Hybrid sequence-based analysis reveals the distribution of bacterial species and genes in the oral microbiome at a high resolution

Masaya Yamaguchi

yamaguchi.masaya.dent@osaka-u.ac.jp

Osaka University Graduate School of Dentistry

Toshihiro Uchihashi

Osaka University Graduate School of Dentistry

Shigetada Kawabata

Osaka University Graduate School of Dentistry

Research Article

Keywords: Microbiome, bacterial single-cell sequencing, 16S rRNA sequencing, metagenome shotgun sequencing

Posted Date: January 3rd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3815264/v1

License: ☑️ ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Biochemistry and Biophysics Reports on July 1st, 2024. See the published version at https://doi.org/10.1016/j.bbrep.2024.101717.
Abstract

Bacteria in the oral microflora are poorly identified owing to the lack of established culture methods for them. Thus, this study aimed to use culture-free analysis techniques, including bacterial single-cell genome sequencing, to identify bacterial species and investigate gene distribution in saliva. Saliva samples from the same individual were classified as inactivated or culturable and then analyzed using 16S rRNA sequencing, metagenomic shotgun sequencing, and bacterial single-cell sequencing. The results of 16S rRNA sequencing revealed similar microbiota structures in both samples, with Streptococcus being the predominant genus. Metagenomic shotgun sequencing showed that approximately 80% of the DNA in the samples was of non-bacterial origin, whereas single-cell sequencing showed an average contamination rate of 10.4% per genome. Single-cell sequencing also yielded genome sequences for 43 out of 48 wells for the inactivated samples and 45 out of 48 wells for the culturable samples. With respect to resistance genes, four out of 88 isolates carried cfxA, which encodes a β-lactamase, and four isolates carried erythromycin resistance genes. Tetracycline resistance genes were found in nine bacteria. Metagenomic shotgun sequencing provided complete sequences of cfxA, ermF, and ermX, whereas other resistance genes, such as tetQ and tetM, were detected as fragments. In addition, virulence factors from Streptococcus pneumoniae were the most common, with 13 genes detected. This study demonstrates that single-cell sequencing can be used to investigate the distribution of resistance genes and virulence factors among individual bacteria in human saliva.

1 Introduction

The oral microbiome is the collection of microorganisms that inhabit the oral cavity and are present in the human saliva. The oral microbiome is influenced by various factors, including diet, oral hygiene, age, and health status [1]. In general, a healthy oral microbiome is dominated by bacteria belonging to the phyla Firmicutes, Actinobacteria, and Proteobacteria, including Streptococcus, Rothia, and Neisseria [1, 2]. The oral microbiome plays an important role in maintaining oral and general health. Some bacteria in the oral microbiome produce antimicrobial compounds that prevent the growth of harmful bacteria in the oral cavity [3]. Oral microbiome dysbiosis is associated with systemic health conditions, such as cardiovascular diseases, diabetes, and respiratory infections [1, 2].

Microbiome analyses have provided additional information regarding the structure of the oral microbiota. However, many unculturable bacteria are present, and the individual bacteria or genes in the oral microbiota remain unidentified [2]. Similar to the gut, the oral cavity may facilitate the accumulation of diverse bacteria and cross-species transfer of genes. Some novel capsular types of Streptococcus pneumoniae and drug resistance genes are derived from oral Streptococcus [4–7]. In addition to cross-species gene transfer, mobile genetic elements, such as plasmids containing antimicrobial resistance (AMR) genes, are transmissible between bacteria. Thus, droplet infection of saliva could be a global source of drug resistance.
Amplicon and metagenomic shotgun sequencing have been used in microflora analysis. Specifically, 16S rRNA amplicon sequencing has been used to analyze the structure of the microflora, but its accuracy is insufficient to estimate phylogeny at the species level [8, 9]. Metagenomic shotgun sequencing allows for the comprehensive analysis of gene structure. However, reconstructing reads in individual bacterial genomes via de novo assembly and binning is challenging [10]. Additionally, in-depth analysis requires a large number of reads, which increases the costs of sequencing and computational resources. Moreover, not all reads obtained can be used to analyze the bacterial layer owing to contamination of the host DNA.

Bacterial single-cell sequencing has also been used for microflora analysis [11, 12]. However, bacteria are approximately 1/10th the size and 1/1000th the amount of DNA in host cells [13–15]. A previous study that used single-cell sequencing only achieved an average genome completeness of approximately 14% in 180 single bacterial cells [16]. Using single-amplified genome (SAG) gel technology, Hosokawa et al. obtained an average genome completeness of 31.8% in 346 isolates [11].

Considering the lack of culture methods for identifying bacterial species in the oral microflora, we aimed to use amplicon, metagenomic shotgun, and single-cell sequencing to identify bacterial species and investigate gene distribution in saliva samples. The bacterial genomes obtained showed a distribution of virulence factors and resistance genes with high accuracy and indicated the possibility of a novel genus.

2 Material and Methods

2.1 Sample preparation

After fasting for > 30 min in the morning, a healthy donor provided 5 mL of saliva. A 1 mL aliquot of the sample was added to OMNIgene ORAL solution (DNA Genotek Inc., Canada), which inactivates but stabilizes bacterial cells, and then stored at room temperature until single-cell sequencing. The presence of live bacteria was investigated as follows. In brief, 3 mL of saliva was centrifuged at 8000 rpm for 10 min, suspended in 800 µL of 50% glycerol/RPMI1640 solution, and then stored at -30°C until single-cell sequencing. The 16S rRNA sequencing, metagenome shotgun sequencing, single-cell isolation, genome amplification, and paired-end genome sequencing of both saliva samples were performed by bitBiome Inc. in Japan.

2.2 16S rRNA gene sequencing and analysis

The V3–V4 hypervariable regions of 16S rRNA genes were analyzed using the Illumina protocol for the preparation of 16S Metagenomic Sequencing Library. Barcoded amplicons, amplified with 341F and 806R primers (forward, 5’-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3’; reverse, 5’-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3’) were sequenced using the Illumina MiSeq 2 × 300 bp platform. The FASTQ data were analyzed using BitBiome Inc. Raw reads were processed with QIIME2 v.2020.2 [17] using DADA2 denoising paired with the options -p-trim-left-f 20, -p-trim-left-r 5, -p-trunc-len-f 0, and -p-trunc-len-r 250 [18]. After quality filtering, bacterial
taxa were estimated using the feature-classifier classify-sklearn and a SILVA reference database [19]. The data were visualized using “taxa collapse” and “taxa barplot.”

2.3 Metagenome shotgun sequencing and analysis

Total DNA was extracted from the saliva samples by using The QIAamp PowerFecal Pro Kit (QIAGEN, Germany). Each sample was sequenced using Illumina MiSeq 2 × 75 bp. FASTQ data were analyzed using BitBiome Inc. The sequence reads were filtered using bbduk.sh v38.79 (https://sourceforge.net/projects/bbmap/) with the parameters qtrim = r, trimq = 10, minlength = 40, maxns = 1, and minavgquality = 15, and human genome contaminations were removed by mapping with bbmap.sh 38.79 with the parameters quickmatch fast untrim minid = 0.95, maxindel = 3, bw = 12, minhits = 2, and human_masked_index qtrim rl trimq = 10 [20]. The processed sequences were assembled with Megahit v1.1.3 using default settings [21]. Metagenome-assembled genomes (MAGs) were constructed by binning with MetaBAT2 in metaWRAP v.1.3.2 using default settings [22, 23]. The coding sequences of the contigs and MAGs were examined with Prokka v1.14.6 using the options --rawproduct and --mincontiglen 200 [24]. The number, total length, and GC content of the assembled contigs and MAGs were evaluated using QUAST v5.0.2 [25]. The assembled contigs were taxonomically classified with MetaPhlAn3 using default settings and mpa_v30_CHOCOPhlAn_201901 [26], whereas the MAGs were taxonomically classified using GTDB-Tk v1.3.0 and database release 95 [27]. The completeness and contamination of the MAGs were calculated using CheckM v1.1.2 [28].

2.4 Bacterial single-cell genome sequencing with SAG-gel

Single cells were isolated into gel beads, and their genomes were amplified using the SAG gel method [11]. Single-gel beads harboring a SAG were sorted into individual wells of a 96-well microplate. Each SAG was sequenced using Illumina MiSeq 2 × 75 bp. The FASTQ data were analyzed using bitBiome Inc. Sequence reads were filtered using bbduk.sh v38.79, and human genome contaminations were removed by mapping with bbmap.sh 38.79 and metagenome shotgun sequencing. The processed sequences were assembled with SPAdes v3.14.0, using the settings --sc, --careful, --disable-rr, and --disable-gzip-output, and contigs > 1000 bp were retained for subsequent analyses. The coding sequences were examined with Prokka v1.14.6 using the options --raw product and --mincontiglen 200 [24]. The contig number, total length, and GC content of the assembled genomes were evaluated using QUAST v5.0.2 [25]. Completeness and contamination were calculated using CheckM v1.1.2 [28]. The assembled genomes were taxonomically classified using GTDB-Tk v1.1.1 and the database release 95 [27].

2.5 Sequencing-based profiling

Sequencing-based profiling was performed as previously described [29]. Quality control and preprocessing of the FASTQ files from the next-generation sequencing were performed using fastp v.0.20.0 [30]. To identify the bacterial species, we performed an average nucleotide identity (ANI) analysis of the assemblies by using Microbial Genomes Atlas MiGA online (http://microbial-genomes.org/) [31]. The AMR and virulence factor profiles were determined using ARIBA 2.14.4, with cleaned sequencing data [32]. We used the Comprehensive Antibiotic Resistance Database (CARD) v.3.0.8 [33] and core and full
datasets of the virulence factor database (VFDB) [34] as reference for ST, AMR, and virulence factor profiling, respectively. The minimum percentage identities for the assemblies were set to 93 and 90 for the CARD and other databases, respectively. The analyzed data were visualized using Phandango [35].

2.6 Ethical statement

The study was conducted with written informed consent from the donor and approved by the Institutional Review Board of Osaka University Graduate School of Dentistry (R4-E4).

3 Results

3.1 Comparison of 16S rRNA, metagenome shotgun, and bacterial single-cell sequencing on the human salivary microbiome

The results of the 16S rRNA sequencing, metagenome shotgun sequencing, 48-well single-cell isolation, and short-read genome sequencing of the inactivated and culturable samples are shown in Figure S1. Taxonomic bar plots at the genus level based on 16S rRNA analysis revealed that the inactivated and culturable samples had similar microbiome structures. In both samples, *Streptococcus* was the predominant genus, followed by *Prevotera*, *Neisseria*, and *Veillonella* (Figure 1A and Table S1). For bacterial single-cell analysis, genome sequences were obtained from 43 out of 48 wells for the OMNiGene-preserved samples and from 45 out of 48 wells for the glycerol stock samples. Genomic completeness greater than 80% was achieved in 17 wells for the OMNiGene-preserved samples and in 24 wells for the glycerol stock samples compared with known genomic sequences (Figure 1B and Table S2). Similar to the 16S rRNA sequencing results, single-cell sequencing results showed that *Streptococcus* was the most abundant genus in the samples, followed by *Prevotella*. By contrast, the percentage of *Neisseria*, which was high in the 16S rRNA sequencing, was low, and *Veillonella* and *Alloprevotella* were not detected. In addition, 60 bacterial genera were detected using 16S rRNA sequencing, whereas only 17 genera were detected using single-cell sequencing (Figure 1A and 1C, and Table S2).

The total raw read counts for the metagenomic shotgun and single-cell analyses were 61,126,868 and 55,918,930, respectively (Tables S2 and S3). Metagenomic shotgun sequencing revealed an average contamination rate of 81.6%, indicating the difficulty in separating bacterial DNA from human saliva-derived specimens (Figure 1D and Table S3). By contrast, bacterial single-cell sequencing obtained a much lower average contamination rate of 10.4% per genome because of the single-cell separation process (Figure 1D and Table S2). Metagenome binning yielded nine bins from metagenome assemblies, of which eight were identified by metagenome shotgun sequencing and GTDBtk analysis, and 44 strains were identified at the species level by bacterial single-cell sequencing and GTDBtk analyses (Tables S2 and 1).

3.2 Detection of antimicrobial resistance genes and virulence factor genes from metagenome shotgun and bacterial single-cell sequencing
Metagenome shotgun sequencing revealed that cfxA encoding a β-lactamase and the erythromycin resistance genes ermF and ermX were complete sequences, whereas the other AMR genes were fragmented (Figure 2). Single-cell sequencing showed that four of the 88 isolates harbored cfxA and four harbored ermF. For tetracycline resistance, nine isolates carried tet32, tetM, tetO, or tetQ (Figure 3).

Both analyses detected fewer genes encoding virulence factors than AMR genes. Metagenome shotgun sequencing revealed intact gapA, an essential gene encoding an enzyme for glycolysis, whose ortholog inhibits the biological effects of C5a on human neutrophils [36]. Bacterial single-cell sequencing showed that in addition to gapA, the two samples contained 12 intact genes encoding virulence factors derived from S. pneumoniae (Figure 4). SPH0456+ (cpsB), cps4J, cps4K, and cps4L are pneumococcal polysaccharide capsule synthesis genes; lmb and pavA encode pneumococcal adhesins; piuA and psaA_1 encode pneumococcal transporters; nanB encodes sialidase; slrA encodes peptidyl-prolyl cis-trans isomerase; srtA encodes sortase A; and tig_ropA_2 encodes a trigger factor. STER1442, STER1444+, epsB+, and tig_ropA_5 were isolated from Streptococcus thermophilus, fbpA_6, psaA_3, and rfbB_1+ from Streptococcus gordonii, and ctrC_1 and lbpA from Neisseria. Concerning pneumococcal virulence factors, genes encoding pneumococcal cell surface proteins such as choline-binding proteins (cbpD, cbpG, lytA, and lytB) and cell wall anchoring proteins (iga, nanA, pavB1/ pfbB, pfbA, zmpB, and zmpC) were also detected in fragments. We previously reported that the orthologs of iga, nanA, pfbA, zmpB, and zmpC are distributed among closely related streptococcal species, including oral Streptococcus, which is consistent with our results [37-39].

Although the host species of several of the detected pneumococcal virulence factors remained unclear, several pneumococcal virulence factors were harbored by oral streptococci. Neisseria ctrC was detected in a single-cell isolate, OSU002-0007, which was predicted to be Neisseria mucosa. Another single-cell isolate containing Neisseria lbpA was not identified in this species. Bacterial single-cell sequencing allowed us to elucidate the level at which bacteria have specific genes, which is difficult to achieve with metagenomic shotgun sequencing.

In the MiGA ANI analysis, 10 of the 88 genomes met >95% of the criteria for species identification (Table 2). Although several genomes were predicted to be S. pneumoniae through GTDBtk taxonomy analysis (Table S2), MiGA ANI analysis showed that S. pneumoniae had the highest ANI value (91.4 %) in only one genome (OSU002-0038; Table 2). Despite the high genomic completeness of 99% determined by CheckM, the ANI value was less than 95%, indicating that no bacteria were virtually identified as S. pneumoniae in the samples. S. pneumoniae belongs to the mitis group of oral Streptococcus and cannot be distinguished from S. mitis or S. oralis through 16S rRNA sequencing, and some strains are difficult to identify even by biochemical tests [29, 40, 41]. Therefore, care must be taken to distinguish bacteria from closely related species at the species level. ANI analysis also suggested five single-cell-isolated bacteria as potential novel species (Table 2). Bacterial single-cell sequencing could be a powerful tool for searching for novel bacterial species in the microbiome.

Discussion
Saliva plays a multifaceted role in digestion, tooth remineralization, and oral cavity cleaning. However, it is also a vector of droplet infections. Daily activities such as talking, coughing, and sneezing produce large amounts of respiratory droplets that are subsequently deposited on dry surfaces [42]. These droplets facilitate the transmission of bacteria such as *S. pneumoniae* [43]. However, many microorganisms in the saliva are unidentified. Identifying the salivary microflora is crucial to understand the mode of human-to-human transmission of microorganisms and/or AMR genes.

*S. pneumoniae* can acquire genetic material from oral streptococcal species, resulting in drug resistance and serotype replacement [4-7]. Jensen *et al.* comprehensively assessed the involvement of homologous recombination between oral streptococci and *S. pneumoniae* in acquiring β-lactam drug resistance. This study focused on the diversity at the DNA and amino acid levels of the transpeptidase region of *pbp2x* in 107 strains, *pbp2b* in 96 strains, and *pbp1a* in 88 strains of oral *Streptococcus* [4]. The findings revealed that polymorphic sites arising from spontaneous mutations in *pbp* accounted for 39% of all polymorphic sites observed in susceptible and resistant strains of *S. mitis*, *S. oralis*, and *Streptococcus infantis*. By contrast, extensive sequence variation was observed exclusively in resistant strains of *S. pneumoniae*. These results suggested that the previously diversified sequence in oral streptococci was imported by *S. pneumoniae* possibly because of the selective pressure exerted by antimicrobial agents.

In 2020, Ganaie *et al.* reported the discovery of the 100th pneumococcal capsular serotype, designated as 10D [7]. This study revealed that the capsular synthesis genes of serotype 10D exhibited three large regions of homology with genes arranged in the same order (syntenic) as those found in serotypes 6C and 39 and the capsular synthesis genes of *S. mitis* SK145. Notably, the syntenic region of 10D with SK145 spanned approximately 6000 bp and included a short fragment of *wciNa* at the 5’ end. The presence of this nonfunctional *wciNa* fragment provided compelling evidence of interspecies gene transfer from oral streptococci to *S. pneumoniae*. Moreover, the sequence of *wcrO*<sub>10D</sub>, a capsular synthesis gene cluster of serotype 10D, displayed low homology (40%–50% amino acid identity) with the *wcrO* genes of serotypes 33C, 34, 35F, and 36 despite the sequencing of the capsular synthesis gene cluster in over 20,000 pneumococcal strains. By contrast, *wcrO*<sub>10D</sub> exhibited surprisingly high homology (94% amino acid identity) with *RS00925* from *S. mitis* SK145. These findings suggest that the 100th capsular serotype 10D arose from the acquisition of the *S. mitis* gene by *S. pneumoniae*. Resistance and diversification of capsular types of *S. pneumoniae* pose significant threats to human health, and understanding the sources of this diversity is crucial. These results indicate that oral streptococci serve as an external genetic pool for pneumococci. Thus, elucidating the genetic diversity of the oral flora is important.

In the present study, saliva samples from the same individuals were classified as inactivated or culturable and then analyzed using 16S rRNA sequencing, metagenomic shotgun sequencing, and bacterial single-cell sequencing. Both inactivated and culturable samples were suitable for analysis, but the inactivated samples were preferred when dealing with samples that may pose a risk to analysts, such as those from COVID-19-infected individuals. Genomic sequencing facilitated the exploration of the
metabolic systems of unculturable bacteria present in the samples, potentially allowing the cultivation of previously unculturable bacteria from culturable samples.

Traditionally, metagenomic shotgun sequencing has been used to study gene functions in the bacterial flora. However, metagenome shotgun analysis involves binning to identify bacterial species and then constructing a metagenome-assembled genome to search for gene distribution [44]. Although the binning technology has improved and contributed to the identification of many bacteria, the high homology of essential genes among species makes complete classification extremely difficult. In other words, identifying which bacteria possess the genes detected by metagenome shotgun sequencing is a future challenge and is expected to be improved by short- and long-read hybrid analysis, long-read deep sequencing combined with Hi-C-seq, and other techniques [45, 46].

The single-cell analysis performed in this study proved to be highly effective, allowing the precise identification of genes present in individual bacteria. In addition, the use of preisolated bacteria eliminated the need for binning, which poses challenges in terms of improving accuracy. Sequencing DNA in a single bacterial cell can reveal the host of mobile genetic elements, such as plasmids and phages, whereas Hi-C-seq provides a community composition profile. Elucidating the dynamics of gene transfer between bacterial species at high frequencies is essential to understand the spread of resistance genes through saliva and develop effective control strategies. Future studies could combine multiple methods to elucidate the human microbiota structure and identify previously unrecognized bacterial species and genetic characteristics.

**Abbreviations**

antimicrobial resistance (AMR), single-amplified genome (SAG), metagenome-assembled genomes (MAGs), average nucleotide identity (ANI), Comprehensive Antibiotic Resistance Database (CARD), virulence factor database (VFDB)

**Declarations**

**Ethics approval and consent to participate**

The study was conducted with written informed consent from the donor and approved by the Institutional Review Board of Osaka University Graduate School of Dentistry (R4-E4).

**Consent for publication**

Not applicable

**Availability of data and materials**

The 16S rRNA amplicon sequencing, metagenome-shotgun sequencing, and bacterial single-cell sequencing data are deposited in the DNA Data Bank of Japan under BioProject PRJDB16375. The DRR
run numbers are DRR502934-DRR503025.

**Competing Interests**

The authors declare that they have no competing interests.

**Funding**

This study was partly supported by AMED (JP20wm0325001), the Japan Society for the Promotion of Science KAKENHI (grant numbers 20KK0210, 21K10137, 22H03262, 22K19618, 22K19619, and 23H03073), Takeda Science Foundation, the Joint Research Program of the Research Center for GLOBAL and LOCAL Infectious Diseases, Oita University (2022B05), and the Drug Discovery Science Division, Open and Transdisciplinary Research Initiatives, Osaka University. This work was conducted as part of “The Nippon Foundation - Osaka University Project for Infectious Disease Prevention.” The funders had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

**Authors’ contributions**

M.Y. and S.K. designed the study. M.Y. and T.U. performed sampling. M.Y. performed bioinformatics analyses. M.Y. wrote the manuscript. T.U., and S.K. contributed to the writing of the manuscript.

**Acknowledgements**

This study was partly achieved using the NIG supercomputer at the ROIS National Institute of Genetics and SQUID at the Cybermedia Center, Osaka University, and supported by the “Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN)” in Japan (Project ID: EX22701 and jh230035).

We acknowledge the Bioinformatics Research Unit of Osaka University Graduate School of Dentistry for the data analysis.

**References**


36. Terao Y, Yamaguchi M, Hamada S, Kawabata S. Multifunctional glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus pyogenes* is essential for evasion from neutrophils. J Biol Chem


**Tables**

Tables are available in the Supplementary Files section.

**Figures**
Figure 1. Human saliva microbiome analysis by 16S rRNA, meta-genome shotgun, and bacterial single cell sequencing.

After fasting for more than 30 minutes in the morning, 5 mL of saliva was collected from a healthy donor. One mL was added to OMNigen ORAL solution, and then stored at room temperature. Another 3 mL of saliva was suspended in 800 μL of 50% glycerol/RPMI1640 solution and stored at -30°C. A. Taxonomic bar plots by 16S rRNA analysis. Individual taxonomy legend and relative frequency values were shown in Table S1. B. Genome completeness and contamination plots of single-cell analysis. Using 48 wells, single cell isolation, genome amplification, and paired-end genome sequencing of both saliva samples were performed by bitBiome. C. Taxonomy in single-cell analysis. Eight samples were excluded since those sequence data were not enough to identify the taxonomy. D. Human DNA contamination rate in metagenome shotgun and single-cell analyses. Two samples in single-cell analyses were excluded since any sequence data were obtained from the samples.

Figure 1

See image above for figure legend
Figure 2. A. Burden of antimicrobial resistance (AMR) genes in the metagenome shotgun sequenced samples. The reference data was obtained from CARD (Comprehensive Antibiotic Resistance Database). Green, light blue, blue, orange, and grey indicate matches to reference, interrupted, fragmented, partial, and lacking genes, respectively. The "interrupted" means that all of reference did not represent in the assembly, the "fragmented" means the gene was assembled to > 2 contigs, and the "partial" means that the gene was not a complete one containing start and stop codons. B. Burden of genes encoding virulence factors in the metagenome shotgun sequenced samples. The reference data was obtained from Virulence Factor Database (VFDB) full dataset. Green, light blue, blue, orange, and grey indicate matches to reference, interrupted, fragmented, partial, and lacked genes, respectively. The clustering trees were generated by ARIBA based on the gene distribution. Graphical data was obtained using Phandango.

Figure 2

See image above for figure legend
Figure 3. Burden of AMR genes in the single-cell isolated bacteria.
The reference data was obtained from CARD. Green, light blue, blue, orange, and grey indicate matches to reference, interrupted, fragmented, partial, and lacking genes, respectively. The clustering tree was generated by ARIBA based on the gene distribution. Graphical data was obtained using Phandango.

Figure 3

See image above for figure legend
Figure 4. Burden of genes encoding virulence factors in the single-cell isolated bacteria. The reference data was obtained from VFDB full dataset. Green, light blue, blue, orange, and grey indicate matches to reference, interrupted, fragmented, partial, and lacked genes, respectively. The clustering tree was generated using ARIBA based on the gene distribution. Graphical data was obtained using Phandango.

Figure 4

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
- FigS1.pdf
- SupplementaryTables.xlsx
- Table1.xlsx
- Table2.xlsx