JOURNAL OF ORAL MICROBIOLOGY, 2018
VOL. 10, 1424455
https://doi.org/10.1080/20002297.2018.1424455

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Acid-resistant genes of oral plaque microbiome from the functional metagenomics

Meng Zhang[a], Yuqiao Zheng[a], Yuhong Li[a], Han Jiang[a], Yuping Huang[b] and Minquan Du[a]

[a]MOST KLOS & KLOBM, School & Hospital of Stomatology, Wuhan University, Wuhan, China; [b]Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China

ABSTRACT

Acid resistance is one of key properties assisting the survival of cariogenic bacteria in a dental caries environment, but only a few genes conferring acid resistance have been identified to data. Functional metagenomics provides a systematic method for investigating commensal DNA to identify genes that encode target functions. Here, the host strain Escherichia coli DH10B and a constructed bidirectional transcription vector pSKII^{+}-lacZ contributed to the construction of a metagenomic library, and 46.6 Mb of metagenomic DNA was cloned from carious supragingival plaque of 8children along with screening for lethal functionality. The screen identified 2 positive clones that exhibited a similar aciduric phenotype to that of the positive controls. Bioinformatic analysis revealed that these two genes encoded an ATP/GTP-binding protein and a malate dehydrogenase. Moreover, we also performed functional screening of Streptococcus mutans, since it is one of the predominant cariogenic strains but was not identified in our initial screening. Five positive clones were retrieved. In conclusion, our improved functional metagenomics screening method helped in the identification of important acid resistance genes, thereby providing new insights into the mechanism underlying caries formation as well as in the prevention and treatment of early childhood caries (ECC).

ARTICLE HISTORY
Received 10 August 2017
Accepted 3 January 2018

KEYWORDS
Early childhood caries; microbiome; functional metagenomics; acid-resistance; dental plaque

Introduction

Early childhood caries (ECC) is one of the most common infectious diseases in children and affects up to 60–90% of children worldwide [1]. ECC can not only cause tremendous pain [2] but can also results in an enormous social cost [3]. Consequently, the prevention and treatment of ECC are extremely important and mandate intensive study [4]. It is well known that microorganisms living in the dental plaque biofilm play a vital and direct role in the occurrence of tooth decay [2]. Survival and propagation of microorganisms in the caries microenvironment inducing tooth decay require that they exhibit certain properties, such as adhesion, acidogenicity and acid resistance. Among them, acid resistance is a key factor for the survival of cariogenic bacteria [5] in the typical carious microenvironment, in which the pH values fluctuate in the range from pH 7 to 3, primarily due to diet [6]. Hence, the acid resistance of oral microbes is taken into account in the study of ECC prevention and treatment, which offers potential new insights into clinical applications and personalized medicine [7].

Currently, only a minor fraction of acid-resistance genes in oral microorganisms have been discovered and investigated given the limitations of genetic and molecular biology techniques primarily based on culture-dependent methods [8]. In the mid-1980s, Loesche et al. [5] reported that *Streptococcus mutans* is a principal causative agent of dental caries, and its prevalence was 70–100% of individual cases of caries. Subsequently, a large number of studies on the acid production and acid-resistant properties of *Streptococcus* were performed [9,10], and some relevant acid-resistant genes, such as *ffh, uvrA,* and *dnaK*, were discovered and studied [11–13]. These achievements can be partially attributed to the ease in propagating this strain in the laboratory. However, methods in propagating generally difficult-to-grow microbial strains have scarcely been established to date. Recent high-throughput sequencing studies based on culture-independent analysis of childhood caries revealed that *S. mutans* is not the only dominant pathogenic bacterium in caries. *Lactobacillus, Selenomonas* and *Neisseria* can also be detected at high levels in severe carious plaques [14,15], and some of these bacteria cannot be cultured (Human Oral Microbiome Database (HOMD), http://www.homd.org/). In addition to *Streptococcus*, our knowledge about the acid resistance and related regulatory genes of cariogenic microorganisms, especially those that are not cultivable, is limited; thus, it offers a fertile field for the current study.

To explore this field, we utilized function-based metagenomics [16], which can identify both novel and previously known genes involved in target functions,

CONTACT Minquan Du duminquan@whu.edu.cn
Supplemental data for this article can be accessed here.

even with unknown genetic information. Currently, function-based screening of metagenomic libraries has been widely deployed in different fields, including the screening of environmental libraries [17], the investigation of mouse gut colonization [18], of human faecal samples [19] and others. However, acid-resistance screening studies of oral plaque samples have not yet been reported.

Therefore, this study attempted to identify genes conferring acid resistance from the supragingival plaque of 4–6-year-old children using improved function-based metagenomics to provide a new entry point for the study of the caries mechanism and ECC prevention. A high-copy vector allowing bidirectional transcription was constructed to enhance library expression efficiency. Total microbial DNA of supragingival plaque samples from carious teeth of children was extracted. The DNA was used to construct a metagenomic library, and quality assessment was done before lethal screening was conducted. Bioinformatics analysis of the resulting positive clones identified two genes that were predicted to, respectively, encode an ATP/GTP-binding protein and malate dehydrogenase.

Materials and methods

Ethics statement

Informed consent forms were signed by the parents of all subjects prior to study enrolment. The entire study design, protocol and informed consents were approved by the Medical Ethics Committee at the School & Hospital of Stomatology, Wuhan University, Wuhan, China. Additionally, we confirmed that all experiments were performed in accordance with relevant guidelines and regulations.

Dental plaque samples

Supragingival dental plaques, derived from carious teeth of 4–6 years-old children, were sampled to construct the metagenomic library. The inclusion criteria were as follows: (1). Healthy medical history (2). No current bacterial/viral infections and no antibiotics within the past 3 months (3). At least three cavitated caries teeth among more than 18 present teeth (code >4, according to the criteria of the International Caries Detection and Assessment System) (4). No detectable enamel and dentin hypoplasia and (5). No prior treatment (including fluoride) for dental decay.

Only supragingival plaque samples around the cavity were collected using a sterile curette that was immersed into a 1.5 ml Eppendorf tube containing 1 ml of autoclaved saline solution. The kids were asked not to eat, drink and brush within 2 h before sampling. Every individual was repeatedly sampled and all samples were stored at −80°C until use.

Bacterial strains, plasmids, media, and culture conditions

Escherichia coli DH10B, DH5α, and JM110 (TaKaRa Bio Inc. Japan) strains were routinely cultured in Luria–Bertani (LB) medium at 37°C. The shaking speed was set at 200 rpm for liquid cultures. If needed, the growth medium was supplemented with 100 mg/ml ampicillin (Amp) to maintain pUC19, pBluescript SKII(+) (pSKII$^+$) and pSKII$^+$-lacZ (pSL, a dual-direction expression vector constructed in this study) vectors. *E. coli* JM110 was mainly used for demethylation.

The standard *S. mutans* UA159 strain (ATCC 700,610) was kindly provided by MOST KLOS & KLOBM, School & Hospital of Stomatology, Wuhan University. The strain was amplified in Brain Heart Infusion (BHI) broth in an anaerobic jar at 37°C.

The construction and verification of the vector pSL allowing bidirectional transcription

To enhance the efficiency of library screening, a bidirectional transcription vector pSKII$^+$-lacZ was constructed. To verify whether both promoters were functional, a reporter gene encoding green fluorescent protein (GFP) was cloned and ligated into the pSL (for a detailed protocol see Supplementary File). Finally, the strains *E. coli* DH10B/pSL-GFP (OD = 1.2–1.5) were observed under excitation fluorescence microscopy to detect the fluorescence intensity.

DNA extraction

The plaque samples were mixed in one sterile tube to obtain total DNA. Two independent aliquots of mixed samples were prepared separately for screening and cross validation. The protocol of total DNA extraction was taken from a previous study [20]. The purity and concentration of DNA were measured using agarose gel electrophoresis and NanoDrop 2000 (Thermo Scientific, Wilmington, DE). All the DNA samples were stored at −80°C before further analysis.

Construction of a metagenomic library

For the construction of the supragingival plaque metagenomic library, total DNA was partially digested with *Sau*3A I (Thermo Scientific) and three DNA fragments of 1–4 kb, 4–7 kb, and 7–10 kb were separately excised from the agarose gel and purified using the OMEGA Gel Extraction Kit (Omega Bio-Tek, Norcross, GA). The DNA

fragments were ligated into a dephosphorylated and *Bam*HI-digested pSL vector at molar ratios of 3:1 to 7:1 (insert/vector) using T4 DNA ligase. Then, 10 μL of ligation mixture was transformed in competent *E. coli* DH10B, which was plated on LB_{Amp} agar and incubated for 14 h at 37°C to amplify the libraries. The detection of metagenomics library quality was completed prior to screening. Per plate 1–2% of the colonies were randomly selected to test the efficiency of self-ligation by extracting the plasmids, digesting with *Eco*RI and *Xba*I and performing agarose gel electrophoresis. All the ligation and transformation reactions were performed sequentially, and the plates were stored at 4°C for less than 5 days to ensure sufficient time to implement the next step. All primers were designed using Primer Premier 5 (PREMIRE Biosoft, Palo Alto, CA) and are listed in Supplementary Table S1. The protocol for library construction is illustrated in Figure 1(a).

In addition, we used two known acid-resistant genes, *orf1* and *ffh*, to construct two positive controls, *E. coli* DH10B/pSL-orf1 and pSL-ffh, the former from the environmental microbes [17], the latter from *S. mutans* [11]. The detailed protocol is described in the Supplementary File.

Screening of the metagenomic library

Colonies of the library were directly incubated in 96-well plates containing 0.7 ml of pH 1.9 LB broth supplemented with 100 μg ml^{-1} Amp for 3 h at 37°C, which is a lethal condition for strain *E. coli* DH10B/pSL. The LB broth was acidified to pH 1.9 using an aqueous solution of 37% (w/v) HCl and sterilized in an autoclave for 30 min at 115°C. After the acid shock, 3 μL of cells were dropped on LB_{Amp} agar and incubated at 37°C for 14 h (Figure 1(b)). Meanwhile, negative (*E. coli* DH10B/pSL) and positive controls were cultured and subjected to acid shock. Colonies that survived like the positive controls were regarded as potential acid resistant clones. Then, the plasmids of these potential acid resistant clones were isolated and transformed into new competent *E. coli* DH10B,

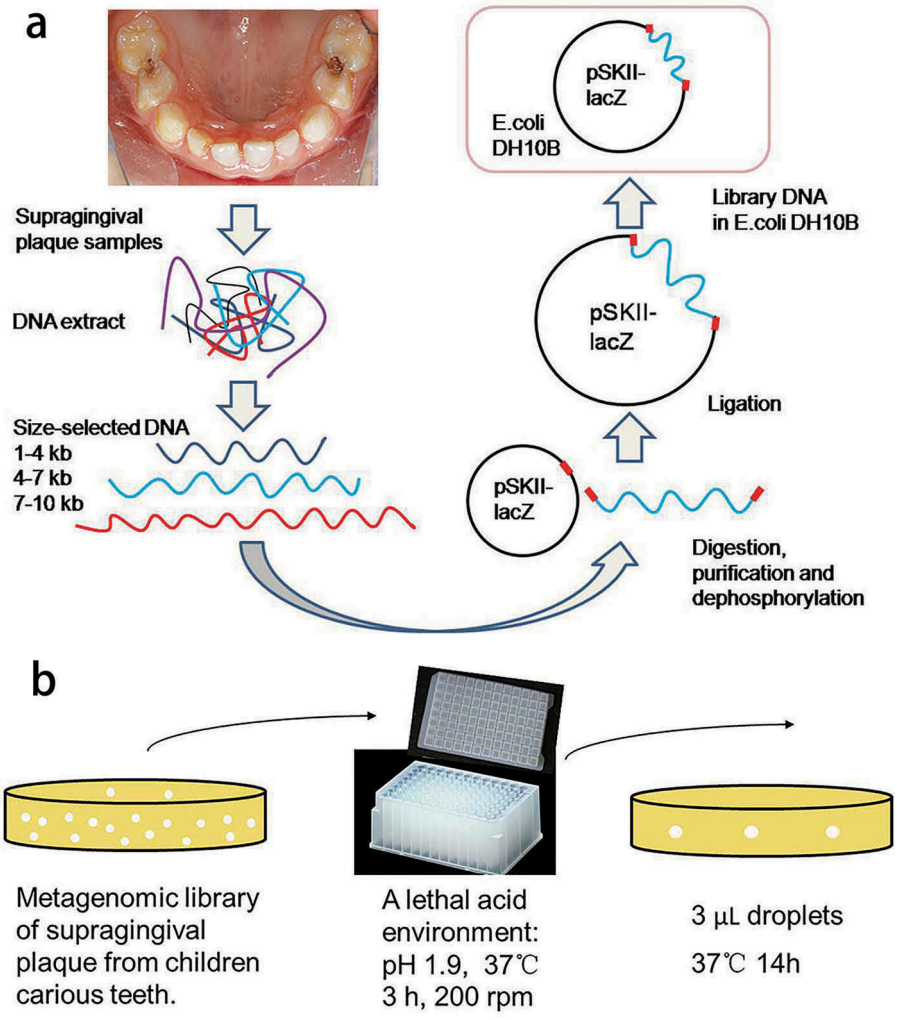

Figure 1. Schematic diagram of (a) library construction and (b) functional screening.

after which the acid shock was repeated. Those clones that continuously exhibited an acid-resistant phenotype were considered to be positive clones. These plasmids were subsequently sequenced using primer P2 (Sangon Biotech Co.) and stored at −80° C until used.

In silico *analysis*

Based on the sequencing data of the positive clones, potential operon reading frames (ORFs) with start (ATG) and stop (TAA, TAG, TGA) codons were predicted and analysed using Vector NTI Advance11.5.1 and DNAman v6.0. In addition, alignments of nucleotide sequences and transcribed amino acid sequences were performed against public databases using BlastN and BlastX (basic local alignment search tool) at the NCBI on-line server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The predicted ORFs similar to known or unknown proteins and longer than 100 bp were obtained and further analysed. The conserved domains of inserts were analysed using the CD search module in NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Their structure and putative function were annotated based on similarities to the sequences in the Clusters of Orthologous Groups (COG) [21], Protein Families (Pfam) [22] (http://xfam.org/), and Blocks [23] (InterPro http://www.ebi.ac.uk/interpro/) databases and based on the results of BLAST searches in UniProt (http://www.uniprot.org/). If the similarity of the protein sequence alignment was less than 30%, it was considered to be an unknown sequence. Trans-membrane helices were forecasted with the TMpred programme (http://www.ch.embnet.org/software/TMPRED_ form.html) and UniProt.

Acid-resistance assay

The acid-resistance assay was based on the library functional screening protocol used in this study. The positive clones and negative/positive controls were activated and inoculated overnight (12 h) in 5 ml LB_{Amp} broth at 37°C, which contained 2–3 × 10^9 cfu (colony-forming units) per millilitre of culture (OD 1.45–1.55). Then, 10 µL of culture was separately transferred into 1 ml of PBS (phosphate-buffered saline, pH 7.2) and 1 ml of LB broth (pH 1.9). The initial cell population was quantified by plating serial dilutions of PBS buffer on the LB agar and incubating for 14 h at 37°C before counting the cfu. The final cell population was determined as described above after the strains were subjected to acid shock (pH 1.9). Subsequently, the percentage survival value was calculated by dividing the final number of cfu ml^{-1} after the acid shock by the initial number of cfu ml^{-1} and the average values of three independent experiments were reported.

Screening and analysis of S. mutans UA159

The library construction, functional screening and sequence analysis of *S. mutans* UA159 were conducted with the above protocol.

Results

Subjects and samples

Thirty-two caries associated supragingival plaque samples were obtained through four consecutive repeat samplings from eight children aged 4–6 (sample information is shown in Supplementary Table S2). All 32 samples were mixed together and placed into two 1.5 ml Eppendorf tubes. The two resulting samples weighted 28.2 and 29.1 mg, respectively.

Validation of the dual-direction expression vector pSL

To improve the efficiency of library screening, a dual-direction expression vector, $pSKII^+$-lacZ (pSL), was constructed (the schematic of pSL is shown in Figure 2(a), for a detailed protocol see Supplementary File). We used green fluorescent protein (GFP) to detect whether the two promoters were effective. Strains of *E. coli* DH10B/pSL-GFP with two different directions of GFP showed green fluorescence (Figure 2(b,c)), indicating that both promoters worked. However, stronger fluorescence was observed with the reverse GFP. Analysis found that this could be attributed to the start codon ATG of the vector-self aiding in the generation of a fusion protein weakening the expression of GFP.

Preparation of samples and construction of libraries

The extraction of the total DNA from plaque samples can be found in a previous study [20], named DNA1 and DNA2; the concentrations are 469.8 and 492.9 ng/µl, respectively. The test results of gel electrophoresis and NanoDrop 2000 are shown in Supplementary Figure S1a and Table S3. Approximately 0.08 mg of high-quality DNA per milligram of sample (wet weight) was obtained. Subsequently, we constructed a plaque library using the DNA1 (see Methods). To ensure the library quality, 1–2% colonies per plate were randomly selected to detect the ligation rate of exogenous fragments. We only selected one plate from the same batch of transformation products for library quality testing. Hence, among the 30 colonies used to perform the plasmid extraction and enzyme-digestion test (Supplementary Figure S1b), 85.5% contained insert DNA. After statistical analyses,

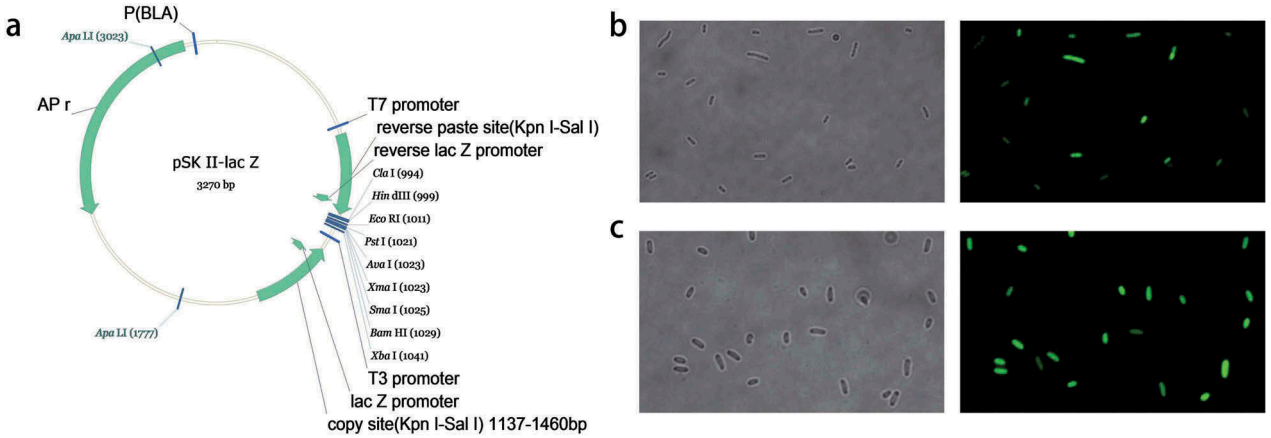

Figure 2. Bidirectional transcription vector pSKII-lacZ (pSL). (a) Illustration of vector pSL with two promoters. (b,c) The reporter gene-GFP (green fluorescent protein) was ligated into *E. coli* DH10B/pSL in a different direction and observed under excitation fluorescence microscopy to test efficacy of ligation. (b) The direction of GFP is the same as that of the original lac Z promoter. (c) The direction of GFP is the same as that of the reverse lac Z promoter (the left is a photo in which fluorescence has not been excited, the right is a photo in which fluorescence has been excited).

approximately 11, 290 colonies were obtained with 46.6 Mb of DNA.

Acid-resistance screening of the library

The lethal screening condition was acid shock for 3 h in the medium of pH 1.9, which was confirmed by pre-experimentation to be fatal to the strain *E. coli* DH10B/pSL, but non-toxic to the positive control (*E. coli* DH10B/pSL-orf1 and pSL-ffh). Both the inserts orf1 and ffh are known aciduric-related genes. The former encodes RNA-binding protein and a separate functional metagenomics study of the environmental sample confirmed its function [17], the latter comes from the *S. mutans* strain [11]. The purpose of construction of the strain *E. coli* DH10B/pSL-ffh and its functional verification was to detect whether the acid-resistant genes from oral bacteria could be heterologously expressed in *E. coli*.

Hence, after the first round of acid shock, seven colonies were identified that could survive in this hostile environment, but only two of the seven clones continuously exhibited acid resistance upon repeated acid shocks (repeated three times). Therefore, these two clones were regarded as positive clones carrying genes conferring acid resistance.

Molecular analysis of acid-resistant clones

The sequences of two positive clones are displayed in the Supplementary File. The genetic organization, genetic similarities, potential ORFs, phylogenetic relationship, putative function, and transmembrane domain predictions of the two clones are analysed and summarized in Table 1. As shown in the table, the two acid-resistant clones (pSL-1–31 and 2–41) contained two complete heterologous ORFs. The

Table 1. Illustration of the ORFs identified in plasmids pSL-1–31 and pSL-2–41 and the bioinformatics analysis of two positive clones.

Code	1–31	2–41
Seq length (bp)	836	2636
Site[a] (Nt range[b])	1–94aa (835–425bp)	1–643aa (402–2324bp)
G + C %	46.9	47.7
Alignment[c]	*Actinomyces* sp. oral taxon 178 str.F0338	*Haemophilus haemolyticus*
Protein (accession No.)/size (aa)[c]	ATP/GTP-binding protein (EFW08722)/137aa	Malate dehydrogenase (WP_005646657.1)/756aa
E-value	2.00E-49	0
Identities %	78/80 (98%)	372/643 (58%)
Accession No. of the similar protein[d]	E8JLB4	H1LMB2
Putative function[d]	Play a pivotal role in various distinct biological processes covering protein export, membrane transport, DNA repair etc.	Malic (pfam00390) and Malic_M (PF03949/smart00919) Play a key role in the central oxidative pathway. It assists *E. coli* to survive in the acid environment.
Possible transmembrane helices[e]	0	0

[a]Protein range (site) of alignment into the known proteins.
[b]Nucleotide range (nt range) of the predicted ORF within insert.
[c]Most similar protein, currently identified with blastp: organism, name, accession no, and protein size.
[d]Most similar protein, currently identified with UniProt: accession number of the similar protein, conserved domain and putative function.
[e]Results from the TMpred server.
bp: base pair; aa: amino acids; score: bit score of the alignment using BLAST.

clones exhibited high similarity (98 and 58%) to known functional proteins, indicating that no novel genes were discovered from this plaque library. The G + C contents of the inserts were 46.9 and 47.7%.

The sequences of pSL-1–31 were aligned with sequences from *Actinomyces* sp. oral taxon 178 str. F0338 (Actinobacteria), which is a gram-positive, anaerobe/facultative anaerobe bacterium that encodes an ATP/GTP-binding protein, belonging to the nucleotide-binding protein family. It is involved in various distinct biological processes. The other positive clone pSL-2–41 was aligned with sequences from *Haemophilus haemolyticus* (Proteobacteria), which is a gram-negative, facultative anaerobe that encodes malate dehydrogenase, and has a known acid-resistant gene (the details are presented in Table 1).

Acid-resistance assay

To quantify the ability to resist acid by the inserts in positive clones, an acid-resistance assay was performed on the two functional clones (pSL-1–31, pSL-2–41) and negative/positive controls (Figure 3). In Figure 3, there were statistically significant differences between all pairs of groups except for group pSL-1–31 and pSL-orf. The test analysis revealed that the two ORFs in the inserts were responsible for substantial improvement in the acid resistance of host cells but to a different degree. Furthermore, pSL-ffh displayed the highest resistance level ($P < 0.0001$), which not only further verified the acid-resistance conferred by ffh, but also provided evidence for the heterologous expression of the genes from *S. mutans* in *E. coli*.

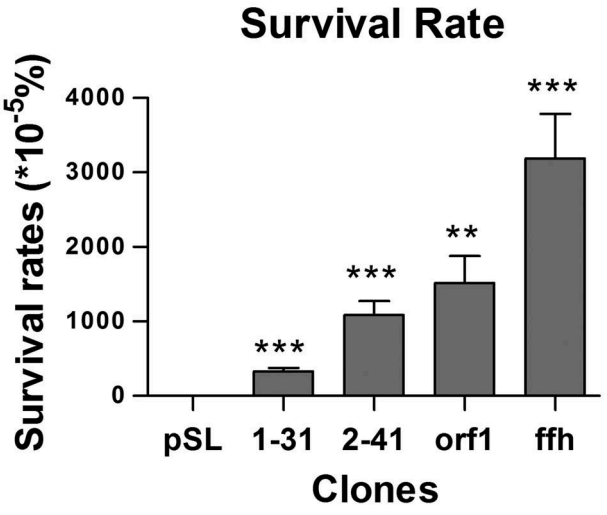

Figure 3. The capacity of acid resistance of positive clones and negative/positive controls using the acid resistance assay. pSL-2–41 displays a similar acid-resistant capability as the positive control pSL-orf1 ($P = 0.3055$); 1–31 vs. pSL $P < 0.0001$; 2–41 vs. pSL $P < 0.0001$; orf1 vs. pSL $P = 0.0007$; ffh vs. pSL $P < 0.0001$; 1–31 vs. 2–41 $P = 0.0011$. Values represent the mean of three independent experiments and vertical bars are the standard deviations. Statistical analysis was performed using t-tests or nonparametric tests. $^{**}P < 0.001$ and $^{***}P < 0.0001$.

Screening and analysis of S. mutans UA159

Based on the above results, we found that, in our study, although *S. mutans* was regarded as the primary cariogenic agent, no functional genes were retrieved from the examined strain. This interesting result prompted us to construct and screen another library from *S. mutans* UA159 as outlined above.

Eventually, five positive clones were obtained (the sequences of these clones are presented in the Supplementary File), which encoded molecular chaperone DnaK (U-57), chaperone protein ClpC (U-38), 23S rRNA-methyltransferase (U-41), acylphosphatase (U-23), rod-shape-determining protein MreD and glucan-binding protein B (U-8) (Table 2). The DnaK molecular chaperone (U-57) is a ubiquitously distributed class I stress protein, which tolerates the low pH potentially through regulating the biogenesis or stabilization of the F-ATPase compound [13,24]. The functions of Clp protease (U-38) and methyltransferase (U-41) have been corroborated by a previous study [17]. Only *SMU_1725* (U-23), *mreD* and *gbpB* (U-8) were newly identified functional genes from *S. mutans* UA159. These genes are predominantly involved in catalyzing the hydrolysis of various acyl phosphate carboxyl-phosphate bonds, regulating cell shape and the formation of biofilm on inanimate substrate.

Discussion

As one of the most determinant properties of cariogenic bacteria in inducing tooth decay in children, acid resistance deserves extensive research. Currently, various meticulous investigations on the molecular and physiological responses to acid shock are limited to cultivable oral bacteria [25], such as *S. mutans*, the gastric inhabitant *E. coli* [26] and *Helicobacter pylori* [27]. However, knowledge regarding the aciduric features of so far uncultivable microorganisms remains scarce. As a crucial technology used to access and explore the genomes and functional information of microorganisms including uncultivable ones, functional metagenomics [16] exhibit high application value. To the best of our knowledge, this is the first work that applied functional metagenomics to explore the acid-resistant genes of dental plaque microbiomes. The improved function-based screening strategy used here allowed the discovery of two genes involved in acid resistance, which could provide new clues for the development of therapeutic agents targeting the microorganisms to attenuate caries disease progression and prevent the occurrence of ECC.

Table 2. Bioinformatics analyses of five positive clones screened from *S. mutans* UA159.

Code	U-57	U-38	U-41	U-23	U-8
Seq length (bp)	1243	710	1172	1050	2332
Site[a] (Seq No.[b])	1–239aa (516–1231bp)	155–368aa (68–709bp)	1–159aa (494–18bp)	1–92aa (373–648bp)	1–168aa (1079–1582bp) / 1–182aa (1695–2330bp)
Gene	*dnaK*	*clpC*	*rlmH*	*SMU_1725*	*mreD* / *gbpB*
Protein/size (aa)[c]	Molecular chaperone DnaK /612	Class III stress response-related ATP-dependent Clp protease, ATP-binding subunit/800	23S rRNA-methyltransferase RlmH /159	Acylphosphatase /92	Rod shape-determining protein MreD/168 / glucan-binding protein B/431
Putative function[d]	Participates in acid-resistant process	Participates in diverse cellular processes including membrane fusion, proteolysis and DNA replication	Specifically methylates the pseudouridine at position 1,915 in 23S rRNA.	Catalyzes the hydrolysis of various acyl phosphate carboxyl-phosphate bonds	Regulate cell shape / single-species biofilm formation on inanimate substrate
E-value	2E-165	3E-103	2E-107	8E-58	2E-93 / 2E-88
Identities	239/239 (100%)	214/214 (100%)	159/159 (100%)	92/92 (100%)	168/168 (100%) / 182/182 (100%)

[a]The protein range (site) that can be aligned with the known protein.
[b]The nucleotide range (nt range) of predicted ORF within insert.
[c]Most similar protein, currently identified with blastp: protein name and protein size.
[d]Putative function of protein, currently identified with UniProt.
bp: base pair; aa: amino acids.

Based on successful use of functional metagenomics in various fields, it is well known that the quality of the metagenomic library plays an essential role in successful identification of functional genes. In this study, the following three measures were taken to ensure the quality of the library: (i) Successful construction of a high-copy dual-direction expression vector pSL. Disregarding frameshift mutations, this vector increases the likelihood that some exogenous fragments will be expressed, especially when the direction of the insert is consistent with the direction of the reverse lacZ promoter, which enhances the efficiency of library screening. (ii) Detection of library quality. To ensure that this library contains a variety of appropriately sized DNA fragments for screening, three different sizes of DNA were separately retrieved, the ligation system between them and the vector was optimized, and the ligation and transformation efficiencies were assessed. Finally, we obtained 11, 290 recombinant clones, containing 46.6 Mb DNA (the inserts of approximately 242 clones covered 1 Mb DNA); 85.5% of the colonies contained extraneous DNA fragments. Compared to other studies, 1 Mb DNA covers about 358 [28] or 157 clones [29], and the quality of our library was sufficient to perform the next step. (iii) The setting up of positive controls, pSL-orf1 and pSL-ffh, which can not only assist us to monitor whether there is a deviation in the course of the entire experiment but also allows us to initially evaluate the aciduric capability of the positive clones by comparing them with the growth of the positive-control colonies. These improvements guarantee the effectiveness and credibility of acid-resistant genes obtained in this study.

The two acid-resistant genes obtained in this study were aligned with sequences from two common cariogenic genera, *Actinomyces* and *Haemophilus* [30], which represent two of seven dominant cariogenic phyla (Actinobacteria and Proteobacteria) [31]. Bioinformatics analysis found that the ORF in pSL-2–41 encodes two domains (the malic and Malic_M domains) that form the N-terminal region of the malate dehydrogenase protein, but the acid-resistance assay indicated that this fragment is sufficient to allow *E. coli* to survive in this hostile environment. Jain's group [32] confirmed that the hosts *E. coli* DH5α and *E. coli* W3110 can tolerate acidic pH with the help of malate dehydrogenase, which mainly functions to pump out protons and restore the cytoplasmic pH back to neutral. Here, the additional heterologous copy of the malate dehydrogenase from *H. haemolyticus* (*E. coli* DH10B/pSL-2–41) can assist the host strain in tolerating the more hostile environment (pH 1.9) compared to the negative control (*E. coli* DH10B/pSL), which indicated that exogenous insert may enhance the acid resistance of host bacteria. Whether the detailed mechanism is consistent with *E. coli* requires further study. In addition, this protein can catalyze the interconversion between malate and oxaloacetate with the help of cofactor NAD^+ or $NADP^+$ and plays a critical role in the central oxidative pathway of living organisms [33]. In gram-negative bacteria, this protein generally forms a dimeric molecule and participates in the tricarboxylic acid (TCA) cycle. In addition, this protein can protect against oxidative stress via a reductive TCA cycle [34]. Thus, malate dehydrogenase plays an important role in helping the host resist acid and oxidative stress.

The other positive clone, pSL-1–31, encodes an ATP/GTP binding protein, which is involved in various biological processes; however, its acid-resistant properties and mechanisms remain unknown. Studies showed that this gene plays a pivotal role in a variety of biology processes in both prokaryotes and eukaryotes, including cell signalling [35], cell division, protein export, membrane transport, DNA repair, and multidrug resistance in tumours [36,37]. The functional range of the nucleotide binding protein family is gradually expanding with the accession of some new members. Harris et al. [38] described a gene involved in axon regeneration from the sciatic nerve of mice, named *Nna I*. Subsequent analysis revealed that it is a novel nucleotide binding protein related to zinc carboxypeptidases. In addition, the proteins encoded by disease-resistant genes in most plants contain ATP/GTP binding motifs, such as kinase-1a, kinase-2, and kinase-3a motifs, which confer resistance on the host tomato against *Fusarium oxysporum*, root-knot nematodes or potato aphids [39]. However, this study is the first to confirm the acid-resistance of the ATP/GTP binding protein. To preliminarily understand its possible mechanisms, the available mechanisms of acid resistance are summarized in Box1 in the Supplementary File. After acid exposure, the intracellular pH decreases, and DNA is damaged. The cells can resist the acid stress through a variety of defence mechanisms, such as regulating the flow of protons inside and outside cells [11,13] and neutralizing intracellular protons through producing alkaline compounds [40]. Further, the DNA repair enzyme system can also assist cells to resist acid stress through repair of the damaged DNA [12]. The acid-resistance property of the ATP/GTP binding protein could be attributed to the above mechanism. In addition, due to previously established consistency between various mechanisms of antibiotic resistance and acid-resistance [41], we speculate that multidrug resistance was probably involved in influencing acid-resistance. Of course, more in-depth studies are required to confirm this speculation.

Only a low number of functional genes were identified in this study, and no genes were obtained from *S. mutans*. One possible explanation for this is that addition of a single gene is often not sufficient to detect a

new phenotype, such as acid-resistance. If a single gene encodes only a subunit of a multi-unit enzyme that is critical for an enhanced acid resistance, no phenotype will result. Further, expression of the gene in question and successful translation may not be efficient in the *E. coli* host since some gram-positive proteins are toxic when translated within *E. coli*. An alternative explanation might be the limitations of the samples and method: (i) Studies have confirmed that *S. mutans* is not the only predominant caries-pathogenic bacterium [14,15]; depending on individual patients, it might not be the principal ECC bacterium [42]. Hence, in this study, the abundance of *S. mutans* might only occupy a small fraction of the collected caries samples, limiting the genetic information in our library and reducing the chance of obtaining functional genes. (ii) Some genetic information could be lost during the library construction process. Alternatively, it is known that the process of ligation of the various inserts into a vector is accompanied by a certain amount of randomness and uncontrollability [16], which can diminish or completely lose the genetic information of *S. mutans* and some other strains.

Conclusion

This is the first investigation that identifies the acid-resistant genes in the oral microbiota through improved functional metagenomics. Several known and unknown acid-resistant genes are obtained from the library of dental plaque and *S. mutans* in this study, which not only provide evidence for the feasibility of the experimental methods and the credibility of the results, but also promote studies to investigate the aciduric mechanism and the prevention and treatment of ECC. A larger library is needed to be constructed to obtain more functional genes in future work.

Acknowledgements

The authors would like to thank each kindergarten, children and their parents for their cooperation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC) [Grant number: 81170954].

Notes on contributors

Dr. Meng Zhang is a PhD fellow at the department of Preventive Dentistry, School & Hospital of Stomatology, Wuhan University. Her research focuses on studying the community composition of oral biofilm, and the pathogenic mechanism of cariogenic microorganisms.

Dr. Yuqiao Zheng is a research fellow with an interest in the study of cariogenic microorganisms.

Prof. Yuhong Li is an expert in dental endodontics and interested in the clinical study of dental caries.

Prof. Han Jiang is an expert in the department of preventive dentistry with an interest in the epidemiological study of dental caries, especially ECC.

Yuping Huang is Associate Professor at the College of Life Sciences, Wuhan University, with an interest in molecular biology.

Prof. Minquan Du is director of the department of preventive dentistry, School & Hospital of Stomatology, Wuhan University. Du's primary research focuses on plaque ecology in relation to dental caries and cariogenic mechanism of microorganisms.

References

[1] Petersen PE, Bourgeois D, Ogawa H, et al. The global burden of oral diseases and risks to oral health. Bull World Health Organ. 2005;83(9):661–669.
[2] Selwitz RH, Ismail AI, Pitts NB. Dental caries. Lancet (London, England). 2007;369(9555):51–59.
[3] Casamassimo PS, Thikkurissy S, Edelstein BL, et al. Beyond the dmft: the human and economic cost of early childhood caries. J Am Dent Assoc (1939). 2009;140(6):650–657.
[4] Editorial. Oral health: prevention is key. Lancet (London, England). 2009;373:1. DOI:10.1016/s0140-6736(08)61933-9
[5] Loesche WJ. Role of *Streptococcus mutans* in human dental decay. Microbiol Rev. 1986;50:353–380.
[6] Jensen ME, Polansky PJ, Schachtele CF. Plaque sampling and telemetry for monitoring acid production on human buccal tooth surfaces. Arch Oral Biol. 1982;27:21–31.
[7] Sonnenburg JL, Fischbach MA. Community health care: therapeutic opportunities in the human microbiome. Sci Transl Med. 2011;3:78ps12.
[8] Ward DM, Weller R, Bateson MM. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature. 1990;345:63–65.
[9] Martín-Galiano AJ, Ferrándiz MJ. de la Campa AG. The promoter of the operon encoding the F0F1 ATPase of *Streptococcus pneumoniae* is inducible by pH. Mol Microbiol. 2001;41:1327–1338.
[10] Quivey RG Jr., Kuhnert WL, Hahn K. Adaptation of oral streptococci to low pH. Adv MicrobPhysiol. 2000;42:239–274.
[11] Gutierrez JA, Crowley PJ, Cvitkovitch DG, et al. *Streptococcus mutans ffh*, a gene encoding a homologue of the 54 kDa subunit of the signal recognition particle, is involved in resistance to acid stress. Microbiology (Reading, England). 1999;145:357–366.
[12] Hanna MN, Ferguson RJ, Li YH, et al. uvrA is an acid-inducible gene involved in the adaptive response to low pH in *Streptococcus mutans*. J Bacteriol. 2001;183(20):5964–5973.
[13] Lemos JA, Luzardo Y, Burne RA. Physiologic effects of forced down-regulation of dnaK and groEL expression

in *Streptococcus mutans*. J Bacteriol. 2007;189:1582–1588.
[14] Gross EL, Leys EJ, Gasparovich SR, et al. Bacterial 16S sequence analysis of severe caries in young permanent teeth. J Clin Microbiol. 2010;48:4121–4128.
[15] Wade WG. The oral microbiome in health and disease. Pharmacol Res. 2013;69:137–143.
[16] Lam KN, Cheng J, Engel K, et al. Current and future resources for functional metagenomics. Front Microbiol. 2015;6:1196.
[17] Guazzaroni ME, Morgante V, Mirete S, et al. Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. Environment Microbiol. 2013;15:1088–1102.
[18] Lee SM, Donaldson GP, Mikulski Z, et al. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature. 2013;501:426–429.
[19] Sommer MO, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. Science (New York, NY). 2009;325:1128–1131.
[20] Zhang M, Chen Y, Xie L, et al. Pyrosequencing of plaque microflora in twin children with discordant caries phenotypes. PloS One. 2015;10:e0141310.
[21] Tatusov RL, Natale DA, Garkavtsev IV, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 2001;29:22–28.
[22] Bateman A, Coin L, Durbin R, et al. The Pfam protein families database. Nucleic Acids Res. 2004;32:D138–S141.
[23] Henikoff JG, Greene EA, Pietrokovski S, et al. Increased coverage of protein families with the blocks database servers. Nucleic Acids Res. 2000;28:228–230.
[24] Lemos JA, Chen YY, Burne RA. Genetic and physiologic analysis of the groE operon and role of the HrcA repressor in stress gene regulation and acid tolerance in *Streptococcus mutans*. J Bacteriol. 2001;183:6074–6084.
[25] Svensater G, Larsson UB, Greif EC, et al. Acid tolerance response and survival by oral bacteria. Oral MicrobiolImmunol. 1997;12:266–273.
[26] Foster JW. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nature Rev Microbiol. 2004;2:898–907.
[27] Baker-Austin C, Dopson M. Life in acid: pH homeostasis in acidophiles. Trends Microbiol. 2007;15:165–171.
[28] Morgante V, Mirete S, De Figueras CG, et al. Exploring the diversity of arsenic resistance genes from acid mine drainage microorganisms. Environment Microbiol. 2015;17:1910–1925.
[29] Torres-Cortés G, Millán V, Ramírez-Saad HC, et al. Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. Environment Microbiol. 2011;13:1101–1114.
[30] Xu X, He J, Xue J, et al. Oral cavity contains distinct niches with dynamic microbial communities. Environment Microbiol. 2015;17:699–710.
[31] Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, et al. The oral metagenome in health and disease. Isme J. 2012;6:46–56.
[32] Jain PK, Jain V, Singh AK, et al. Evaluation on the responses of succinate dehydrogenase, isocitrate dehydrogenase, malate dehydrogenase and glucose-6-phosphate dehydrogenase to acid shock generated acid tolerance in *Escherichia coli*. Adv Biomed Res. 2013;2:75.
[33] Takahashi-Íñiguez T, Aburto-Rodríguez N, Vilchis-González AL, et al. Function, kinetic properties, crystallization, and regulation of microbial malate dehydrogenase. J Zhejiang Univ Sci B. 2016;17:247–261.
[34] Oh TJ, Kim IG, Park SY, et al. NAD-dependent malate dehydrogenase protects against oxidative damage in *Escherichia coli* K-12 through the action of oxaloacetate. Environ ToxicolPharmacol. 2002;11:9–14.
[35] Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. Cell. 2010;141:1117–1134.
[36] Higgins CF, Gallagher MP, Mimmack ML, et al. A family of closely related ATP-binding subunits from prokaryotic and eukaryotic cells. BioEssays: News Rev Mole Celldev Biol. 1988;8:111–116.
[37] Higgins CF, Hiles ID, Salmond GP, et al. A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature. 1986;2-8(323):448–450.
[38] Harris A, Morgan JI, Pecot M, et al. Regenerating motor neurons express Nna1, a novel ATP/GTP-binding protein related to zinc carboxypeptidases. Mol Cell Neurosci. 2000;16:578–596.
[39] Tameling WI, Elzinga SD, Darmin PS, et al. The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. Plant Cell. 2002;14:2929–2939.
[40] Lu P, Ma D, Chen Y, et al. L-glutamine provides acid resistance for *Escherichia coli* through enzymatic release of ammonia. Cell Res. 2013;23:635–644.
[41] Lin J, Nishino K, Roberts MC, et al. Mechanisms of antibiotic resistance. Front Microbiol. 2015;6:34.
[42] Corby PM, Lyons-Weiler J, Bretz WA, et al. Microbial risk indicators of early childhood caries. JClin Microbiol. 2005;43:5753–5759.