nature communications

Article

https://doi.org/10.1038/s41467-025-58280-w

Large-scale biosynthetic analysis of human microbiomes reveals diverse protective ribosomal peptides

Received: 5 November 2024

Accepted: 14 March 2025

Published online: 28 March 2025

Check for updates

Jian Zhang[1,3], Dengwei Zhang[1,3], Yi Xu[2], Junliang Zhang[1], Runze Liu[1], Ying Gao[1], Yuqi Shi[1], Peiyan Cai[1], Zheng Zhong[1], Beibei He[1], Xuechen Li[1], Hongwei Zhou[2], Muxuan Chen[2] ✉ & Yong-Xin Li[1] ✉

The human microbiome produces diverse metabolites that influence host health, yet the chemical landscape of ribosomally synthesized and post-translationally modified peptides (RiPPs)—a versatile class of bioactive compounds—remains underexplored. Here, we conduct a large-scale biosynthetic analysis of 306,481 microbial genomes from human-associated microbiomes, uncovering a broad array of yet-to-be-discovered RiPPs. These RiPPs are distributed across various body sites but show a specific enrichment in the gut and oral microbiome. Big data omics analysis reveals that numerous RiPP families are inversely related to various diseases, suggesting their potential protective effects on health. For a proof of principle study, we apply the synthetic-bioinformatic natural product (syn-BNP) approach to RiPPs and chemically synthesize nine autoinducing peptides (AIPs) for in vitro and ex vivo assay. Our findings reveal that five AIPs effectively inhibit the biofilm formation of disease-associated pathogens. Furthermore, when ex vivo testing gut microbiota from mice with inflammatory bowel disease, we observe that two AIPs can regulate the microbial community and reduce harmful species. These findings highlight the vast potential of human microbial RiPPs in regulating microbial communities and maintaining human health, emphasizing their potential for therapeutic development.

Humans and their symbiotic microbes form a complex and diversified holobiont. Perturbation of the human microbiota, known as dysbiosis, has been implicated in the pathogenesis of various diseases[1,2], which is largely attributed to the chemical interactions mediated by metabolites de novo biosynthesized or transformed by the colonized microbiota[3–6]. As a result, extensive efforts have been made to decipher bioactive molecules secreted by the human microbiota[7,8], revealing a rather diverse array of metabolites such as polyketides (PKs), non-ribosomal peptides (NRPs), and ribosomally synthesized and post-translationally modified peptides (RiPPs). RiPPs have particularly attracted increasing attention because of their diverse structures[9], various biological functions, and manifold roles in physiology and ecology[10,11].

RiPPs are a class of diverse and widespread natural products produced by a multitude of producers, including bacteria, fungi, and plants[10]. They often exhibit diverse activities[12], allowing them to mediate the communications between microbes or between microbes and human hosts. Antibacterial RiPPs can enable commensal producers to combat pre-existing or foreign pathogens[13]. For example, lactocilin, a thiopeptide from vaginal *Lactobacillus gasseri*, demonstrates

[1]Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China. [2]Microbiome Medicine Center, Department of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China. [3]These authors contributed equally: Jian Zhang, Dengwei Zhang. ✉e-mail: muxuanchen@126.com; yxpli@hku.hk

potent antibacterial effects against vaginal pathogens such as *Corynebacterium aurimucosum*[14]. Nisin-like lantibiotics identified in gut microbiome against both pathogens and human gut commensals could shed light on the future development of therapeutics[15]. Additionally, immunomodulatory RiPPs[16–19] can interact with the human immune system, exemplified by tonsillar microbiome-derived Salivaricins, which can block the binding of IL-6 and IL-21 to their receptors to disrupt the immune response[16]. Beyond impacting the microbial community or host directly, RiPPs can also act as signaling molecules to regulate human microbiome responses, as evidenced by autoinducing peptides (AIPs) from the skin microbiome, which maintain skin barrier homeostasis by disrupting the *agr* system of the *Staphylococcus aureus* pathogen[20]. These observations underscore the diverse chemical nature of RiPPs and their varied protective roles to human health. Nonetheless, our current understanding in this area is primarily limited to sporadic reports, highlighting the urgent imperative for a more comprehensive investigation.

While over 40 subclasses of RiPPs are differentiated by their unique structures, their biosynthesis follows a standard three-step process controlled by biosynthetic gene clusters (BGCs)[9,21]. The biosynthesis process includes expressing a precursor peptide, installing various structural features to the core peptide through post-translational modifications (PTM), and finalizing with leader cleavage and export to produce active RiPP products. The conserved biosynthetic logic allows the development of computational approaches for RiPP discovery, rendering diverse tools poised for RiPP identification from the human microbiome[22,23]. These tools typically are established on precursor-centric genome mining strategies (e.g., DeepRiPP[24] and TrRiPP[25]) or primarily targeting the PTM enzymes (e.g., antiSMASH[26] and DeepBGC[27]). Leveraging various bioinformatics tools, researchers are now focusing on understanding the biosynthetic potential of RiPPs within the complex microbiota. Some studies have concentrated on the biosynthetic potential of specific strains, as illustrated by the discovery of bioactive RiPPs like Ruminococcin C[28] and Streptosactin[29] from human commensal bacteria *Ruminococcus gnavus* and *Streptococcus* spp., respectively. Others have centered on in silico investigation of RiPP from a limited number of reference genomes[14,30–32]. However, our understanding of the biosynthetic landscape of RiPPs from human microbiomes is still limited, hindered by a lack of explored reference genomes and imperfect prediction tools. As such, a more thorough examination of RiPPs is needed to improve our understanding of their biosynthesis, laying the groundwork for understanding the crucial biological role of RiPPs within the complex microbiota.

The exponential increase in human microbial genome data provides an opportunity for systematically exploring RiPPs biosynthesis in the human microbiome and their potential role in human health. In this study, using precursor-centric and PTM-centric approaches, we systematically examined the RiPP biosynthetic potential of 306,481 genomes of the human microbiota to identify 12,076 yet-to-be-discovered RiPP families (the clusters of RiPP precursors with a threshold of 50% identity), largely expanding the chemical landscape of RiPPs. For a proof of principle study, we have found 30 RiPP families, through comparative meta-omic analyses, that are negatively associated with multiple diseases such as inflammatory bowel disease (IBD) and colorectal cancer (CRC). While synthetic-bioinformatic natural product (syn-BNP) strategies have historically focused on NRPs due to their modular biosynthetic logic and compatibility with bioinformatic prediction tools[33–35], we sought to expand this framework to RiPPs. Specifically, we targeted autoinducing peptides (AIPs), considering their biological significance, synthetic feasibility, and well-established enzymology. Experimental validation of biosynthesis-guided chemically synthesized AIPs demonstrated their potential to counter biofilm formation of disease-related pathogens. Using an ex vivo culture to simulate the gut microbiota's response to AIPs, we identified two AIPs capable of modifying the gut microbial community and reducing pathogenic species from IBD-affected mice. These findings offer valuable insights into the chemical diversity of RiPPs and their protective roles in human health, sparking growing interest in the therapeutic potential of RiPPs released by our microbiota.

Results

Mining human microbiomes reveals the untapped biosynthetic potential of RiPPs

Current genome mining for RiPPs can be approached through two main methods: tailoring enzymes-oriented approach (e.g., antiSMASH 6.0[26]) and precursors-centric approach (e.g., DeepRiPP[24] and TrRiPP[25]). The enzymes-oriented approach is reliable in identifying known RiPPs but may miss out on discovering uncharacterized ones. On the other hand, the precursors-centric approach is better suited for identifying uncharacterized RiPPs, especially from highly fragmented metagenome-assembled genomes but may have a slightly lower accuracy rate. To comprehensively explore the biosynthetic potential of RiPPs in the human microbiomes, we combined both types of tools to analyze a dataset consisting of 306481 reference microbial genomes from diverse human body sites[36–38], including the gut, oral cavity, skin, airways, vagina, and nasal cavity (Fig. 1, Supplementary Fig. 1). Results showed that most genomes (78.3% identified by antiSMASH and 95.3% identified by DeepRiPP and TrRiPP) could encode RiPPs.

The analysis of antiSMASH detected 410,487 RiPP BGCs in 239,927 genomes, accounting for 74.5% of the predicted secondary metabolite BGCs (Fig. 1b, Supplementary Fig. 1, Supplementary Data 1). The other two precursor-centric tools identified 1,777,091 RiPP precursors in 292,041 genomes, with 359,815 precursors overlapping between the two tools (Fig. 1c, Supplementary Fig. 2, Supplementary Data 2). Through examining the adjacent genes of RiPP precursors identified by precursor-centric approaches, we found that the genomic contexts of 32.6% of precursors contained genes associated with known RiPP biosynthesis (Fig. 1c, Supplementary Fig. 3, Supplementary Tables 1-2), while the genomic contexts of 58.4% of precursors contained uncharacterized RiPP biosynthetic enzymes collected in decRiPPter[39] (Fig. 1c). These findings indicated that the precursor-centric approach offered a higher level of uncharacterized biosynthetic potential. The analysis also showed that antiSMASH-defined RiPP BGCs were more abundant in 25 phyla, expanding to 29 phyla with the precursor-centric approaches (Fig. 1d, Supplementary Fig. 1). Notably, the most abundant phyla in the human gut[40], Bacteroidota and Firmicutes, harbored the most abundant RiPPs. This widely distributed RiPPs across diverse human microbiomes underscores human microbes' robust RiPP biosynthetic potential.

Broad diversity and novelty of RiPP precursor peptides in the human microbiome

Given the tremendous potential of RiPP biosynthesis in the human microbiome, we next sought to examine their chemical diversity and novelty. To balance the accuracy and novelty of genome mining in our analysis, we only retained 423,831 RiPP precursors confirmed by at least two mining strategies for downstream analysis, which could effectively reduce the number of false positives while preserving uncharacterized RiPPs. The 423,831 RiPP precursors were found either located within RiPP BGC regions[41] identified by antiSMASH (62.4%) or by both precursors-centric approaches (37.6%) (Fig. 2a, Supplementary Data 3), with rSAM-modified RiPPs being the most abundant (Fig. 2b). Furthermore, we also examined whether these orphan precursors are associated with other antiSMASH-defined RiPP BGCs within the same genome. We found 109,581 out of 159,475 precursors co-exist with other RiPP BGCs within the genome, with 3473 pairs being situated in the same contig (Supplementary Data 3). This suggests that those orphan precursors might be the precursors of other far-distant BGCs. These RiPPs were strongly enriched in 21

https://doi.org/10.1038/s41467-025-58280-w

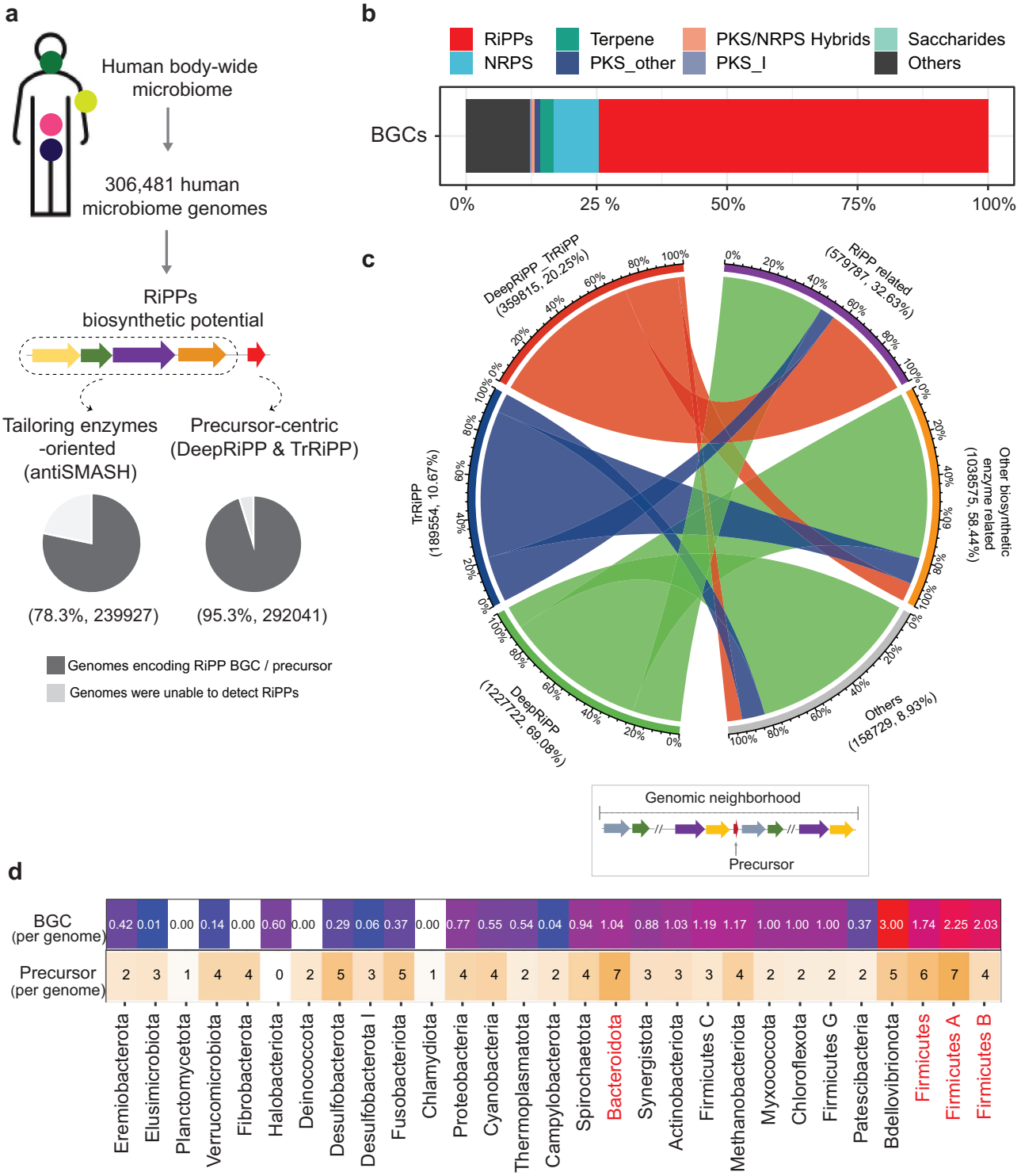
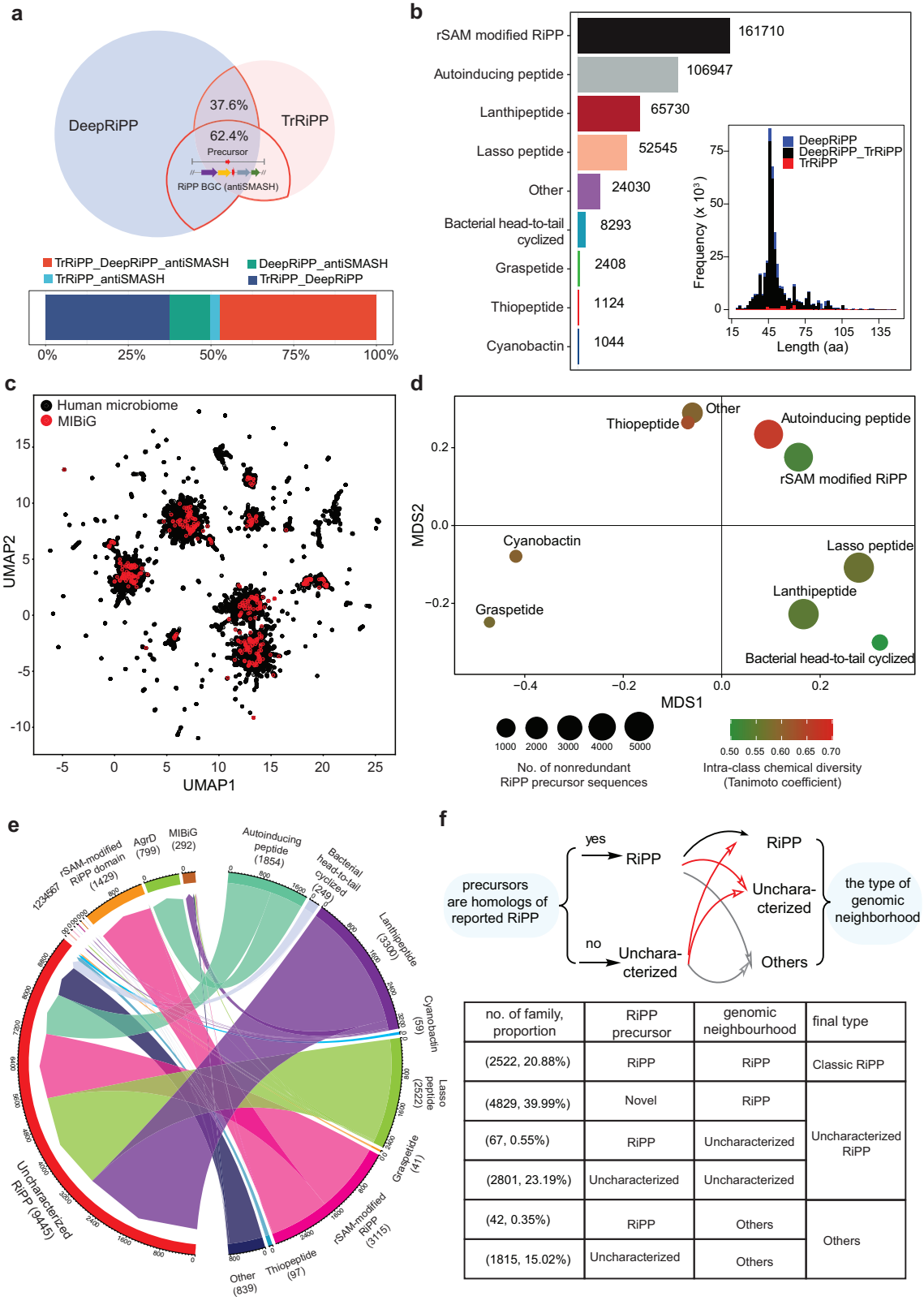


Fig. 1 | Comprehensive survey of RiPPs in the human microbiome. a The workflow of identifying RiPP BGCs and RiPP precursors from 30,6481 human-associated microbial genomes collected from wide body sites. Tailoring enzymes-oriented tool (e.g., antiSMASH) and precursors-centric tools (e.g., DeepRiPP and TrRiPP) are adopted for predicting RiPP BGCs and RiPP precursors, respectively. The pie charts show the percentage of genomes that encode RiPPs. **b** BGCs identified by antiSMASH are grouped into eight BGC classes, including RiPPs (410,487, 74.5%), Terpene (14,753, 2.7%), PKS/NRPS Hybrids (4067, 0.7%), Saccharides (17, 0.003%), NRPS (47,281, 8.6%), PKS_others (5210,0.9%), PKS_I (1506,0.3%), and Others (67,328, 12.2%). **c** Genomic context of identified RiPP precursors by DeepRiPP and TrRiPP. The left panel of the figure displays the scales representing the proportion of precursors identified by either DeepRiPP or TrRiPP, or both approaches (DeepRiPP_TrRiPP). The right panel of the figure illustrates the genomic context associated with the identified precursors. This includes: (1) RiPP related: Genes in the genomic context that are associated with known RiPP biosynthesis. (2) Other biosynthetic enzyme related: Precursors that co-occur with potentially uncharacterized RiPP biosynthetic enzyme(s) from a broader PTM enzyme dataset collected by decRiPPter[39]. (3) Others: Precursors located in the biosynthetic gene cluster (BGC) region of other secondary metabolite BGC classes or under other conditions. **d** The top section of the bar chart represents the counts of BGCs per genome in each phylum, while the bottom section represents the counts of RiPP precursors per genome in each phylum. Taxonomic classification was determined based on annotations from the Genome Taxonomy Database (GTDB). Phyla with higher RiPP biosynthetic potential are highlighted in red.

https://doi.org/10.1038/s41467-025-58280-w



no. of family, proportion	RiPP precursor	genomic neighbourhood	final type
(2522, 20.88%)	RiPP	RiPP	Classic RiPP
(4829, 39.99%)	Novel	RiPP	Uncharacterized RiPP
(67, 0.55%)	RiPP	Uncharacterized	
(2801, 23.19%)	Uncharacterized	Uncharacterized	
(42, 0.35%)	RiPP	Others	Others
(1815, 15.02%)	Uncharacterized	Others	

phyla, 1014 genera, and 3369 species. At the genus level, which is an appropriate taxonomic rank for assessing secondary metabolite biosynthetic diversity[42], we found that RiPP biosynthetic potential considerably varied from 1 to 33 RiPP precursors per genome, with *Elizabethkingia* ($n = 33$), *Chryseobacterium* ($n = 18$) and *Tissierella* ($n = 15$) having the highest number of RiPP precursors (Supplementary Fig. 4, Supplementary Information).

To evaluate RiPP sequence diversity, we calculated the Extended-connectivity fingerprints (ECFPs)[43] of each RiPP precursor, a topological fingerprint for molecular characterization utilized in two-dimensional chemical space. Comparing them with experimentally validated RiPP precursors in the MIBiG database (3.0)[6], we observed a notably expanded chemical space (Fig. 2c, Supplementary Data 4), suggesting that human microbiome-derived RiPPs could significantly

Fig. 2 | RiPP precursor peptides present a hypervariable chemical diversity and novelty. a RiPP precursors retained for downstream analysis. Upper: Two circles in different colors represent the RiPP precursor peptides identified by DeepRiPP and TrRiPP. The red-circled area highlights the precursors that were retained for downstream analysis. Bottom: The stacked barplot illustrates the proportion of RiPP precursors retained for analysis. **b** The outer barplot displays the count of precursor sequences for nine RiPP classes, while the inner stacked barplot represents the distribution of precursor length. **c** Uniform Manifold Approximation and Projection (UMAP) plot showing the chemical space of RiPP precursors obtained from the human microbiome (black dots) and experimentally validated RiPP precursors deposited in the MIBiG 3.0 database (red dots). **d** Multi-Dimensional Scaling (MDS) plot displays the chemical diversity of predicted mature precursors within and between RiPP classes. Dot size signifies the count of unique precursor sequences per class, color indicates median Tanimoto coefficient reflecting class similarity, and distance between dots represents similarity among different RiPP classes. **e**, **f** Classify the novelty of RiPP families based on their precursor and genomic neighborhood (Supplementary Information). **e** The chord diagram illustrates the novelty of identified precursor families (left panel) for nine different RiPP classes (right panel). The novelty of RiPP precursor families into "MIBiG" (homologous to characterized precursors in MIBiG), families with RiPP-associated domains (1, Graspetide (3 families), 2, Lanthipeptide (31 families), 3, Lassopeptide (26 families), 4, Thiopeptide (5 families), 5, RiPP-like (25 families), 6, LAP (19 families), and 7, other known RiPPs (2 families)), and "Uncharacterized RiPP" families. Numbers in brackets indicate the count per category. **f** RiPP families are classified as Classic RiPP (known precursor homology and defined genomic neighborhood), Uncharacterized RiPP (no known precursor homology or novel genomic neighborhood), and Others (remaining families).

enlarge the chemical diversity within the RiPP superfamily. Next, we explored the chemical diversity of nine RiPP classes based on the core precursor sequences predicted by the cleavage prediction module of DeepRiPP[24] (Fig. 2d, Supplementary Data 4). Pairwise Tanimoto coefficient was used to measure the similarity between precursors. Notably, autoinducing peptides exhibited the highest median Tanimoto coefficient, suggesting that their sequences were comparatively conserved within the class. In contrast, bacterial head-to-tail cyclized precursors had the lowest Tanimoto coefficient, indicating high intra-class sequence diversity. Moreover, we found that different RiPP classes were far apart, indicating their sequence uniqueness. Overall, RiPP precursor peptides in the human microbiome exhibit a remarkably diverse chemical landscape characterized by both intra-class and inter-class variations.

To further assess the novelty of RiPPs in the human microbiome, we examined their precursor sequences and tailoring enzymes, which can dictate the chemical structure of mature RiPP products (Fig. 2e, f and Supplementary Figs. 5-6, Supplementary Information). When evaluating the chemical novelty of RiPPs based on their precursors, we clustered 423,831 RiPP precursors into 12,076 families with a threshold of 50% identity (Fig. 2e, Supplementary Data 5). Through querying these families against known RiPP precursors deposited in the MIBiG database, we surprisingly found that only 292 (2.41%) families were homologous (*e*-value < 0.05 and coverage ≥ 90%). For the remaining families, we used RPS-BLAST to search protein domains against the Conserved Domains Database (CDD), detecting 2339 families (19.37%) containing well-defined RiPP precursor domains, such as AgrD domain and rSAM-modified RiPP domain (Supplementary Information). Notably, most families (9445, 78.21%) remained uncharacterized or exhibited significant distinctions from known RiPPs (Fig. 2e). We examined the adjacent genomic context of RiPP precursor families when evaluating the novelty based on tailoring enzymes. We found that the genomic contexts of 60.9% RiPP precursor families harbored classical RiPP biosynthetic-related genes, while the genomic contexts of 23.8% RiPP precursor families contained uncharacterized RiPP biosynthetic-related enzymes defined by DeepBGC[27] (Fig. 2f). Considering both aspects, 63.7% of RiPP precursor families have the potential to produce uncharacterized RiPP products (Fig. 2f). This finding suggests the possibility of uncharacterized enzymology and chemistry in these RiPPs families.

RiPPs are niche-specific and actively transcribed in the healthy human microbiome

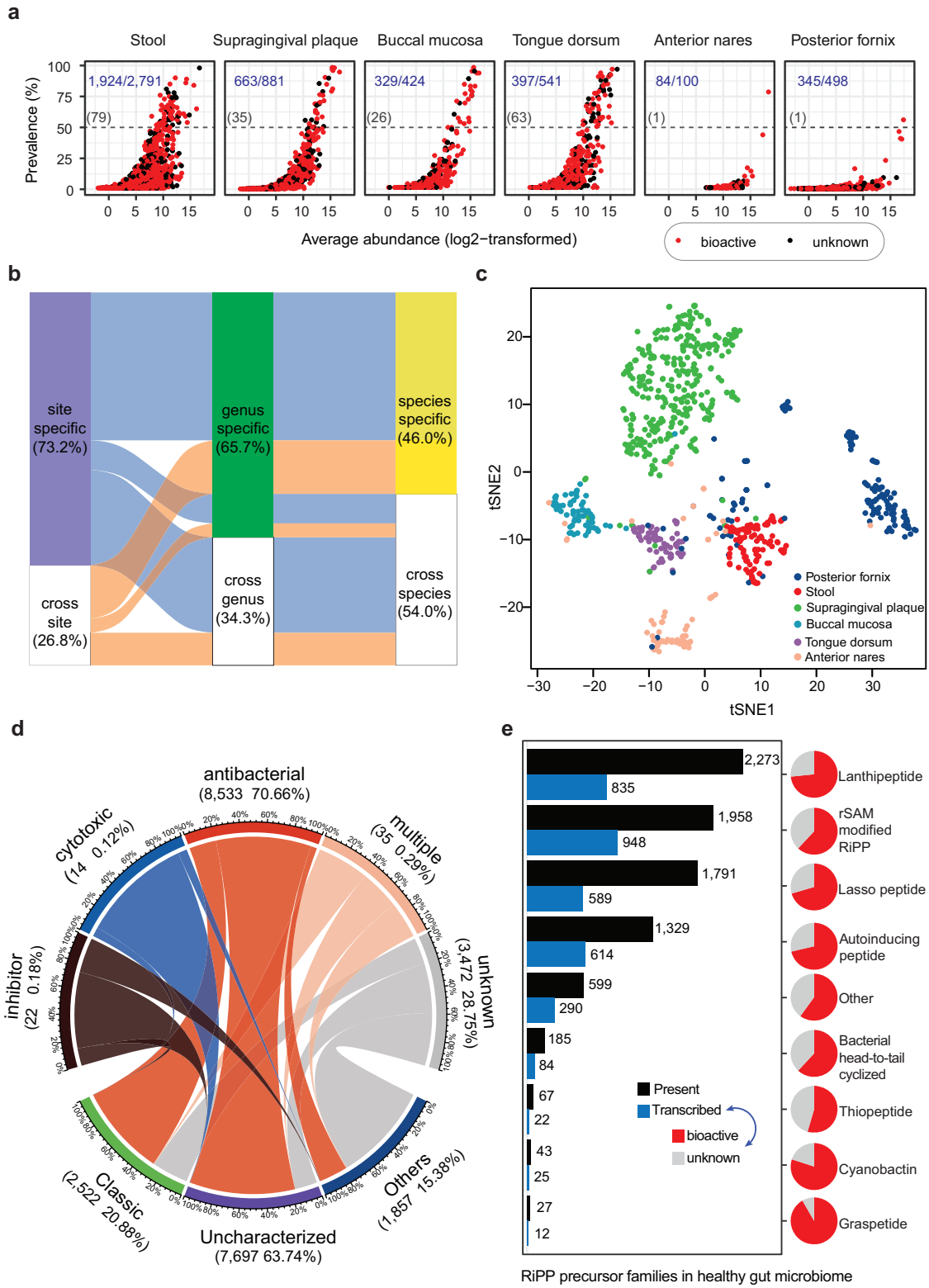
We then attempted to assess RiPP distribution in the human microbiome by examining their profile in healthy human microbiomes. We analyzed 748 metagenomic samples from the Human Microbiome Project (HMP)[44], covering six human body sites: gut, buccal mucosa (oral), supragingival plaque (oral), tongue dorsum (oral), anterior nares (skin), and posterior fornix (vaginal) (Supplementary Data 6). Metagenomic analysis revealed 28.63% (3457/12,076) of RiPP precursor families were detected in healthy individuals, varying considerably from 100 families in anterior nares to 2791 families in the stool (Fig. 3a, Supplementary Data 5). Accumulation curves showed a steep increase, indicating that more RiPP precursor families could be identified in each body site with larger sample sizes (Supplementary Fig. 7a). Compared to the sporadic presence in the skin and vagina, RiPP precursor families exhibited higher diversity, prevalence, and abundance in the oral and gut (Fig. 3a and Supplementary Fig. 7b). Among these detected families, 73.2% (2530/3457) were niche-specific, while 0.9% (31/3457) were present across six body sites in at least one sample (Supplementary Fig. 7c). This could be largely attributed to the fact that the producers of niche-specific RiPPs were habitat-specific (Fig. 3b). Furthermore, two-dimensional visualization demonstrated distinct patterns of RiPP precursor families among different body sites, indicating their niche-specificity (Fig. 3c). Activity prediction showed that most RiPPs (70.7%, 8533/12,076) were antibacterial (Fig. 3d), possibly enabling their producers for niche adaption.

To further confirm whether RiPPs are potentially active in the human microbiome, we looked into their transcription in 281 pairs of metagenomic and metatranscriptomic data[45] (Supplementary Data 6). Among 8272 (68.50%, 8272/12,076) RiPP precursor families detected in metagenome samples, 3419 (28.31%, 3419/12,076) were found to be actively transcribed in metatranscriptome data. Furthermore, a significant proportion of them in each class (ranging from 54.5% of thiopeptide to 91.7% of graspetide) were predicted to exhibit bioactivity (Fig. 3e). These findings necessitate investigating the potential role of bioactive RiPPs in human health.

Differentially transcribed RiPP families are associated with multi-diseases

Inspired by numerous functional RiPPs from human microbiota linked to mediate microbe-microbe interactions or microbe-host interactions, we aimed to explore the potential impact of RiPPs on human health. For a proof of principle study, we conducted comparative metatranscriptome analyses to identify underlying bioactive RiPPs associated with multiple diseases, including type 1 diabetes mellitus (T1D), obesity (OB), liver cirrhosis with hepatitis C infection (LC), colorectal cancer (CRC), Parkinson's disease (PD), and inflammatory bowel disease (IBD, encompassing Crohn's disease (CD) and ulcerative colitis (UC)) (Supplementary Data 6). The metatranscriptomic analysis showed significant differences in RiPP diversity between the health and disease groups (Supplementary Fig. 8). Specifically, the healthy groups had a lower diversity than patients with OB. Permutational multivariate analysis of variance (PERMANOVA) based on Bray-Curtis dissimilarity showed a significant difference ($p < 0.05$) in the overall composition of transcribed RiPP precursor families between health and disease in CRC and IBD case-control cohorts. In total, we identified 195 precursor families depleted while 146 enriched in disease groups within these disease case-control cohorts (Fig. 4a, Supplementary Data 7). Remarkably, 23 families exhibited depletion across multiple disease groups, while 7 families were enriched in multiple disease groups (Fig. 4b, Supplementary Fig. 9).

https://doi.org/10.1038/s41467-025-58280-w



RiPP precursor families in healthy gut microbiome

Some families (e.g., BF_63 and BF_94) significantly differentiated CD patients from healthy individuals in a random forest classification model based on the abundance of transcribed precursor families (Supplementary Fig. 10), suggesting their important roles in maintaining human health. Although the causal relationship and specific mechanisms remain to be determined, these RiPPs may be contributors to host health, and future experimental studies warrant establishing the exact roles of differentially transcribed RiPPs (e.g., exacerbate disease pathogenesis or promote human health[16,46]).

Autoinducing peptides negatively associated with diseases exhibit anti-pathogenic biofilm and anti-inflammatory activity

Although many differentially expressed uncharacterized RiPPs were discovered in our analysis of healthy/disease states (Fig. 4), their

Fig. 3 | Identified RiPP families are present and transcribed in the healthy human microbiome. a The data represents RiPP precursor families' prevalence and average abundance across six human body sites. Each dot signifies a family, with bioactive ones in red and unknown ones in black. Blue numbers denote the count of bioactive and total families per site. Dark grey number in the brackets indicates the number of RiPPs that are present in at least 50% of individuals at each body site. **b** The Sankey diagram illustrates the distribution of RiPP families in the healthy human microbiome, categorizing them as either niche-specific (present in one site) or cross-niche (present in more than one site), and further differentiating them as genus/species-specific or genus/species-cross. **c** The t-SNE plot showcases the distinct profiles of RiPP precursor families in different body sites. The clustering of RiPP families within each body site indicates conservation. Each dot on the plot represents one metagenome sample. **d** Predicted activity of RiPP families. The scale represents the proportion of each predicted activity (upper) or uncharacterized RiPP families (bottom). The number in brackets indicates the count and percentage of the RiPP family relative to all families. The term "multiple" indicates multi-functional RiPPs. Different colors are used to highlight the connections between RiPP types and their predicted activities. **e** Presence and transcription of all RiPP classes in 281 paired metagenomic and metatranscriptomic data from fecal samples of healthy individuals. The pie charts present the percentage of transcribed RiPP families in each class that predicted bioactivity (red) or unknown (grey).

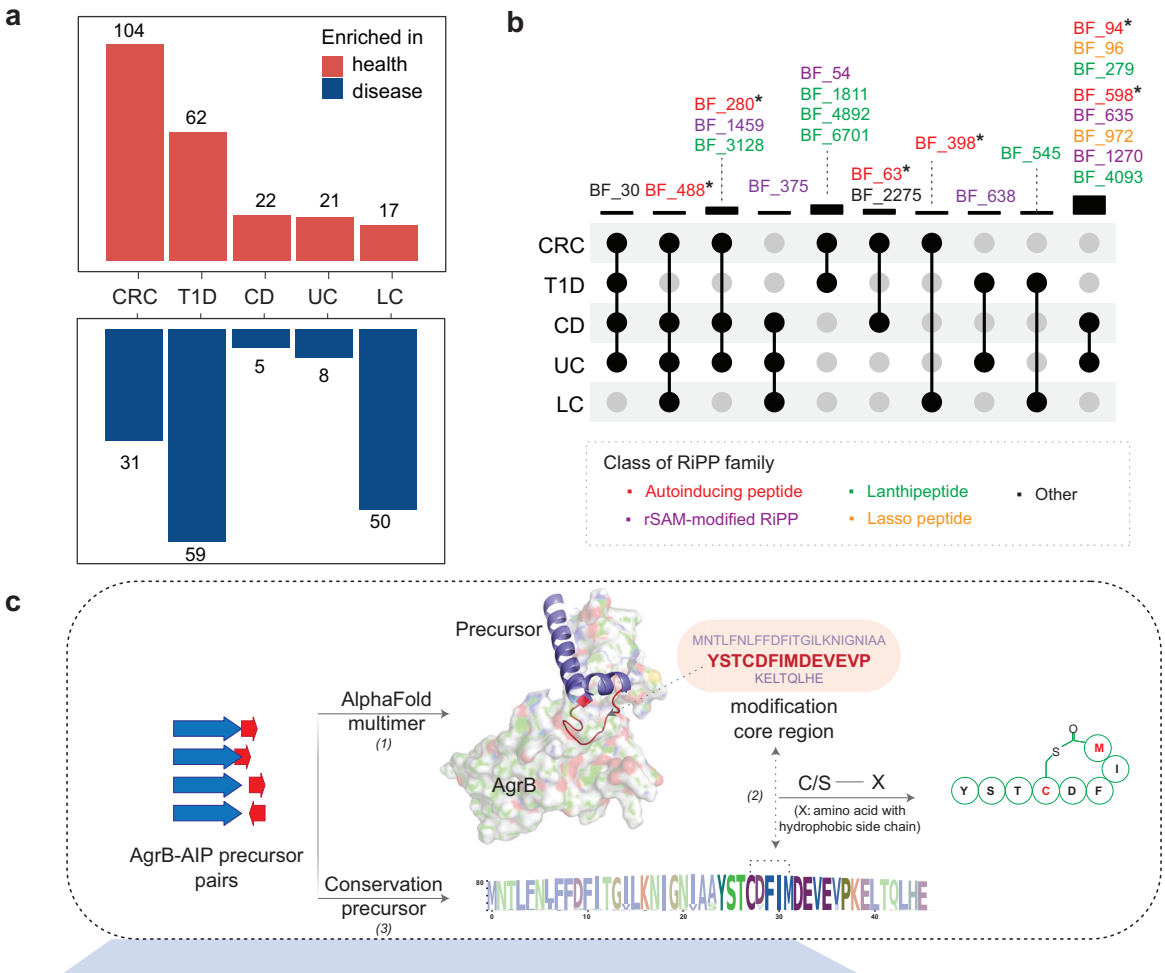
structural complexity and incompletely comprehensively mapped enzymology hindered biosynthetic pathway reconstruction, mature peptide isolation, and functional characterization. Considering the comparative conciseness and predictability of the AIP products, we next sought to experimentally validate the potential functions of six AIP families (BF_63, BF_94, BF_280, BF_398, BF_488, and BF_598), which were enriched in the healthy microbiome (Fig. 4b, Supplementary Figs. 11-16, Supplementary Table 3). The AIP biosynthesis involves the removal of a C-terminal follower and modifications, which are achieved by the AgrB enzyme. The resulting intermediate then undergoes a second cleavage in the N-terminus, ultimately producing mature AIP with variable lengths of the exotail[47]. Although we intended to carry out heterologous expression of six AIPs, this proved unsuccessful. This might be because the protease involved in the second cleavage of AIPs' biosynthetic pathway is still not fully understood, and the native producers of these six AIPs are unavailable. We, therefore, alternatively attempted to chemically synthesize the candidate mature chemicals of AIPs. Here, we extended syn-BNP strategies[33–35], which historically focused on NRPs, to RiPPs. Specifically, we employed a combination of structural prediction and conservation analysis of RiPP precursor to deduce the matured structure of AIPs (Fig. 4c). To do so, we initially established this approach on a well-studied AIP-I from *Staphylococcus aureus* (Supplementary Information). Specifically, we utilized AlphaFold-Multimer to predict the interaction between precursor peptide and AgrB protein and found that the modification core region precisely docks into the catalytic pocket. In the core region, thiolactone or lactone will be installed between Cysteine or Serine and another amino acid containing a hydrophobic side chain by AgrB. Additionally, we can deduce the exocyclic tail region of the final AIPs, as the proteolysis step recognizes and occurs at conserved residues. AIPs containing thiolactone, without the exocyclic region, may rearrange to form homodetic cyclopeptides (cAIPs)[48,49]. Therefore, we could infer the structure of mature AIPs by combining the prediction of core region, modified residues, and exocyclic tail. The accuracy and reliability of this approach were further validated by another two reported AIPs (Supplementary Fig. 17).

According to the biosynthetic logics of AIPs, we deduced the potential mature chemicals of six AIP families (Fig. 4b, Supplementary Fig. 18) and chemically synthesized 9 AIP chemicals, including 3 variants. These were confirmed through high-resolution MS/MS and NMR (Fig. 4d, Supplementary Figs. 19-36, Supplementary Information). After observing that none of them exhibited significant antibacterial activity against a panel of 17 human pathogens (Supplementary Data 8), even at concentrations of 100 or 50 μg/mL, we proceeded to evaluate their inhibitory activities against the biofilm formation of these pathogens (Supplementary Figs. 37-38). We observed that the 5 AIPs (BF_94_et_free, BF_280_t, BF_280_c, BF_398_c, and BF_598_c) demonstrated significant inhibition against at least one pathogenic biofilm (Fig. 4e, Supplementary Fig. 38, Supplementary Data 9), 4 of which showed extended-spectrum activities against pathogenic biofilm across different species. For example, BF_398_c and BF_94_et_free from *Lachnospira* genus, showed relatively potent anti-biofilm activities to *Staphylococcus aureus*, *Listeria monocytogenes*, and *Peptostreptococcus stomati*. We further explored their immunomodulatory effects in vitro in lipopolysaccharide (LPS)-induced mouse macrophage RAW264.7 cells. Using an enzyme-linked immunosorbent assay (ELISA), we observed that BF_94_et_L1 could significantly decrease two pro-inflammatory cytokines, tumor necrosis factor-α (TNF-α) and interleukins 6 (IL-6), at a physiologically relevant low concentration of 0.1 μg/mL (Fig. 4f, Supplementary Data 10). Collectively, their multiple functions implied that human microbiome-derived AIPs potentially provide a multifaceted protective role in human health.

AIPs affect the composition of IBD mouse fecal-derived ex vivo microbial community

Efforts to investigate the role of AIPs in disease contexts would greatly benefit from studying a system that mimics disease-derived microbial communities. We, thus, adopted an IBD mouse fecal-derived ex vivo microbial community to model how the gut microbiota responds to AIPs (Fig. 5a, Supplementary Data 11). We selected two functional AIPs (i.e., BF_280_c and BF_398_c) for assay, given their potent inhibition against biofilm formation, which is a major virulence factor in IBD[50,51]. Antibiotic vancomycin was used as a positive control due to its clinical use in IBD treatment[52]. After incubating fecal-derived microbiota with AIPs or vancomycin for 48 hours, we found, except for vancomycin (10 μg/mL), no marked difference in major taxa when compared with the blank group (Fig. 5b, Supplementary Fig. 39, Supplementary Data 11). Nevertheless, despite their lack of antibacterial activity in our tests, both two AIPs (10 μg/mL) significantly reduced the microbial diversity, as reflected by the Shannon diversity index (Fig. 5c, Supplementary Data 11). Beta diversity analysis utilizing principal coordinate analysis based on Bray-Curtis distances revealed distinct clustering between blank and treatment, with notable differences observed in groups of AIPs (10 μg/mL) and vancomycin (1 μg/mL and 10 μg/mL) (Fig. 5d, Supplementary Fig. 40, Supplementary Data 11). These results indicated that BF_280_c and BF_398_c could affect the overall microbial community, possibly by (1) manipulating the production of other metabolites such as antibiotics[53], which could inhibit the growth of pathogens and reconstitution of the protective gut microbiota; (2) regulating (stimulating or inhibiting) the quorum sensing pathway[54,55], thus affecting the communication between microbes.

Using MaAsLin2[56] (microbiome multivariable associations with linear models), we found variable number of significantly differentially abundant species (FDR-adjusted p value < 0.05) in groups of vancomycin (1 μg/mL and 10 μg/mL), BF_280_c (10 μg/mL), and BF_398_c (0.1 μg/mL and 10 μg/mL) when compared with blank group (Supplementary Fig. 41, Supplementary Data 11). At the same concentration of 10 μg/mL, vancomycin affected more species ($n = 73$) than the other two AIPs ($n = 16$ and 40). Nevertheless, most differentially abundant species were affected in a similar pattern by vancomycin or AIPs at 10 μg/mL concentration (Supplementary Fig. 41c, Supplementary Data 11), some of which are closely associated with IBD. As depicted in Fig. 5e, they inhibited the growth of pathobiont species from genera *Paramuribaculum*[57] (*Paramuribaculum intestinale*), and *Oscillospiraceae*[58,59] (*Oscillospiraceae bacterium*), which are known to



Pathogen	AIPs (μg/mL) (lowest concentration for significantly inhibit biofilm formation)				
	BF_94_et_free	BF_280_t	BF_280_c	BF_398_c	BF_598_c
Staphylococcus aureus ATCC 43300	25	1.56	1.56	1.56	25
Listeria monocytogenes ATCC 19115	50	>100	25	25	50
Peptostreptococcus stomatis DSM17678	25	>50	>50	25	>50
Streptococcus gallolyticus subsp. Gallolyticus DSM 16831	12.5	> 50	> 50	> 50	25
Candida albicans ATCC 10231	100	>100	>100	>100	25

be enriched in patients with IBD. Another species, *Heminiphilus faecis*[60,61], which could produce IBD pathogenic factors, was also found to be reduced. Additionally, both AIPs decreased the abundance of pathobiont species from the *Prevotella* genus[62,63] (*Prevotella sp MGM1*), which are linked to increased susceptibility to mucosal inflammation in IBD. These findings implied that AIPs could modulate the fecal microbial composition and reduce potential pathogens in IBD, justifying their possibly protective effects on human health.

Discussion

Extensive studies have highlighted the profound functional impact of the host-microbiome interaction on human health, as well as the

Fig. 4 | Differentially transcribed RiPP families in multi-disease case-control cohorts. a The bar plot displays the number of RiPP precursor families that are significantly enriched in the healthy group (top) or the disease group (bottom). Statistical significance was determined using criteria of |log2 fold change | ≥1 and adjusted p values ≤ 0.05. **b** The intersection of differentially transcribed RiPP precursor families that are enriched in healthy groups compared to multiple disease groups. The top bar plot and connecting lines represent the differential RiPP precursor families identified in the corresponding disease case-control cohorts. Precursor families are highlighted in different colors based on their RiPP classes.*Families chosen to tested bioactivity in this study. **c** The workflow for predicting the mature AIP uses the reported AIP-I from *Staphylococcus aureus* as an illustrative example (Method). **d** Chemically synthesized AIPs with antibiofilm or anti-inflammatory activity in this study. Here, the compound name consists of the family name and a specific suffix denoting its characteristics. For example, "et_L1" indicates that the AIP has an exotail with a length of 1 amino acid, "et_free" signifies an AIP without an exotail, "c" denotes cyclopeptides, and "t" represents exotail-free AIPs with a thiolactone group. **e, f** $n = 3$ biological samples for each experiment, 2 independent experiments. Significance was determined using one-way ANOVA test. **e** Anti-biofilm activity of AIPs against pathogenic biofilm by pathogens. $n = 3$ biologically independent samples. **f** Anti-inflammatory activity of BF_94_et_L1 in RAW264.7 cell lines. Bars represent mean ± standard error. For all p values, $p < 0.05$ mean significant difference compared with the control group. Exact adjusted p-values from left to right are: <1e−15, 1.248e−5, 1.866e−5, and 2.706e−12 in TNFα (upper) and <1e−15, 0.0089, 0.0006, and 6.055e−7 in IL-6 (lower).

etiology and progression of various diseases[64]. These research findings underscore the crucial role that human microbiota plays in health maintenance by producing a wide variety of unique metabolites[65]. These metabolites possess various biological activities and mediate interactions between microbes or between microbes and the host. Yet, a significant knowledge gap persists in understanding microbial ribosomal peptides' diversity, abundance, prevalence, and role in maintaining microbial and human homeostasis. In this study, we investigated the biosynthetic potential of RiPPs within the human-associated microbiota, demonstrating their remarkable chemical diversity, novelty, and potential bioactivity. However, it is important to recognize that the scope of RiPP discovery is constrained by the availability of microbial genome datasets, predominantly from the gut microbiome, particularly bacteria. Concurrently, while utilizing a combination of deep learning-based and rule-based methods can help strike a balance between precision and novelty in studying RiPP biosynthesis, this strategy may result in a somewhat constrained depiction of the entire RiPP family from the human microbiome, as part of actual RiPPs may be excluded from the analysis. Nonetheless, our study thoroughly uncovers the profile of RiPP biosynthesis in the human microbiome and compiles an atlas of RiPPs for future research on innovative antimicrobial strategies and therapeutic interventions targeting microbiome[66,67].

The ecological dynamics of the human microbiome are greatly shaped by specialized microbial metabolites that play protective roles and can potentially impact human health. Recent analyses of the human microbiome have revealed hidden potential for antibiotics, with RiPPs being the most commonly predicted compounds across various microbial environments within the body[13]. Concurrently, peptides with potentially harmful effects were also identified, including a lasso peptide from the oral bacterium *Rothia aeria* and a potentially toxic lantibiotic from translocating *Streptococcus*[68,69]. Our study observed that RiPPs are enriched in the human gut and oral microbiome, most of which are niche-specific bioactive molecules and actively transcribed in healthy human gut microbiomes. In our analysis of activity predictions, over 70% of RiPPs were predicted to be antibacterial, which should be interpreted with caution. DeepBGC was trained primarily on antimicrobial RiPPs from MIBiG, inevitably leading to a potential bias towards antibacterial prediction. Additionally, the RiPPs analyzed in this study might fall outside the algorithm's applicability domain. Recent research has revealed that common RiPP classes possess more diverse functions[16–20] than previously recognized, including roles in microbial communications that are not included in DeepBGC's training data. Therefore, the functions of RiPPs might be more diverse than our predictions suggest. RiPPs with protective roles can mediate microbe-microbe interaction (e.g., antibacterial activities[13], quorum-sensing or quorum-quenching[20]), or microbiome-host interaction (e.g., immunomodulatory activities[16–19], and cytotoxicity[70]). These interactions could influence the balance of the human-microbe holobiont and potentially substantially impact human health. We believe that these protective RiPPs, sourced from the human microbiome, represent innovative antimicrobial tactics and therapeutic interventions that are yet to be fully explored.

In microbe-microbe interactions, quorum-sensing is used for social coordination, often via peptide-mediated mechanisms in Gram-positive bacteria. RiPPs, particularly AIPs, can act as signaling molecules, regulate their production, and respond to quorum-sensing signals[10]. Our study identified six AIP families enriched in healthy microbiomes, most exhibiting significant antibiofilm activity against IBD- or CRC-related pathogens. Polymicrobial biofilms, particularly in IBD and CRC[71], complicate treatment as antimicrobials need to target all biofilm pathogens. The combination of AIPs with antibiotics, specifically AIP-I, has shown potential in enhancing biofilm infection treatment[72,73], as it can trigger MRSA biofilm dispersal and increase the susceptibility of detached cells to antibiotics. Our study revealed five AIPs with anti-pathogenic biofilm activity and found that two anti-biofilm AIPs could adjust the gut microbial community in an ex vivo assay and hereby reduced pathogenic species linked to IBD, suggesting an alternative treatment option. These AIPs may assist in microbial communication through quorum-sensing or indirectly influencing the microbiome by controlling the production of antibacterial metabolites[53]. Considering their increased prevalence in healthy individuals compared to patients with IBD or CRC, our findings suggest that they could provide protective roles by inhibiting the growth or biofilm formation of pathogens[74]. While exploring AIPs' impact on harmful biofilms and microbiome homeostasis could open up new therapeutic avenues, their exact mechanisms remain to be understood, and their bioactivity needs to be confirmed in vivo animal models.

Our study provides a comprehensive analysis of the biosynthetic landscape of RiPPs in the largely unexplored human microbiome, using (meta)genome mining and extensive omics analysis. In a proof of principle study, we linked RiPP profiles with various human diseases and pinpointed several RiPP candidates that may potentially impact human health. However, several important caveats should be considered when interpreting the biological significance of these differentially expressed RiPPs between healthy and diseased states. For example, the observed differences in RiPP diversity between health/disease states could potentially be explained by underlying variations in microbial species composition. We also need to acknowledge that observed associations could be due to co-occurrence with other functional genes within the same species, rather than direct effects of the RiPPs themselves. To establish causality and biological relevance, future studies should systematically investigate the interplay between RiPP diversity, microbial community dynamics, and disease progression using both observational and mechanistic approaches.

The success of our approach in elucidating and optimizing AIP analogues underscores the transformative potential of combining computational tools like AlphaFold with biosynthetic pathway enzymology. While traditional RiPP discovery pipelines often stall when the heterologous expression fails, our methodology demonstrates that structural prediction and chemical synthesis can circumvent these

https://doi.org/10.1038/s41467-025-58280-w

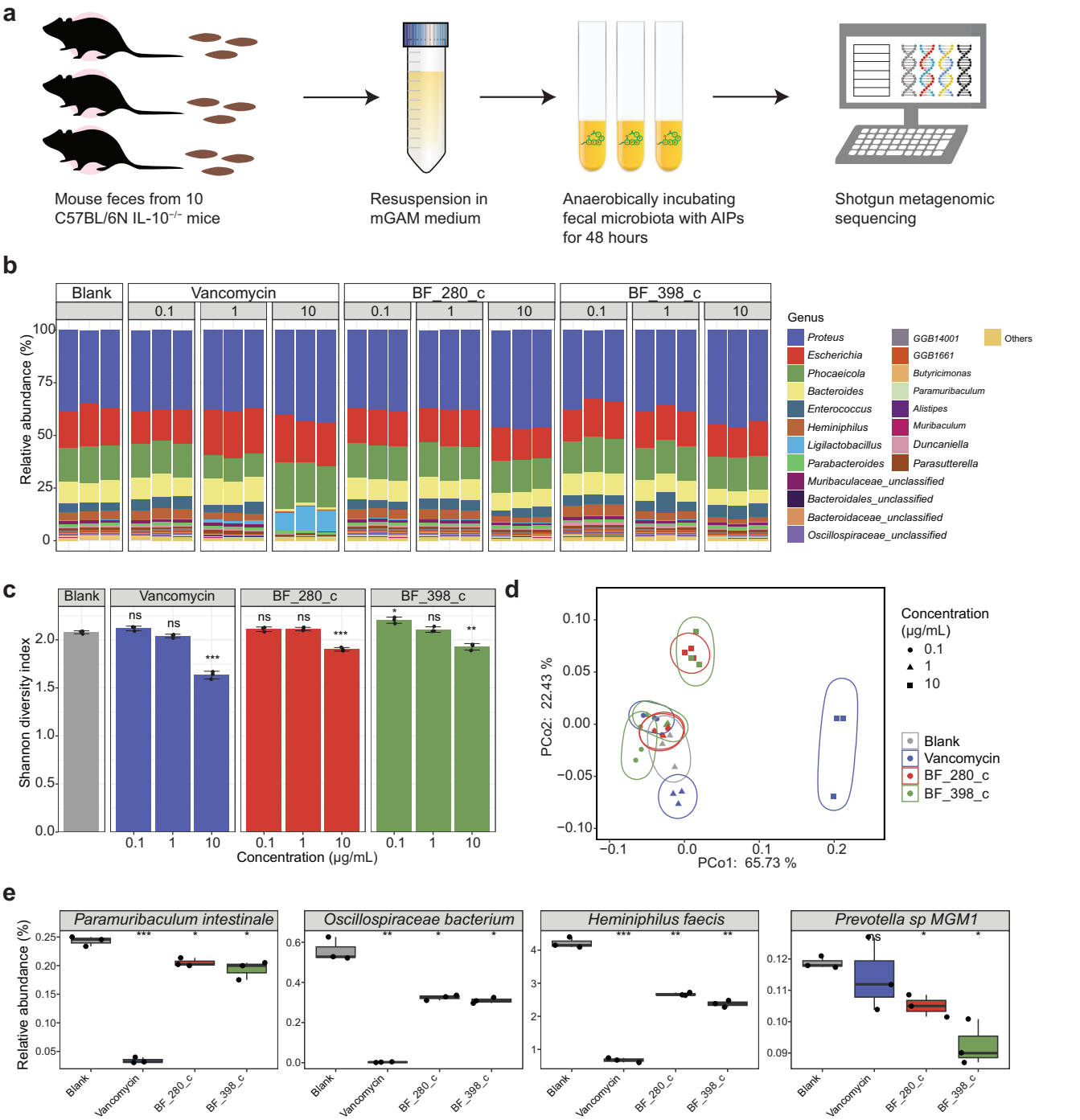


Fig. 5 | AIPs regulate IBD mouse fecal-derived ex vitro microbial community. **a** Graphical depiction of the ex vivo experimental setup. **b** Relative abundance of microbial genera after 48 h of treatment with vancomycin, BF_280_c, and BF_398_c at concentrations of 0.1, 1, and 10 μg/mL. DMSO was used as a blank control. Only 20 abundant species are displayed. **c** Bar plots show the alpha diversity of the microbial community at the species level, as represented by the Shannon diversity index. Data ($n = 3$) are mean ± standard deviation. Significances between treatment groups and blank group were indicated by using two-sided Welch's *t*-test. The exact *p* values are as follows (order as 0.1, 1, 10 μg/mL): vancomycin, 0.082, 0.054, 0.85e−3; BF_280_c, 0.13, 0.067, 0.25e-3; BF_398_c, 0.010, 0.23, 0.76e-2. Three biological replicates ($n = 3$) were included for each group. **d** Principal coordinate analysis (PCoA) of microbial community based on the Bray-Curtis dissimilarity at the species level. The enclosing ellipses are estimated using the Khachiyan algorithm by R function "geom_mark_ellipse", representing the distinct clustering of groups. **e** Relative abundance of four representative taxa. All box plots include center lines representing the median, box limits representing upper and lower quartiles, whiskers representing the 1.5x interquartile range, and points representing outliers. The exact *p* values are as follows (order as vancomycin, BF_280_c, BF_398_c): *Paramuribaculum intestinale*, 0.41e−3, 0.036, 0.039; *Oscillospiraceae bacterium*, 0.0020, 0.021, 0.013; *Heminiphilus faecis*, 0.35e−3, 0.0044, 0.0076; *Prevotella sp MGM1*, 0.55, 0.036, 0.027. MaAsLin2, a tool relying on general linear models to find multivariable association, was used to compare the species between vancomycin/AIP and the control group. The Benjamini-Hochberg ("BH") method was adopted to adjust *p* values for multiple comparisons. Three biological replicates ($n = 3$) were included for each group. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; ns, not significant.

limitations. However, these strategies may need reconsidering when applied to complex RiPP classes. For instance, lanthipeptide biosynthesis often involves multiple enzymes with unclear recognition rules, making structure determination difficult without heterologous expression. Lassopeptides' complex structures require precise modeling of peptide folding, which is a substantial challenge for prediction accuracy. These limitations reveal a persistent challenge in RiPP bioinformatics: the development of enzyme-aware predictive tools that holistically incorporate precursor peptide sequences and the biochemical context of modifying enzymes—including their substrate specificities, catalytic mechanisms, and reaction dynamics. Our analysis of differentially expressed RiPPs in healthy and disease states (Fig. 4) revealed numerous uncharacterized candidates, including putative lanthipeptides and rSAM-modified peptides. While their experimental characterization fell outside this study's scope, these candidates still warrant further investigation. Advances in predictive tools (e.g., AlphaFold 3's improvement in handling of multimeric complexes) and growing structural databases (e.g., RiPP-PRISM, RODEO) will enhance our ability to model complex RiPPs.

Despite certain limitations, our study offers valuable insights into the diversity and potential functions of RiPPs in the human microbiome. We also identified protective RiPPs that can combat pathogenic biofilms and possess the ability to maintain the balance of microbial communities, thus offering alternative antimicrobial tactics or therapeutic interventions that target the microbiome.

Methods

Biosynthetic gene cluster analysis and RiPP precursor identification from human microbiome reference genomes

We retrieved reference genomes of body-wide human microbes from two available datasets: 289,232 genomes from Unified Human Gastrointestinal Genome dataset (UHGG, v2.0[37,38]), and 17,249 genomes from CIBIO[36] (Supplementary Data 12). To normalize taxonomic annotations for all the genomes, we re-annotated all of them using GTDB-Tk[75] (v2.0.0) against GTDB (rev207 version). All genomes were analyzed by antiSMASH 6.0[26] for BGC detection using default parameters. To identify the RiPP precursor peptide, all open reading frames (ORFs) in each genome were annotated by prodigal-short[76]. Subsequently, small ORFs (≤150 amino acids) containing a start and stop codon and ribosome-binding site motifs were subjected to TrRiPP[25] and DeepRiPP[24] for RiPP precursor identification with default parameters. Notably, DeepRiPP can identify RiPP products even when the precursor genes are distant from the tailoring enzymes, while TrRiPP can detect RiPPs from highly fragmented metagenomes. Combining these tools can effectively identify more potential RiPP precursor peptides in metagenome-assembled genomes, which are predominant in our collection.

Analysis of genomic context of predicted RiPP precursor peptides

For all identified RiPP precursors, we first checked whether they were within the region of antiSMASH-defined BGCs. For those precursors outside antiSMASH-defined BGC region, we defined 10 genes of the precursor upstream and downstream as genomic neighborhoods. We used two methods to examine the characteristics of these genomic neighborhoods associated with RiPP. First, we analyzed the protein domain of precursors and their 10 flanking genes downstream and upstream by a domain-based approach: RPS-BLAST. A domain was significantly assigned with a default CDD e-value threshold (0.01), and the protein aligns to at least 80% of the PSSM's length. The assigned domain belonging to a dataset of known RiPP precursor domains and RiPP-related biosynthetic enzyme domains was noted as RiPP-related domains (Supplementary Tables 1 and 2). This database was collected and expanded based on the hmm-rule-parser of antiSMASH and previous studies[9,21,26,77]. For the remaining precursor, the domains were paired with a broader dataset of biosynthetic enzyme domains, which were used in the decRiPPter[39] pipeline for exploring uncharacterized biosynthetic enzyme candidates for RiPP.

Chemical space comparison of RiPP precursor peptides

Known RiPP precursor peptides were collected from The Minimum Information about a Biosynthetic Gene cluster database (MIBiG 3.0)[78]. Extended-connectivity fingerprints (ECFPs) are used to predict and gain insight into RiPPs precursors' chemical diversity[43]. Then ECFP6 fingerprints for each unique RiPP precursor were compared with each other to generate Tanimoto coefficient matrices[79] and visualize the chemical space by uniform manifold approximation and projection (UMAP). Specifically, The ECFP6 chemical fingerprints of the domain representatives were calculated using The Chemistry Development Kit[80]. Subsequently, we generated a Jaccard distance matrix based on the chemical fingerprints and performed dimension reduction using densMAP[81] with n_neighbors= 15, implemented in the UMAP Python package. Furthermore, we calculated the pairwise Tanimoto coefficient of each precursor using NumPy and scikit-learn. The average Tanimoto coefficient was obtained by averaging all of these coefficients. Next, core precursor sequences predicted by the cleavage predictions module of DeepRiPP were used to explore the chemical diversity of nine RiPP classes. Pairwise Tanimoto coefficient was further adopted for measuring the similarity between precursors. The precursor with the highest median within-class Tanimoto coefficients was chosen as representative structures to generate Tanimoto coefficient matrices for intra-classes and further calculate the diversity for intra-classes. Tanimoto score ranges from 0 to 1. The higher Tanimoto score means higher chemical similarity between two precursors or two subclasses.

Novelty examination of RiPP precursor families

To trade off the novelty and accuracy, 423831 RiPP precursors that were either within RiPP BGCs region[41] identified by enzymes-oriented approach or identified by both two precursors-centric approaches were retained for further analyses. These precursor peptides were grouped by MMseqs2 (13.45111)[82] with the following parameters: easy-cluster clusterRes tmp --min-seq-id 0.5 --single-step-clustering --cluster-mode 2 --cov-mode 2 -c 0.95. The precursors within a family are more likely to share a similar function[83]. Each RiPP precursor family was classified based on precursor homology and genomic context (Supplementary Information). The novelty of RiPP families was further classified into three types: (1) "classic RiPP families", which are identified by having precursors that exhibit similarity to known RiPP precursors and are located in a genomic context associated with typical RiPP biosynthetic gene clusters. (2) "Uncharacterized RiPP families", which consist of uncharacterized precursors and/or genomic neighborhoods containing uncharacterized biosynthetic genes. (3) "Others".

Bioactivity prediction of RiPP families

We utilized the DeepBGC (https://github.com/Merck/deepbgc)[27] tool to identify the potential bioactivity of RiPP precursors. This approach allowed us to assess and classify the potential bioactivity of RiPPs based on the predicted functional genes within their genomic neighborhoods. Our approach involved collecting gene regions within 10 genes upstream and downstream of RiPP precursor genes, which we referred to as genomic neighborhoods. We then utilized DeepBGC to predict the potential function of each RiPP. This tool could account for the four most common compound activity classes: antibacterial, cytotoxic, inhibitor, and antifungal. Precursors exhibiting multiple predicted activities were labeled as "multiple", while the activity of each RiPP family was determined by considering the activities of over 50% of the precursors within the family. It is important to acknowledge that the training set's bias toward antimicrobial RiPPs from MIBiG may limit the prediction accuracy of DeepBGC when analyzing RiPPs from

the human microbiome, as these peptides likely possess a broader range of uncharacterized biological functions[16–20] besides antimicrobial activity.

Analysis of metagenomic and metatranscriptomic data

The metagenomic and metatranscriptomic samples were downloaded from sequence read archive (SRA) of the NCBI (Supplementary Data 6). Raw metagenome and metatranscriptome sequencing reads were quality filtered using *bbduk.sh* with the following parameters: qtrim=rl ktrim=r mink=11 trimq=10 minlen=40. The resulting metatranscriptomic reads were subjected to *bbmap.sh* for removing reads derived from ribosomal RNAs. To remove the human host contamination, the high-quality metagenomic and metatranscriptomic sequencing reads were further searched against the human reference genome (GRCh38.p13) from NCBI, and unmapped reads were mapped to microbe reference genomes using BWA (0.7.17-r1188)[84] with default parameters. The reads mapped to RiPP precursor gene were counted by featurecounts[85] with the following parameters: -f -t CDS -M -O -g transcript_id -F GTF -s 0 -p --fracOverlap 0.25 -Q 10 -primary. In metatranscriptomic datasets, the count file with the absolute abundance for each family was imported, and differential gene abundance between healthy and diseased subjects was normalized and analyzed by implementing DESeq2 pipeline in R. Of note, the abundance of each family was calculated by the sum of the abundances of all genes in the family. Additionally, MetaPhlAn 4 (version 4.1.1)[86] was used to compute the relative abundance of microbial species. In addition, accumulation curves, alpha-diversity, beta diversity and PERMANOVA were performed using the R package *vegan*. Beta diversity was performed to quantify the relative abundance differences in the overall composition of RiPP precursor families between the disease and the control groups. PERMANOVA was performed to show the encoding and expression profile differences of RiPP precursor families between disease and control groups. Specifically, families with a prevalence ≥ 5% in a cohort were subjected to further analysis (Beta diversity, differential analysis, and classifier) for investigating potential causality in human disease. In metagenomic datasets, we calculated the significance of prevalence using a two-sided Fisher's exact test, $p < 0.05$. BiG-SCAPE (version 1.1.5)[87] was used to visualize the genomic context of differential RiPP precursors families. Besides, only 50 members with formative genomic context (larger gene sizes) in RiPP families with larger members (≥ 50 genes) were chosen for analysis. Each representation biosynthetic gene cluster was chosen to show the conserved domain and products in each family. Multiple sequence alignment of all precursor families was conducted using MAFFT (v7.508)[88], followed by Jalview (version: 2.11.4.1) for visualization.

Chemical synthesis of AIPs

To begin, we initiated the calculation and analysis of the precursor-AgrB complex (Supplementary Table 3). The calculation was performed using the following command: colabfold_batch --amber --templates --num-recycle 3 --use-gpu-relax --model-type alphafold2_multimer_v3 input_path output_path. The resulting structures were visualized using PyMol (version 2.5.3). During our analysis of the reported AIP precursor and paired AgrB, we observed that the binding of the Precursor-AgrB complex can be partially buried. In other words, the core peptide of the precursor should be fully accommodated within the catalytic pocket of AgrB. Based on this observation, we manually examined the predicted complexes. Specifically, we focused on two aspects: 1. We inspected whether the conserved C-terminal core region is captured and effectively buried within the catalytic pocket of AgrB. By conducting this analysis, we aimed to gain insights into the interaction between the precursor and AgrB, shedding light on the potential core peptide of AIP precursors. Next, we aimed to comprehensively assess the conservation patterns of precursor sequences within each AIP family, incorporating both our study data and relevant sequences available in public databases. Thus, we first collected non-redundant precursor sequences from our study. Additionally, we included similar sequences (> 50% similarity) obtained from NCBI as additional sources of AIPs found in nature. Next, Logo sequences for each clustered short peptide family were generated by makelogo.py, which is available at https://github.com/yxllab-hku/cluser_to_logo/blob/main/makelogo.py. With the bioinformatic prediction, we deduced the potential mature chemicals of six AIP families and chemically synthesized 9 potential AIPs (Supplementary Information).

Antibiofilm assay

Biofilm formation was assessed using the crystal violet method. The strains listed in Supplementary Data 8 were employed as the tested strains. To initiate the experiment, the inocula from overnight cultures were diluted at a ratio of 1:100 using the corresponding culture medium. To perform dose-response assays, the chosen peptide was prepared as a stock solution and serially diluted in DMSO. Following this, 1 μL of each concentration was carefully dispensed into the corresponding wells. Subsequently, 99 μL of the bacterial culture, diluted accordingly, was added to the respective wells. Additionally, vehicle controls consisting of 1 μL of DMSO were included to establish the baseline biofilm formation of the strains. The microtiter plate was then incubated under appropriate conditions. Following the incubation period, the biofilm was gently washed with PBS to remove any unattached cells and then allowed to air dry. Next, 100 μL of a 0.1% crystal violet solution was added to each well to stain the biofilm. The plate was further incubated at room temperature for 15 minutes to ensure proper staining. After the incubation, each well was washed three times with water to completely remove the dye. The plate was then inverted and allowed to air-dry at ambient temperature throughout the night. To quantify the biofilm formation, 100 μL of a 30% acetic acid solution was added to each well to dissolve the crystal violet. Following an additional incubation period of 10-15 minutes at room temperature, the absorbance was measured at 550 nm. Three replicated wells were used for each group to ensure accuracy and reproducibility.

Anti-inflammatory activity by ELISA experiment

The RAW264.7 cells were seeded at a density of 15×10^4 cells/well in 24-well plates and cultured at 37 °C and 5% CO2 overnight. Cells were treated with AIPs (1 μg/mL and 0.1 μg/mL) or 25 μg/mL of Dexamethasone for 3 hours and then stimulated with 1 μg lipopolysaccharides (LPS) for 24 hours. The secretion of TNFα and IL6 was measured according to the ELISA kit (Elabscience). ELISA methods were applied according to the manufacturer's instructions without any modifications by using mouse IL-6 (Interleukin 6) ELISA Kit (E-EL-M0044), mouse TNF-α (tumor necrosis factor alpha) ELISA Kit (E-EL-M3063). Optical densities were read on a plate reader set at 450 nm a microplate reader (BioTerk, Winooski, VT, USA). The concentration of each parameter in the samples was calculated from the standard curve, multiplied by the dilution factor and was expressed as mean ± standard error of the mean (SEM).

Ex vivo screening study

Mice. Mouse studies were performed in accordance with all relevant ethical regulations and were approved by the Ethics Committee of the Animal Experimental Center of ZhuJiang Hospital, Southern Medical University (LAEC-2022-059). IL-10-deficient (IL-$10^{-/-}$) male mice with C57BL/6 N background were purchased from Cyagen Biosciences (Guangzhou, China). All the mice were housed in specific pathogen free (SPF) conditions with a 12 hours light/dark cycle, and were provided sterilized water and food ad libitum (temperature 23 ± 2°C, humidity 45 ± 5%). IL-10-deficient mice were used to model spontaneous chronic colitis, which closely mimics human inflammatory bowel disease (IBD)[89].

Feces collection. Fecal samples were collected from 10 IL-10-deficient C57BL/6N male mice (16 weeks old), with an average of 7-8 droppings per mouse. The collected pellets were combined and resuspended in 65 mL of a rich medium mGAM. Following resuspension, the tubes were gently centrifuged at 112 g for 2 minutes, and the supernatants were then retained for further assay.

Two selected AIPs were pre-dissolved in DMSO to final concentrations of 0.1 g/mL, 1 g/mL, and 10 g/mL. Subsequently, 1 μL of AIPs with varying concentrations and 49 μL of fecal supernatant were inoculated into 950 μL of mGAM media, resulting in a final volume of 1 mL. The mixtures were then anaerobically incubated at 37 °C for 48 hours. Following the incubation period, the bacterial growth media were centrifuged at 16,099 g for 10 minutes. The supernatants were carefully removed, and the resulting pellets were washed twice with 1 mL of PBS before being collected for DNA extraction. The pelleted samples were extracted using the QIAamp® DNA Micro Kit (Qiagen) according to the manufacturer's instructions. Finally, all DNA samples were prepared for shotgun metagenomics sequencing.

Metagenomic sequencing. All DNA samples were submitted to Novogene for shotgun metagenomics sequencing using a 150 bp paired-end protocol. Initially, 1 μg of DNA per sample was fragmented via sonication to achieve a size of 350 bp. Subsequently, the fragmented DNA underwent end-polishing, A-tailing, and ligation with a full-length adaptor for library construction. The constructed libraries were assessed for quality and quantity using Qubit and real-time PCR for quantification, as well as a bioanalyzer for size distribution analysis. The quantified libraries were then pooled and sequenced on the Illumina NovaSeq X Plus platform, generating approximately 10.7 Gb of data per sample.

Metagenomic analysis. Raw reads from metagenomic sequencing were processed using Fastp v0.21.1 for quality control, with parameter of "--detect_adapter_for_pe -l 50 -5 3 -3 3". High-quality metagenomic sequencing reads were further subjected to KneadData (https://github.com/biobakery/kneaddata) for detecting and removing reads belonging to the human genome, though searching against the mouse reference genome (GRCm39) from GENCODE[90]. MetaPhlAn (version 4.1.1)[86] was used to generate taxonomic profiles of metagenomes.

Alpha diversity, as measured by Shannon diversity, and beta diversity, evaluated through Bray-Curtis dissimilarity, were computed for species-level taxonomic profiles utilizing the Vegan package in R[91]. Differential abundance analysis between groups was conducted using MaAsLin2 (version 1.12.0)[56] (microbiome multivariable associations with linear models). Species with a false discovery rate (FDR)-adjusted *p*-value of less than 0.05 were deemed significantly different between the two groups.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw read sequences of the shotgun metagenomic sequences were deposited at Sequence Read Archive (SRA, NCBI) under accession of PRJNA1166984. The BGC and RiPP precursors identified in this study are available in Supplementary dataset 1-3, which have been deposited in Zenodo (https://doi.org/10.5281/zenodo.14994355)[92]. Other data supporting the findings of this study are available in supplementary information, supplementary datasets, and source data files. The raw data for omics data analysis were collected from NCBI datasets through the accession numbers provided in Supplementary Data 6. Reference genomes of body-wide human microbes from two available datasets: 289232 genomes from Unified Human Gastrointestinal Genome dataset (UHGG, v2.0, https://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/human-gut/v2.0/), and 17249 genomes from CIBIO. The genomes used in this study are provided in Supplementary Data 12. Source data are provided with this paper.

Code availability

Codes related to the analyses in this study are available at https://github.com/ZHANGJianArya/RiPPs_human-microbiome. The source code used in the paper has also been assigned a citable DOI through Zenodo (https://doi.org/10.5281/zenodo.14866373)[93].

References

1. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med* **24**, 392–400 (2018).
2. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol* **19**, 55–71 (2021).
3. Wang, L., Ravichandran, V., Yin, Y., Yin, J. & Zhang, Y. Natural Products from Mammalian Gut Microbiota. *Trends Biotechnol.* **37**, 492–504 (2019).
4. Rebuffat, S. Ribosomally synthesized peptides, foreground players in microbial interactions: recent developments and unanswered questions. *Nat. Prod. Rep.* **39**, 273–310 (2022).
5. Wilson, M. R., Zha, L. & Balskus, E. P. Natural product discovery from the human microbiome. *J. Biol. Chem.* **292**, 8546–8552 (2017).
6. Silpe, J. E. & Balskus, E. P. Deciphering Human Microbiota-Host Chemical Interactions. *ACS Cent. Sci.* **7**, 20–29 (2021).
7. Zhang, Y., Li, P., Ma, Y., Wang, J. & Chen, Y. Artificial intelligence accelerates the mining of bioactive small molecules from human microbiome. *Clin. Transl. Med* **12**, e1011 (2022).
8. Cully, M. Microbiome therapeutics go small molecule. *Nat. Rev. Drug Discov.* **18**, 569–572 (2019).
9. Montalban-Lopez, M. et al. New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **38**, 130–239 (2021).
10. Li, Y. & Rebuffat, S. The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology. *J. Biol. Chem.* **295**, 34–54 (2020).
11. Sugrue, I., Ross, R. P. & Hill, C. Bacteriocin diversity, function, discovery and application as antimicrobials. *Nat. Rev. Microbiol.* **22**, 556–571 (2024).
12. Zhong, G., Wang, Z. J., Yan, F., Zhang, Y. & Huo, L. Recent Advances in Discovery, Bioengineering, and Bioactivity-Evaluation of Ribosomally Synthesized and Post-translationally Modified Peptides. *ACS Bio Med Chem. Au* **3**, 1–31 (2023).
13. Mousa, W. K., Athar, B., Merwin, N. J. & Magarvey, N. A. Antibiotics and specialized metabolites from the human microbiota. *Nat. Prod. Rep.* **34**, 1302–1331 (2017).
14. Donia, M. S. et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).
15. Zhang, Z. J. et al. Activity of Gut-Derived Nisin-like Lantibiotics against Human Gut Pathogens and Commensals. *ACS Chem. Biol.* **19**, 357–369 (2024).
16. Li, J. et al. Tonsillar Microbiome-Derived Lantibiotics Induce Structural Changes of IL-6 and IL-21 Receptors and Modulate Host Immunity. *Adv. Sci. (Weinh.)* **9**, e2202706 (2022).
17. Wang, S. et al. Enhancement of Macrophage Function by the Antimicrobial Peptide Sublancin Protects Mice from Methicillin-Resistant Staphylococcus aureus. *J. Immunol. Res* **2019**, 3979352 (2019).
18. Li, J., Chen, J., Yang, G. & Tao, L. Sublancin protects against methicillin-resistant Staphylococcus aureus infection by the combined modulation of innate immune response and microbiota. *Peptides* **141**, 170533 (2021).
19. Barbour, A. et al. Discovery of phosphorylated lantibiotics with proimmune activity that regulate the oral microbiome. *Proc. Natl Acad. Sci. USA* **120**, e2219392120 (2023).

20. Williams, M. R. et al. Quorum sensing between bacterial species on the skin protects against epidermal injury in atopic dermatitis. *Sci. Transl. Med.* **11**, eaat8329 (2019).
21. Arnison, P. G. et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
22. Mullowney, M. W. et al. Artificial intelligence for natural product drug discovery. *Nat. Rev. Drug Discov.* **22**, 895–916 (2023).
23. Zhong, Z., He, B., Li, J. & Li, Y. X. Challenges and advances in genome mining of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Synth. Syst. Biotechnol.* **5**, 155–172 (2020).
24. Merwin, N. J. et al. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl Acad. Sci. USA* **117**, 371–380 (2020).
25. Gao, Y., Zhong, Z., Zhang, D., Zhang, J. & Li, Y. X. Exploring the roles of ribosomal peptides in prokaryote-phage interactions through deep learning-enabled metagenome mining. *Microbiome* **12**, 94 (2024).
26. Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**, W29–W35 (2021).
27. Hannigan, G. D. et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res* **47**, e110 (2019).
28. Balty, C. et al. Ruminococcin C, an anti-clostridial sactipeptide produced by a prominent member of the human microbiota Ruminococcus gnavus. *J. Biol. Chem.* **294**, 14512–14525 (2019).
29. Bushin, L. B., Covington, B. C., Rued, B. E., Federle, M. J. & Seyedsayamdost, M. R. Discovery and Biosynthesis of Streptosactin, a Sactipeptide with an Alternative Topology Encoded by Commensal Bacteria in the Human Microbiome. *J. Am. Chem. Soc.* **142**, 16265–16275 (2020).
30. Lin, X. et al. The genomic landscape of reference genomes of cultivated human gut bacteria. *Nat. Commun.* **14**, 1663 (2023).
31. Aleti, G. et al. Identification of the Bacterial Biosynthetic Gene Clusters of the Oral Microbiome Illuminates the Unexplored Social Language of Bacteria during Health and Disease. *mBio* **10**, e00321-19 (2019).
32. Zhang, D. et al. A systematically biosynthetic investigation of lactic acid bacteria reveals diverse antagonistic bacteriocins that potentially shape the human microbiome. *Microbiome* **11**, 91 (2023).
33. Chu, J. et al. Synthetic-Bioinformatic Natural Product Antibiotics with Diverse Modes of Action. *J. Am. Chem. Soc.* **142**, 14158–14168 (2020).
34. Chu, J., Vila-Farres, X. & Brady, S. F. Bioactive Synthetic-Bioinformatic Natural Product Cyclic Peptides Inspired by Non-ribosomal Peptide Synthetase Gene Clusters from the Human Microbiome. *J. Am. Chem. Soc.* **141**, 15737–15741 (2019).
35. Nelson, S. & Parkinson, E. I. Synthetic-bioinformatic natural product-inspired peptides. *Nat. Prod. Rep.* **42**, 50–66 (2025).
36. Pasolli, E. et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649–662 e620 (2019).
37. Almeida, A. et al. A unified catalog of 204938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* **39**, 105–114 (2021).
38. Gurbich, T. A. et al. MGnify Genomes: A Resource for Biome-specific Microbial Genome Catalogues. *J. Mol. Biol.* **435**, 168016 (2023).
39. Kloosterman, A. M. et al. Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLoS Biol.* **18**, e3001026 (2020).
40. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
41. Blin, K. et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81–W87 (2019).
42. Gavriilidou, A. et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol* **7**, 726–735 (2022).
43. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model* **50**, 742–754 (2010).
44. Human Microbiome Project, C. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
45. Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol* **3**, 356–366 (2018).
46. Duan, Y. et al. Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease. *Nature* **575**, 505–511 (2019).
47. Zhao, A. et al. Reconstitution of the S. aureus agr quorum sensing pathway reveals a direct role for the integral membrane protease MroQ in pheromone biosynthesis. *Proc. Natl Acad. Sci. USA* **119**, e2202661119 (2022).
48. Gless, B. H. et al. Rearrangement of Thiodepsipeptides by S -> N Acyl Shift Delivers Homodetic Autoinducing Peptides. *J. Am. Chem. Soc.* **143**, 10514–10518 (2021).
49. Molloy, E. M. et al. Enzyme-Primed Native Chemical Ligation Produces Autoinducing Cyclopeptides in Clostridia. *Angew. Chem. Int Ed. Engl.* **60**, 10670–10679 (2021).
50. Tytgat, H. L. P., Nobrega, F. L., van der Oost, J. & de Vos, W. M. Bowel Biofilms: Tipping Points between a Healthy and Compromised Gut? *Trends Microbiol* **27**, 17–25 (2019).
51. Baumgartner, M. et al. Mucosal Biofilms Are an Endoscopic Feature of Irritable Bowel Syndrome and Ulcerative Colitis. *Gastroenterology* **161**, 1245–1256 e1220 (2021).
52. Ricciuto, A. et al. Oral vancomycin is associated with improved inflammatory bowel disease clinical outcomes in primary sclerosing cholangitis-associated inflammatory bowel disease (PSC-IBD): A matched analysis from the Paediatric PSC Consortium. *Aliment Pharm. Ther.* **59**, 1236–1247 (2024).
53. Duerkop, B. A. et al. Quorum-sensing control of antibiotic synthesis in Burkholderia thailandensis. *J. Bacteriol.* **191**, 3909–3918 (2009).
54. Wu, L. & Luo, Y. Bacterial Quorum-Sensing Systems and Their Role in Intestinal Bacteria-Host Crosstalk. *Front Microbiol* **12**, 611413 (2021).
55. Piewngam, P. et al. Pathogen elimination by probiotic Bacillus via signalling interference. *Nature* **562**, 532–537 (2018).
56. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol.* **17**, e1009442 (2021).
57. Feng, P. et al. Crocetin Prolongs Recovery Period of DSS-Induced Colitis via Altering Intestinal Microbiome and Increasing Intestinal Permeability. *Int. J. Mol. Sci.* **23**, 3832 (2022).
58. Huang, P. et al. Effects of hydroxychloroquine on the mucosal barrier and gut microbiota during healing of mice colitis. *Am. J. Transl. Res* **16**, 4144–4153 (2024).
59. Yang, J. Y., Chen, S. Y., Wu, Y. H., Liao, Y. L. & Yen, G. C. Ameliorative effect of buckwheat polysaccharides on colitis via regulation of the gut microbiota. *Int J. Biol. Macromol.* **227**, 872–883 (2023).
60. Sanidad, K. Z. et al. Gut bacteria-derived serotonin promotes immune tolerance in early life. *Sci. Immunol.* **9**, eadj4775 (2024).
61. Di Giovangiulio, M. et al. The Neuromodulation of the Intestinal Immune System and Its Relevance in Inflammatory Bowel Disease. *Front Immunol.* **6**, 590 (2015).
62. Lo Presti, A. et al. Phylogenetic analysis of Prevotella copri from fecal and mucosal microbiota of IBS and IBD patients. *Therap Adv Gastroenterol* **16**, 17562848221136328 (2023).

63. Iljazovic, A. et al. Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation. *Mucosal Immunol.* **14**, 113–124 (2021).
64. Ogunrinola, G. A., Oyewale, J. O., Oshamika, O. O. & Olasehinde, G. I. The Human Microbiome and Its Impacts on Health. *Int J. Microbiol* **2020**, 8045646 (2020).
65. Donia, M. S. & Fischbach, M. A. HUMAN MICROBIOTA. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).
66. Garber, K. Drugging the gut microbiome. *Nat. Biotechnol.* **33**, 228–231 (2015).
67. Sharma, S., Hegde, P., Panda, S., Orimoloye, M. O. & Aldrich, C. C. Drugging the microbiome: targeting small microbiome molecules. *Curr. Opin. Microbiol* **71**, 102234 (2023).
68. King, A. M. et al. Systematic mining of the human microbiome identifies antimicrobial peptides with diverse activity spectra. *Nat. Microbiol.* **8**, 2420–2434 (2023).
69. Jia, B., Kim, K. H., Ruan, W., Kim, H. M. & Jeon, C. O. Lantibiotic-encoding Streptococcus in the human microbiome are underlying risk factors for liver diseases. *J. Infect.* **84**, e70–e72 (2022).
70. Rahman, I. R., Sanchez, A., Tang, W. & van der Donk, W. A. Structure-Activity Relationships of the Enterococcal Cytolysin. *ACS Infect. Dis.* **7**, 2445–2454 (2021).
71. Muniz Pedrogo, D. A., Sears, C. L. & Melia, J. M. P. Colorectal Cancer in Inflammatory Bowel Disease: A Review of the Role of Gut Microbiota and Bacterial Biofilms in Disease Pathogenesis. *J. Crohns Colitis* **18**, 1713–1725 (2024).
72. Lauderdale, K. J., Malone, C. L., Boles, B. R., Morcuende, J. & Horswill, A. R. Biofilm dispersal of community-associated methicillin-resistant Staphylococcus aureus on orthopedic implant material. *J. Orthop. Res* **28**, 55–61 (2010).
73. Koo, H., Allan, R. N., Howlin, R. P., Stoodley, P. & Hall-Stoodley, L. Targeting microbial biofilms: current and prospective therapeutic strategies. *Nat. Rev. Microbiol* **15**, 740–755 (2017).
74. Sturme, M. H. et al. Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie Van. Leeuwenhoek* **81**, 233–243 (2002).
75. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
76. Santos-Aberturas, J. et al. Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Res* **47**, 4624–4637 (2019).
77. Benjdia, A. & Berteau, O. Radical SAM Enzymes and Ribosomally-Synthesized and Post-translationally Modified Peptides: A Growing Importance in the Microbiomes. *Front Chem.* **9**, 678068 (2021).
78. Terlouw, B. R. et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res* **51**, D603–D610 (2023).
79. Bajusz, D., Racz, A. & Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform* **7**, 20 (2015).
80. Willighagen, E. L. et al. The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminform* **9**, 33 (2017).
81. Narayan, A., Berger, B. & Cho, H. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nat. Biotechnol.* **39**, 765–774 (2021).
82. Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
83. Sangar, V., Blankenberg, D. J., Altman, N. & Lesk, A. M. Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics* **8**, 294 (2007).
84. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314-324 (2019).
85. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
86. Blanco-Miguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
87. Navarro-Munoz, J. C. et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
88. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
89. Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K. & Muller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
90. Frankish, A. et al. GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res* **51**, D942–D949 (2023).
91. Oksanen, J. Vegan: community ecology package. http://vegan.r-forge.r-project.org/ (2010).
92. Zhang, J. Large-scale biosynthetic analysis of human microbiomes reveals diverse protective ribosomal peptides [Data set]. *Zenodo* https://doi.org/10.5281/zenodo.14994355 (2025).
93. Zhang, J. & Zhang, D. W. Large-scale biosynthetic analysis of human microbiomes reveals diverse protective ribosomal peptides. *Zenodo* https://doi.org/10.5281/zenodo.14866373 (2025).

Acknowledgements

This work is partially funded by Hong Kong Research Grants Council Collaborative Research Fund (C7014-24G) and General Research Fund (27107320, 17115322, and 17102123) to Y.X.L. The authors would like to thank Zhiman Song and Cunlei Cai for their help in MS and NMR analysis. The authors also thank Vecteezy.com for distributing free vector art at https://www.vecteezy.com, which is partially adopted and modified in Fig. 5a.

Author contributions

Conceptualization: Y.X. L., J.Z., D.W.Z.; Bioinformatics analysis: J.Z., D.W.Z., Y.G., Z.Z., B.B.H.; Experimental validation: J.Z., Y.X., J.L.Z., R.Z.L., Y.G., Y.Q.S., P.Y.C.: Visualization: J.Z., D.W.Z.; Writing—original draft: J.Z.; Writing—review and editing: Y.X.L., D.W.Z., J.Z., X.C.L., and H.W.Z.; Investigation: Y.X.L. and D.W.Z.; Supervision: M.X.C., Y.X.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58280-w.

Correspondence and requests for materials should be addressed to Muxuan Chen or Yong-Xin Li.

Peer review information *Nature Communications* thanks Marnix Medema, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025