

SCIENTIFIC REPORTS



OPEN

Genome mining unveils widespread natural product biosynthetic capacity in human oral microbe *Streptococcus mutans*

Received: 06 September 2016

Accepted: 28 October 2016

Published: 21 November 2016

Liwei Liu^{1,*}, Tingting Hao^{1,2,*}, Zhoujie Xie¹, Geoff P. Horsman³ & Yihua Chen^{1,2}

Streptococcus mutans is a major pathogen causing human dental caries. As a Gram-positive bacterium with a small genome (about 2 Mb) it is considered a poor source of natural products. Due to a recent explosion in genomic data available for *S. mutans* strains, we were motivated to explore the natural product production potential of this organism. Bioinformatic characterization of 169 publically available genomes of *S. mutans* from human dental caries revealed a surprisingly rich source of natural product biosynthetic gene clusters. Anti-SMASH analysis identified one nonribosomal peptide synthetase (NRPS) gene cluster, seven polyketide synthase (PKS) gene clusters and 136 hybrid PKS/NRPS gene clusters. In addition, 211 ribosomally synthesized and post-translationally modified peptides (RiPPs) clusters and 615 bacteriocin precursors were identified by a combined analysis using BAGEL and anti-SMASH. *S. mutans* harbors a rich and diverse natural product genetic capacity, which underscores the importance of probing the human microbiome and revisiting species that have traditionally been overlooked as “poor” sources of natural products.

Over 60% of drugs on the market are derived from natural products, and 38% are of microbial origin, primarily from actinobacteria and fungi^{1,2}. The last three decades have witnessed tremendous progress in elucidating the biosynthetic mechanisms of varied microbial natural products including compounds synthesized by polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), and the ribosomally synthesized and post-translationally modified peptides (RiPPs). This genes-to-molecules foundational knowledge has enabled the identification of cryptic natural product gene clusters from microbial genomes and facilitated genomics guided discovery of novel natural products.

PKSs are usually classified into three types: type I PKSs are large modular proteins containing various functional domains; type II PKSs are single or bifunctional proteins used in an iterative fashion; and type III PKSs employ coenzyme A-tethered substrates instead of the acyl carrier proteins (ACP) recruited by types I and II PKS systems. A typical type I PKS module minimally contains acyltransferase (AT), ACP (also called thiolation domain, T) and ketosynthase (KS) domains. Additional PKS domains include ketoreductase (KR), dehydratase (DH), enol reductase (ER) as well as some rare functional domains that together generate an extraordinary diversity of polyketide skeletons^{3,4}. Similar to type I PKSs, NRPSs are also modular proteins minimally comprising adenylation (A), peptidyl carrier protein (PCP, also called thiolation domain T) and condensation (C) domains. Examples of additional functional domains include cyclase (Cy), epimerase (E) and methyltransferase (MT) domains⁵. A thioesterase (TE) domain or a terminal reductase (R) domain releasing the mature polyketide or polypeptide chain from the biosynthetic assembly line is commonly found at the carboxyl terminus⁶. In addition, a phosphopantetheinyl transferase (PPTase) is necessary to type I and II PKS and NRPS proteins, which converts them from apo- to functional holo-form by transferring a phosphopantetheinyl arm from coenzyme A to the Ser residue of their thiolation domains⁷.

RiPPs represent a large family of natural products derived from ribosomally synthesized polypeptide precursors, which usually consist of an N-terminal leader peptide and a core peptide. The leader peptide is essential

¹State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China. ²University of Chinese Academy of Sciences, Beijing, 100049, China. ³Department of Chemistry and Biochemistry, Wilfrid Laurier University, Waterloo, ON, N2L3C5, Canada. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Y.C. (email: chenjihua@im.ac.cn)

for guiding post-translational modification enzymes and transporters acting on the precursor peptide, and the core peptide becomes the final product after removal of the leader peptide by proteases⁸. The bacteriocins are an important group of RiPPs that can be classified as either modified or unmodified based on post-translational modifications to their core peptides. To date, a number of bacteriocins with diverse structures and bioactivities are generated by the action of various modification enzymes such as Ser/Thr dehydratases, methyltransferases, and macrocyclases^{9–12}.

As microbial genome sequence space rapidly expands, a genomics guided approach is increasingly important for natural product discovery. Analysis of diverse bacterial genomes demonstrated statistically that the number of natural product gene clusters in a strain increases with the size of its genome, implying that bacteria with genome sizes less than 2 Mb rarely contain polyketide and nonribosomal peptide biosynthetic gene clusters and are poor sources of natural products^{13,14}. Consequently, natural product genome mining efforts have focused almost exclusively on bacteria with large genomes such as cyanobacteria and actinomycetes from terrestrial or aquatic environments^{15–17}. Until recently the human microbiome has seldom been considered as a source of natural products. However, *in silico* analysis of thousands of bacterial genomes from the human microbiome revealed that the gut and oral cavity are the two richest habitats for bacteria harboring numerous natural product gene clusters. These microbes are distributed across diverse bacterial families, including some with small genome sizes, which suggests that some human-associated bacteria with small genome sizes could represent an overlooked source of natural products¹⁸.

Streptococcus mutans is a Gram-positive facultative anaerobe considered as a major contributor to dental cavities in humans and animals. Acid resistance, biofilm formation, and production of bacteriocins are some of the characteristics that contribute to successful colonization and proliferation of *S. mutans* in the human oral cavity. Several bioactive compounds have recently been isolated from *S. mutans*, including mutanobactins and mutacins I–IV^{19–24}. Mutanobactins are compounds of hybrid PKS/NRPS origin that inhibit the morphological transition of *Candida albicans*, an opportunistic human pathogen¹⁹. Mutacins I–IV are bacteriocins exhibiting antibacterial activity against closely related species and some other Gram-positive bacteria²⁵.

Despite the isolation of several compounds from *S. mutans*, this species has seldom been considered as a potential source of natural products due to its relatively small genome size (about 2 Mb). Recent genome sequencing projects have provided a great deal of genomic data for many strains of *S. mutans* that enables a comprehensive survey of its potential natural product biosynthetic capacity. Herein we report the identification of 355 natural product gene clusters from 169 *S. mutans* genomes, which reveal this species as a rich source of natural products in the human microbiome and serves as an important reminder to re-examine the natural product biosynthetic capacity of bacteria that have been previously neglected due to small genome sizes.

Results

An overview of natural product biosynthetic gene clusters in *S. mutans*. A total of 169 *S. mutans* genomes were analyzed with anti-SMASH and BAGEL, which identified 355 natural product gene clusters distributed among 138 *S. mutans* strains (Supplementary Tables SI, SII and SIII). The identified natural product gene clusters belonged to NRPS (1), PKS (7), PKS/NRPS hybrid (136) and RiPP (211) families (Supplementary Tables SI and SII). The number of gene clusters in each positive *S. mutans* genome varied from one to five, with 66 (39%) strains containing more than two (Supplementary Table SI).

Natural product gene clusters possessing modular proteins (including NRPS, PKS and hybrid PKS/NRPS) were found in 114 (67.4%) strains, 29 (17.2%) of which harbored no less than two gene clusters. For example, the six strains *S. mutans* 4SM1, *S. mutans* 4VF1, *S. mutans* 15VF2, *S. mutans* 1SM1, *S. mutans* NLML1 and *S. mutans* SA38 contained one PKS gene cluster and one hybrid PKS/NRPS gene cluster, and three clusters (1 PKS, 1 NRPS and 1 hybrid PKS/NRPS) were discovered in *S. mutans* B24Sm2 occupying 5.4% of its 2.28 Mb genome.

In addition, 211 RiPPs gene clusters were also detected, including 32 modified bacteriocin (lantibiotic) gene clusters from 28 strains and 179 unmodified bacteriocin gene clusters from 116 strains (Supplementary Table SII). All of the modified bacteriocin gene clusters were typical lantibiotic gene clusters, and three strains, *S. mutans* R221, *S. mutans* ST1 and *S. mutans* B107SM-B had more than one lantibiotic gene clusters. At least two unmodified bacteriocin gene clusters were identified in 46 (27.2%) strains.

The hybrid PKS/NRPS gene clusters in *S. mutans*. Based on gene composition and organization, 136 identified hybrid gene clusters ranging from 12.6 kb to 36.2 kb in length were classified into six groups shown in Fig. 1, Supplementary Tables SI and SIII. Group I consists of 58 homologous hybrid PKS/NRPS gene clusters sharing greater than 78% nucleotide sequence identity. The gene clusters in this group contain 12 genes including one loading domain, one PKS module, six NRPS modules and a set of tailoring enzymes at the end of this cluster. The only known products of group I are the aforementioned mutanobactins isolated from *S. mutans* UA159^{19,20,26}, implying that similar compounds are encoded by group I gene clusters.

In total, 22 gene clusters belonging to group II were identified, and these possessed five genes encoding proteins with more than 50% identity to homologs in the reutericyclin gene cluster from *Lactobacillus reuteri* TMW1.106, TMW1.112 and TMW1.656 (Fig. 2). Genes *rtcN* and *rtcK* were proposed to respectively encode an NRPS and a PKS that catalyze installation of the core structure of reutericyclin; genes *rtcABC* encode an acyltransferase cassette responsible for the addition of an acetyl group²⁷. The high similarity of sequence and gene organization suggests that the group II gene clusters encode a compound similar to reutericyclin (Fig. 2). Furthermore, four extra genes in group II clusters encode two regulatory proteins, one transporter, and one protein of unknown function that might be involved in regulation, resistance or modification of the end product.

The third group consists of 29 hybrid PKS/NRPS gene clusters. Three modular proteins encoded by group III clusters comprise two PKS modules, three NRPS modules and a terminal reductase domain for a reductive release of the thioester-tethered product. Group III clusters also encode six transporters, an O-methyltransferase,

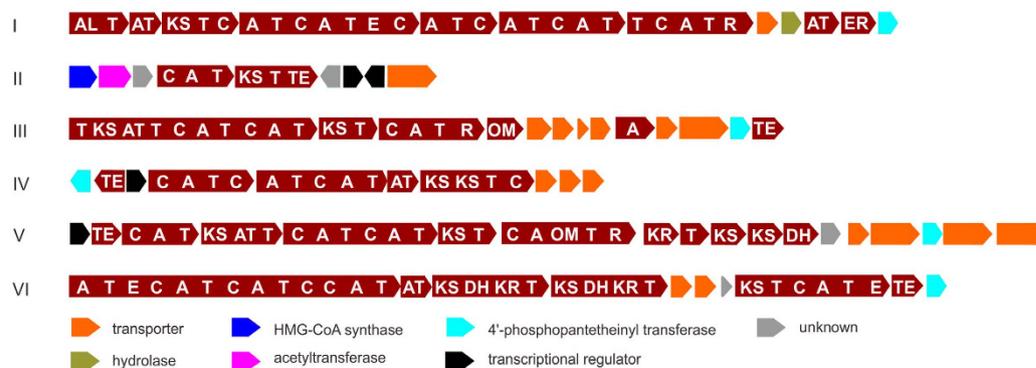


Figure 1. The 136 hybrid PKS/NRPS gene clusters identified in this study. The functional domains in NRPS and PKS proteins are indicated in bold: A, adenylation domain; AL, loading domain; AT, acyl-transferase; C, condensation domain; DH, dehydratase; E, epimerization; KR, ketoreductase; KS, ketosynthase; OM, O-methyltransferase; R, terminal reductase; T, thiolation domain (including both peptidyl carrier protein and acyl carrier protein); TE, thioesterase. The other functional enzymes and transporter proteins are indicated by arrows with different colors.

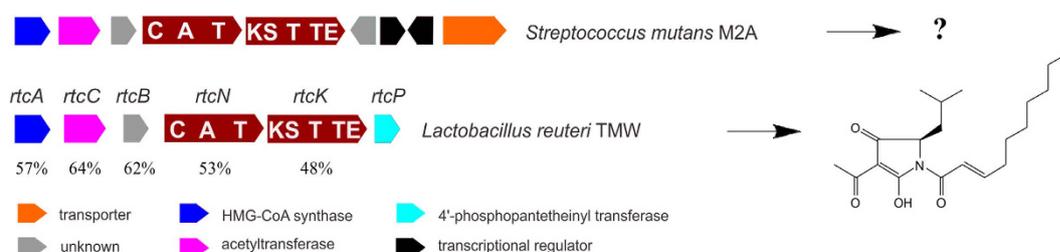


Figure 2. The biosynthetic gene cluster of reutericyclin in *Lactobacillus reuteri* TMW and one example of the group II hybrid PKS/NRPS gene clusters from *S. mutans* M2A. The amino acid identities between two gene clusters are shown under the genes of the reutericyclin cluster. The functional domains in PKSs and NRPSs are indicated in bold: A, adenylation domain; C, condensation domain; KS, ketosynthase; T, thiolation domain; TE, thioesterase. The other functional enzymes and transporters are indicated by arrows with different colors.



Figure 3. *S. mutans* B24Sm2 possesses one NRPS gene clusters identified by anti-SMASH. NRPS functional are indicated in bold: A, adenylation domain; C, condensation domain; E, epimerase domain; T, thiolation domain; X, HXXPF repeat domain. The other functional enzymes and transporter proteins are indicated by arrows with different colors.

an adenylation domain, a PPTase and a type II TE that may function in an editing role to hydrolyze improperly loaded elongation units from the thiolation domains.

Groups IV and V hybrid PKS/NRPS gene clusters are always present concomitantly and were identified in four strains: *S. mutans* NMT4863, *S. mutans* N29, *S. mutans* 5SM3 and *S. mutans* B112SM-A. Group VI hybrid PKS/NRPS gene clusters have 19 members, and each is composed of four genes encoding modular proteins with totally five NRPS modules and three PKS modules. Genes encoding AT, TE, transporter and PPTase proteins are also found in the group VI clusters. Significantly, hybrid PKS/NRPS gene cluster groups III-VI are unlike any known clusters with identified products, indicating that they produce novel compounds that await characterization.

The NRPS gene clusters in *S. mutans*. Only one NRPS gene clusters were identified from 169 *S. mutans* genomes, and it was discovered in *S. mutans* B24Sm2 (Fig. 3). This new NRPS gene cluster comprises seven genes, two of which encode large NRPS complexes possessing four modules. Specifically, the third NRPS module is missing of the adenylation domain, but has a functionally unknown HxxPF repeat domain in front of two repeated T domains. A condensation domain is found at the C-terminus instead of a TE or a reductase domain, suggesting that the peptide product may be released from the NRPS machinery by a condensation reaction as reported previously²⁸.



Figure 4. Detection by anti-SMASH of seven PKS gene clusters from 169 *S. mutans* genomes. Two complete gene clusters were found from *S. mutans* A38 and B24Sm2. The other five gene clusters, which are cut off by gaps marked with a green rectangle, were found in *S. mutans* 4VF1, 4SM1, 15VF2, 1SM1 and NLML1. The HMGS cassette found in the PKS gene clusters is enclosed by a pink rectangle. The functional domains in PKS proteins are indicated in bold: AL, loading domain; AT, acyltransferase; ATd, AT docking domains (ATd) that serve as locations for interaction with the free-standing AT that loads building blocks to the thiolation domain²⁹. Surprisingly, a methyltransferase domain replaces the typical TE or R domain at the C-terminus of the PKS, indicating an unusual polyketide chain release mechanism. In addition to the modular PKS genes, a PPase gene was found at the 5'-end followed by a putative hydrolase and HMG-CoA synthase cassette for installing branched polyketide chains³⁰. Three transporters encoded by genes at the 3'-end of the cluster may export the product out of the cell. The other five strains (*S. mutans* 4VF1, *S. mutans* 4SM1, *S. mutans* 15VF2, *S. mutans* NLML-1 and *S. mutans* 1SM1) harbor PKS clusters that are split into two parts by a gap (green box in Fig. 4) that normally contains the ACP (T) and KS domains of the second modular protein and the ATd and DH domains of the third modular protein. Because only draft sequence data are available for these five strains and the gaps are all at the ends of the sequencing contigs, it is most likely that the five PKS clusters are split artificially.

The PKS gene clusters in *S. mutans*. Only one group of type I PKS gene cluster was identified, which distributes in 7 of the 169 analyzed strains and has two versions (Fig. 4). In *S. mutans* A38 and *S. mutans* B24Sm2, the PKS gene clusters span about 70 kb and contain 8 genes encoding proteins with 14 AT-less (or trans-AT) PKS modules. In AT-less PKS systems, the PKS modular proteins possess AT docking domains (ATd) that serve as locations for interaction with the free-standing AT that loads building blocks to the thiolation domain²⁹. Surprisingly, a methyltransferase domain replaces the typical TE or R domain at the C-terminus of the PKS, indicating an unusual polyketide chain release mechanism. In addition to the modular PKS genes, a PPase gene was found at the 5'-end followed by a putative hydrolase and HMG-CoA synthase cassette for installing branched polyketide chains³⁰. Three transporters encoded by genes at the 3'-end of the cluster may export the product out of the cell. The other five strains (*S. mutans* 4VF1, *S. mutans* 4SM1, *S. mutans* 15VF2, *S. mutans* NLML-1 and *S. mutans* 1SM1) harbor PKS clusters that are split into two parts by a gap (green box in Fig. 4) that normally contains the ACP (T) and KS domains of the second modular protein and the ATd and DH domains of the third modular protein. Because only draft sequence data are available for these five strains and the gaps are all at the ends of the sequencing contigs, it is most likely that the five PKS clusters are split artificially.

Streptococcal distribution of NRPS and PKS genes. A careful analysis of the A-T-C/KS-AT-T module existing in NRPS, PKS, and hybrid PKS/NRPS gene clusters in the genomes of 10,038 *Streptococcus* strains from the Entrez Protein database, which is limited to genome-annotated *Streptococcus spp.* from the NCBI database (up to April 18, 2016), showed that about 12.1% of the A-T-C modules and 10.5% KS-AT-T module are found in 169 *S. mutans* genomes (Supplementary Table SIV). The result revealed an uneven distribution of NRPS, PKS and hybrid PKS/NRPS clusters in different genera of *Streptococcus*; *S. mutans* is about 7-fold enriched in A-T-C modules and 6-fold enriched KS-AT-T modules compared to the *Streptococcus* average. This implies that *S. mutans* may be an exceptional natural product producer among a plethora of relatively barren *Streptococci*.

The lantibiotic gene clusters in *S. mutans*. We identified 32 gene clusters encoding modified RiPPs that all belonged to the lantibiotic family and could be classified into three groups based on the modification genes *LanBC* and *LanM*, which catalyze thioether ring formation (Fig. 5 and Supplementary Table SII)³¹. The first group contains the *lanBC* genes and an independent protease, in contrast to the transporter protease domain found in the other two groups. The five modified lantibiotic gene clusters belonging to the first group are highly conserved in both gene composition and organization, with the exception that *S. mutans* B88SM-A and *S. mutans* 66-2 A have only one precursor gene instead of two. The first group gene clusters are very similar to those encoding mutacin I and III from *S. mutans* CH43 and UA789, but possess more different precursor peptides (Fig. 5 and Supplementary Figure SI)^{21,22}. Aligning 18 of the group I precursor peptides with Weblogo 3.0 revealed that the characteristic motif “F(N/D)LD” of class I lantibiotic leader peptides exists as a variant “FDVQ” motif here³². Notably, it was proposed that the “FDVQ” motif of the mutacin 1140 (S3) leader peptide is not required for its maturation, but a new “EDLF” box following “FDVQ” was important for recognition by post translational modification enzymes³³. Interestingly, both the “FDVQ” and “EDLF” motifs and the cleavage site after the “PDT(R/H)” box, which was characterized in mutacin 1140, are well conserved in all leader peptides of the first group lantibiotic precursors, indicating a conserved and potentially unique post translational modification.

The second group of lantibiotics, possessing one *lanM* for thioether formation instead of *lanBC*, was identified in 20 gene clusters. Among them, five complete gene clusters containing all 13 genes were identified, while the other 15 clusters are comprised of only the 6 genes at the 3' end of the cluster (Fig. 5, underlined in black). Once again, we cannot rule out truncation due to poor quality of the draft genome sequences, since all the truncated clusters are at the ends of the contigs. The gene organization of the complete clusters in the second group are very similar to the mutacin K8 gene cluster (Supplementary Figure SI) with the exception of the *S. mutans* B082SM-A cluster possessing only two precursor genes instead of four³⁴. An alignment of all 34 precursor peptides revealed a conserved “GG” cleavage site that is characteristic of this group (Fig. 5).

The third group consists of 7 clusters that encode modified RiPPs and possess two putative *lanM* genes. The cluster from *S. mutans* GS-5 is known to produce mutacin Smb by the action of two components, SmbA and SmbB³⁵. The remaining six pairs of precursors are identical to SmbA and SmbB, revealing strong conservation among this group of modified bacteriocins.

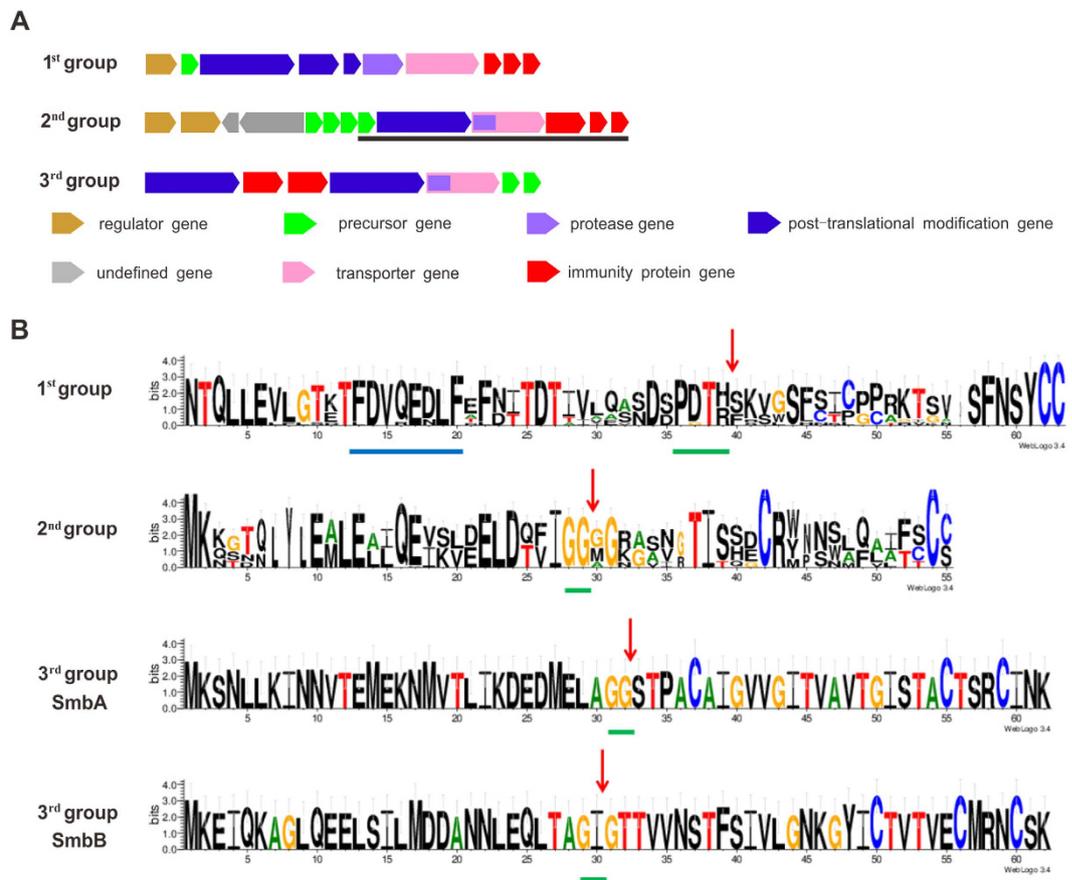


Figure 5. Composition of the lantibiotic gene clusters in *S. mutans* (A) and alignment of the lantibiotic precursors (B). (A) The putative lantibiotic gene clusters classified into three groups based on thioether-forming genes, *LanBC* and *LanM*. The genes are indicated by arrows with different colors. (B) The conserved motifs in class I lantibiotic leader peptides are underlined in blue. The green underline indicates the conserved sites in front of the cleavage sites, which are marked by red arrows. This figure was generated by web-based software WebLogo 3³¹.

Previous phylogenetic analysis of cyanobacteria proposed that the C39 domain-containing ABC transporters could reveal the evolutionary history of different C39 domain-containing transporters and their corresponding gene clusters³⁶. We employed a similar analysis to dissect the phylogenetic relationship among the lantibiotic gene clusters. The independent C39 protease (encoding by gene P) from the first group was artificially joined with the following ABC transporter (encoding by gene T) and aligned with the C39 domain-containing ABC transporters from the other two groups. The constructed phylogenetic tree (Fig. 6) reveals three distinct groups that correspond well with the above classification based on gene organization, with the first group more distantly related than the other two.

Unmodified bacteriocin gene clusters in *S. mutans*. Anti-SMASH and BAGEL analysis of the 169 *S. mutans* genomes identified 179 gene clusters encoding unmodified bacteriocins that were classified into five groups based on gene organization (Fig. 7 and Supplementary Table SII). The first group contains 63 members, each containing two adjacent precursor genes flanked by an upstream transpose gene and a downstream immunity protein gene³⁷. Mutacin IV from the unsequenced strain *S. mutans* UA140 represents the best-studied example of this group, and this putative two-peptide (mutacin IV-1 and IV-2) bacteriocin has potent antibiotic activity against the mitis group of oral *streptococci*²³. Analysis of precursor genes of this group revealed that most (>85%) encode identical precursor peptides (named as mutacin IV-1 and mutacin IV-2 in Supplementary Figure SI), while the others have minor variants in their leader peptides, indicating that mutacin IV biosynthesis in *S. mutans* is highly conserved (data not shown).

Twenty-four unmodified bacteriocin gene clusters belonging to the 2nd group were identified that each possessed a simple and invariant gene organization of one precursor gene, two unknown genes and two ABC transporter gene with one containing a protease domain (Fig. 7). The two ABC transporters in *S. mutans* UA159, named NImTE, were found to transport mutacin IV out of the cell³⁸. Interestingly, no protease gene was identified in the 1st group unmodified bacteriocin gene clusters, and the high similarity between leader peptides of the precursors in the first and second groups (Fig. 8) suggests that the two groups might share transporters like NImTE in strains such as *S. mutans* UA159 that harbor both groups.

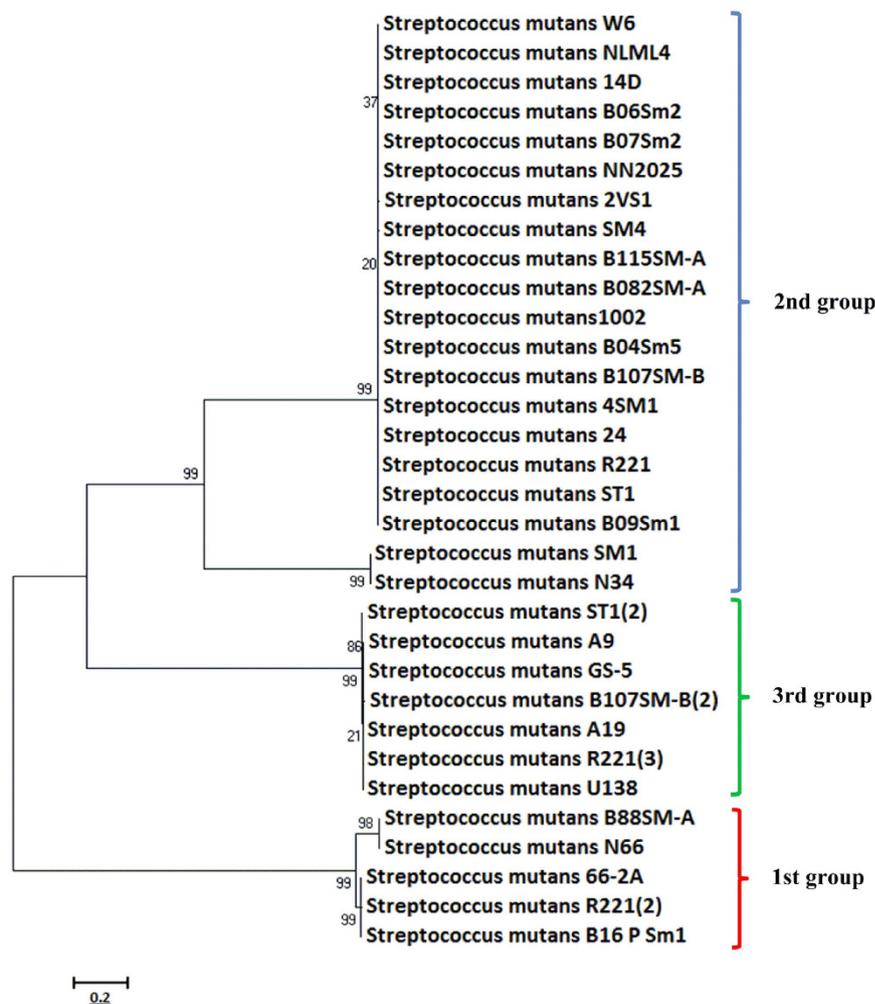


Figure 6. Evolutionary relationships among peptidase C39 domain-containing proteins in lantibiotic gene clusters from *S. mutans*. The phylogenetic tree was made by MEGA 7.0 using the Maximum likelihood method and with 1000 bootstrap replications for each branch. The bootstrap value of nodes is indicated. Each taxon is composed of a single strain and the major clades are marked with different colors (blue, red and green) and corresponding family classification. The number in brackets following the strain names indicates the number of lantibiotic gene clusters identified if more than one.

There are 73 unmodified bacteriocin gene clusters belonging to the third group, which has the most complicated gene organization with six precursor genes, three transporter genes, one immunity gene, one protease gene and several other genes with varied functions, such as serine-tRNA ligase, transposase and unknown functions (Fig. 7). The only knowledge about this group of clusters was from *S. mutans* UA159, in which deletion of the “f” precursor gene (SMU.1914c) changed the strain’s antibacterial profile, and the unidentified product of the precursor “f” gene in *S. mutans* UA159 was called mutacin V³⁹.

The fourth group of unmodified bacteriocin gene cluster contains 16 members, which appears to be a simplified version of the third group gene cluster. Thirteen genes comprising the left half of the third group cluster between the serine-tRNA ligase gene and the thioredoxin gene are replaced by three new precursor genes and a gene with unknown function in the fourth group clusters. The remaining precursors (precursor “d”, “e”, and “f”) on the right side of the fourth group clusters were almost identical to those in the third group, implying a close evolutionary relationship between these two groups of unmodified bacteriocin gene clusters.

The fifth group consists of three unmodified bacteriocin gene clusters with a precursor gene and two transporter genes. Interestingly, an uncommon gene that encodes a putative recombinase is located next to the precursor gene, indicating that the precursor gene may have originated from another location in the genome or from foreign DNA.

An alignment of representative precursors of unmodified bacteriocins (Fig. 8) reveals conserved leader peptides that harbor a “GG” motif that is usually located before the proteolytic cleavage site. On the contrary, the diversity of core peptide sequences demonstrates strong potential for the discovery of new bacteriocins from *S. mutans*.

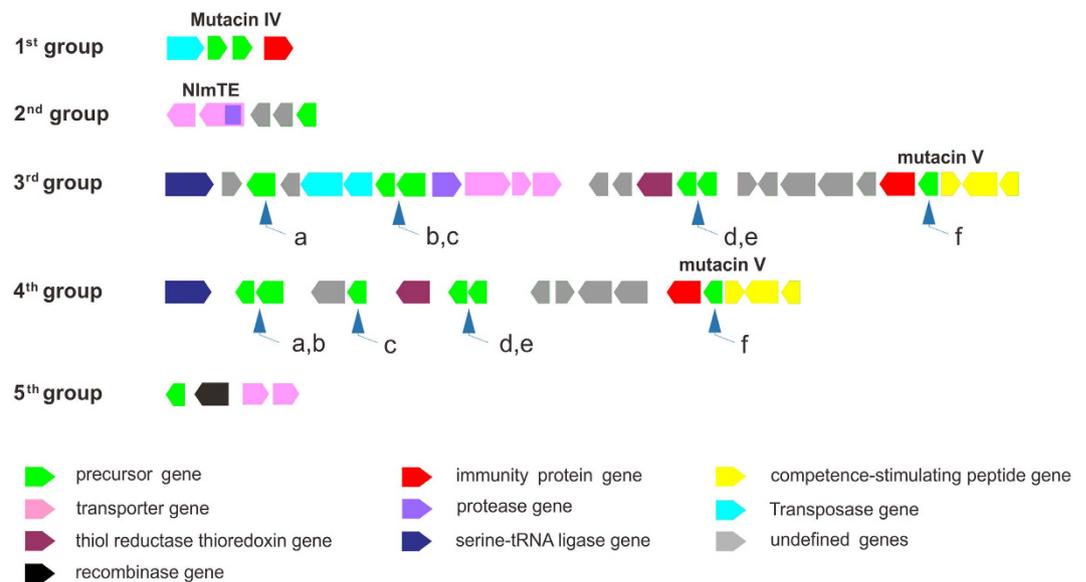


Figure 7. Composition of unmodified bacteriocin gene clusters in *S. mutans*. Functions of all genes are indicated by different colors. NlmTE, mutacin IV and mutacin V labels are located above the genes encoding homologs of them.

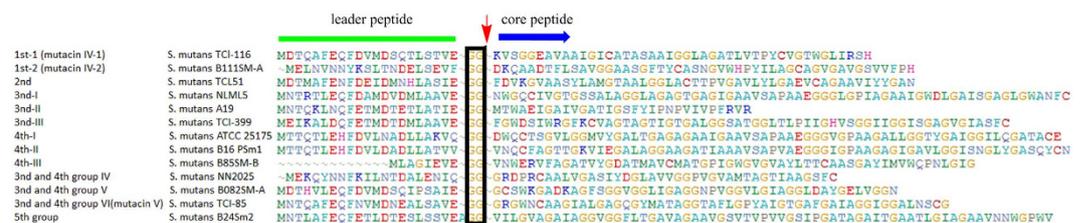


Figure 8. Alignment of unmodified bacteriocin precursors in *S. mutans*. The names of representative bacteriocin precursors and their host strains on the left. The conserved “GG” motif is boxed; the cleavage site is indicated by an arrow; conserved leader peptide amino acids are marked with asterisks.

In addition to the unmodified bacteriocin gene clusters mentioned above, we also found 173 putative bacteriocin precursor genes in *S. mutans*, which even contain 20 mutacin VI homologous precursor genes. Because no related genes (e.g. protease, immunity and transporter genes) could be identified nearby, we treated them as orphan precursor genes that may encode novel bacteriocins and summarized them with the representatives in Supplementary Figure SII.

Discussion

We have identified 355 natural product gene clusters from the genomes of 169 strains of the human oral microbe *S. mutans*, and 39% of these strains contained more than 2 clusters. Since only five genomes (until November 19th, 2015) were sequenced in complete level with the others in draft, the number of natural product gene clusters is undoubtedly much higher. Although bacteria with small genomes have been considered a poor source of PKS- and NRPS-synthesized natural products^{13,14}, our discovery of 144 PKS, NRPS and hybrid PKS/NRPS clusters and 211 RiPPs clusters in the 169 strains suggests bacteria with small genome sizes like *S. mutans* may represent important and overlooked sources of natural products.

That *S. mutans* houses so many NRPS and hybrid PKS/NRPS secondary metabolite gene clusters in its small genome is surprising and suggests *S. mutans* may dispatch a battery of antibiotic weapons to fight, communicate and thrive in its particular niche. The relative paucity of A-T-C domains encoded by other species of oral *Streptococcus*, such as *Streptococcus oligofermentans* (Supplementary Table SIV), suggests that the environment of dental caries has selected for a robust genetic complement of natural product capacity in *S. mutans*. Similarly, the wide spread of bacteriocin gene clusters in *S. mutans* may also be due to the selective pressure from the complicated oral environment. It has been shown that *S. mutans* apply some mutacins to build up competitive advantage in biofilm formation and in colonization in new host^{40,41}. In addition, a number of mutacins can influence the competence capacity of *S. mutans*, e.g. the small peptide ComC produced by an unmodified bacteriocin gene cluster can induce *S. mutans* UA150 to a high competence state efficiently; and, an apparent lower transformation efficiency was observed in the mutacin IV or V mutant strains^{42,43}. Therefore, mutacins might be functional not only

as antibiotics against the competitive microbes, but also as genetic tools prompting the DNA exchange frequency of *S. mutans*, which all render a better fitness of this species in its oral habitat.

Historically, microbial natural product chemists have intensively investigated microbes from soil, water and other natural environments, but have generally overlooked the microorganisms hosted by our own bodies. Recent human microbiome research has not only highlighted the importance of the human microbiome to health, but has also revealed a new scientific frontier of deciphering the precise chemical language of communication among and between microbes and the human host^{44–47}. Because secondary metabolites frequently exhibit antimicrobial and signaling activities, characterization of compounds from human commensal or pathogenic microbes can enlighten our understanding of the host-microbe relationship. As a causative agent of human tooth caries, understanding *S. mutans* secondary metabolism can help us understand the biology of tooth decay and provide therapeutic targets for future preventative treatments. The discovery of a rich source of natural product gene clusters from the 169 genomes of *S. mutans* from human caries sets the stage for isolating these compounds and elucidating their biological roles. Significantly, compounds from *S. mutans* are already active against gram positive pathogenic bacteria (e.g. mutacin 1140) and eukaryotic pathogens (e.g. mutanobactins), and the abundance of natural product gene clusters in *S. mutans* genomes predicts that many more fascinating compounds await discovery.

Methods

Data and strains. Complete and draft genomes of 169 *S. mutans* were collected from public genome database of NCBI (until November 19th, 2015) and saved in FASTA format. Their detailed information is listed in Supplementary Table SIII. The protein information of transporters with protease domains or with flanking protease genes from lantibiotic gene clusters was provided in Supplementary Table SV.

Detection of putative natural product gene clusters. AntiSMASH (Antibiotics & Secondary Metabolite Analysis SHell) was used with default settings to identify NRPS and PKS gene clusters⁴⁸. Functional domain prediction was performed with NRPSpredictor2 and NaPDoS (Natural Product Domain Seeker)^{49,50}. Annotations were made using Conserved Domain search and BLASTp (Basic Local Alignment Search Tool)⁵¹. The borders of these gene clusters were manually refined by checking synteny among protein sequences within the automatically predicted gene clusters from antiSMASH.

RiPPs precursors were identified through a combination of antiSMASH and BAGEL⁵². The results from both tools were compared, and only those gene clusters positive in both methods were analyzed in this study. The lantibiotic gene clusters were refined by visualizing all genes required for lantibiotic peptide biosynthesis. The unmodified bacteriocin gene clusters were identified if precursor-encoding genes were located together with transporters, immunity proteins or proteases.

Phylogenetic study. The phylogenetic study was conducted on protein sequences of transporters having a protease domain or a protease gene in the lantibiotic gene clusters using MEGA7.0 software employing the Maximum likelihood method. The protein sequences were retrieved from NCBI in FASTA format and aligned by clustalW in MEGA7.0.

References

- Berdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot* **65**, 385–395 (2012).
- Molinari, G. Natural products in drug discovery: present status and perspectives. *Adv Exp Med Biol* **655**, 13–27 (2009).
- Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**, 380–416 (2001).
- Yu, D. Y., Xu, F. C., Zeng, J. & Zhan, J. X. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* **64**, 285–295 (2012).
- Strieker, M., Tanovic, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* **20**, 234–240 (2010).
- Du, L. C. & Lou, L. L. PKS and NRPS release mechanisms. *Nat Prod Rep* **27**, 255–278 (2010).
- Lambalot, R. H. *et al.* A new enzyme superfamily-The phosphopantetheinyl transferases. *Chem Biol* **3**, 923–936 (1996).
- 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**, 1568 (2013).
- Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins—a viable alternative to antibiotics? *Nat Rev Microbiol* **11**, 95–105 (2013).
- Sahl, H. G. & Bierbaum, G. Lantibiotics: Biosynthesis and biological activities of uniquely modified peptides from gram-positive bacteria. *Annu Rev Microbiol* **52**, 41–79 (1998).
- van Belkum, M. J., Martin-Visscher, L. A. & Vederas, J. C. Structure and genetics of circular bacteriocins. *Trends Microbiol* **19**, 411–418 (2011).
- Hou, Y. P. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org Lett* **14**, 5050–5053 (2012).
- Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *P Natl Acad Sci USA* **111**, 9259–9264 (2014).
- Jenke-Kodama, H., Sandmann, A., Muller, R. & Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* **22**, 2027–2039 (2005).
- Guo, C. J. & Wang, C. C. Recent advances in genome mining of secondary metabolites in *Aspergillus terreus*. *Front Microbiol* **5**, 717 (2014).
- Calteau A. *et al.* Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genomics* **15**, 977 (2014).
- Ju, K. S. *et al.* Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *P Natl Acad Sci USA* **112**, 12175–12180 (2015).
- 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).
- Joyner, P. M. *et al.* Mutanobactin A from the human oral pathogen *Streptococcus mutans* is a cross-kingdom regulator of the yeast-mycelium transition. *Org Biomol Chem* **8**, 5486–5489 (2010).
- Wang, X. R., Du, L., You, J. L., King, J. B. & Cichewicz, R. H. Fungal biofilm inhibitors from a human oral microbiome-derived bacterium. *Org Biomol Chem* **10**, 2044–2050 (2012).

21. Qi, F. X., Chen, P. & Caufield, P. W. Purification of mutacin III from group III *Streptococcus mutans* UA787 and genetic analyses of mutacin III biosynthesis genes. *Appl Environ Microbiol* **65**, 3880–3887 (1999).
22. Qi, F. X., Chen, P. & Caufield, P. W. Purification and biochemical characterization of mutacin I from the group I strain of *Streptococcus mutans*, CH43, and genetic analysis of mutacin I biosynthesis genes. *Appl Environ Microbiol* **66**, 3221–3229 (2000).
23. Qi, F. X., Chen, P. & Caufield, P. W. The group I strain of *Streptococcus mutans*, UA140, produces both the lantibiotic mutacin I and a nonlantibiotic bacteriocin, mutacin IV. *Appl Environ Microbiol* **67**, 15–21 (2001).
24. Novak, J. *et al.* Detection of modified amino acids in lantibiotic peptide mutacin II by chemical derivatization and electrospray ionization mass spectroscopic analysis. *Anal Biochem* **236**, 358–360 (1996).
25. Kamiya, R. U., Hoffling, J. F. & Goncalves, R. B. Frequency and expression of mutacin biosynthesis genes in isolates of *Streptococcus mutans* with different mutacin-producing phenotypes. *J Med Microbiol* **57**, 626–635 (2008).
26. Wu, C. G. *et al.* Genomic island TnSmu2 of *streptococcus mutans* harbors a nonribosomal peptide synthetase-polyketide synthase gene cluster responsible for the biosynthesis of pigments involved in oxygen and H₂O₂ tolerance. *Appl Environ Microbiol* **76**, 5815–5826 (2010).
27. Lin, X. B. *et al.* Genetic determinants of reutericyclin biosynthesis in *Lactobacillus reuteri*. *Appl Environ Microbiol* **81**, 2032–2041 (2015).
28. Keating, T. A. *et al.* Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. *ChemBiochem* **2**, 99–107 (2001).
29. Lohman, J. R. *et al.* Structural and evolutionary relationships of “AT-less” type I polyketide synthase ketosynthases. *P Natl Acad Sci USA* **112**, 12693–12698 (2015).
30. Buchholz, T. J. *et al.* Polyketide beta-Branching in Bryostatin Biosynthesis: Identification of Surrogate Acetyl-ACP Donors for BryR, an HMG-ACP Synthase. *Chem Biol* **17**, 1092–1100 (2010).
31. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).
32. Plat, A., Kluskens, L. D., Kuipers, A., Rink, R. & Moll, G. N. Requirements of the engineered leader peptide of nisin for inducing modification, export, and cleavage. *Appl Environ Microbiol* **77**, 604–611 (2011).
33. Escano, J., Stauffer, B., Brennan, J., Bullock, M. & Smith, L. The leader peptide of mutacin 1140 has distinct structural components compared to related class I lantibiotics. *Microbiol Open* **3**, 961–972 (2014).
34. Robson, C. L., Wescombe, P. A., Klesse, N. A. & Tagg, J. R. Isolation and partial characterization of the *Streptococcus mutans* type All lantibiotic mutacin K8. *Microbiology* **153**, 1631–1641 (2007).
35. Yonezawa, H. & Kuramitsu, H. K. Genetic analysis of a unique bacteriocin, Smb, produced by *Streptococcus mutans* GS5. *Antimicrob Agents Chemother* **49**, 541–548 (2005).
36. Wang, H., Fewer, D. P. & Sivonen, K. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLoS One* **6**, e22384 (2011).
37. Hossain, M. S. & Biswas, I. SMU.152 Acts as an Immunity Protein for Mutacin IV. *J Bacteriol* **194**, 3486–3494 (2012).
38. Hale, J. D. F., Heng, N. C. K., Jack, R. W. & Tagg, J. R. Identification of nlmTE, the locus encoding the ABC transport system required for export of nonlantibiotic mutacins in *Streptococcus mutans*. *J Bacteriol* **187**, 5036–5039 (2005).
39. Hale, J. D. F., Ting, Y. T., Jack, R. W., Tagg, J. R. & Heng, N. C. K. Bacteriocin (mutacin) production by *Streptococcus mutans* genome sequence reference strain UA159: elucidation of the antimicrobial repertoire by genetic dissection. *Appl Environ Microbiol* **71**, 7613–7617 (2005).
40. Kreth, J., Merritt, J., Shi, W. & Qi, F. Competition and coexistence between *Streptococcus mutans* and *Streptococcus sanguinis* in the dental biofilm. *J Bacteriol* **187**, 7193–7203 (2005).
41. Hillman, J. D. Genetically modified *Streptococcus mutans* for the prevention of dental caries. *Antonie Van Leeuwenhoek* **82**, 361–366 (2002).
42. Perry, J. A., Jones, M. B., Peterson, S. N., Cvitkovitch, D. G. & Levesque, C. M. Peptide alarmone signalling triggers an auto-active bacteriocin necessary for genetic competence. *Mol Microbiol* **72**, 905–917 (2009).
43. Dufour, D., Cordova, M., Cvitkovitch, D. G. & Levesque, C. M. Identification of a novel role associated with a streptococcal bacteriocin: regulation of the competence pathway. *J Bacteriol* **193**, 6552–6559 (2011).
44. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).
45. Challinor, V. L. & Bode, H. B. Bioactive natural products from novel microbial sources. *Ann N Y Acad Sci* **1354**, 82–97 (2015).
46. Koppel, N. & Balskus, E. P. Exploring and understanding the biochemical diversity of the human microbiota. *Cell Chem Bio* **23**, 18–30 (2016).
47. Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516 (2016).
48. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* **43**, W237–W243 (2015).
49. Ziemert, N. *et al.* The natural product domain seeker NaPDoS, a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, e34064 (2012).
50. Rottig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* **39**, W362–W367 (2011).
51. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**, D225–D229 (2011).
52. van Heel, A. J., de Jong, A., Montalban-Lopez, M., Kok, J. & Kuipers, O. P. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* **41**, W448–W453 (2013).

Acknowledgements

This work was supported in part by the Ministry of Science and Technology of China (2015CB150600, 2013CB734000) and the National Natural Science Foundation of China (31522001, 31570050). Y. Chen is a scholar of ‘the 100 Talents Project’ of the Chinese Academy of Sciences.

Author Contributions

The gene cluster screening was designed by L.L., T.H. and Z.X. Analysis of results was performed by L.L., T.H., Z.X., G.H. and Y.C. The paper was written by L.L., G.H. and Y.C. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, L. *et al.* Genome mining unveils widespread natural product biosynthetic capacity in human oral microbe *Streptococcus mutans*. *Sci. Rep.* **6**, 37479; doi: 10.1038/srep37479 (2016).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016