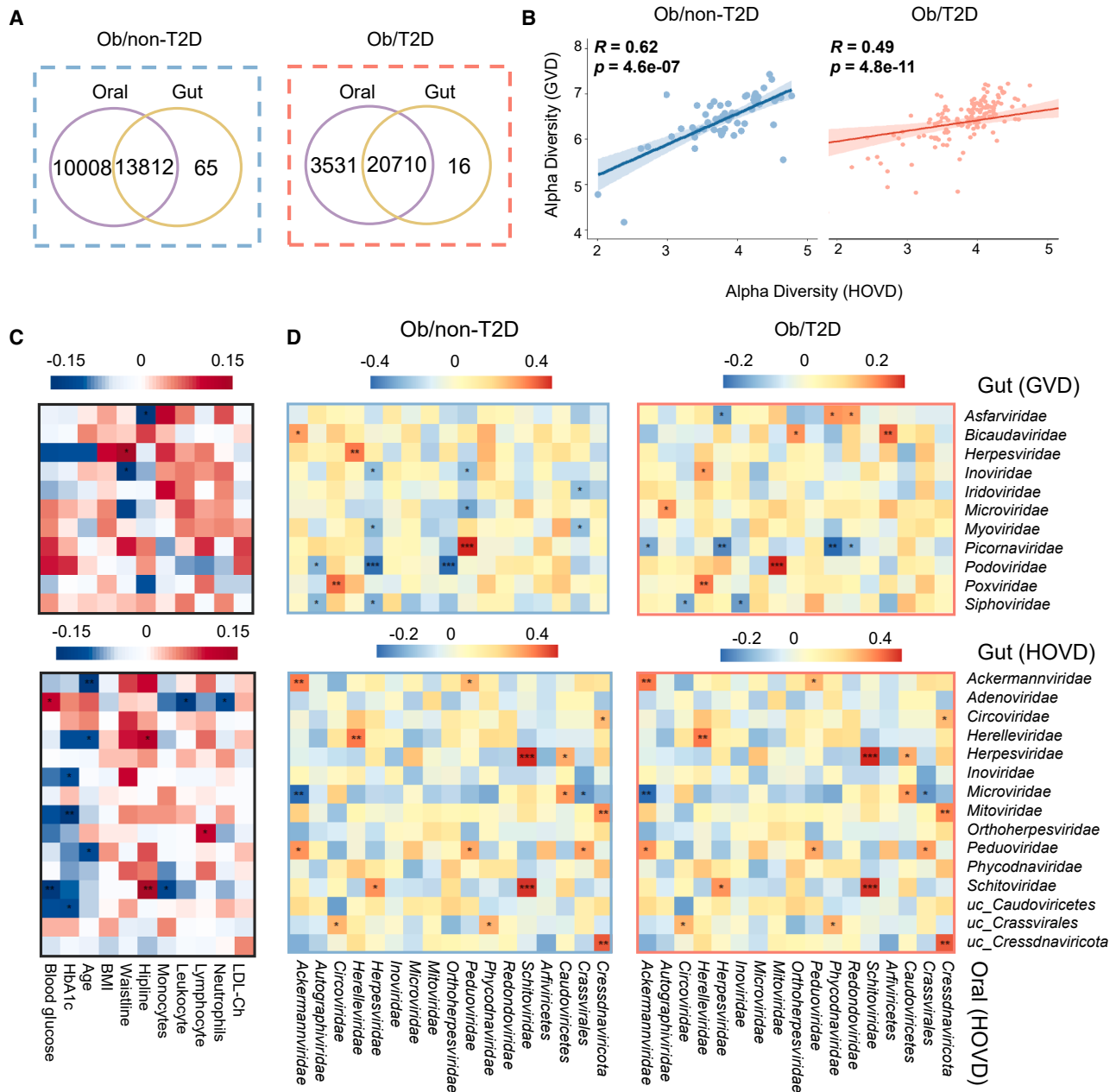


Figure 6. Relevance between the oral and gut virome in obese patients with or without T2D

(A) The shared vOTUs between the gut virome annotated by GVD and HOVD in the Ob/non-T2D group (left, in the blue dashed box) and the Ob/T2D group (right, in the red dashed box).

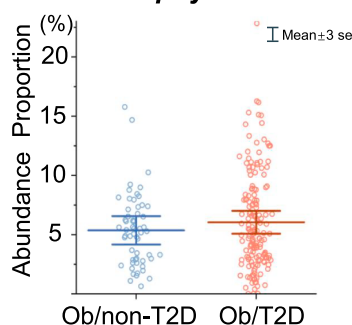
(B) Correlation between alpha diversity between the oral virome and gut virome in obese individuals with and without T2D. The p values were analyzed in linear regression, and the shaded area depicts an estimated 95% confidence interval.

(C) Heatmaps of Spearman correlation coefficients and p value between clinical parameters and relative abundances of gut viral families annotated by GVD (upper) and HOVD (lower). Clinical parameters contained age, blood glucose, height, weight, body mass index (BMI), waistline, hipline, glycated hemoglobin (HbA1c), monocyte, leukocyte, lymphocyte, neutrophil, and low-density lipoprotein-cholesterol (LDL-Ch).

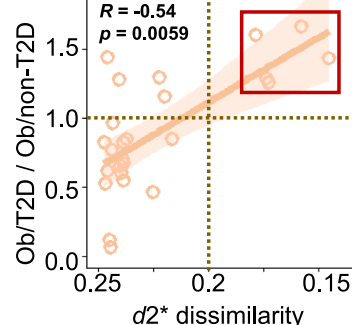
(D) Heatmaps showing correlation coefficients and p value between oral and gut viral families annotated by GVD (upper) and HOVD (lower) in the Ob/non-T2D group (left, in the blue dashed box) and the Ob/T2D group (right, in the red dashed box).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

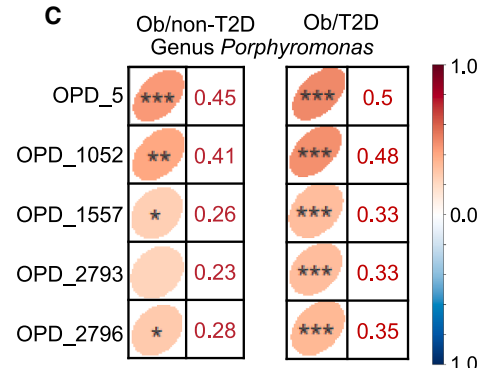
A Genus *Porphyromonas*



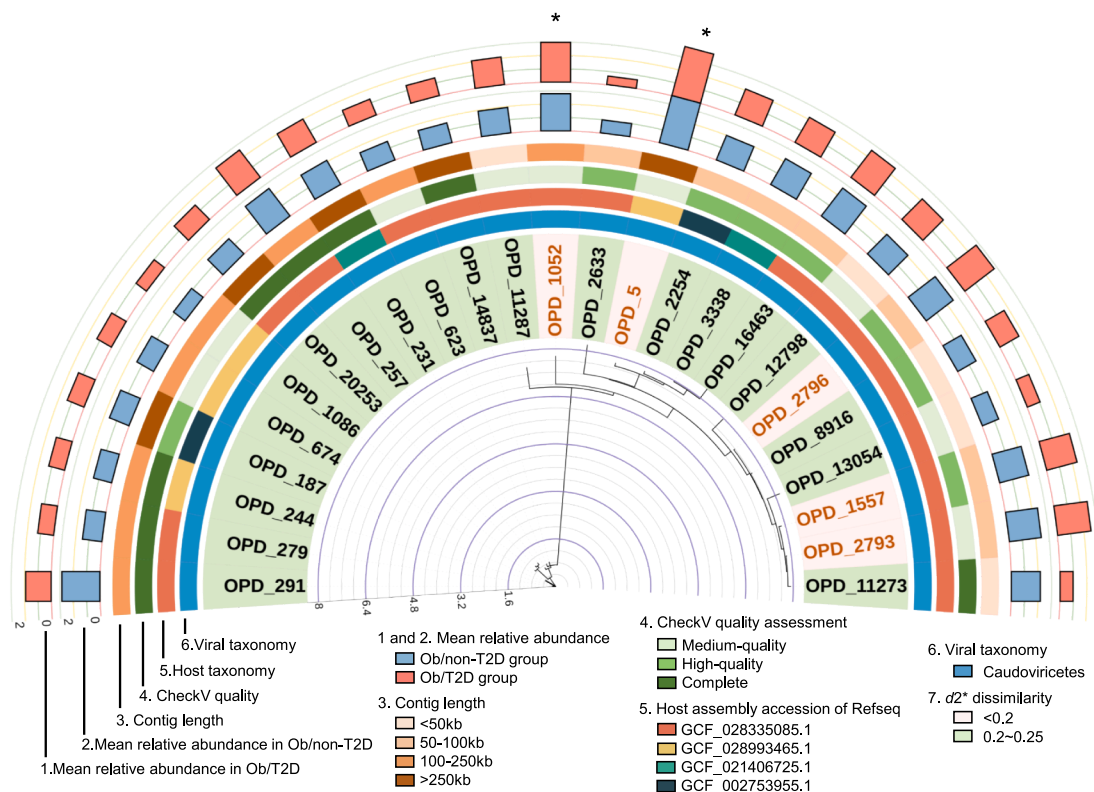
B



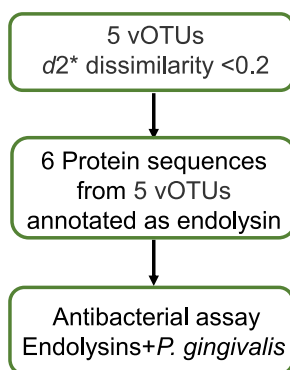
C



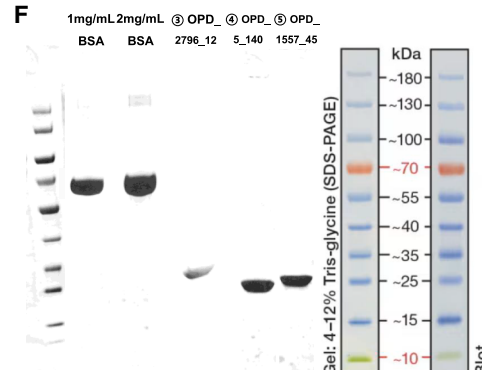
D



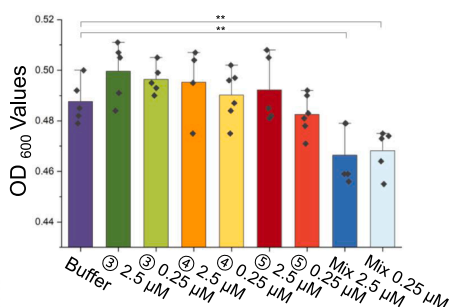
E



F



G



(legend on next page)

integrity and supporting their classification as high-quality viral genomes. Notably, only two vOTUs with significantly increased abundance in the Ob/T2D group had a $d2^*$ dissimilarity <0.2 (Figure 7D). These 5 vOTUs may represent specific viral targets for designing strategies to treat and manage periodontitis in patients with diabetes mellitus.

Endolysins inhibiting the growth of *P. gingivalis*

Next, we aim to explore potential clues for inhibiting *P. gingivalis* from these five vOTUs. We focused on the identification of endolysin proteins as potential therapeutic candidates against *P. gingivalis*, given their well-documented antibacterial activity.⁵⁶ Endolysins are phage-derived enzymes that degrade bacterial cell walls, triggering osmotic lysis and bacterial death. Due to their potency, we prioritized endolysins as functional elements for phage-based interventions. To this end, we searched *P. gingivalis* phage genomes for endolysin-annotated sequences by aligning predicted open reading frames (ORFs) against the KEGG, TIGRFAM, Pfam, and VOGDB database. Due to their potent antibacterial activity, we prioritized endolysins as candidate functional elements for phage-based interventions. As a result, six distinct endolysin sequences were identified for further characterization as potential candidates for therapeutic applications. We performed a search using the NCBI BLASTP tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)⁵⁷ and compared the five sequences most similar to these six (Figure S9A). The sequences of these six proteins are similar to proteins with functions such as N-acetylmuramoyl-L-alanine amidase and phage-associated cell wall hydrolase autolysin, found in *Porphyromonas* and Caudoviricetes (Figure S9B). Additionally, we used ColabFold-AlphaFold2 to predict the protein structures (Figure S9C).⁵⁸ The six sequences were cloned into plasmids, expressed in *Escherichia coli*, and the proteins were purified, yielding Protein OPD_2796_12 (Pro.3), OPD_5_140 (Pro.4), and OPD_1557_45 (Pro.5) (Figures 7E and 7F). The proteins, individually and in mixtures, were incubated with *P. gingivalis* in anaerobic bags at concentrations of 0.25 and 2.5 μ M for 120 min. Among all the groups, mixture group at 0.25 μ M most effectively inhibited the growth of *P. gingivalis*, with the effect being closest to the positive control group (lysozyme group), suggesting that it could be a promising therapy for treating periodontitis, potentially controlling blood glucose levels in T2D patients with periodontitis (Figure 7G).

DISCUSSION

We integrated oral viral genomes from OVD, CHVD, Genbank, and our own metagenomic data to generate HOVD. To our knowledge, this is the most unified oral virome database so far. The oral metagenomes were from nine countries across four continents, and we identified 24,440 oral phage operational taxonomic units and 83 oral eukaryotic virus genomes, all with completeness greater than 50%. Through the established human oral virome catalog, we have focused on items involved in the composition and structure of the human oral virome related to clinical factors associated with obesity and T2D and interactions between the human oral virome, the bacteriome, and the gut virome. We identified *P. gingivalis*-targeting bacteriophages and validated their therapeutic potential, demonstrating that three of six computationally screened endolysins effectively inhibit *P. gingivalis* growth as the latent therapeutic strategies for patients with diabetes periodontitis, based on oral and gut metagenomes obtained from obese patients with or without T2D.

Compared with the oral virome database published previously,^{16,17} HOVD addressed several advancements and advantages. It included a much higher proportion of high-quality oral viral genomes and integrated more oral metagenomic samples. HOVD comprised two modules: OPD for phage genomes and OED for eukaryotic virus genomes. All of the aforementioned suggest an exquisite and high-quality oral virome database. Additionally, 99.4% of oral phage genomes were annotated in OPD, which is higher than 45.8% in OVD and 93.7% in CHVD. Oral precise sample locations were important and closely relevant to oral epidemiological, clinical, and etiological parameters.⁵⁹ We provided more detailed labeling of oral metagenome sites, such as distinguishing plaque samples as subgingival plaque and supragingival plaque, closely connected with diseases. For example, in patients with periodontitis, the total density of species associated with the condition, such as *P. gingivalis* and *Bacteroides forsythus*, was found to be higher subgingivally compared to supragingivally.⁶⁰ Supragingival metagenomic linkage groups were reported to be more interconnected in the healthy control group, while the salivary groups were highly interconnected in patients with rheumatoid arthritis.⁹ Moreover, HOVD is a freely accessible online resource equipped with interactive analytical tools, enabling researchers to perform precise

Figure 7. Phylogenomics of oral vOTUs hosted by *P. gingivalis*

- (A) Comparisons of the abundance proportion of genus *Porphyromonas* between the Ob/T2D group and the Ob/non-T2D group. Data are presented as mean $\pm$ 3 $\times$ SEM; t test was applied for statistical analysis.
- (B) Correlations between $d2^*$ dissimilarity and the ratio of mean relative abundance in 25 vOTUs in the Ob/T2D group to that in the Ob/non-T2D group. Values with the $d2^*$ dissimilarity of 0.2 and the ratio of 1 are labeled with the dashed line. The red box in the upper right indicates each vOTU with $d2^*$ dissimilarity <0.2 increased in the Ob/T2D group.
- (C) Correlations between the proportion of genus *Porphyromonas* and mean relative abundance of each vOTU with $d2^*$ dissimilarity <0.2 in the Ob/T2D group and the Ob/non-T2D group. The correlation coefficient and p value are shown in the corresponding square.
- (D) Phylogeny of vOTUs hosted by *P. gingivalis*. The cladogram was depicted with Interactive Tree Of Life (iTOL). The background color of each vOTU indicates that the $d2^*$ dissimilarity is less than 0.25 (green) or less than 0.2 (pink). Outer rings display metadata for each vOTU; t test was applied for statistical analysis.
- (E) The pipeline selected 3 endolysin protein sequences.
- (F) Western blot showing purified OPD_2796_12 (Pro.3), OPD_5_140 (Pro.4), and OPD_1557_45 (Pro.5).
- (G) Determination of the endolysin activity of Pro.3, 4, 5, Mix proteins (0.25 and 2.5 μ M), and buffer control at 37°C. Boxplots represent the interquartile range (IQR), with the center line indicating the median. Whiskers extend to 1.5 $\times$ IQR, and individual data points beyond this range are shown as outliers. Statistical significance was assessed using two-sided t tests for all pairwise comparisons.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

analyses of oral virome data. We believe that HOVD provides a valuable platform for advancing oral microbiome research.

Due to the intimate relationship between chronic diseases and the oral microbiome,⁶¹ we focused on the oral microbiome in obese participants with and without T2D. Compared to the Ob/non-T2D group, the Ob/T2D group exhibited a reduction in oral viral diversity. Notably, the alpha diversity of oral bacterial communities remained similar between the two groups. Furthermore, we observed a disruption in phage-bacteria interactions in the Ob/T2D group, as evidenced by weaker correlations in the phage-bacteria abundance network. A similar trend in the interaction between the virome and bacteriome has been observed in the gut microbiome of Ob/T2D patients.⁶² The disruption of phage-bacteria interactions has also been documented in several other disease states, including diabetic nephropathy and autism spectrum disorder.^{63,64} In disease states like T2D, the weakening of phage-bacteria interactions may be due to changes in microbial communities, human host metabolism, and inflammatory responses. These factors disrupt the balance between phages and their bacterial hosts, leading to a less connected and more uniform interaction network. Another possible explanation might be cross-infective phages, which were involved in one-to-many or many-to-many relationships and complicated correspondence between hosts and phage communities.^{46,65} The first oral phage-host interaction network computed by Wang et al. demonstrated that a regulatory function had been identified for a small portion of cross-infective phages that infected both periodontal pathogens and commensal bacteria.⁴⁶ The cross-infective phages that potentially modulate the bacterial dysbiosis prevalent in oral diseases still need to be further explored.¹⁸ Future research could investigate the mechanisms underlying the dysregulation of phage-bacteria interactions in various diseases and explore the development of diagnostic tools and therapeutic strategies based on these interactions.

Periodontitis might contribute to an increase in the incidence of cases of T2D. Due to the tremendous global burden of diabetes, it was urgent to prevent and treat periodontal disease to mitigate the influence of periodontitis on diabetes and its complications.^{66,67} Several studies have suggested an indicated link between altered glucose metabolism in prediabetes and diabetes and a shift in the periodontal microbiome.⁶⁸ Among the oral pathogens, *P. gingivalis* was a Gram-negative, rod-shaped, obligate anaerobe distributed worldwide and a causative agent of periodontitis.⁶⁹ Oral administration of *P. gingivalis* could lead to alveolar bone loss and promote glucose metabolism disorders in mice.⁷⁰ In this study, we used 16S rRNA gene sequencing to profile *Porphyromonas* in Ob/T2D and Ob/non-T2D groups, as it offers accurate genus-level resolution and reduces biases from host and viral DNA present in metagenomic data. The combined use of 16S rRNA and metagenomic sequencing enabled cross-validation across bacterial and viral communities, strengthening the robustness of our findings. Then, we computationally identified several phages that are closely related to *P. gingivalis*. Phages may play multiple roles in modulating the microbiome. They were able to infect bacteria directly and reduce bacterial abundance, a concept known as “phage therapy.”^{71,72} Although we were unable to isolate phages targeting

P. gingivalis (data not shown), this may be due to the low phage titer in saliva samples from patients with periodontitis used for isolation. However, we successfully synthesized endolysins, identified from oral virome data, which are capable of inhibiting *P. gingivalis* growth. These findings suggest that exploring phages and phage-encoded enzymes targeting *P. gingivalis* could offer a promising strategy to inhibit this bacterium, potentially reducing the incidence of periodontitis and T2D.

Enhanced phage activity has been suggested to play an important role in the progression of periodontal disease.⁷³ Therefore, the phages identified as being hosted by *P. gingivalis* could be therapeutic targets for these diseases. Additionally, phages may encode genes that modify bacterial physiology and interactions. Prophages have been shown to contribute to the acquisition and dissemination of antibiotic resistance genes, as well as the evolution of pathogenicity in *P. gingivalis*.⁷⁴ With the global rise in antibiotic-resistant bacterial infections, the use of phages as alternative antimicrobial strategies has gained popularity.^{75,76} In the future, we would further verify the possibility of screening phages in the treatment of diabetes and periodontitis.

In summary, the established human oral virome catalog integrated the largest number of oral metagenomic samples to date. It included a greater number of high-quality oral viral genomes, distinguished between oral eukaryotic viruses and bacteriophages, and provided more detailed labeling of oral sample sources. This advancement broadened our understanding of the human oral virome and uncovered more unknowns related to the functions and potential hosts of oral bacteriophages. Besides, our study provided evidence that obesity with T2D was characterized by decreased oral viral diversity, lower correlations with clinical features, disrupted viral-bacterial interactions, and an enhanced oral-gut virome connection, compared to obese individuals without T2D. We identified potential *P. gingivalis* phages and validated their therapeutic potential by screening six putative endolysins, demonstrating that three endolysins effectively inhibit *P. gingivalis* growth as possible therapeutic strategies for individuals with diabetes and periodontitis. Our results indicated a crucial role of the oral virome in the development of T2D, explored the oral-gut transmission in a hyperglycemic state, and offered avenues for developing therapeutic strategies for T2D and periodontitis.

Limitations of the study

Although our study has deepened the understanding of the human oral virome and its relevance to the oral bacteriome and gut virome in obese participants with T2D, several limitations still need to be addressed in the future. First, most oral phage genomes could be annotated at the class level, while almost 97% (23,610/24,440) of genomes could not be distinguished at the order level, indicating many unknown aspects of the human oral virome that await exploration. Besides, we did not analyze all the oral metagenomes with iVirP, potentially resulting in insufficient excavation of viral genomes. Also, only approximately 30% of viral genomes could be assigned to bacterial hosts, which reflects the current limitations of host prediction methods arising from incomplete reference databases and the complex nature of virus-host interactions. Furthermore, we did not

analyze statistically significant correlations between oral viromes and clinical indices collected. Given the well-established association between type 2 diabetes and periodontitis, more clinical features related to the diagnosis of obesity, diabetes, and periodontitis should be included, and more effective models should be developed to evaluate the correlation between the human oral virome and clinical manifestations. Additionally, limited analysis was performed, including the relationship between the oral bacteriome and virome and between the oral and gut virome. It needs to be further analyzed and discussed, including specific viruses shared by the oral and intestinal microbiota and particularly remarkable alterations in correlations between the gut virome and bacteriome, which might be involved in the development of T2D.^{63,64}

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Guanxiang Liang (guanxiangliang@tsinghua.edu.cn).

Materials availability

All related reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- The authors declared that all the data supporting the findings of this study were available within the article and its supplementary information files or from the corresponding author upon reasonable request. The gut data presented in the study were published previously⁴⁵ and were deposited in the China National GeneBank Sequence Archive (<https://db.cngb.org/cnsa/>), accession number CNP0002048. The oral data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (<https://ngdc.cncb.ac.cn/gsa/>), accession number CRA028123.
- Also, we developed a web tool (<https://hovd.org/>) to introduce our oral virome dataset. The remaining data are available in the manuscript, supplemental information, or Source Data file. The study protocol is provided in the supplemental information file.
- Supporting code can be accessed via <https://hovd.org/code-availability/> and <https://github.com/willmaxu/Human-oral-virome-database>, and name and version of the software were included in the method.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all the subjects who participated in the study possible. The study was supported by grants from the National Natural Science Foundation of China (81991501, 82341116, and 32200036); National Key R&D Program of China (2022YFA1206100 and 2022YFE0118300); Research Foundation of Peking University School and Hospital of Stomatology (PKUSS20230109 and PKUSS-2024CRFG03); PKUMed-Wisbiom Joint Laboratory for Human Microbiome Research (88451Y0148); the Start-up Foundation of Tsinghua University (400-53332101822), Tsinghua University Initiative Scientific Research Program, Tsinghua University Dushi Program (20241080062), and Tsinghua-Peking Center for Life Sciences (045-61000100122); the China Postdoctoral Science Foundation (2023M740139); Jiangsu Provincial Medical Key Discipline (ZDXK202252); Jiangsu Provincial Health Commission Key Medical Projects (K2023014); and Suzhou Gusu Health Talent Plan Talent Research Project (GSWS2022068).

AUTHOR CONTRIBUTIONS

T.X. conceived of the study, analyzed the data, and wrote the manuscript. X.J., G. Liu, X. Chen, Q.L., G.Z., and B.L. contributed to data analysis and interpretation. Y.Z., X.L., X. Chai, Y.H., H.W., and F.D. participated in saliva sample collection. Y.C. developed the website tools and supported data analysis algorithms. G. Liang and F.C. conceived and designed the study, revised the manuscript, and supervised the project. G. Liang coordinated the study design and provided critical revisions. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

F.C., G. Liang, and T.X. are the inventors on a pending patent application related to phage-derived endolysins targeting *P. gingivalis* described in this study. The application is assigned to Tsinghua University.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Participants enrollment
- **METHOD DETAILS**
 - Sample collection
 - DNA extraction and metagenomic sequencing
 - Preprocessing and assembly
 - Human oral viral database collection
 - Human oral eukaryotic viruses and phage database identification
 - Taxonomic annotation
 - Host prediction
 - Functional annotation and protein clustering
 - 16S rRNA amplicon sequencing analysis
 - Oral, gut virome of obese patients with T2D analysis
 - Prediction the contribution of oral virome and bacteriome to the disorder state
 - Prediction vOTUs hosted by *Porphyromonas gingivalis*
 - Alignment of similar protein sequences of endolysin
 - Protein structure prediction using ColabFold-AlphaFold2
 - Construction of the recombinant his-tag fusion proteins
 - Protein purification
 - Endolysin activity assay
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **ADDITIONAL RESOURCES**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2025.102325>.

Received: April 7, 2025

Revised: July 4, 2025

Accepted: August 5, 2025

REFERENCES

1. Sepich-Poore, G.D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J.A., and Knight, R. (2021). The microbiome and human cancer. *Science* 371, eabc4552. <https://doi.org/10.1126/science.abc4552>.
2. Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. <https://doi.org/10.1038/nature11234>.

3. Dominguez-Bello, M.G., Godoy-Vitorino, F., Knight, R., and Blaser, M.J. (2019). Role of the microbiome in human development. *Gut* 68, 1108–1114. <https://doi.org/10.1136/gutjnl-2018-317503>.
4. Young, V.B. (2017). The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* 356, j831. <https://doi.org/10.1136/bmj.j831>.
5. Baker, J.L., Mark Welch, J.L., Kauffman, K.M., McLean, J.S., and He, X. (2024). The oral microbiome: diversity, biogeography and human health. *Nat. Rev. Microbiol.* 22, 89–104. <https://doi.org/10.1038/s41579-023-00963-6>.
6. Gao, L., Xu, T., Huang, G., Jiang, S., Gu, Y., and Chen, F. (2018). Oral microbiomes: more and more importance in oral cavity and whole body. *Protein Cell* 9, 488–500. <https://doi.org/10.1007/s13238-018-0548-1>.
7. Escapa, I.F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F.E., and Lemon, K.P. (2018). New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems* 3, e00187–18. <https://doi.org/10.1128/mSystems.00187-18>.
8. Fan, X., Alekseyenko, A.V., Wu, J., Peters, B.A., Jacobs, E.J., Gapstur, S.M., Purdue, M.P., Abnet, C.C., Stolzenberg-Solomon, R., Miller, G., et al. (2018). Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. *Gut* 67, 120–127. <https://doi.org/10.1136/gutjnl-2016-312580>.
9. Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., Wu, X., Li, J., Tang, L., Li, Y., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* 21, 895–905. <https://doi.org/10.1038/nm.3914>.
10. Liang, G., and Bushman, F.D. (2021). The human virome: assembly, composition and host interactions. *Nat. Rev. Microbiol.* 19, 514–527. <https://doi.org/10.1038/s41579-021-00536-5>.
11. Hitch, G., Pratten, J., and Taylor, P.W. (2004). Isolation of bacteriophages from the oral cavity. *Lett. Appl. Microbiol.* 39, 215–219. <https://doi.org/10.1111/j.1472-765X.2004.01565.x>.
12. Camargo, A.P., Nayfach, S., Chen, I.M.A., Palaniappan, K., Ratner, A., Chu, K., Ritter, S.J., Reddy, T.B.K., Mukherjee, S., Schulz, F., et al. (2023). IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* 51, D733–D743. <https://doi.org/10.1093/nar/gkac1037>.
13. Abeles, S.R., Robles-Sikisaka, R., Ly, M., Lum, A.G., Salzman, J., Boehm, T.K., and Pride, D.T. (2014). Human oral viruses are personal, persistent and gender-consistent. *ISME J.* 8, 1753–1767. <https://doi.org/10.1038/ismej.2014.31>.
14. Robles-Sikisaka, R., Ly, M., Boehm, T., Naidu, M., Salzman, J., and Pride, D.T. (2013). Association between living environment and human oral viral ecology. *ISME J.* 7, 1710–1724. <https://doi.org/10.1038/ismej.2013.63>.
15. Gao, C., Li, X., Zhao, X., Yang, P., Wang, X., Chen, X., Chen, N., and Chen, F. (2022). Standardized studies of the oral microbiome: From technology-driven to hypothesis-driven. *iMeta* 1, e19. <https://doi.org/10.1002/imt2.19>.
16. Tisza, M.J., and Buck, C.B. (2021). A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc. Natl. Acad. Sci. USA* 118, e2023202118. <https://doi.org/10.1073/pnas.2023202118>.
17. Li, S., Guo, R., Zhang, Y., Li, P., Chen, F., Wang, X., Li, J., Jie, Z., Lv, Q., Jin, H., et al. (2022). A catalog of 48,425 nonredundant viruses from oral metagenomes expands the horizon of the human oral virome. *iScience* 25, 104418. <https://doi.org/10.1016/j.isci.2022.104418>.
18. Martínez, A., Kuraji, R., and Kapila, Y.L. (2021). The human oral virome: Shedding light on the dark matter. *Periodontol* 2000 87, 282–298. <https://doi.org/10.1111/prd.12396>.
19. Ho, S.X., Min, N., Wong, E.P.Y., Chong, C.Y., and Chu, J.J.H. (2021). Characterization of oral virome and microbiome revealed distinctive microbiome disruptions in paediatric patients with hand, foot and mouth disease. *NPJ Biofilms Microbiomes* 7, 19. <https://doi.org/10.1038/s41522-021-00190-y>.
20. Ye, H.L., Zhi, M.F., Chen, B.Y., Lin, W.Z., Li, Y.L., Huang, S.J., Zhou, L.J., Xu, S., Zhang, J., Zhang, W.C., et al. (2024). Alterations of oral and gut viromes in hypertension and/or periodontitis. *mSystems* 9, e0116923. <https://doi.org/10.1128/msystems.01169-23>.
21. Gao, L., Kang, M., Zhang, M.J., Reza Sailani, M., Kuraji, R., Martinez, A., Ye, C., Kamarajan, P., Le, C., Zhan, L., et al. (2020). Polymicrobial periodontal disease triggers a wide radius of effect and unique virome. *NPJ Biofilms Microbiomes* 6, 10. <https://doi.org/10.1038/s41522-020-0120-7>.
22. Guo, R., Li, S., Zhang, Y., Zhang, Y., Wang, G., Ullah, H., Ma, Y., and Yan, Q. (2022). Dysbiotic Oral and Gut Viromes in Untreated and Treated Rheumatoid Arthritis Patients. *Microbiol. Spectr.* 10, e0034822. <https://doi.org/10.1128/spectrum.00348-22>.
23. Guidry, J.T., Birdwell, C.E., and Scott, R.S. (2018). Epstein-Barr virus in the pathogenesis of oral cancers. *Oral Dis.* 24, 497–508. <https://doi.org/10.1111/odi.12656>.
24. Ly, M., Abeles, S.R., Boehm, T.K., Robles-Sikisaka, R., Naidu, M., Santiago-Rodriguez, T., and Pride, D.T. (2014). Altered oral viral ecology in association with periodontal disease. *mBio* 5, e01133-14. <https://doi.org/10.1128/mBio.01133-14>.
25. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* (London, England) 396, 1223–1249. [https://doi.org/10.1016/S0140-6736\(20\)30752-2](https://doi.org/10.1016/S0140-6736(20)30752-2).
26. O'Hearn, M., Lara-Castor, L., Cudhea, F., Miller, V., Reedy, J., Shi, P., Zhang, J., Wong, J.B., Economos, C.D., Micha, R., et al. (2023). Incident type 2 diabetes attributable to suboptimal diet in 184 countries. *Nat. Med.* 29, 982–995. <https://doi.org/10.1038/s41591-023-02278-8>.
27. Zhou, B., Stevens, G.A., Bjerregaard, P., Onat, A., Aekplakorn, W., Aggemyang, C., Ahmadvand, A., Alkerwi, A.A., Anjana, R.M., Balakrishna, N., et al. (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* (British edition) 387, 1513–1530. [https://doi.org/10.1016/S0140-6736\(16\)00618-8](https://doi.org/10.1016/S0140-6736(16)00618-8).
28. International Diabetes Federation (2021). IDF Diabetes Atlas, 10th Edition (International Diabetes Federation). <http://www.diabetesatlas.org>.
29. Ahmad, E., Lim, S., Lamprey, R., Webb, D.R., and Davies, M.J. (2022). Type 2 diabetes. *Lancet* 400, 1803–1820. [https://doi.org/10.1016/S0140-6736\(22\)01655-5](https://doi.org/10.1016/S0140-6736(22)01655-5).
30. Kitamoto, S., Nagao-Kitamoto, H., Hein, R., Schmidt, T.M., and Kamada, N. (2020). The Bacterial Connection between the Oral Cavity and the Gut Diseases. *J. Dent. Res.* 99, 1021–1029. <https://doi.org/10.1177/0022034520924633>.
31. Kitamoto, S., Nagao-Kitamoto, H., Jiao, Y., Gilliland, M.G., Hayashi, A., Imai, J., Sugihara, K., Miyoshi, M., Brazil, J.C., Kuffa, P., et al. (2020). The Intermucosal Connection between the Mouth and Gut in Commensal Pathobiont-Driven Colitis. *Cell* 182, 447–462.e14. <https://doi.org/10.1016/j.cell.2020.05.048>.
32. Kitamoto, S., and Kamada, N. (2022). Periodontal connection with intestinal inflammation: Microbiological and immunological mechanisms. *Periodontol* 2000 89, 142–153. <https://doi.org/10.1111/prd.12424>.
33. Li, B., Jiao, X., and Liang, G.J.B. (2024). iVirP: An Integrative, Efficient, and User-Friendly Pipeline to Annotate Viral Contigs from Raw Reads of Metagenome or VLP Sequencing. Preprint at bioRxiv. <https://doi.org/10.1101/2024.01.21.576577>.
34. Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A., et al. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* 37, 29–37. <https://doi.org/10.1038/nbt.4306>.
35. Nayfach, S., Páez-Espino, D., Call, L., Low, S.J., Sberro, H., Ivanova, N.N., Proal, A.D., Fischbach, M.A., Bhatt, A.S., Hugenholtz, P., and Kyrpides, N.

- N.C. (2021). Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat. Microbiol.* 6, 960–970. <https://doi.org/10.1038/s41564-021-00928-6>.
36. Ahlgren, N.A., Ren, J., Lu, Y.Y., Fuhrman, J.A., and Sun, F. (2017). Alignment-free \$d_2^A\$ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* 45, 39–53. <https://doi.org/10.1093/nar/gkw1002>.
37. Zhu, J., Tian, L., Chen, P., Han, M., Song, L., Tong, X., Sun, X., Yang, F., Lin, Z., Liu, X., et al. (2022). Over 50,000 Metagenomically Assembled Draft Genomes for the Human Oral Microbiome Reveal New Taxa. *Genom. Proteom. Bioinform.* 20, 246–259. <https://doi.org/10.1016/j.gpb.2021.05.001>.
38. Warinner, C., Rodrigues, J.F.M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y., Fiddyment, S., et al. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nat. Genet.* 46, 336–344. <https://doi.org/10.1038/ng.2906>.
39. Tierney, B.T., Yang, Z., Lubner, J.M., Beaudin, M., Wibowo, M.C., Baek, C., Mehlenbacher, E., Patel, C.J., and Kostic, A.D. (2019). The Landscape of Genetic Content in the Gut and Oral Human Microbiome. *Cell Host Microbe* 26, 283–295.e8. <https://doi.org/10.1016/j.chom.2019.07.008>.
40. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>.
41. Haft, D.H., Selengut, J.D., and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31, 371–373. <https://doi.org/10.1093/nar/gkg128>.
42. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. <https://doi.org/10.1093/nar/gky995>.
43. Chen, B.Y., Lin, W.Z., Li, Y.L., Bi, C., Du, L.J., Liu, Y., Zhou, L.J., Liu, T., Xu, S., Shi, C.J., et al. (2023). Roles of oral microbiota and oral-gut microbial transmission in hypertension. *J. Adv. Res.* 43, 147–161. <https://doi.org/10.1016/j.jare.2022.03.007>.
44. Matsha, T.E., Prince, Y., Davids, S., Chikte, U., Erasmus, R.T., Kengne, A.P., and Davison, G.M. (2020). Oral Microbiome Signatures in Diabetes Mellitus and Periodontal Disease. *J. Dent. Res.* 99, 658–665. <https://doi.org/10.1177/0022034520913818>.
45. Jia, R., Huang, M., Qian, L., Yan, X., Lv, Q., Ye, H., Ye, L., Wu, X., Chen, W., Chen, Y., et al. (2021). The Depletion of Carbohydrate Metabolic Genes in the Gut Microbiome Contributes to the Transition From Central Obesity to Type 2 Diabetes. *Front. Endocrinol.* 12, 747646. <https://doi.org/10.3389/fendo.2021.747646>.
46. Wang, J., Gao, Y., and Zhao, F. (2016). Phage-bacteria interaction network in human oral microbiome. *Environ. Microbiol.* 18, 2143–2158. <https://doi.org/10.1111/1462-2920.12923>.
47. Pargin, E., Roach, M.J., Skye, A., Papudeshi, B., Inglis, L.K., Mallawaarachchi, V., Grigson, S.R., Harker, C., Edwards, R.A., and Giles, S.K. (2023). The human gut virome: composition, colonization, interactions, and impacts on human health. *Front. Microbiol.* 14, 963173. <https://doi.org/10.3389/fmicb.2023.963173>.
48. Li, Y., Liu, Y., Cui, J., Zhu, M., Wang, W., Chen, K., Huang, L., and Liu, Y. (2024). Oral-gut microbial transmission promotes diabetic coronary heart disease. *Cardiovasc. Diabetol.* 23, 123. <https://doi.org/10.1186/s12933-024-02217-y>.
49. Park, S.Y., Hwang, B.O., Lim, M., Ok, S.H., Lee, S.K., Chun, K.S., Park, K.K., Hu, Y., Chung, W.Y., and Song, N.Y. (2021). Oral-Gut Microbiome Axis in Gastrointestinal Disease and Cancer. *Cancers* 13, 2124. <https://doi.org/10.3390/cancers13092124>.
50. Gregory, A.C., Zablocki, O., Zayed, A.A., Howell, A., Bolduc, B., and Sullivan, M.B. (2020). The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe* 28, 724–740.e8. <https://doi.org/10.1016/j.chom.2020.08.003>.
51. Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motono, I., Narushima, S., Kiguchi, Y., Yasuma, K., Watanabe, E., Tanoue, T., et al. (2017). Ectopic colonization of oral bacteria in the intestine drives T(H)1 cell induction and inflammation. *Science* 358, 359–365. <https://doi.org/10.1126/science.aan4526>.
52. Liao, C., Rolling, T., Djukovic, A., Fei, T., Mishra, V., Liu, H., Lindberg, C., Dai, L., Zhai, B., Peled, J.U., et al. (2024). Oral bacteria relative abundance in faeces increases due to gut microbiota depletion and is linked with patient outcomes. *Nat. Microbiol.* 9, 1555–1565. <https://doi.org/10.1038/s41564-024-01680-3>.
53. Nibali, L., Gkrani, N., Mainas, G., and Di Pino, A. (2022). Periodontitis and implant complications in diabetes. *Periodontol 2000* 90, 88–105. <https://doi.org/10.1111/prd.12451>.
54. Lalla, E., and Papapanou, P.N. (2011). Diabetes mellitus and periodontitis: a tale of two common interrelated diseases. *Nat. Rev. Endocrinol.* 7, 738–748. <https://doi.org/10.1038/nrendo.2011.106>.
55. Jain, S., and Darveau, R.P. (2010). Contribution of *Porphyromonas gingivalis* lipopolysaccharide to periodontitis. *Periodontol* 54, 53–70. <https://doi.org/10.1111/j.1600-0757.2009.00333.x>.
56. Rahman, M.U., Wang, W., Sun, Q., Shah, J.A., Li, C., Sun, Y., Li, Y., Zhang, B., Chen, W., and Wang, S. (2021). Endolysin, a Promising Solution against Antimicrobial Resistance. *Antibiotics (Basel, Switzerland)* 10, 1277. <https://doi.org/10.3390/antibiotics10111277>.
57. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. <https://doi.org/10.1093/nar/gkn201>.
58. Kim, G., Lee, S., Levy Karin, E., Kim, H., Moriwaki, Y., Ovchinnikov, S., Steinegger, M., and Mirdita, M. (2025). Easy and accurate protein structure prediction using ColabFold. *Nat. Protoc.* 20, 620–642. <https://doi.org/10.1038/s41596-024-01060-5>.
59. Simón-Soro, A., Tomás, I., Cabrera-Rubio, R., Catalan, M.D., Nyvad, B., and Mira, A. (2013). Microbial geography of the oral cavity. *J. Dent. Res.* 92, 616–621. <https://doi.org/10.1177/0022034513488119>.
60. Ximénez-Fyvie, L.A., Haffajee, A.D., and Socransky, S.S. (2000). Comparison of the microbiota of supra- and subgingival plaque in health and periodontitis. *J. Clin. Periodontol.* 27, 648–657. <https://doi.org/10.1034/j.1600-051x.2000.027009648.x>.
61. Docktor, M.J., Paster, B.J., Abramowicz, S., Ingram, J., Wang, Y.E., Correll, M., Jiang, H., Cotton, S.L., Kokaras, A.S., and Bousvaros, A. (2012). Alterations in diversity of the oral microbiome in pediatric inflammatory bowel disease. *Inflamm. Bowel Dis.* 18, 935–942. <https://doi.org/10.1002/ibd.21874>.
62. Yang, K., Niu, J., Zuo, T., Sun, Y., Xu, Z., Tang, W., Liu, Q., Zhang, J., Ng, E.K.W., Wong, S.K.H., et al. (2021). Alterations in the Gut Virome in Obesity and Type 2 Diabetes Mellitus. *Gastroenterology* 161, 1257–1269.e13. <https://doi.org/10.1053/j.gastro.2021.06.056>.
63. Fan, G., Cao, F., Kuang, T., Yi, H., Zhao, C., Wang, L., Peng, J., Zhuang, Z., Xu, T., Luo, Y., et al. (2023). Alterations in the gut virome are associated with type 2 diabetes and diabetic nephropathy. *Gut Microbes* 15, 2226925. <https://doi.org/10.1080/19490976.2023.2226925>.
64. Wan, Y., Zhang, L., Xu, Z., Su, Q., Leung, T.F., Chan, D., Wong, O.W.H., Chan, S., Chan, F.K.L., Tun, H.M., and Ng, S.C. (2024). Alterations in fecal virome and bacteriome virome interplay in children with autism spectrum disorder. *Cell Rep. Med.* 5, 101409. <https://doi.org/10.1016/j.xcrm.2024.101409>.
65. Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B., and Hochberg, M.E. (2013). Phage-bacteria infection networks. *Trends Microbiol.* 21, 82–91. <https://doi.org/10.1016/j.tim.2012.11.003>.
66. Genco, R.J., Graziani, F., and Hasturk, H. (2020). Effects of periodontal disease on glycemic control, complications, and incidence of diabetes mellitus. *Periodontol* 83, 59–65. <https://doi.org/10.1111/prd.12271>.

67. Tomic, D., Shaw, J.E., and Magliano, D.J. (2022). The burden and risks of emerging complications of diabetes mellitus. *Nat. Rev. Endocrinol.* 18, 525–539. <https://doi.org/10.1038/s41574-022-00690-7>.
68. Polak, D., Sanui, T., Nishimura, F., and Shapira, L. (2020). Diabetes as a risk factor for periodontal disease—plausible mechanisms. *Periodontology* 83, 46–58. <https://doi.org/10.1111/prd.12298>.
69. Reyes, L. (2021). *Porphyromonas gingivalis*. *Trends Microbiol.* (Regular ed.) 29, 376–377. <https://doi.org/10.1016/j.tim.2021.01.010>.
70. Kang, N., Zhang, Y., Xue, F., Duan, J., Chen, F., Cai, Y., and Luan, Q. (2022). Periodontitis induced by *Porphyromonas gingivalis* drives impaired glucose metabolism in mice. *Front. Cell. Infect. Microbiol.* 12, 998600. <https://doi.org/10.3389/fcimb.2022.998600>.
71. Strathdee, S.A., Hatfull, G.F., Mutalik, V.K., and Schooley, R.T. (2023). Phage therapy: From biological mechanisms to future directions. *Cell* 186, 17–31. <https://doi.org/10.1016/j.cell.2022.11.017>.
72. Pires, D.P., Costa, A.R., Pinto, G., Meneses, L., and Azeredo, J. (2020). Current challenges and future opportunities of phage therapy. *FEMS Microbiol. Rev.* 44, 684–700. <https://doi.org/10.1093/femsre/fuaa017>.
73. Matrishin, C.B., Haase, E.M., Dewhirst, F.E., Mark Welch, J.L., Miranda-Sanchez, F., Chen, T., MacFarland, D.C., and Kauffman, K.M. (2023). Phages are unrecognized players in the ecology of the oral pathogen *Porphyromonas gingivalis*. *Microbiome* 11, 161. <https://doi.org/10.1186/s40168-023-01607-w>.
74. Gu, B.L., She, Y., Pei, G.K., Du, Y., Yang, R., Ma, L.X., Zhao, Q., and Gao, S.G. (2023). Systematic analysis of prophages carried by *Porphyromonas gingivalis*. *Infect. Genet. Evol.* 113, 105489. <https://doi.org/10.1016/j.meegid.2023.105489>.
75. Uytendaele, S., Chen, B., Onsea, J., Ruythooren, F., Debaveye, Y., Devolder, D., Spriet, I., Depypere, M., Wagemans, J., Lavigne, R., et al. (2022). Safety and efficacy of phage therapy in difficult-to-treat infections: a systematic review. *Lancet Infect. Dis.* 22, e208–e220. [https://doi.org/10.1016/s1473-3099\(21\)00612-5](https://doi.org/10.1016/s1473-3099(21)00612-5).
76. Hatfull, G.F., Dedrick, R.M., and Schooley, R.T. (2022). Phage Therapy for Antibiotic-Resistant Bacterial Infections. *Annu. Rev. Med.* 73, 197–211. <https://doi.org/10.1146/annurev-med-080219-122208>.
77. Quast, C., Priesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. <https://doi.org/10.1093/nar/gks1219>.
78. Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks.
79. Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. <https://doi.org/10.1093/nar/gkz239>.
80. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9, 286–298. <https://doi.org/10.1093/bib/bbn013>.
81. Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 1091. <https://doi.org/10.1038/s41587-019-0252-6>.
82. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
<i>Porphyromonas gingivalis</i>	Central Laboratory of Peking University School of Stomatology	W83
Chemicals, peptides, and recombinant proteins		
VAHTS Universal DNA Library Prep Kit for Illumina HEPES	Vazyme Amresco	Cat# ND607-01; Cas: 7365-45-9
Critical commercial assays		
VAHTS Universal DNA Library Prep Kit for Illumina	Vazyme	Cat# ND607-01
Deposited data		
Analyzed data	Previously published paper and this paper	The gut data: https://doi.org/10.3389/fendo.2021.747646 ; CNSA: CNP0002048 (https://db.cngb.org/cnsa/); The oral data: CRA028123 (https://ngdc.cncb.ac.cn/gsa)
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Human Oral Virome Database website	This paper	https://hovd.org/
Oral virome database from Cenote Human Virome Database (CHVD)	Tisza et al. ¹⁶	https://zenodo.org/record/4498884
Oral Virus Database (OVD)	Li et al. ¹⁷	https://github.com/RChGO/OVD/
Oral virome selected from Genebank	Genebank	https://www.ncbi.nlm.nih.gov/labs/virus/ , Downloaded in October 31 st , 2023
Metagenomic Gut Virus (MGV) protocol	Nayfach et al. ³⁵	https://github.com/snayfach/MGV
MMSeqs2 virus protein database 2022-05-19	N/A	https://zenodo.org/records/6574914
Oral microbial genome catalog	Zhu et al. ³⁷	https://db.cngb.org/microbiome/genomecatalog/human_oral/
VOGDB	N/A	http://vogdb.org
Reference sequence (RefSeq) database	NCBI	https://www.ncbi.nlm.nih.gov/refseq/ , Downloaded in April 11 th 2024
Analyzed data	Previously published paper	https://doi.org/10.3389/fendo.2021.747646 ; CNSA: CNP0002048
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Human Oral Virome Database website	This paper	https://hovd.org/
Oral virome database from Cenote Human Virome Database (CHVD)	Tisza et al. ¹⁶	https://zenodo.org/record/4498884
Oral Virus Database (OVD)	Li et al. ¹⁷	https://github.com/RChGO/OVD/
Oral virome selected from Genebank	Genebank	https://www.ncbi.nlm.nih.gov/labs/virus/ , Downloaded in October 31 st , 2023
Metagenomic Gut Virus (MGV) protocol	Nayfach et al. ³⁵	https://github.com/snayfach/MGV
MMSeqs2 virus protein database 2022-05-19	N/A	https://zenodo.org/records/6574914
Oral microbial genome catalog	Zhu et al. ³⁷	https://db.cngb.org/microbiome/genomecatalog/human_oral/
VOGDB	N/A	http://vogdb.org
Reference sequence (RefSeq) database	NCBI	https://www.ncbi.nlm.nih.gov/refseq/ , Downloaded in April 11 th 2024
Software and algorithms		
iVirP	Li et al. ³³	https://github.com/li-bw18/iVirP

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FastQC 0.12.1	N/A	https://github.com/s-andrews/FastQC
VSEARCH 2.23.0	N/A	https://github.com/torognes/vsearch
SPAdes 3.15.5	N/A	https://github.com/ablab/spades
BLAST 2.14.0	N/A	https://github.com/enormandeau/ncbi_blast_tutorial
CheckV 1.0.1	N/A	https://bitbucket.org/berkeleylab/checkv/src/master/
MMSeqs2 13.451111	N/A	https://github.com/soedinglab/MMseqs2
VirHostMatcher 1.0	Ahlgren et al. ³⁶	https://github.com/jessieren/VirHostMatcher
Prodigal 2.6.3	N/A	https://github.com/hyattpd/Prodigal
HMMER 3.3.2	N/A	https://github.com/EddyRivasLab/hmmer
QIIME2 2021.4	Bolyen et al. ⁷⁶	https://github.com/qiime2/qiime2
Gephi 0.9.2	Bastian et al. ⁷⁷	https://gephi.org/
FastTree 2.1.11	N/A	https://github.com/morgannprice/fasttree
iToL	Letunic et al. ⁷⁸	https://itol.embl.de/
MAFFT 7.505	Katoh et al. ⁷⁹	https://github.com/GSLBiotech/mafft
Jalview 2.11.4.1	Waterhouse et al. ⁸⁰	https://www.jalview.org/
Origin 2021b	Origin Lab	https://www.originlab.com/
R 4.0.3	CRAN	https://www.r-project.org/
iVirP	Li et al. ³³	https://github.com/li-bw18/iVirP
FastQC 0.12.1	N/A	https://github.com/s-andrews/FastQC
VSEARCH 2.23.0	N/A	https://github.com/torognes/vsearch
SPAdes 3.15.5	N/A	https://github.com/ablab/spades
BLAST 2.14.0	N/A	https://github.com/enormandeau/ncbi_blast_tutorial
CheckV 1.0.1	N/A	https://bitbucket.org/berkeleylab/checkv/src/master/
MMSeqs2 13.451111	N/A	https://github.com/soedinglab/MMseqs2
VirHostMatcher 1.0	Ahlgren et al. ³⁶	https://github.com/jessieren/VirHostMatcher
Prodigal 2.6.3	N/A	https://github.com/hyattpd/Prodigal
HMMER 3.3.2	N/A	https://github.com/EddyRivasLab/hmmer
QIIME2 2021.4	Bolyen et al. ⁷⁶	https://github.com/qiime2/qiime2
Gephi 0.9.2	Bastian et al. ⁷⁷	https://gephi.org/
FastTree 2.1.11	N/A	https://github.com/morgannprice/fasttree
iToL	Letunic et al. ⁷⁸	https://itol.embl.de/
MAFFT 7.505	Katoh et al. ⁷⁹	https://github.com/GSLBiotech/mafft
Jalview 2.11.4.1	Waterhouse et al. ⁸⁰	https://www.jalview.org/
Origin 2021b	Origin Lab	https://www.originlab.com/
R 4.0.3	CRAN	https://www.r-project.org/
Other		
NovaSeq platform	Illumina	https://www.illumina.com/
NovaSeq platform	Illumina	https://www.illumina.com/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The study was approved by the Ethics Committee of Suzhou Municipal Hospital (K-2019-016-H01). Participants in our study were individually informed about the aim and details of research and signed the informed consent.

Participants enrollment

This study included a total of 220 obese participants, comprising 162 obese patients with T2D [Ob/T2D; females ($n = 99$) and males ($n = 63$)] and 58 obese patients without T2D [Ob/non-T2D; females ($n = 31$) and males ($n = 27$)] as controls who underwent medical examinations simultaneously. Gender information was extracted from medical records and reflects physician-documented gender at enrollment, as reported by patients or based on standard identification documents. The diagnosis criteria for T2D were based on the criteria of weight for Chinese adults was according to the instructions of the National Health Commission of the People's Republic of China) and met one of the following conditions: (a) waist circumference, males ≥ 90 cm; females ≥ 85 cm; (b) fasting plasma glucose

level >7 mmol/L; (c) without treatment medicines. Exclusion criteria were: (a) subjects who are under 18 years old; (b) pregnant or lactating women; presence of major gastrointestinal surgery and any known chronic disorders (such as pulmonary, cardiovascular, gastrointestinal, hepatic, and renal diseases); (c) intake of antibiotics, steroids, or probiotics in the past 2 months; (d) family history of diabetes or other related diseases.

METHOD DETAILS

Sample collection

Oral saliva and feces samples were collected at the same follow-up. Participants were not allowed to eat, drink, or brush their teeth for 2 h before sampling. Feces were freshly obtained and gathered with specific containers. And unstimulated saliva was taken with drooling into a 50 mL sterile centrifuge tube within 15 min. Samples were immediately frozen at 4°C and transferred to the refrigerator at −80°C for DNA extraction.

DNA extraction and metagenomic sequencing

The genomic DNA was extracted from the clinical samples and the concentration and quality of the purified DNA was determined by the Qubit dsDNA HS assay kit. The oral and feces samples were sequenced using both metagenomics and 16sRNA, simultaneously. The metagenomic libraries were established using the protocol of the manufacturer (Vazyme, VAHTS Universal DNA Library Prep Kit for Illumina, cat. ND607-01). Metagenomic sequencing was performed using whole genome shotgun sequencing on an Illumina NovaSeq platform with 150 bp paired-end reads based on the instructions of the manufacturer (Illumina, San Diego, CA, USA). And the V3-V4 region of the 16S rRNA gene was amplified using the forward primers containing the sequence “CCTACGGRRBG CASCAGKVRVGAAT” and reverse primers containing the sequence “GGACTACNVGGGTWTCTAATCC”.

Preprocessing and assembly

Raw reads of oral and feces samples were fed to iVirP (<https://github.com/li-bw18/iVirP>) to preprocess and assemble.³³ The reads were qualified via FastQC (v 0.12.1), and removed human reads by matching the human genome GRCh38. The clean reads from all the oral samples were assembled into 2,199 contigs following the default parameters. After clustering with VSEARCH (v 2.23.0), 2,155 contigs with ≥1 kbp sequence length (average length 59,613.1 bp) in total were obtained. Also, the contigs with ≥1 kbp sequence length from all the oral samples after using SPAdes (v 3.15.5) were annotated based on their best matches against the Nucleotide Sequence Database (NT database) using BLAST (v 2.14.0). The contigs identified as viruses with >90 query coverage, and e-value <1e−15 were estimated with CheckV (v 1.0.1) to select genomes with >50% genome completeness. Then, a total of 706 contigs (average length 47,637.4 bp) were used to further analyze.

Human oral viral database collection

We collected several studies containing human oral viral genome databases publicly available until November 2023, including Oral Virus Database (OVD),¹⁷ oral virome database from Cenote Human Virome Database (CHVD),¹⁶ and oral virome selected from Genbank (downloaded from <https://www.ncbi.nlm.nih.gov/labs/virus/>, in October 31st, 2023).

Human oral eukaryotic viruses and phage database identification

Based on the taxonomic annotation of viral sequences in the previous database, we distinguished human oral eukaryotic viral and human oral phage databases. The viral sequences identified as unknown were classified into the human oral phage database.

In the analysis of phage contigs, 64,802 contigs was extracted and clustered into 60,134. After assessment with CheckV, 29,033 contigs was reserved with >50% genome completeness.

All phage genomes were clustered into species-level vOTUs on the basis of 95% average nucleotide identity (ANI) and 85% alignment fraction (AF) of the shorter sequence according to the protocol by Nayfach et al. (<https://github.com/snayfach/MGV>).³⁵ 24,440 viral genomes were recognized as human oral phage database after clustering into species-level. Also, at the family level, we filtered the connections between vOTUs with >20% AAI or >10% genes shared and employed an inflation factor of 1.2. At the genus level, we reserved the connections between vOTUs with >50% AAI or >20% genes shared and applied an inflation factor of 2.0. We achieved 1,240 genus-level vOTUs and 342 family-level vOTUs in total.

In the analysis of eukaryotic virus, we kept the contigs between 90% and 150% subject sequence length/query sequence length matched with NT database. Then, 83 eukaryotic viral contigs were identified as human oral eukaryotic virome database.

Taxonomic annotation

Viral taxonomic annotation was performed using MMSeqs2 (v13.45111). Query databases were created for viral contigs using “mmseqs created”, and annotated by aligning to MMSeqs2 virus protein database (download from <https://zenodo.org/records/6574914>, v 2022-05-19). The “mmseqs taxonomy” module was used with the “-lca-ranks” parameter set to species, genus, family, order, class, phylum, and kingdom. Annotation results were exported using mmseqs createtsv. Default parameters were applied unless otherwise noted.

Host prediction

We used an oral microbial genome catalog containing 3,589 publicly available species-level genome bins by Zhu et al. (https://db.cngb.org/microbiome/genomecatalog/human_oral/).³⁷ The potential host of oral 24,440 viral genomes was predicted by VirHostMatcher (<https://github.com/jessieren/VirHostMatcher>), which calculates $d2^*$, an oligonucleotide frequency dissimilarity measure that quantifies sequence composition differences between viruses and potential hosts. A smaller $d2^*$ value suggests greater genomic similarity and a higher likelihood of host-virus association.³⁶ For each viral genome, we selected the host genome with the lowest $d2^*$ value as its most probable host.

Functional annotation and protein clustering

1,364,888 protein-coding genes were identified from all the contigs of the Human oral virus database with Prodigal (v 2.6.3) and were annotated based on HMM searches using protein family databases, including KEGG,⁴⁰ TIGRFAM,⁴¹ Pfam,⁴² VOGDB (<http://vogdb.org>). HMMER (v 3.3.2) was applied to performed searches using the hmmsearch utility with default parameters. We clustered all the proteins with 30% AAI and 70% alignment coverage using MMseqs2 (v 13.451111).

16S rRNA amplicon sequencing analysis

Raw sequencing reads of 16S rRNA sequences were qualified and calculated with QIIME2 (v 2021.4).⁸¹ Silva reference databases were used to annotate oral bacterial genomes.⁷⁷ A reference feature table from the QIIME2 was applied to the future analysis.

Oral, gut virome of obese patients with T2D analysis

Alpha and beta diversities were calculated by the diversity and PCoA function with R package vegan (v 2.6-2) and R package ape (v 5.6-2). Data processing was visualized by R package ggplot2 (v 3.4.4), and R package pheatmap (v 1.0.12). Correlations between the oral, gut virome, oral bacterial, and clinical parameters were performed using the Spearman correlation coefficient with R package corplot (v 0.90). Network diagrams of microbial communities were calculated and visualized by the “Gephi” interactive platform (v 0.9.2).⁷⁸

Prediction the contribution of oral virome and bacteriome to the disorder state

We applied eXtreme Gradient Boosting (R package xgboost, v 1.7.7.1) method to predict how oral virome and bacteriome might be contribute to the disorder state. We defined the disorder state, including obesity with T2D (represented as “1”) and obesity without T2D (represented as “0”). Clinical parameters contained age, height, weight, body mass index (BMI), waistline, hipline, monocyte, leukocyte, lymphocyte, neutrophil and low density lipoprotein-cholesterol (LDL-Ch).

Every prediction model was trained on a training set (70% of the samples) and tested on a test set (30% of the samples). SHapley Additive exPlanations (SHAP) values were used to quantify the contributions of each feature to the model. The following prediction models were measured:

- Model 1. Disorder state $\sim$ clinical parameters;
- Model 2. Disorder state $\sim$ clinical parameters + oral viral abundance;
- Model 3. Disorder state $\sim$ clinical parameters + oral bacterial abundance;
- Model 4. Disorder state $\sim$ all combined.

Prediction vOTUs hosted by *Porphyromonas gingivalis*

To seek the contigs whose host with VirHostMatcher was *P. gingivalis*, we downloaded *P. gingivalis* sequencing from the Reference sequence (RefSeq) database at NCBI in April 11th 2024. The contigs were identified as *P. gingivalis* phage as the following criterion: (1) the $d2^*$ dissimilarity of *P. gingivalis* was less than 0.25; (2) the $d2^*$ dissimilarity of oral microbial genome catalog was determined as the genus *Porphyromonas* and species unknown. A total of 462 vOTUs were hosted by Genus *Porphyromonas* with Human oral genome database, including *Porphyromonas somerae* ($n = 182$), *Porphyromonas uenonis* ($n = 3$), and unknown species ($n = 277$). And, 336 oral vOTUs were predicted to be hosted by *P. gingivalis* with $d2^*$ dissimilarity < 0.25 , and 28 vOTUs were with $d2^*$ dissimilarity < 0.2 . Finally, 25 genomes with < 0.25 dissimilarity and 5 genome with < 0.2 dissimilarity were kept as *P. gingivalis* phage. FastTree(v.2.1.11) was used to establish phylogenetic trees with default options and visualized using iTOL (<https://itol.embl.de/>).⁷⁹

Alignment of similar protein sequences of endolysin

To identify similar protein sequences of six target proteins, NCBI BLASTP tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was utilized.⁵⁷ Each of the six target sequences was input into BLASTP, and the top five most similar sequences for each query were downloaded. Redundant sequences were removed to ensure distinctiveness. The sequence alignments were then performed using the MAFFT program (v 7.505) in auto mode.⁸⁰ After performing the alignment, the phylogenetic tree was visualized using iTOL.⁷⁹ Finally, the sequence alignment was visualized in Jalview (v 2.11.4.1).⁸²

Protein structure prediction using ColabFold-AlphaFold2

Protein structures were predicted using ColabFold-AlphaFold2 (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>). Target protein sequences in FASTA format were submitted, and MSAs were generated using MMseqs2. AlphaFold2 predicted 3D structures, with the highest confidence score model selected for further analysis.⁵⁸

Construction of the recombinant his-tag fusion proteins

The six gene encoding were tagged with 6xHis sequence at the N-terminus or C-terminus using primers. PCR purified fragments were digested with NdeI and HindIII restriction enzymes and ligated within pET30a vector. After isolation, recombinant plasmids were confirmed by Sanger sequencing and introduced into *E. coli* for protein production and purification.

Protein purification

Endolysin were purified using nickel-based immobilized metal ion affinity chromatography (IMAC). *E. coli* BL21(DE3) were grown at 37°C in 1 L LB media supplemented with kanamycin (50 µg/mL). At OD₆₀₀ ~ 0.8, cells were induced with 0.5 mM IPTG for 14 h at 16°C. After collecting the cells by centrifugation, they were resuspended in lysis buffer (20 mM HEPES-NaOH, pH 7.4, 200 mM NaCl, and 5% glycerol), and were subjected to sonication. Sonicated samples were centrifuged at 16,000 rpm for 50 min. Cell-free supernatant was incubated with his-tagged protein purification resin (equilibrated with lysis buffer) for 2 h at 4°C with gentle shaking. The resin was collected and washed with gradient wash buffer (lysis buffer supplemented with 20 mM or 50 mM imidazole). Protein was eluted using elution buffer (lysis buffer supplemented with 200 mM imidazole). Eluted fractions were concentrated and loaded onto a HiTrap desalting column that was pre-equilibrated with lysis buffer. The fractions were pooled, concentrated, frozen and stored at –80°C for future use.

Endolysin activity assay

Porphyromonas gingivalis was obtained from the Central Laboratory of Peking University School of Stomatology. Cultures were grown anaerobically in Brain Heart Infusion (BHI) medium supplemented with 0.5% hemin and 0.5% vitamin K at 37°C until reaching an optical density.

For the hydrolytic activity assay, Pro.3, Pro.4 and Pro.5 were diluted to 2.5µM and 0.25µM respectively using elution buffer. Then mixed the diluted endolysin with *P. gingivalis* and anaerobically cultured at 37°C for 100 min with the following groups.

- (1) Group 1: *P. gingivalis* + elution buffer (negative control).
- (2) Groups 2 & 3: *P. gingivalis* + 2.5 µM or 0.25 µM Pro.3 endolysin.
- (3) Groups 4 & 5: *P. gingivalis* + 2.5 µM or 0.25 µM Pro.4 endolysin.
- (4) Groups 6 & 7: *P. gingivalis* + 2.5 µM or 0.25 µM Pro.5 endolysin.
- (5) Groups 8 & 9: *P. gingivalis* + 2.5 µM or 0.25 µM of an equimolar mixture of Pro.3, Pro.4, and Pro.5 endolysins.

Buffer was included as negative controls. After incubation, bacterial lysis was quantified by measuring the absorbance at 600 nm of the supernatant, reflecting hydrolysis of endolysin on *P. gingivalis*. OD 600nm was detected by Microplate reader (SpectraMax 190, Molecular Devices, San Jose, CA, USA).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data were presented as the means ± SE. *p* value was calculated by two-tailed Student's *t* tests, Wilcoxon test or Kruskal–Wallis rank-sum test according to the data types using R package rstatix (v 0.7.0). Statistical methods were depicted in figure legends. Statistical analyses were described with Origin Software (2021b) and R software (v 4.0.3).

ADDITIONAL RESOURCES

Human Oral Virome Database (HOVD): <https://hovd.example.org>.