Elucidating the Diversity and Potential Function of Ribosomally Synthesized and Post-translationally Modified Peptides in the Human Microbiome

Jian Zhang[1,#], Dengwei Zhang[1,#], Junliang Zhang[1], Ying Gao[1], Yuqi Shi[1], Peiyan Cai[1], Zheng Zhong[1], Beibei He[1], Xuechen Li[1], Yong-Xin Li[1,*]

[1] Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China

[#] These authors contributed equally to this work.

[*] To whom correspondence may be addressed.

Email: yxpli@hku.hk

Abstract

The human microbiome possesses a vast potential for producing ribosomally synthesized and post-translationally modified peptides (RiPPs) that can impact human health and disease. Understanding the RiPPs-mediated microbe-microbe and microbe-host interactions holds significant implications for human health. However, the RiPP biosynthetic potential in human microbiome and their associations with human disease remain largely uncharacterized. In this study, we systematically analyzed 306,481 human microbiota-associated genomes, revealing a wide diversity of RiPPs that are mostly unknown. RiPP biosynthesis is found in various body sites and exhibits niche-specific enrichment in the gut and oral microbiome. Through a comparative metatranscriptomic analysis, 30 RiPP families with potentially antibacterial and signaling activities are found to be related to multiple diseases. Nine RiPPs, namely autoinducing peptides (AIPs), negatively associated with multiple diseases are chemically synthesized and experimentally validated for their bioactivity. Five AIPs can effectively inhibit biofilm formation by disease-associated pathogens such as *Clostridioides difficile*. These findings highlight the vast potential of human microbial RiPPs in regulating microbial communities and maintaining human health, emphasizing their potential for therapeutic development.

Introduction

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[1-4]. To explore microbial biochemistry and its interactions with the host[5-8], extensive research has been conducted. It is firmly established that chemical interactions mediated by metabolites *de novo* biosynthesized or transformed by the colonized microbiota shape the homeostasis of holobionts[5,7,9-14]. As a result, there is growing interest in the therapeutic potential of molecules secreted by the human microbiota[15,16], particularly in the context of antibiotic discovery and resistance and its impact on human health. Guided by these molecular mechanisms, enormous bioactive small molecules have been identified in the human microbiome[9,17-19]. Among them, ribosomally synthesized and post-translationally modified peptides (RiPPs) have attracted increasing attention because of their diverse structures[20,21], various biological functions[22,23], and manifold roles in physiology and ecology[23,24].

RiPPs are a class of diverse and widespread natural products produced by a multitude of producers, including bacteria, fungi, plants, and marine organisms[24]. Despite the existence of over 40 subclasses of RiPPs based on their distinct structures[21], the biosynthesis of RiPPs follows a unified three-step process governed by biosynthetic gene clusters (BGCs) [20,21]. This involves encoding the RiPP precursor peptide, comprising a leader peptide, a core peptide, and a follower peptide in some cases, followed by installing various structural features onto the core peptide through post-translational modification (PTM) enzymes. The biosynthesis concludes with leader cleavage to generate the biologically active RiPP products. The conserved biosynthetic features serve as a foundation for developing various computational tools for RiPP discovery. These strategies typically involve precursor-centric genome mining strategies (e.g., DeepRiPP[25] and TrRiPP[26]) or primarily targeting the PTM enzymes involved in RiPP BGCs (e.g., antiSMASH[27] and DeepBGC[28]) using canonical rule-based annotation or machine learning algorithms[29]. Notably, antiSMASH has been widely used to discover RiPPs derived from the human microbiota[30-34].

RiPPs often exhibit antibacterial, immunomodulatory, and cytotoxic activities, or serve as signaling molecules[22,23,35] . These attributes contribute to their role as “foreground players”

in the complex human microbial community[11,36], and endow their promising applications in the medical domain, including potential use as treatments or targets for various diseases. The molecular mechanisms of RiPPs in human-microbial interactions can be broadly categorized into four main aspects. Firstly, RiPPs exhibit antibacterial activities that enable commensal producers to combat pre-existing or foreign pathogens[36]. For example, lactocilin, a thiopeptide from vaginal *Lactobacillus gasseri*, demonstrates potent antibacterial effects against vaginal pathogens such as *Corynebacterium aurimucosum*[31]. Secondly, RiPPs possess immunomodulatory activities that allow producers to interact with the human immune system[37-40], exemplified by tonsillar microbiome-derived Salivaricins. They inhibit the binding of IL-6 and IL-21 to their receptors, thereby disrupting the downstream immune response[37]. Additionally, a newly discovered phosphorylated lanthipeptide, salivaricin 10, produced by *Streptococcus salivarius*, exhibits unique proimmune properties that contribute to the homeostasis of the human oral microbiome[40]. Thirdly, the cytotoxicity of RiPPs might exacerbate disease progression. For instance, *Enterococcus faecalis*-produced cytolysin has been associated with the severity of liver disease[41]. Fourthly, beyond directly impacting the microbial community or host, RiPPs can act as signal molecules to regulate quorum-sensing/quenching responses in the human microbiome. For example, autoinducing peptides (AIPs) produced by *Staphylococci* species from the skin microbiome could interrupt the *agr* system of pathogen *Staphylococcus aureus* to maintain the epithelial barrier homeostasis[42]. The multifaceted functionalities exhibited by RiPPs implicate their potential involvement in human health and diseases. Nonetheless, our current understanding in this area is primarily limited to sporadic reports, highlighting the urgent imperative for a more comprehensive investigation.

The diverse activities of RiPPs indicate their significant roles in the human microbiome, thereby attracting ongoing efforts on their discovery. With advancements in sequencing technology and computational tools, (meta)genome-based mining approaches have gained prominence in RiPP discovery in recent years. Some research has emphasized the RiPP biosynthetic potential of specific strains, exemplified by the discovery of bioactive RiPPs Ruminococcin C[43] and Streptosactin[44] from human commensal bacteria *Ruminococcus gnavus* and *Streptococcus* spp., respectively. Other researchers have emphasized the RiPP biosynthetic potential of a specific species or higher taxa level (such as Bacteroidales[45], lactic acid bacteria[32], and human oral *Streptococcus mutans*[34]), unveiling the diversity and distribution in the human microbiome[32,45]. Additionally, biosynthetic analyses have been conducted to identify RiPP BGCs from a small number of reference genomes, such as 2,430 reference genomes of the isolated human microbiota[31], 461 oral bacterial genomes[33], and 3,324 cultivated human gut bacterial reference genomes[30]. While these studies have undoubtedly expanded our understanding of the biosynthetic capacity of RiPPs in the human microbiome, our current comprehension of RiPP biosynthesis remains constrained, due to the tendency to focus exclusively on individual bacterial species, even strains, or a relatively small subset of genomes for analysis. Therefore, there is a compelling need for a more comprehensive inquiry into RiPPs to provide a better understanding of their biosynthetic capabilities, which can serve as a foundation for elucidating their potential connections to human health.

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, we systematically examined the RiPP biosynthetic potential of 306,481 genomes of the human microbiota from the CIBIO[46] and UHGG[47] databases. Our comprehensive analysis of RiPP precursor peptides, encompassing 12,076 distinct precursor families, has revealed a remarkable diversity of bioactive RiPPs, many of which remain uncharacterized. Notably, these RiPPs exhibit wide distribution within the human microbiomes across various body sites, with particular enrichment in the gut and oral cavity. Through

function prediction and comparative meta-omic analyses conducted across different disease case-control cohorts, we have identified 30 multi-disease-related RiPP families exhibiting potential antibacterial, signal molecule, and other underexplored activities, with notable implications in inflammatory bowel disease (IBD) and colorectal cancer (CRC). Guided by structure prediction and sequence conservation, mature products of nine AIPs that are negatively associated with multiple diseases were predicted and successfully chemically synthesized. Experimental validation showed that five AIPs exhibited significant antibiofilm activity against human disease-associated pathogens. These findings highlight the significant potential of human microbial RiPPs in regulating the balance of microbial communities. Our study offers valuable insights into the possible roles of RiPPs in contributing to human health and well-being.

Results

Genome mining of human microbes reveals the untapped biosynthetic potential of RiPPs

The current process of genome mining for RiPPs can be approached through two main methods: tailoring enzymes-oriented approach (e.g., antiSMASH 6.0[27]) and precursors-centric approach (e.g., DeepRiPP[25] and TrRiPP[26]). The enzymes-oriented approach is reliable in identifying known RiPPs but may miss out on discovering novel ones. On the other hand, the precursors-centric approach is better suited for identifying RiPPs 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 microbiome, we have combined both types of tools (**Figure 1**). Our approach analyzed a dataset consisting of 306,481 reference microbial genomes from diverse human body sites[46-48], including the gut, oral cavity, skin, airways, vagina, and nasal cavity (**Supplementary Figure 1**).

The analysis of antiSMASH detected 410,487 RiPP BGCs in 239,927 genomes, accounting for 74.5% of the predicted secondary metabolite BGCs (**Figures 1B, Supplementary Figure 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 (**Figure 1C, Supplementary Figure 2, Supplementary Data 2**). Examining the genes surrounding the RiPP precursors from the precursor-centric approaches, we found that the genomic contexts of 32.6% of precursors contained genes associated with known RiPP biosynthesis (**Figure 1C, Supplementary Figure 3, Supplementary Tables 1-2**), while the genomic contexts of 58.4% of precursors contained potentially novel RiPP biosynthetic enzymes collected in decRiPPter[49] (**Figure 1C**). These findings indicate that the precursor-centric approach offers a higher level of biosynthetic novelty, surpassing traditional tailoring enzymes-oriented strategies. The analysis also showed that antiSMASH-defined RiPP BGCs were more abundant in 25 phyla, expanding to 29 phyla with the precursor-centric approaches (**Figure 1D, Supplementary Figure 1**). Notably, the most abundant phyla in the human gut[50], Bacteroidota and Firmicutes, also harbored the most abundant RiPPs. This extensive distribution of RiPP BGCs and precursor peptides across diverse human microbiomes underscores human microbes' robust RiPP biosynthetic potential.

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

Realizing 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 that were confirmed by at least two mining strategies for downstream analysis, which could effectively reduce the number of false positives while preserving new and novel RiPPs. The 423,831 RiPP precursors were found either within RiPP BGCs region[51] identified by an enzymes-oriented approach (62.4%) or by both precursors-centric approaches (37.6%) (**Figure 2A, Supplementary Data 3**), with rSAM-modified RiPPs being the most abundant (**Figure 2B**). Those retained RiPPs were strongly enriched in 21 phyla, 1,014 genera and 3,369 species. At the genus level, 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 Figure 4)**.

To evaluate RiPP sequence diversity, we here adopted the Extended-connectivity fingerprints (ECFPs)[52] 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 (**Figure 2C**), suggesting that RiPPs derived from the human microbiome could significantly enlarge the chemical diversity within the RiPP superfamily. We next explored the chemical diversity of nine RiPP classes based on the core precursor sequences predicted by the cleavage prediction module of DeepRiPP[25] **(Figure 2D, Supplementary Data 4)**. Pairwise Tanimoto coefficient was used to measure the similarity between precursors. Notably, autoinducing peptides exhibited the highest median Tanimoto coefficient, indicating their conserved intra-class sequences. 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 next examined the precursor sequences and tailoring enzymes, which can dictate the chemical structure of RiPPs (**Figure 3**). Regarding RiPP chemical novelty assessed by precursors, we first clustered 423,831 RiPP precursors into 12,076 families with a threshold of 50% identity (**Figure 3, Supplementary Figure 5, 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. For the remaining families, we used RPS-BLAST to search domains against the Conserved Domains Database (CDD), detecting 2,339 families (19.37%) containing well-defined RiPP precursor domains, such as AgrD domain and rSAM-modified RiPP domain. Notably, most families (9,445, 78.21%) remained uncharacterized or exhibited significant distinctions from known RiPPs **(Figure 3A)**. Regarding RiPP novelty assessed by tailoring enzymes, we examined the adjacent genomic context of RiPP precursor families **(Supplementary Figure 6)**. 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 novel RiPP biosynthetic-related enzymes defined by DeepBGC[28] (**Figure 3**). Combining both factors, 63.7% of RiPP precursor families have the

potential to produce novel RiPP products (**Figure 3C**). This finding suggests the possibility of novel enzymology, chemistry, and bioactivity in these RiPPs families.

RiPPs exhibit niche-specificity in the healthy human microbiome

To further assess RiPP distribution in the human microbiome, we attempted to examine their profile in healthy human microbiomes. We analyzed 748 metagenomic samples from the Human Microbiome Project (HMP)[53], covering six human body sites: gut, buccal mucosa (oral), supragingival plaque (oral), tongue dorsum (oral), anterior nares (skin), and posterior fornix (vaginal). Metagenomic analysis revealed 28.63% (3,457/12,076) of RiPP precursor families were detected in healthy individuals, varying considerably from 100 families in anterior nares to 2,791 families in stool **(Figure 4A)**. Accumulation curves showed a steep increase, indicating that more RiPP precursor families could be identified in each body site with larger sample sizes **(Supplementary Figure 7A)**. Compared to sporadic presence in the skin and vagina, RiPP precursor families exhibited higher diversity, prevalence, and abundance in the oral and gut **(Figure 4A** and **Supplementary Figure 7B)**. Among these detected families, 73.2% (2,530/3,457) were niche-specific, while 0.9% (31/3,457) were present across six body sites in at least one sample **(Supplementary Figure 7C)**. This could be largely attributed to the fact that the producers of niche-specific RiPPs were habitat-specific **(Figure 4B)**. Furthermore, two-dimensional visualization demonstrated distinct patterns of RiPP precursor families among different body sites, indicating their niche-specificity **(Figure 4C)**. Activity prediction showed that a majority of RiPPs (70.7%, 8,533/12,076) were antibacterial **(Figure 4D)**, which thus enables their producers for niche adaption. However, their host translocation from the oral to the gut sometimes contributes to disease development such as liver cirrhosis (LC) and atherosclerotic cardiovascular disease (ACVD)[54,55]. We further profiled RiPP precursor families in gut microbial metagenomes from LC[56] and ACVD[57] disease case-control cohorts **(Supplementary Table 3, Supplementary Data 6)**. Interestingly, we observed that the prevalence of 32 *Streptococcus*-derived RiPP families significantly increased in the disease microbiome compared to healthy individuals **(Supplementary Figure 8)**. However, these RiPP families are predominantly present in oral microbiomes but are rare in the gut microbiome of healthy individuals in HMP **(Supplementary Figure 8)**. This observation may lead to investigating how RiPPs assist oral microbes in colonizing the gut and causing adverse effects on human health.

Although diverse RiPP families could be detected in the human microbiome, whether they are actively transcribed remains questioned. Therefore, we next sought to look for evidence for their transcription by analyzing the 281 paired metagenomic and metatranscriptomic data from faecal samples of healthy individuals[58] **(Supplementary Table 4)**. Among 8,278 (68.55%) RiPP precursor families detected in metagenome samples, 3,525 (29.19%) were found to be actively transcribed in metatranscriptome data (**Figure 4E**). Lanthipeptide precursors were the most abundant among the detected precursor families. Notably, a significant proportion of transcribed RiPP precursor families in each class (ranging from 54.5% of thiopeptide to 91.7% of graspetide) were predicted to exhibit bioactivity (**Figure 4E**). Hence, we found that diverse RiPP precursor families are present and actively transcribed in the healthy gut microbiome.

Differentially transcribed RiPP families are associated with multiple diseases

Inspired by numerous functional RiPPs from human microbiota linked to mediate microbe-microbe interaction (e.g., antibacterial activities[36], quorum-sensing or quorum-quenching[42]), or microbiome-host interaction (e.g., immunomodulatory activities[37-40], cytotoxicity[41]), we aimed to explore potential roles of RiPPs in human health. Accordingly, we conducted comparative metatranscriptome analyses to identify underlying bioactive RiPPs associated with multiple diseases. These diseases included type 1 diabetes mellitus (T1D)[59], obesity (OB)[60], liver cirrhosis with hepatitis C infection (LC)[61], colorectal cancer (CRC)[62], Parkinson's disease (PD)[63], and inflammatory bowel disease (IBD: Crohn's disease (CD) and ulcerative colitis (UC))[64,65] **(Supplementary Table 4)**. Through the metatranscriptomic analysis, we observed significant differences in RiPP diversity between the health and disease groups. Specifically, the healthy groups had a higher diversity than patients with CD but lower diversity than patients with OB **(Supplementary Figure 9A)**. According to permutational multivariate analysis of variance (PERMANOVA) based on Bray-Curtis dissimilarity, a significant difference ($p < 0.05$) in the overall composition of transcribed RiPP precursor families was observed between health and disease in CRC and IBD case-control cohorts **(Supplementary Figure 9B)**. Additionally, we identified 195 precursor families depleted while 146 enriched in disease groups within these disease case-control cohorts **(Figure 5A**, **Supplementary Data 7)**. Remarkably, 23 families exhibited depletion across multiple disease groups **(Figure 5B)**, while 7 families were found to be enriched in multiple disease groups **(Figure 5C)**.

Considering the larger sample size of the IBD cohort (n = 406) compared to other pilot studies, we aimed to further investigate the potential contribution of RiPPs associated with IBD as a representative disease. To achieve this, we constructed a random forest classification model based on transcribed precursor families. The 10-fold cross-validation results showed the area under the receiver operating characteristic curve (AUROC) being 0.79 and 0.76 for CD and UC, respectively **(Figure 5D)**. Among the top 10 contributors to the model, five RiPP families (BF_63, BF_2275, BF_1459, BF_94, and BF_96) that were negatively associated with multiple diseases were found to be depleted in patients with CD by more than two times, suggesting a potential protective role in human health. As most RiPPs were predicted to be antibacterial, they might contribute to human microbiome homeostasis by mediating other microbes. We, therefore, conducted a correlation analysis between RiPP precursor families and microbial species in the IBD case-control cohort. Specifically, three RiPP families (BF_280, BF_90, and BF_638) were potentially involved in mediating the microbial community **(Supplementary Figures 10-12)**. For example, an autoinducing peptide family BF_280 (**Supplementary Figures 10-11**), which was depleted in CD and UC patients, was negatively associated with the abundance of *Enterocloster bolteae*. This species is increased in IBD patients and is considered functional in IBD disease[65]. Another case was a lasso peptide family BF_90, which was depleted in CD patients and positively associated with an important contributor, lasso peptide BF_96 **(Figure 5D)**, was negatively associated with *Flavonifractor plautii*, an important mediator in the microbiome community of IBD patients[66] (**Supplementary Figure 12**).

AIPs inhibit the formation of pathogenic biofilms

We next sought to experimentally validate the potential functions of RiPPs. Given the significant impact of gut microbial biofilm on both disease pathogenesis[67], as in the case of pathogenic biofilm, and human health[68], as in the case of probiotic biofilm, we have directed our attention toward the study of AIPs that often involve in mediating biofilm formation. We

thus aimed to validate six AIP families (BF_63, BF_94, BF_280, BF_398, BF_488, and BF_598) with an enrichment in the healthy microbiome compared to the diseased microbiome, particularly in CRC and IBD **(Figure 5B)**. As native producers of six AIPs are not available, and the protease involved in the second cleavage of AIPs' biosynthetic pathway is still not fully understood currently, which hinders the usage of the heterologous expression approach, we alternatively attempted to chemically synthesize the mature chemicals of six AIPs.

Here, we employ a combination of structural prediction and evolutionary conservatism to forecast the ultimate structure of AIPs **(Figure 6)**, which are subsequently chemically synthesized to test their potential bioactivity. 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[69] **(Figure 6)**. Using AlphaFold-Multimer, we first examined the complex of precursor peptide and AgrB protein of three known AIPs (**Figure 6, Supplementary Figure 14**) and found that core peptide can perfectly insert into the catalytic pocket of AgrB enzyme, suggesting the core region of AIP precursor was predictable. Meanwhile, AgrB facilitates the formation of a thiolactone or lactone between Cysteine (Cys, C) or Serine (Ser, S) and another amino acid containing a hydrophobic side chain, leading to the modification on conserved residues. Furthermore, the exocyclic tail region of AIPs is conserved due to the highly specific nature of the second proteolysis step[70]. By considering these factors, we could accurately predict the mature structures of AIPs. The accuracy and reliability of this approach were validated by the three reported AIPs (**Figure 6, Supplementary Figure 14**).

With the bioinformatic prediction, we deduced the potential mature AIPs for six AIP families that have been linked to IBD and/or CRC **(Figure 5, Supplementary Table 5)**. Specifically, these substances typically give rise to a lactone (BF_63, BF_94, and BF_488) or thiolactone (BF_398, BF_280, BF_598) between a Ser or Cys residue and another amino acid with a hydrophobic side chain. In addition, in the case of exocyclic region-free thiolactone-containing AIPs, their thiolactone could potentially undergo rearrangement through S → N acyl shift, resulting in the formation of homodetic cyclopeptides (cAIPs)[71,72]. As such, it is plausible that mature products of BF_280, BF_598, and exocyclic region-free BF_398 families could lead to the creation of cAIPs (**Figure 6C-F**). Furthermore, BF_94 exhibits sequence similarity to the previously reported exocyclic region-free AIP ([SFFIF]) from *C. thermocellum* Clo1313_2818[73]. We therefore synthesized a homologous variant, BF_94_et_free, for the BF_94 family to further investigate the potential bioactivity of AIPs. Finally, we successfully chemically synthesized nine potential AIPs and validated their structure using high-resolution MS/MS and NMR **(Supplementary Figures 15-23)**.

With the synthesized AIPs in hand, we next sought to investigate the bioactivity of these compounds. We assessed whether these synthesized AIPs could inhibit the biofilm formation by human pathogens (**Table 1**) such as *Peptostreptococcus stomatis*, *Clostridioides difficile, Clostridium symbiosum*, *Streptococcus gallolyticus subsp. Gallolyticus*, *Enterocloster bolteae*, *Staphylococcus aureus*, *Listeria monocytogenes*, *Enterococcus faecalis*, *Bacteroides fragilis*, *Escherichia coli*, and *Candida albicans*. We observed that the 5/9 AIPs demonstrated significant inhibition against at least one pathogenic biofilm, albeit at relatively higher concentrations in some cases **(Table 1)**. Specifically, these AIPs showed extended-spectrum activities against pathogenic biofilm across different species. For example, BF_598_c showed relatively potent anti-biofilm activities to *Clostridioides difficile, Staphylococcus aureus*, *Listeria monocytogenes*, and *Candida albicans*. Similarly, BF_398_c, BF_280_c, and

BF_94_et_free could inhibit biofilm formation of at least three pathogens. Notably, BF_598_c displays a unique antifungal biofilm, which sets it apart from other known AIPs. Furthermore, BF_280_c, BF_280_t, and BF_398_c, which are derived from the clostridia class, significantly reduced the biofilm formation by *Clostridioides difficile* at a low inhibitory concentration of 0.781, 0.781, and 6.25 μg/mL, respectively **(Table 1)**. Moreover, we observed that homodetic cyclopeptide BF_280_c displayed a similarly broad and potent inhibition of pathogenic biofilm formation with thiolactone-containing BF_280_t. This finding differs from the recent report on AIPs from *L. monocytogenes*, where the short-lived thiolactone-containing pentamer was more effective than the final homodetic peptides[74]. Collectively, we demonstrated that AIPs (BF_94_et_free, BF_280_t, BF_280_c, BF_398_c, and BF_598_c) could inhibit pathogenic biofilm formation. These findings suggest that human microbiome-derived AIPs potentially serve a protective function in human health.

Table 1 Anti-biofilm activity of AIPs against pathogenic biofilm by pathogens

Pathogen		AIPs (μg/mL) (*lowest concentration for significantly inhibit biofilm formation)								
		BF_94_et_L1	BF_94_et_free	BF_280_t	BF_280_c	BF_398_et_L1	BF_398_c	BF_488_et_free	BF_598_c	BF_63
G+	*Peptostreptococcus stomatis* DSM17678	> 50	25	>50	>50	>50	25	>50	>50	>50
	Clostridioides difficile ATCC BAA-1870	> 25	> 25	0.781	0.781	> 25	6.25	> 25	12.5	> 25
	Clostridioides difficile ATCC 43255	> 25	> 25	> 25	> 25	> 25	> 25	> 25	25	> 25
	Clostridium symbiosum ATCC 14940	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25
	Enterocloster bolteae DSM15670	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25
	Streptococcus gallolyticus subsp. Gallolyticus DSM 16831	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50
	Listeria monocytogenes ATCC 19115	>100	50	>100	25	> 100	25	> 100	50	> 100
	Staphylococcus aureus ATCC 43300	>100	12.5	6.25	6.25	> 100	6.25	> 100	50	> 100
	Enterococcus faecalis ATCC 29212	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50	> 50
G-	*Fusobacterium nucleatum* ATCC 25586	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25
	Bacteroides fragilis ATCC 43860	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25
	Bacteroides fragilis ATCC 25285	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25
	Porphyromonas gingivalis ATCC 33277	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25	> 25
	Escherichia coli ATCC 35401	>100	>100	>100	>100	>100	>100	>100	>100	>100
fungi	*Candida albicans* ATCC 10231	>100	100	>100	>100	>100	>100	>100	25	>100
	Candida albicans SC 5314	> 50	> 50	> 50	> 50	> 50	> 50	> 50	50	> 50

*n = 3 biologically independent samples. Significance was determined using one-way ANOVA test. Bars represent mean ± standard error. For all p values, $p < 0.05$ mean significant difference compared with the control group.

Discussion

Extensive research has highlighted the profound functional impact of the host-microbiome interaction on the etiology and progression of various diseases[75]. Human microbiota has tremendous potential to produce large amounts of structurally distinct metabolites with various biological activities, whereby mediating microbe-microbe or microbe-host interactions[17]. However, there remains a significant gap in our mechanistic understanding of the role played by microbial biomolecules, particularly microbial ribosomal peptides, in shaping both microbial and human homeostasis.

In this research, we conducted a comprehensive investigation into the potential for RiPP biosynthesis within the human-associated microbiota. This involved integrating multiple prediction methods, leading to the identification of 410,487 RiPP BGCs and 1,777,098 RiPP precursors originating from 29 microbial phyla. In contrast to previous studies, which identified 36,583 BGCs[76], our analysis revealed an expansive repertoire of 550,649 BGCs. Among the most abundant BGCs are RiPPs. Leveraging the rule-based prediction tool antiSMASH, our investigation comprehensively characterized molecules derived from the human microbiota. Complementing two deep learning-based prediction tools, our study upgraded the overall landscape of RiPP biosynthesis and disclosed previously under-appreciated RiPP biosynthetic capacity, for example, remarkable abundance of un-detected lanthipeptides, a large group of RiPPs widely detected from human microbiota[77-80] (**Figures 1 and 3, Supplementary Figures 1-2**). Furthermore, we observed human microbiome-derived RiPP expand the chemical diversity of the RiPP superfamily **(Figure 2)**, and exhibit higher novelty **(Figure 3)** and potential bioactivity **(Figure 4D),** with the majority remaining uncharacterized to date. Our research efforts culminate in the provision of an atlas cataloging RiPPs derived from the human microbiota. This resource is poised to serve as a comprehensive reference for forthcoming investigations to target therapeutic interventions within the microbiome[81,82].

RiPPs are significant players in the ecological landscape due to their diverse bioactivity. A variety of biological activities allow RiPPs to participate in human-microbiota and microbiota-microbiota interactions. One of the most well-known activities of RiPPs is antibacterial properties. Nathan A. Magarvey *et al* [36] conducted an analysis specifically aimed at uncovering the concealed potential for antibiotics within the human microbiome. They revealed that RiPPs were the most frequently predicted compounds across diverse microbial habitats throughout the body. Moreover, RiPPs are prevalent in the metagenome of the gastrointestinal tract. In our study, we observed that some predicted antibacterial RiPPs were actively transcribed in healthy human gut microbiomes **(Figure 4E)**. Besides, 30 of them are significantly associated with multiple diseases in metatranscriptomic data **(Figure 5B, C)**. The antibacterial properties of RiPPs can assist their host organisms in either establishing or disrupting barriers between different environmental niches. However, bioactive RiPPs may serve as mechanisms that assist in colonizing bacteria and disrupting barriers between environmental niches. Andrew M. King *et al*[83] identified specific lanthipeptides produced by oral *Streptococcus* that selectively target healthy gastrointestinal microbes, as well as a lasso peptide from oral *Rothia aeria* that specifically targets *B. adolescentis* from the gut. Moreover, a lantibiotic produced by translocating *Streptococcus* has demonstrated potential hepatocyte toxicity[84]. Viable microbes or specific microbial products that deviate from the expected composition within a given environment are key markers of bacterial translocation. However, crucial metagenomic data from paired saliva and stool samples are very few, strain profiling tools still need to be improved, and bacterial molecular mechanisms are still under-explored.

Microbes can contribute to disease pathogenesis by modulating the homeostasis of the overall microbial community. This communication among microbial communities is often

facilitated through quorum sensing, part of which is achieved by autoinducing peptides[85]. In disease treatment, the presence of polymicrobial biofilms presents an additional challenge, as it requires antimicrobials effective against all pathogens within the biofilm. Combining AIPs with antibiotics shows promise in enhancing efficacy against biofilm infections[86]. For instance, research has demonstrated that the AIP-I effectively triggers dispersal within MRSA biofilms located on titanium disks. This dispersal process enhances the vulnerability of detached MRSA cells to treatment with rifampin and levofloxacin. In our study, we identified six AIP families enriched in healthy microbiomes and obtained synthetic-bioinformatic natural products[87] exhibiting significant anti-pathogenic biofilm activity **(Table 1)**. Furthermore, AIPs enriched in healthy gut microbiomes contribute to the formation of probiotic biofilms in the healthy gut[88]. These biofilms play a vital role in maintaining intestinal immune homeostasis by prolonging bacterial residence time, preventing the colonization of pathogens, and facilitating the exchange of nutrients between the host and the microbiota[68]. Collectively, investigating the role of quorum sensing in biofilm development may offer an avenue for developing innovative therapeutics. Further investigation is required to elucidate the underlying mechanisms.

To summarize, our study has presented a comprehensive overview of RiPP biosynthesis in the underexplored human microbiome. We have identified a wide range of RiPP precursor peptides, showcasing their remarkable chemical diversity, novelty, and bioactivity, with a significant portion of these peptides remaining uncharacterized. Furthermore, we have endeavored to establish associations between RiPP profiles and various human diseases, focusing on identifying RiPP candidates that may have either protective or detrimental impact on human health. Nevertheless, it is crucial to recognize that our research has certain limitations that must be acknowledged. Firstly, despite our extensive efforts, the scope of RiPP discovery is still limited, as the 306,481 reference genomes analyzed in our study represent only a small fraction of human microbes, which are predominated by the gut microbiome, especially bacteria. Secondly, while deep learning-based approaches have enabled broader detection of RiPPs, they may also introduce a higher rate of false positives compared to rule-based approaches. This tradeoff remains a challenge that cannot be fully addressed with the currently available tools. Thirdly, despite our rigorous efforts to establish associations between RiPPs and human diseases, it is crucial to emphasize the need for future experimental validations to solidify these associations. Nonetheless, our study provides valuable insights into the diversity and potential roles of RiPPs in the human 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[47,48]), and 17,249 genomes from CIBIO[46]. To normalize taxonomic annotations for all the genomes, we re-annotated all of them using GTDB-Tk[89] (v2.0.0) against GTDB (rev207 version). All genomes were analyzed by antiSMASH 6.0[27] 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[90]. Subsequently, small ORFs ($\leqslant$150 amino acids) containing a start and stop codon and ribosome-binding site motifs were subjected to TrRiPP[26] and DeepRiPP[25] 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[20,21,27,91]. For the remaining precursor, the domains were paired with a broader dataset of biosynthetic enzyme domains, which were used in the decRiPPter[49] pipeline for exploring any novel 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)[35]. Extended-connectivity fingerprints (ECFPs) are used to predict and gain insight into RiPPs precursors' chemical diversity[52]. Then ECFP6 fingerprints for each unique RiPP precursor were compared with each other to generate Tanimoto coefficient matrices[92] 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[93]. Subsequently, we generated a Jaccard distance matrix based on the chemical fingerprints and performed dimension reduction using densMAP[94] 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, 423,831 RiPP precursors that were either within RiPP BGCs region[51] 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[95] 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[96]. The classification of each RiPP precursor family was further determined based on the novelty of their precursor and their genomic contexts **(Supplementary Figure 5)**. (1) Novelty of precursors. The known RiPP precursors collected from MIBiG 3.0[35] were quired against 12,076 representative precursor sequences by BLASTp. A significant hit was considered with an e-value < 0.05, the alignment spans ≥ 90% of the peptide, and the length of the hit was 90%–110% of the length of the known RiPP precursor peptide. For the remaining families, each family was assigned a "domain representative" if at least 80% of its members contained a specific RiPP precursor domain. Otherwise, the family was classified as a novel

family. (2) Novelty of genomic context of precursors. To assess the novelty of the genomic context surrounding RiPP precursors, we employed antiSMASH or DeepBGCs[28] to identify the presence of biosynthetic genes. Our analysis began with individual precursors, examining their genomic neighborhoods. If a precursor's genomic context was found to be associated with a known RiPP BGC, it was categorized as "RiPP". If the genomic context was novel recognized by DeepBGC, it was categorized as "Novel". Alternatively, if the genomic neighborhood showed evidence of hybridization with other secondary metabolite BGCs or exhibited other conditions, it was placed in the "Others" category. Furthermore, to provide a more comprehensive classification, we grouped and classified the types of genomic neighborhoods for each RiPP precursor family. If over 50% of the family members belonged to a specific type of genomic neighborhood, the entire family was assigned to that respective type. This allowed for a more refined understanding of the genomic diversity within each RiPP precursor family. (3) Novelty of RiPP families. The novelty of RiPP families was further classified into three types: The first type is referred to as "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. The second type is "Novel RiPP families", which consist of novel precursors and/or genomic neighborhoods containing novel biosynthetic genes. Lastly, the third type is "Others".

Bioactivity prediction of RiPP families. We utilized the DeepBGC[28] 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.

Analysis of metagenomic and metatranscriptomic data. The metagenomic and metatranscriptomic samples were downloaded from sequence read archive (SRA) of the NCBI (**Supplementary Tables 3-4**). 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)[97] with default parameters. The reads mapped to RiPP precursor gene were counted by featurecounts[98] 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[99] 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[100] 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[101], followed by Jalview for visualization.

Statistical modeling for classifier. To further investigate whether the abundance of RiPP precursor families can discriminate patients from health control, the relative/normalized abundance was further subjected to the cross-validated random forest classification model, and the receiver operating characteristic (ROC) curve for the predictive disease model was calculated. The average area under the ROC curve value was used to evaluate the performance of the classifier. A higher average AUROC represents the stronger predictive ability of the model to distinguish one group from another group. Following we further rank individual translated RiPP precursor peptide families' contribution to the predictive accuracy of the model[65].

Correlation network analysis. Paired metagenomic and metatranscriptomic samples (180 CD samples, 92 UC samples, and 95 noonIBD samples) were collected. Spearman's correlations between the abundance of RiPP precursor families and microbial species were analyzed using the function *corr.test* in R package psych. All *p*-values were adjusted with the "fdr" method. The correlations with adjusted *p*-values < 0.05 and absolute coefficient ≥ 0.3 were considered significant. Resulting correlation network was visualized by Cytoscape 3.9.0.

Phylogenetic tree construction. The representative genome of each genus was used to generate maximum-likelihood trees via GTDB-Tk (v2)[89] with the following parameters (refer to the GTDB-Tk user guide: https://ecogenomics.github.io/GTDBTk/commands/index.html): 1. GTDB-Tk reference data release 207 was used. 2. Parameters of classify: gtdbtk classify_wf --genome_dir --out_dir --force. 3. Parameters of infer tree from multiple sequence alignment: gtdbtk infer --msa_file MSA_FILE --out_dir OUT_DIR. 4. Using the convert_to_itol command to make the tree suitable for visualization in iTOL: gtdbtk convert_to_itol --input_tree --output_tree. Finally, we annotate their phylogenetic trends using iTOL[102].

Chemical synthesis of AIPs. To begin, we initiated the calculation and analysis of the precursor-AgrB complex. 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. 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 nonredundant 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. Finally, BF_94_et_L1, BF_280_c, BF_398_c, BF_598_c were chemically synthesized by Sangon Biotech (Shanghai, China) and the five predicted AIPs of BF_63_et_L1, BF_94_et_free, BF_280_t, BF_398_et_L1, BF_488_et_free were chemically synthesized by Yuan peptide (Nanjing, China). Their exact mass and MS/MS were confirmed by mass spectrometry, and their required purity was ≥ 90%, determined by high-performance liquid chromatography. The synthesized peptide powder was stored at − 80 °C and dissolved in sterilized DMSO to 10 mg/mL upon use.

LC-HRMS analysis of synthesized peptides. For LC-HRMS analysis, the samples were desalted and prepared at a final concentration of 0.01 mg/mL. LC-MS analysis was carried out using a Waters ACQUITY UPLC BEH C18, 130Å, 1.7 μm column, coupled with a Thermo Scientific UltiMate 3000 UHPLC system and Bruker impact Mass Spectrometer. The LC method employed a gradient for chromatographic separation, starting at 5% B for 2 minutes, followed by 5% to 95% B for 15 minutes, and finally held at 95% for 4 minutes. The column was re-equilibrated to 5% B for 1 minute before the next run started. The column was maintained at 40˚C, and the flow rate was set at 0.2 mL/min, using 0.1% formic acid in H_2O as solvent A and 0.1% formic acid in acetonitrile as solvent B. The MS system was tuned using a sodium formate standard, and all samples were analyzed in positive polarity with m/z range from 150 to 1500 Th, using data-dependent acquisition mode.

Antibiofilm assay. Biofilm formation was assessed using the crystal violet method. The strains listed in **Supplementary Table 6** 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. 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.

Acknowledgments

We thank Zhiman Song, Runze Liu, and Cunlei Cai for their help in MS and NMR analysis.

Funding

This work is partially funded by three Hong Kong Research Grants Council General research grants (HKU27107320, HKU17115322, and HKU17102123).

Author contributions

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

Competing interests

The authors declare that they have no competing interests.

References

1 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat Med* **24**, 392-400 (2018).

2 Baker, J. L., Mark Welch, J. L., Kauffman, K. M., McLean, J. S. & He, X. The oral microbiome: diversity, biogeography and human health. *Nat Rev Microbiol* **22**, 89-104 (2024).

3 Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* **19**, 55-71 (2021).

4 Wong, C. C. & Yu, J. Gut microbiota in colorectal cancer development and therapy. *Nat Rev Clin Oncol* **20**, 429-452 (2023).

5 Krautkramer, K. A., Fan, J. & Backhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nat Rev Microbiol* **19**, 77-94 (2021).

6 Balskus, E. P. Elucidating the Chemistry and Biology of the Human Microbiome. *Biochemistry* **61**, 2777-2778 (2022).

7 Silpe, J. E. & Balskus, E. P. Deciphering Human Microbiota-Host Chemical Interactions. *ACS Cent Sci* **7**, 20-29 (2021).

8 Zepeda-Rivera, M. *et al.* A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature* (2024).

9 Wang, L., Ravichandran, V., Yin, Y., Yin, J. & Zhang, Y. Natural Products from Mammalian Gut Microbiota. *Trends Biotechnol* **37**, 492-504 (2019).

10 Ansari, M. H. R. *et al.* Role of gut microbiota metabolism and biotransformation on dietary natural products to human health implications with special reference to biochemoinformatics approach. *J Tradit Complement Med* **13**, 150-160 (2023).

11 Rebuffat, S. Ribosomally synthesized peptides, foreground players in microbial interactions: recent developments and unanswered questions. *Nat Prod Rep* **39**, 273-310 (2022).

12 Wilson, M. R., Zha, L. & Balskus, E. P. Natural product discovery from the human microbiome. *J Biol Chem* **292**, 8546-8552 (2017).

13 Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* **336**, 1262-1267 (2012).

14 Balskus, E. P. The Human Microbiome. *ACS Infect Dis* **4**, 1-2 (2018).

15 Cully, M. Microbiome therapeutics go small molecule. *Nat Rev Drug Discov* **18**, 569-572 (2019).

16 Beemelmanns, C., Keller, A. & Muller, R. Mining the microbiota for antibiotics. *Nat Microbiol* **9**, 13-14 (2024).

17 Donia, M. S. & Fischbach, M. A. HUMAN MICROBIOTA. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).

18 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).

19 Dai, H. *et al.* Recent advances in gut microbiota-associated natural products: structures, bioactivities, and mechanisms. *Nat Prod Rep* (2023).

20 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).

21 Montalban-Lopez, M. *et al.* New developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* **38**, 130-239 (2021).

22 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).

23 Sugrue, I., Ross, R. P. & Hill, C. Bacteriocin diversity, function, discovery and application as antimicrobials. *Nature Reviews Microbiology* (2024).

24 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).

25 Merwin, N. J. *et al.* DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc Natl Acad Sci U S A* **117**, 371-380 (2020).

26 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** (2024).

27 Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**, W29-W35 (2021).

28 Hannigan, G. D. *et al.* A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res* **47**, e110 (2019).

29 Mullowney, M. W. *et al.* Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov* (2023).

30 Lin, X. *et al.* The genomic landscape of reference genomes of cultivated human gut bacteria. *Nat Commun* **14**, 1663 (2023).

31 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).

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 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** (2019).

34 Liu, L., Hao, T., Xie, Z., Horsman, G. P. & Chen, Y. Genome mining unveils widespread natural product biosynthetic capacity in human oral microbe Streptococcus mutans. *Sci Rep* **6**, 37479 (2016).

35 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).

36 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).

37 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).

38 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).

39 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).

40 Barbour, A. *et al.* Discovery of phosphorylated lantibiotics with proimmune activity that regulate the oral microbiome. *Proc Natl Acad Sci U S A* **120**, e2219392120 (2023).

41 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).

42 Williams, M. R. *et al.* Quorum sensing between bacterial species on the skin protects against epidermal injury in atopic dermatitis. *Sci Transl Med* **11** (2019).

43 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).

44 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).

45 Fernandez-Cantos, M. V. *et al.* Bioinformatic mining for RiPP biosynthetic gene clusters in Bacteroidales reveals possible new subfamily architectures and novel natural products. *Front Microbiol* **14**, 1219272 (2023).

46 Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649-662 e620 (2019).

47 Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* **39**, 105-114 (2021).

48 Gurbich, T. A. *et al.* MGnify Genomes: A Resource for Biome-specific Microbial Genome Catalogues. *J Mol Biol*, 168016 (2023).

49 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).

50 Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214 (2012).

51 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87 (2019).

52 Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J Chem Inf Model* **50**, 742-754 (2010).

53 Human Microbiome Project, C. A framework for human microbiome research. *Nature* **486**, 215-221 (2012).

54 Jin, S., Wetzel, D. & Schirmer, M. Deciphering mechanisms and implications of bacterial translocation in human health and disease. *Current Opinion in Microbiology* **67** (2022).

55 Boopathi, S. *et al.* Expanding germ-organ theory: Understanding non-communicable diseases through enterobacterial translocation. *Pharmacol Res* **194**, 106856 (2023).

56 Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59-64 (2014).

57 Jie, Z. *et al.* The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun* **8**, 845 (2017).

58 Abu-Ali, G. S. *et al.* Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat Microbiol* **3**, 356-366 (2018).

59 Heintz-Buschart, A. *et al.* Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol* **2**, 16180 (2016).

60 Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541-546 (2013).

61 Ali, R. O. *et al.* Longitudinal multi-omics analyses of the gut-liver axis reveals metabolic dysregulation in hepatitis C infection and cirrhosis. *Nat Microbiol* **8**, 12-27 (2023).

62 Lamaudiere, M. T. F., Arasaradnam, R., Weedall, G. D. & Morozov, I. Y. The Colorectal Cancer Microbiota Alter Their Transcriptome To Adapt to the Acidity, Reactive Oxygen Species, and Metabolite Availability of Gut Microenvironments. *mSphere*, e0062722 (2023).

63 Li, P. *et al.* Gut Microbiota Dysbiosis Is Associated with Elevated Bile Acids in Parkinson's Disease. *Metabolites* **11** (2021).

64 Schirmer, M. *et al.* Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat Microbiol* **3**, 337-346 (2018).

65 Lloyd-Price, J. *et al.* Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655-662 (2019).

66 Pisani, A. *et al.* Dysbiosis in the Gut Microbiota in Patients with Inflammatory Bowel Disease during Remission. *Microbiol Spectr* **10**, e0061622 (2022).

67 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* (2024).

68 Deng, Z., Luo, X. M., Liu, J. & Wang, H. Quorum Sensing, Biofilm, and Intestinal Mucosal Barrier: Involvement the Role of Probiotic. *Front Cell Infect Microbiol* **10**, 538077 (2020).

69 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 U S A* **119**, e2202661119 (2022).

70 Wang, B. & Muir, T. W. Regulation of Virulence in Staphylococcus aureus: Molecular Mechanisms and Remaining Puzzles. *Cell Chem Biol* **23**, 214-224 (2016).

71 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).

72 Molloy, E. M. *et al.* Enzyme-Primed Native Chemical Ligation Produces Autoinducing Cyclopeptides in Clostridia. *Angew Chem Int Ed Engl* **60**, 10670-10679 (2021).

73 Verbeke, T. J. *et al.* Pentose sugars inhibit metabolism and increase expression of an AgrD-type cyclic pentapeptide in Clostridium thermocellum. *Sci Rep* **7**, 43355 (2017).

74 Bejder, B. S. *et al.* (bioRxiv, 2023).

75 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).

76 Zou, H., Sun, T., Jin, B. & Wang, S. sBGC-hm: an atlas of secondary metabolite biosynthetic gene clusters from the human gut microbiome. *Bioinformatics* **39** (2023).

77 Hatziioanou, D. *et al.* Discovery of a novel lantibiotic nisin O from Blautia obeum A2-162, isolated from the human gastrointestinal tract. *Microbiology (Reading)* **163**, 1292-1305 (2017).

78 Dabard, J. *et al.* Ruminococcin A, a new lantibiotic produced by a Ruminococcus gnavus strain isolated from human feces. *Appl Environ Microbiol* **67**, 4111-4118 (2001).

79 Haas, W., Shepard, B. D. & Gilmore, M. S. Two-component regulator of Enterococcus faecalis cytolysin responds to quorum-sensing autoinduction. *Nature* **415**, 84-87 (2002).

80 Lee, J. H. *et al.* Comparative genomic analysis of the gut bacterium Bifidobacterium longum reveals loci susceptible to deletion during pure culture growth. *BMC Genomics* **9**, 247 (2008).

81 Garber, K. Drugging the gut microbiome. *Nat Biotechnol* **33**, 228-231 (2015).

82 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).

83 King, A. M. *et al.* Systematic mining of the human microbiome identifies antimicrobial peptides with diverse activity spectra. *Nature Microbiology* (2023).

84 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* (2022).

85 Miller, M. B. & Bassler, B. L. Quorum sensing in bacteria. *Annu Rev Microbiol* **55**, 165-199 (2001).

86 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).

87 Chu, J. *et al.* Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat Chem Biol* **12**, 1004-1006 (2016).

88 Sturme, M. H. *et al.* Cell to cell communication by autoinducing peptides in gram-positive bacteria. *Antonie Van Leeuwenhoek* **81**, 233-243 (2002).

89 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).

90 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).

91 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).

92 Bajusz, D., Racz, A. & Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* **7**, 20 (2015).

93 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).

94 Narayan, A., Berger, B. & Cho, H. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nat Biotechnol* **39**, 765-774 (2021).

95 Steinegger, M. & Soding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026-1028 (2017).

96 Sangar, V., Blankenberg, D. J., Altman, N. & Lesk, A. M. Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics* **8** (2007).

97 Vasimuddin, M., Misra, S., Li, H. & Aluru, S. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314-324 (2019).

98 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).

99 Blanco-Miguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol* (2023).

100 Navarro-Munoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol* **16**, 60-68 (2020).

101 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).

102 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021).

Figure legends

Figure 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 RiPP BGCs and RiPP precursors. **(B)** BGCs identified by antiSMASH are grouped into eight BGC classes according to the classification grouped into eight BGC classes according to the classification rules of BiG-SCAPE, including RiPPs (410487, 74.5%), Terpene (14753, 2.7%), PKS/NRPS Hybrids (4067, 0.7%), Saccharides (17, 0.003%), NRPS (47281, 8.6%), PKS_others (5210,0.9%), PKS_I (1506,0.3%), and Others (67328, 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 novel RiPP biosynthetic enzyme(s) from a broader PTM enzyme dataset collected by decRiPPter[49]. (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.

Figure 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. Among the retained precursors, 47.3% (200,340) were identified by both DeepRiPP and TrRiPP and located in a RiPP BGC region, 12.1% (51,207) were predicted by DeepRiPP and located in a RiPP BGC region, 3.0% (12,809) were predicted by TrRiPP and located in a RiPP BGC region, and 37.6% (159,475) were detected by both TrRiPP and DeepRiPP. Of note, the RiPP BGC region was defined by antiSMASH. (B) The barplot on the outside shows the number of precursor sequences for nine different RiPP classes. The stacked barplot on the inside 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). Each dot represents a precursor sequence. (D) Multi-Dimensional Scaling (MDS) plot visualizes the hypervariable chemical space of predicted mature precursors from nine different RiPP classes. The size of each dot is proportional to the number of unique precursor sequences within the respective class. The color of each dot represents the class-wise median Tanimoto coefficient, which indicates the similarity between precursor sequences within the class. The distance between dots reflects the similarity between different RiPP classes.

Figure 3 Classify the novelty of RiPP families based on their precursor and genomic neighborhood.

(A) Novelty of the RiPP precursor in each family. The chord diagram illustrates the novelty of identified precursor families (left panel) for nine different RiPP classes (right panel). Left panel: The "MIBiG" category represents precursor families that exhibit homology to experimentally characterized RiPP precursors deposited in the MIBiG database. Precursor families that contain well-defined RiPP-associated domains are highlighted in black. These domains include AgrD, rSAM-modified RiPPs-associated domains, and other domains such as Graspetide-associated (3 families), Lanthipeptide-associated (31 families), Lassopeptide-associated (26 families), Thiopeptide-associated (5 families), RiPP like-associated (25 families), LAP-associated (19 families), and other known RiPPs-associated (2 families). Those precursor families that do not belong to the above categories as "Novel RiPP". The numbers in brackets represent the number of precursor families within each category. (B) Workflow for classifying the novelty of RiPP families based on the novelty of precursor and genomic context **(Methods, Supplementary Figure 5)**. (C) The number and proportion of RiPP families under each condition integrating RiPP precursor and genomic neighborhood. The novelty of RiPP families was further classified into classic RiPP, Novel RiPP, and Others. The classic RiPP category includes families with known precursor homology and a well-defined genomic neighborhood. The Novel RiPP category includes families with no known precursor homology and/or with a novel genomic neighborhood. The Others category includes families that do not fit into either of the above categories.

Figure 4. Identified RiPP families are present and transcribed in the healthy human microbiome.

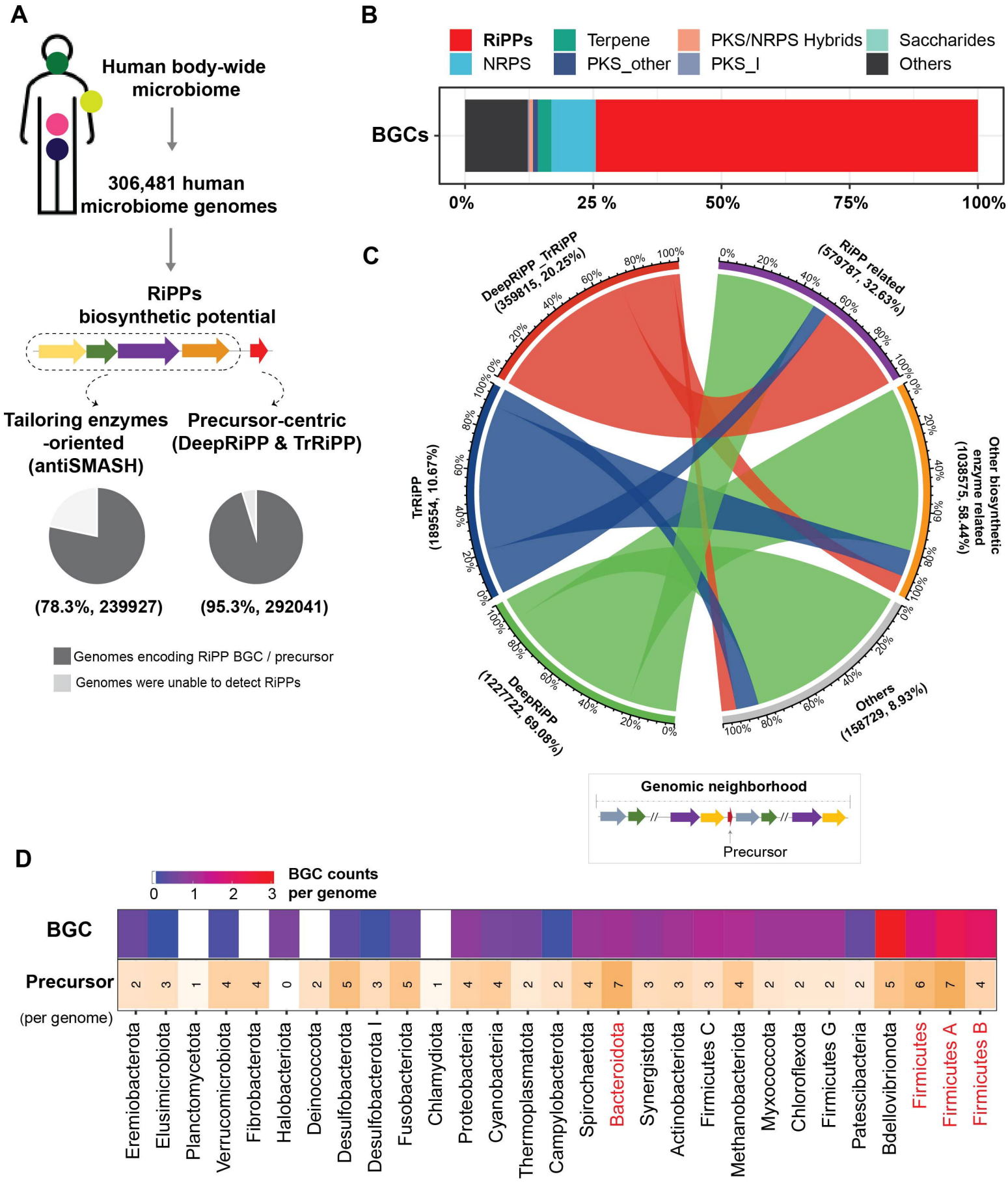
(A) The prevalence and average abundance of RiPP precursor families in six different human body sites. Each dot represents one RiPP precursor family, with predicted bioactive families shown in red and families of unknown function shown in black. The numbers in red represent the number of bioactive families and the total number of families detected in the specific site. (B &C) Niche-specificity of identified RiPP families in the healthy human microbiome. (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) Forecasted activity of RiPP families. The scale represents the proportion of each predicted activity (upper) or RiPP type based on novelty (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 healthy human gut microbiome. The bar plot illustrates the number of RiPP precursor families identified in the metagenome (black bars) and the metatranscriptome (blue bars). The pie charts present the percentage of transcribed RiPP families in each class that predicted bioactivity (red) or unknown (grey).

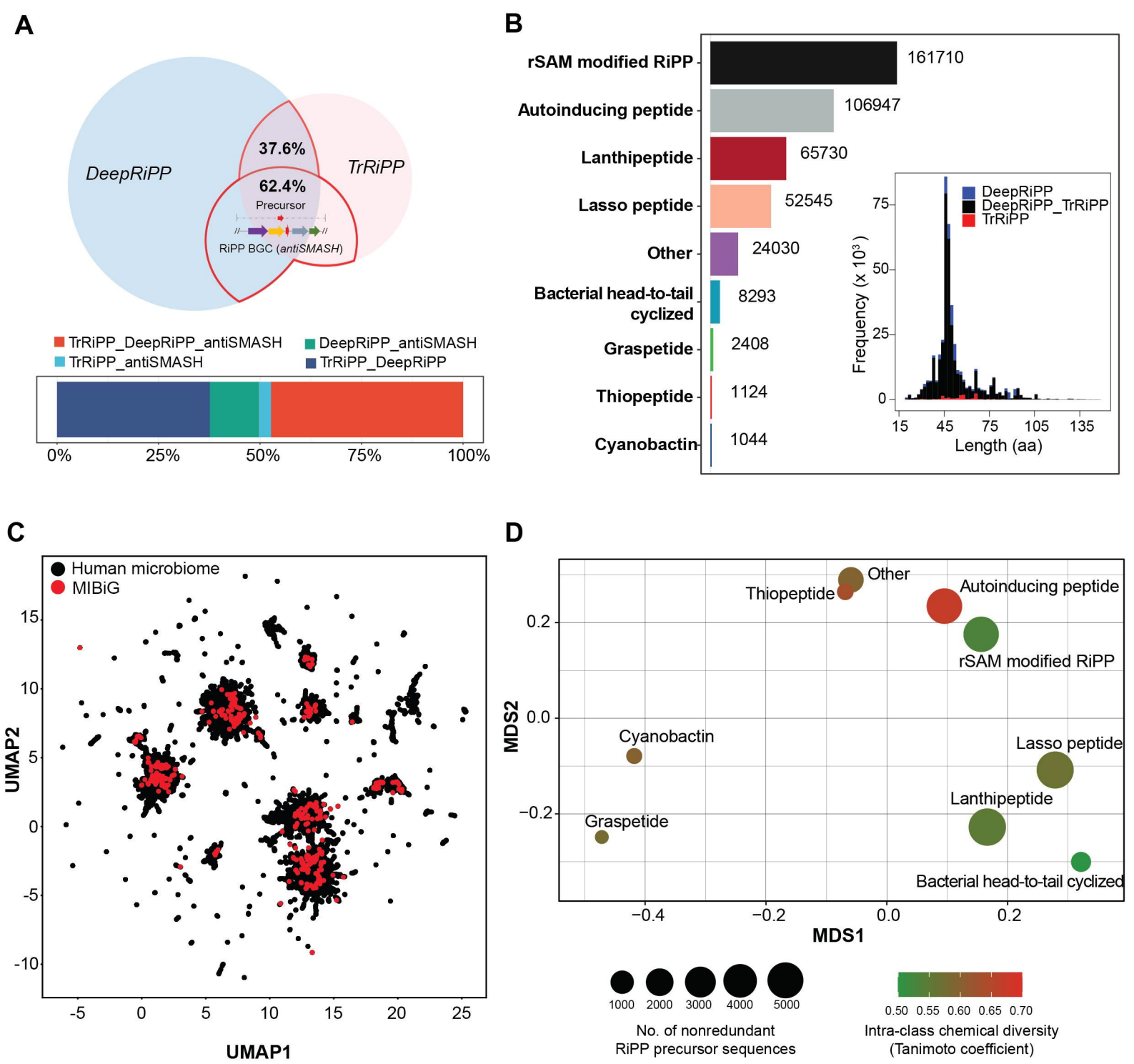
Figure 5 Differentially transcripted RiPP families in multi-disease case-control cohorts.

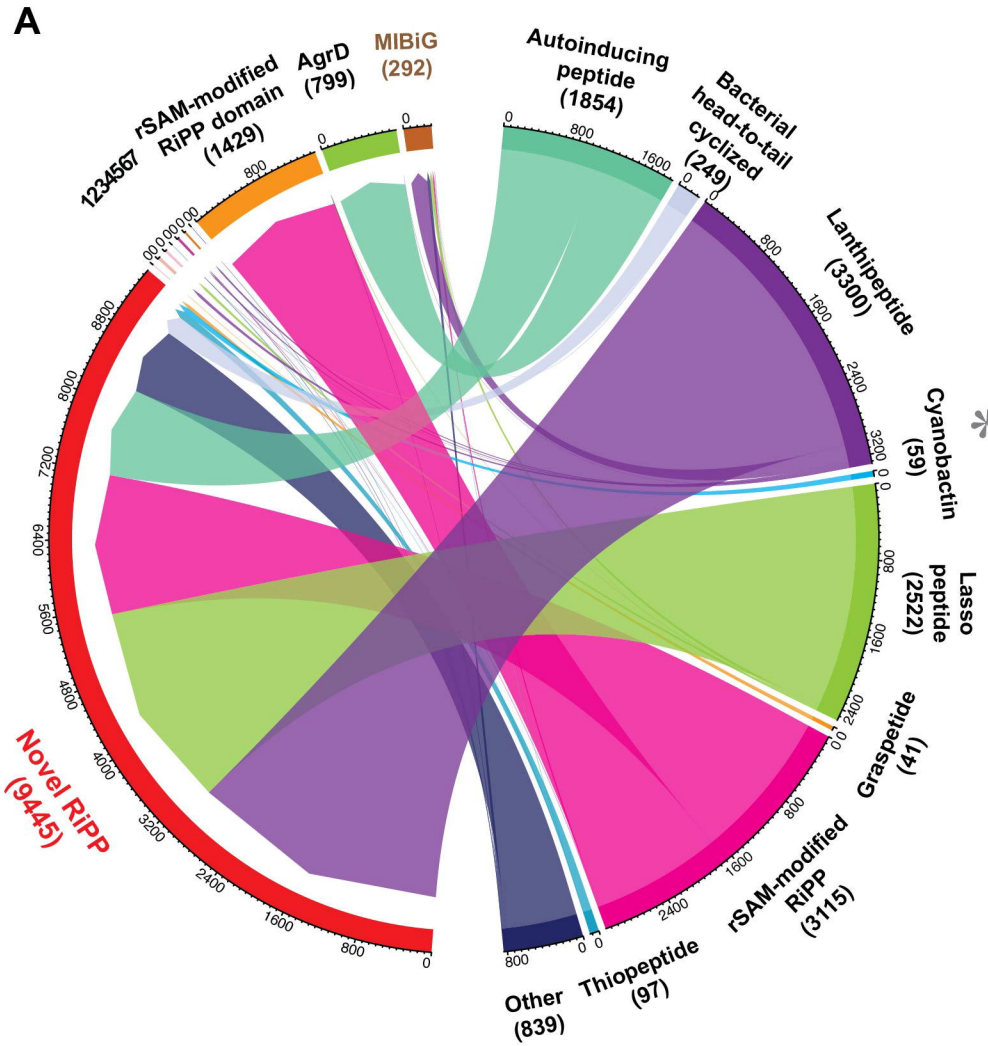
(A) The bar plot displays the number of RiPP precursor families that are significantly enriched in the healthy group (top) or the disease group (bottom). Statistical significance was determined using criteria of |log2 fold change| ≥1 and adjusted *p* values ≤ 0.05. (B & C) The intersection of differentially transcribed RiPP precursor families that are either depleted (B) or enriched (C) in multiple disease groups compared to healthy groups is shown. The bar plot on the top, along with connecting lines, represents the number of RiPP precursor families identified in the corresponding disease cohorts. Precursor families are highlighted in different colors based on their RiPP classes. (D) Biomarker families in IBD disease. Left panel: The performance of the random forest classifiers based on RiPP abundance in discriminating the health group from the disease group. Right panel: The box plot on the left side illustrates the top 10 families with the greatest contribution to the discriminative model. The corresponding log2 fold change is depicted in the bar plot on the right side. The red bar represents precursor families enriched in the health group, while the blue bar represents RiPP families enriched in the CD group.

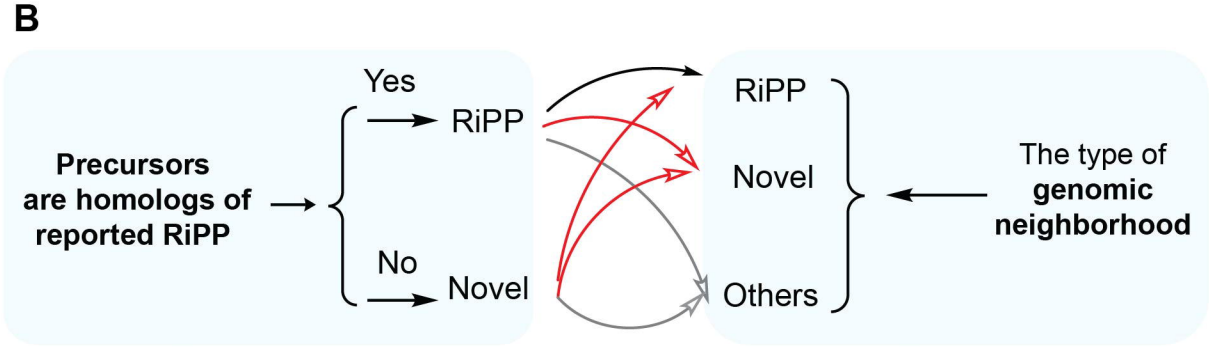
Figure 6 3D-structure of precursor-AgrB complex and precursor sequence conservation guided mature AIP chemical synthesis.

Upper: The workflow about how to predict the mature AIP using the reported AIP-I from *Staphylococcus aureus* as an illustrative example. (1) The depicted complex represents the predicted interaction between the precursor peptide and AgrB enzyme. (2) This step focuses on identifying potentially modified residues between cysteine or serine and another amino acid with a hydrophobic side chain. (3) To determine the exocyclic region of mature AIP, we analyzed the conservation patterns of targeted precursor families. This analysis can provide insights into the structure of the exocyclic region. **Bottom:** Putative mature AIPs chemically synthesized in this study. The circled region within the complex corresponds to the precursor sequence, the potentially modified core region is highlighted in red. Sequence logo analysis reveals a high level of conservation within this core region across the precursor family, which aids in predicting the exocyclic amino acids. The structures displayed on the right side represent the putative mature AIP products.


A
Human body-wide microbiome
306,481 human microbiome genomes
RiPPs biosynthetic potential
Tailoring enzymes -oriented (antiSMASH)
Precursor-centric (DeepRiPP & TrRiPP)
(78.3%, 239927)
(95.3%, 292041)
Genomes encoding RiPP BGC / precursor
Genomes were unable to detect RiPPs
B
RiPPs
NRPS
Terpene
PKS_other
PKS/NRPS Hybrids
PKS_I
Saccharides
Others
BGCs
0%
25 %
50%
75%
100%
C
DeepRiPP_TrRiPP (359815, 20.25%)
RiPP related (579787, 32.63%)
Other biosynthetic enzyme related (1038575, 58.44%)
Others (158729, 8.93%)
DeepRiPP (1227722, 69.08%)
TrRiPP (189554, 10.67%)
Genomic neighborhood
Precursor
D
BGC counts per genome
0 1 2 3
BGC
Precursor
(per genome)
Eremiobacterota 2
Elusimicrobiota 3
Planctomycetota 1
Verrucomicrobiota 4
Fibrobacterota 4
Halobacteriota 0
Deinococcota 2
Desulfobacterota 5
Desulfobacterota I 3
Fusobacteriota 5
Chlamydiota 1
Proteobacteria 4
Cyanobacteria 4
Thermoplasmatota 2
Campylobacterota 2
Spirochaetota 4
Bacteroidota 7
Synergistota 3
Actinobacteriota 3
Firmicutes C 3
Methanobacteriota 4
Myxococcota 2
Chloroflexota 2
Firmicutes G 2
Patescibacteria 2
Bdellovibrionota 5
Firmicutes 6
Firmicutes A 7
Firmicutes B 4

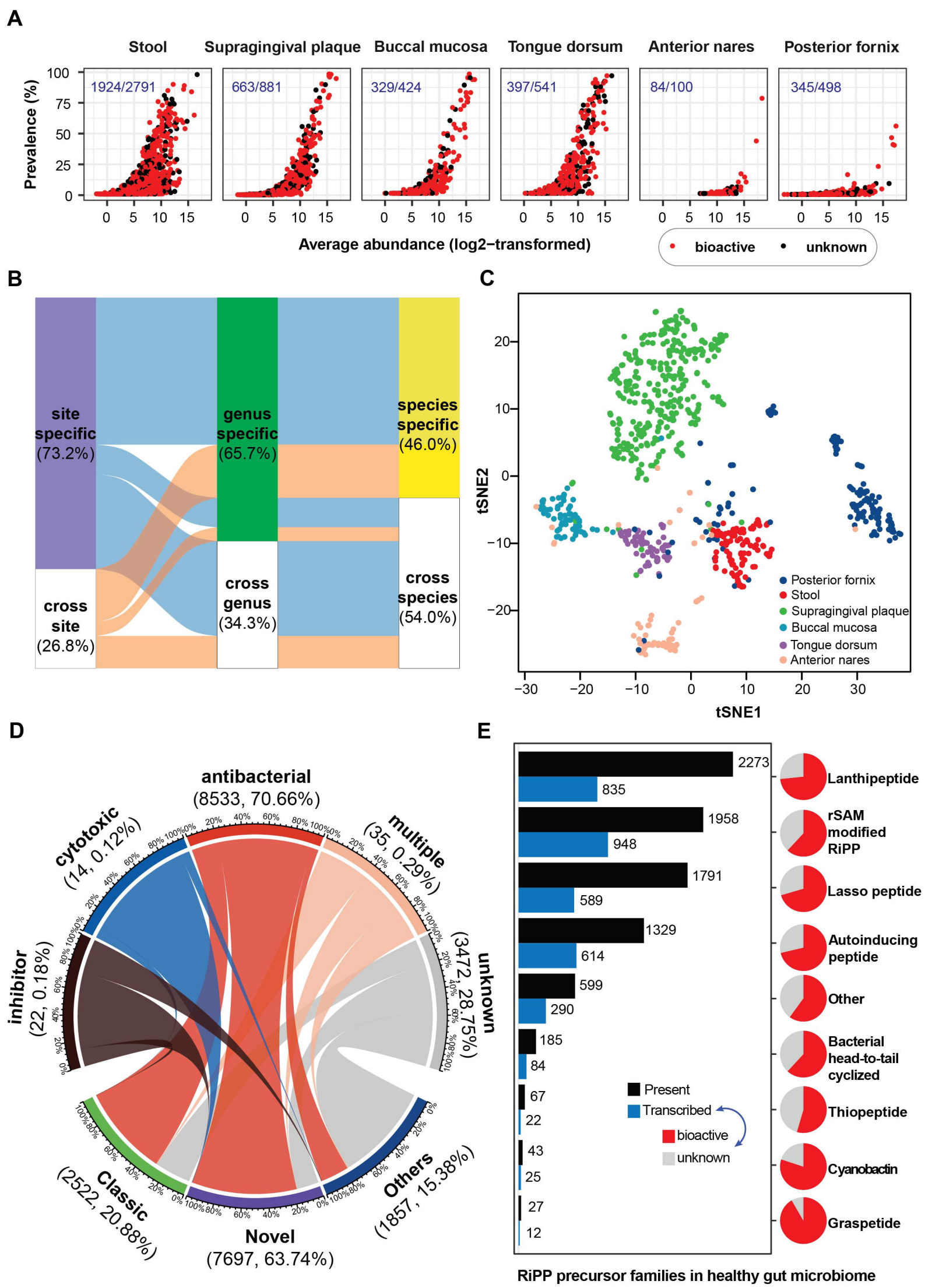

A
DeepRiPP
TrRiPP
37.6%
62.4%
Precursor
RiPP BGC (antiSMASH)
TrRiPP_DeepRiPP_antiSMASH
DeepRiPP_antiSMASH
TrRiPP_antiSMASH
TrRiPP_DeepRiPP
0%
25%
50%
75%
100%
B
rSAM modified RiPP
161710
Autoinducing peptide
106947
Lanthipeptide
65730
Lasso peptide
52545
Other
24030
Bacterial head-to-tail cyclized
8293
Graspetide
2408
Thiopeptide
1124
Cyanobactin
1044
DeepRiPP
DeepRiPP_TrRiPP
TrRiPP
Frequency (x 10³)
Length (aa)
C
Human microbiome
MIBiG
UMAP2
UMAP1
D
Other
Thiopeptide
Autoinducing peptide
rSAM modified RiPP
Cyanobactin
Lasso peptide
Lanthipeptide
Graspetide
Bacterial head-to-tail cyclized
MDS2
MDS1
1000
2000
3000
4000
5000
No. of nonredundant RiPP precursor sequences
0.50
0.55
0.60
0.65
0.70
Intra-class chemical diversity (Tanimoto coefficient)

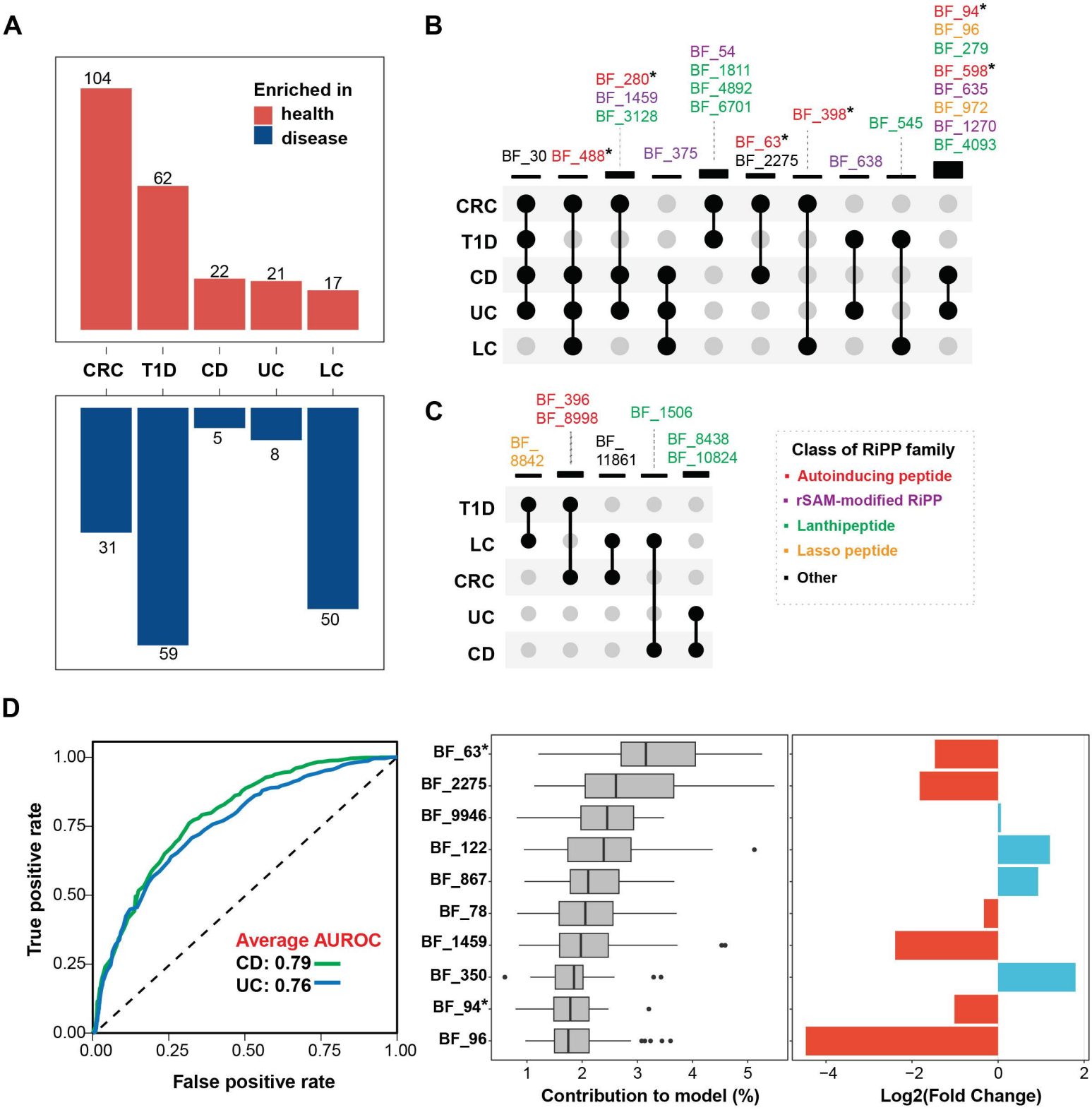


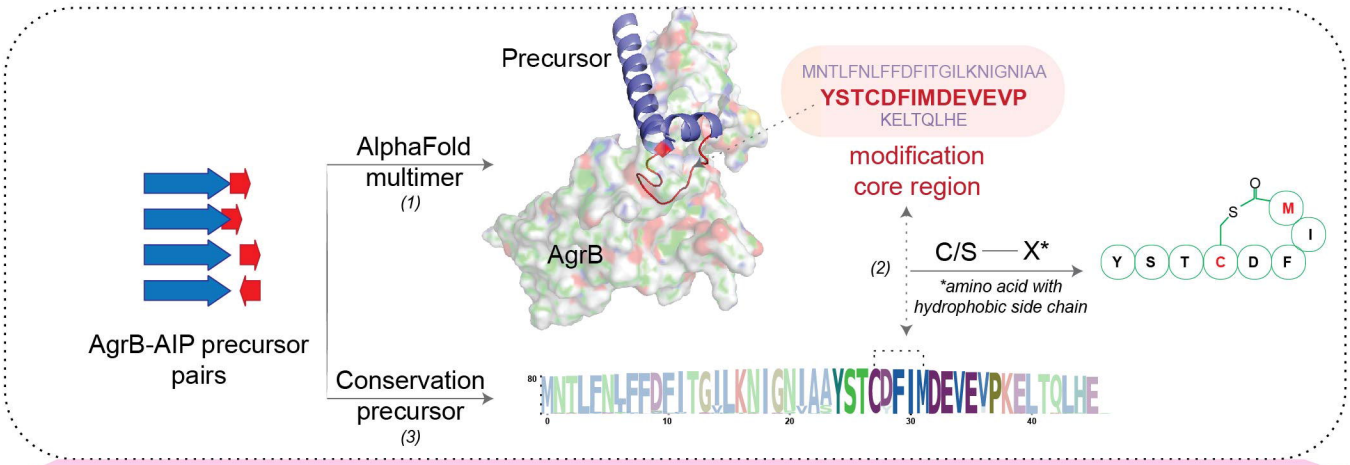


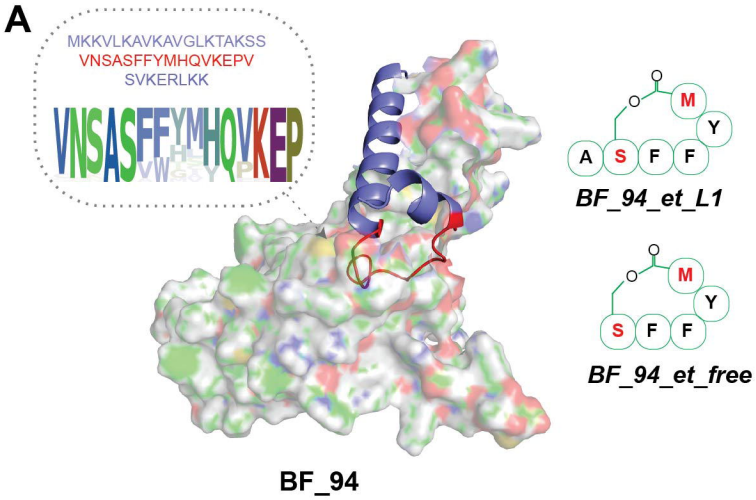
C

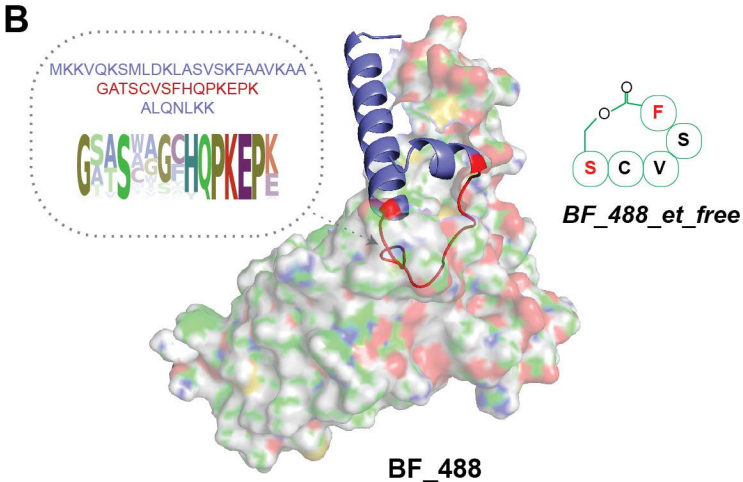
Number,proportion	RiPP precursor	Genomic neighbourhood	Final type
(2522, 20.88%)	RiPP	RiPP	Classic RiPP
(4829, 39.99%)	Novel	RiPP	Novel RiPP
(67, 0.55%)	RiPP	Novel	
(2801, 23.19%)	Novel	Novel	
(42, 0.35%)	RiPP	Others	Others
(1815, 15.02%)	Novel	Others	

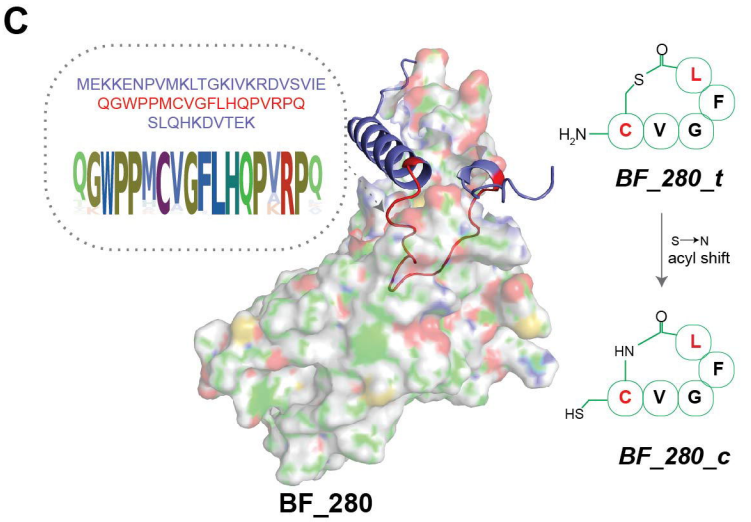

A
Stool
Supragingival plaque
Buccal mucosa
Tongue dorsum
Anterior nares
Posterior fornix
Prevalence (%)
1924/2791
663/881
329/424
397/541
84/100
345/498
Average abundance (log2−transformed)
bioactive
unknown
B
site specific (73.2%)
cross site (26.8%)
genus specific (65.7%)
cross genus (34.3%)
species specific (46.0%)
cross species (54.0%)
C
tSNE1
tSNE2
Posterior fornix
Stool
Supragingival plaque
Buccal mucosa
Tongue dorsum
Anterior nares
D
antibacterial (8533, 70.66%)
multiple (35, 0.29%)
unknown (3472, 28.75%)
Others (1857, 15.38%)
Novel (7697, 63.74%)
Classic (2522, 20.88%)
inhibitor (22, 0.18%)
cytotoxic (14, 0.12%)
E
Present
Transcribed
bioactive
unknown
Lanthipeptide
rSAM modified RiPP
Lasso peptide
Autoinducing peptide
Other
Bacterial head-to-tail cyclized
Thiopeptide
Cyanobactin
Graspetide
RiPP precursor families in healthy gut microbiome

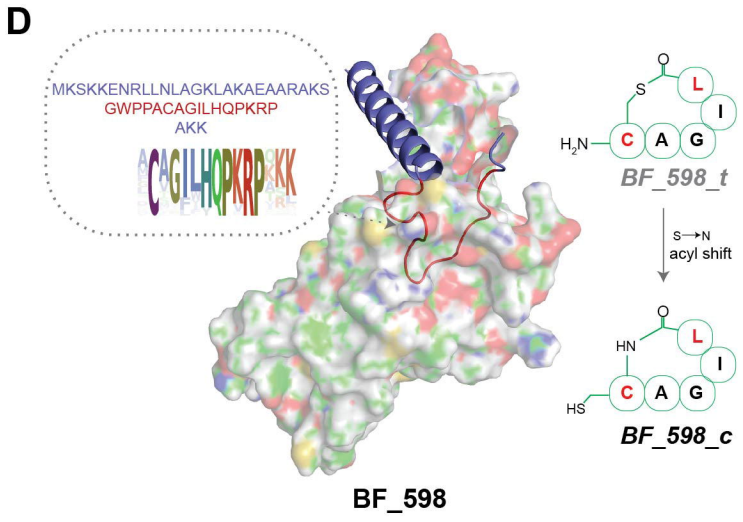

A
Enriched in
health
disease
104
62
22
21
17
CRC
T1D
CD
UC
LC
31
59
5
8
50
B
BF_30
BF_488*
BF_280*
BF_1459
BF_3128
BF_375
BF_54
BF_1811
BF_4892
BF_6701
BF_63*
BF_2275
BF_398*
BF_638
BF_545
BF_94*
BF_96
BF_279
BF_598*
BF_635
BF_972
BF_1270
BF_4093
C
BF_396
BF_8998
BF_8842
BF_11861
BF_1506
BF_8438
BF_10824
Class of RiPP family
Autoinducing peptide
rSAM-modified RiPP
Lanthipeptide
Lasso peptide
Other
D
True positive rate
False positive rate
Average AUROC
CD: 0.79
UC: 0.76
BF_63*
BF_2275
BF_9946
BF_122
BF_867
BF_78
BF_1459
BF_350
BF_94*
BF_96
Contribution to model (%)
Log2(Fold Change)

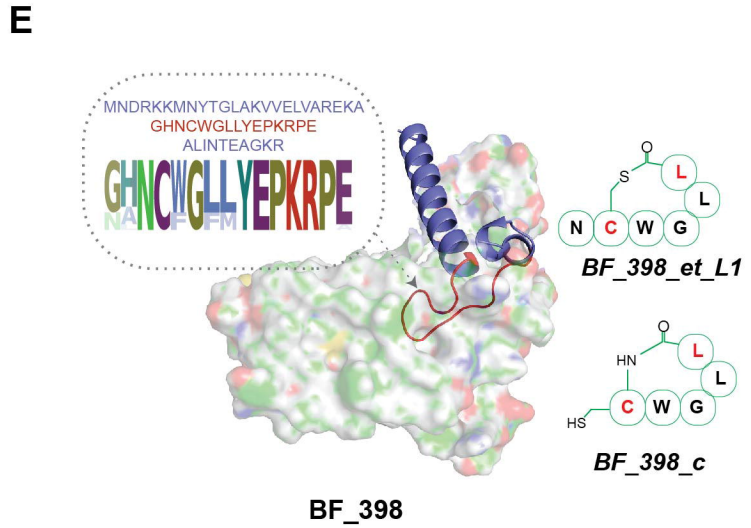

AgrB-AIP precursor pairs
AlphaFold multimer
(1)
Precursor
AgrB
MNTLFNLFFDFITGILKNIGNIAA
YSTCDFIMDEVEVP
KELTQLHE
modification core region
(2)
C/S — X*
*amino acid with hydrophobic side chain
Conservation precursor
(3)

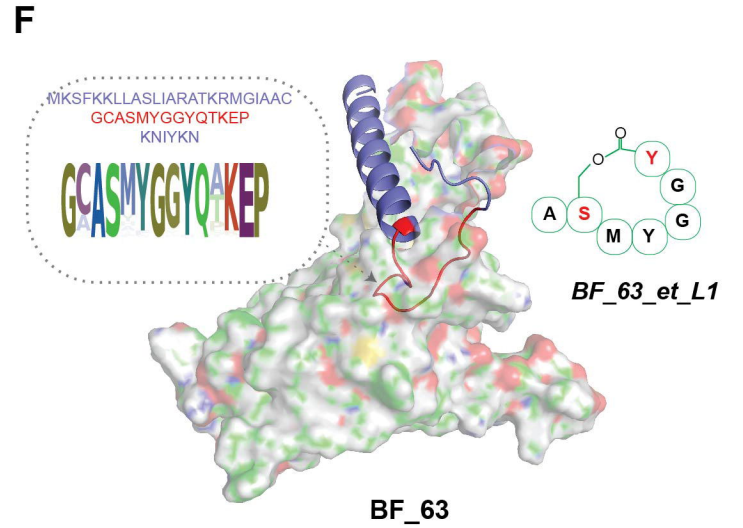

A
MKKVLKAVKAVGLKTAKSS
VNSASFFYMHQVKEP
SVKERLKK
BF_94_et_L1
BF_94_et_free
BF_94


B
MKKVQKSMLDKLASVSKFAAVKAA
GATSCVSFHQPKEPK
ALQNLKK
BF_488_et_free
BF_488


C
MEKKENPVMKLTGKIVKRDVSVIE
QGWPPMCVGFLHQPVRPQ
SLQHKDVTEK
BF_280_t
S→N
acyl shift
BF_280_c
BF_280


D
MKSKKENRLLNLAGKLAKAEAARAKS
GWPPACAGILHQPKRP
AKK
BF_598_t
S→N
acyl shift
BF_598_c
BF_598


E
MNDRKKMNYTGLAKVVELVAREKA
GHNCWGLLYEPKRPE
ALINTEAGKR
BF_398_et_L1
BF_398_c
BF_398


F
MKSFKKLLASLIARATKRMGIAAC
GCASMYGGYQTKEP
KNIYKN
BF_63_et_L1
BF_63