

Genomic Features and Comparative Genomic Analysis of *Streptococcus* sp. v1. nov., Isolated from an Endophthalmitis patient

Meiqin Zheng

Wenzhou Medical University <https://orcid.org/0000-0002-6772-7162>

Yutong Kang

Wenzhou Medical University

Yi Xu

Wenzhou Medical University

Fangyi Xu

Wenzhou Medical University

Yanjiong Chen (✉ chenyanjiong@126.com)

Xian Jiaotong University: Xi'an Jiaotong University

Research Article

Keywords: Endophthalmitis, metagenome-assembled genome, *Streptococcus* sp. v1. nov., comparative genomics

Posted Date: August 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-285829/v1>

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

[Read Full License](#)

Abstract

Endophthalmitis is an acute inflammatory intraocular condition that can cause permanent vision loss. The treatment strategy and visual outcome partly depend on the identification of the pathogenic agents. In this study, metagenomic sequencing was conducted to investigate the microbiome of the vitreous of an endophthalmitis patient whose disease progressed rapidly and was accompanied by severe pain. Metagenomic sequencing data revealed that the vitreous sample was predominated by *Streptococcus*, with a low-diversity microbiome. The main antibiotic resistance gene (ARG) types included beta-lactam, macrolide-lincosamide-streptogramin, and multidrug resistance genes. Additionally, the metagenome-assembled genome sequence of *Streptococcus* sp. v1. nov. was isolated. The average nucleotide identities (ANI) and DNA–DNA hybridization (DDH) analysis between the genomes of *Streptococcus* sp. v1. nov and related species revealed that *Streptococcus* sp. v1. nov belongs to a new species. Pan/core genome analysis for *Streptococcus* sp. v1. nov. and tetra correlation search (TCS) top 25 hit strains revealed that most unique genes of *Streptococcus* sp. v1. nov. were linked to the ABC transporter system, which could indicate the unique virulence and pathogenic potential of *Streptococcus* sp. v1. nov. In addition, 7 virulence factors were identified in total, and most of them were classified as “offensive virulence factors”. The high pathogenicity of *Streptococcus* sp. v1. nov. could explain the patient's rapid disease progression. Our study first isolated an ocular pathogen with high virulence based on metagenomic sequencing and bioinformatics analysis, which has important reference value for revealing the composition and genome characteristics of pathogens in endophthalmitis patients in the future.

1. Introduction

Endophthalmitis is a severe sight-threatening intraocular infection following eye trauma or surgery (Rejdak et al. 2016). Pathogens may be introduced from the patient's own microbiota or through contaminated solutions or instruments used during ophthalmic surgery (Bannerman et al. 1997)(Buchta et al. 2014). In the majority of cases, the detection of pathogens is accomplished by culture of vitreous and/or aqueous. However, traditional culture methods have some limitations in identifying hard-to-culture and slow-growing microorganisms. Culture-negative endophthalmitis is not uncommon, with an incidence between 20% and 30% (Durand 2017). Therefore, the spectrum of microorganisms involved in endophthalmitis is not fully described or characterized. A recent metagenomic sequencing study of the vitreous of patients with endophthalmitis after cataract surgery or intravitreal injection showed that the vitreous of patients with endophthalmitis contained a variety of microorganisms (Kirstahler et al. 2018).

Due to the low pathogen detection rate of traditional culture methods and polymerase chain reaction (PCR) (McCannel 2011)(Yao et al. 2013), clinically differentiating infectious endophthalmitis and sterile endophthalmitis is difficult. For sterile endophthalmitis, steroid treatment may yield better results than antibiotic treatment. Metagenome sequencing technology can identify the entire microbial community, including archaea, bacteria, fungi, and viruses, and is expected to be the most important tool for improving the treatment and diagnosis of infectious diseases. In addition, advances in sequencing

throughput and computing technology make it possible to recover culture-independent genomes from metagenomes (Parks et al. 2017).

Here, metagenomic sequencing was conducted on a vitreous sample of a posttraumatic endophthalmitis patient to reveal the composition of its pathogens and antibiotic resistance genes. We recovered a high-quality draft genome of "*Streptococcus* sp. v1. nov." by metagenome assembly and genome binning and performed subsequent in silico determinations of its taxonomic affiliation, phylogenetic relationship and potential virulence factors. In the future, using metagenomic sequencing and bioinformatics analysis to more quickly identify eye pathogens and antibiotic resistance will contribute to precision clinical treatment and reduce blindness incidence. High-quality genomes of unculturable or difficult-to-culture eye pathogens can also be obtained from metagenomic binning.

2. Material And Methods

2.1 Sample collection, DNA extraction, and metagenomic sequencing

The vitreous (intraocular body fluid) sample was obtained from a patient suffering from endophthalmitis and attending the Optometry and Eye Hospital Affiliated with Wenzhou Medical University. The patient was a 62-year-old man who presented with trauma to the right eye due to an iron splinter. On examination, there was a y-shaped, full-thickness corneal perforation wound. He had received debridement, suturing, anterior chamber flushing, emergent eye globe magnet aspiration, vitreous cavity injection (vancomycin, 1 mg/0.1 ml), systemic anti-inflammatory therapy (intravenous cefazolin and hydrocortisone) and topical antibiotic therapy (tobramycin and dexamethasone combination eye drops, levofloxacin ophthalmic solution, gatifloxacin ophthalmic gel, and pranopfen eye drops). However, the effect is unsatisfactory. Microscopic examination of smears found gram-positive cocci from the vitreous aspirate and vitreous humor. Because of progressing rapidly inflammation, severe pain and poor visual acuity, the patient decided to undergo surgery again, and then the right eye was enucleated.

Genomic DNA was extracted using a MicroElute Genomic DNA Kit (Omega, Guangzhou, China). Library construction and whole-metagenome shotgun sequencing were performed using a 2 × 150 bp paired-end protocol on the HiSeq X10 platform (Novogene, Beijing, China).

2.2 Bioinformatics analyses

The quality of raw sequencing reads was examined using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequences in the paired-end reads were removed with Cutadapt (<http://code.google.com/p/cutadapt/>). Afterward, low-quality reads were trimmed with Trim Galore (Wang et al. 2013). Clean metagenomic reads were visualized with SplicingViewer (Liu et al. 2012). Bowtie2 was used to map trimmed high-quality processed reads to the human reference genome (hg19) to obtain clean nonhuman sequences (Langmead and Salzberg 2012). Taxonomic annotation was executed using Metaphlan2 with default parameters (Truong et al. 2015).

ARG abundance was determined using the ARG-OAP online pipeline (<http://smile.hku.hk/SARGs>) (Yin et al. 2018). The number of copies of ARGs per copy of 16S rRNA was calculated and used as ARG abundance.

Metagenome-assembled genomes (MAGs) were reconstructed using the metaWRAP pipeline (Uritskiy et al. 2018). The clean nonhuman sequences were assembled with the metaWRAP Assembly module (Nurk et al. 2017). Initial binning was performed with the metaWRAP Binning module (`-concoct, -metabat2, and -maxbin2`) (Kang et al. 2019)(Wu et al. 2016)(Alneberg et al. 2014). Refined bins were obtained with the metaWRAP `bin_refinement` module (`-c 70 -x 5`) (Song and Thomas 2017) and subsequently reassembled using the metaWRAP `Reassemble_bins` module to improve assembly quality (Bankevich et al. 2012). The completion and contamination of the bins were evaluated with CheckM (Parks et al. 2015). The taxonomy of MAGs was estimated with the metaWRAP `classify_bins` module.

Repetitive sequences were identified using the RepeatMasker and RepeatModeler pipelines (Tarailo-Graovac and Chen 2009). Digital DDH was calculated using the Genome-to-Genome Distance Calculator (GGDC) (<https://ggdc.dsmz.de/>). The ANI was determined using FastANI software (Jain et al. 2018). Core-pan-genome analysis was performed using the Bacterial Pan Genome Analysis (BPGA) pipeline (Chaudhari et al. 2016). Then, we used Evolview to visualize a phylogenetic tree of the core gene sequences, which was constructed based on the unweighted pair-group method with arithmetic means (UPGMA) (Subramanian et al. 2019). We used BPGA software to access the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Cluster of Orthologous Groups (COG) database.

For variation analysis, we used *Streptococcus mitis* SK321 as a reference strain. Single-nucleotide polymorphism (SNP) and small insertion/deletion (InDel) variants were called using the GATK software suite (McKenna et al. 2010). Low-confidence variants were filtered using the VariantFiltration tool in GATK and `varFilter` option in SAMTOOLS (Li et al. 2009). Structural variants (SVs) were detected using Breakdancer (Fan et al. 2014). In addition, we applied CNVnator to identify copy number variation (CNV) (Abyzov et al. 2011).

2.3 GenBank accession numbers

The sequence data of the whole genome of *Streptococcus* sp. v1. nov. have been deposited in GenBank under the accession number SAMN15860715.

3. Results And Discussion

3.1 Microbial community revealed by metagenomic sequencing

Metagenomic shotgun sequencing of the endophthalmitis sample revealed a diverse microbial community. After trimming low-quality reads and filtering human sequences, a total of 1.78 million microbial reads were obtained for subsequent analysis. At the kingdom level, only bacteria were

identified. The bacterial sequences were classified as 2 phyla, 2 genera, and 4 species. The microbial community was dominated by *Streptococcus_mitis_oralis_pneumoniae* (97.98%), followed by *Pseudomonas_unclassified* (0.96%), *Helicobacter pylori* (0.72%) and *Neisseria_unclassified* (0.33%) (Fig. 1A). *Streptococcus* is a common pathogen causing severe infectious endophthalmitis (Yospaiboon et al. 2018). The virulence of pathogens is one of the factors affecting the prognosis of endophthalmitis (Sadaka et al. 2012). Compared with coagulase-negative *Staphylococcus* endophthalmitis, *Streptococcus* endophthalmitis has a worse visual prognosis (Lemley and Han 2007)(Durand 2017).

3.2 Abundance and categories of ARG types and subtypes

In summary, 6 ARG types comprising 14 ARG subtypes were identified in the endophthalmitis sample. Beta-lactam resistance genes were the dominant ARG type (4.41×10^{-1} copies/16S rRNA gene), followed by ARGs for macrolide-lincosamide-streptogramin (2.19×10^{-1} copies/16S rRNA gene) and multidrug resistance (1.79×10^{-1} copies/16S rRNA gene) (Fig. 1B). The high abundance of multidrug resistance may be one of the reasons why the patient was ineffectively treated with systemic and topical antibiotics. The top 5 ARG subtypes with the highest abundance were *mefA*, PBP-2X, multidrug transporter, PBP-1A, and PBP-1B (Additional Fig. 1). These results were valuable for guiding rational clinical medication prescription.

3.3 *Streptococcus* genome recovered from metagenomes

Metagenome assembly and genome binning based on the MetaWRAP pipeline resulted in one near-complete genomic bin with 23 scaffolds. The genome size of the obtained bin was 2 Mbp, encoding 1974 genes with a mean gene length of 907 bp. The genome was estimated at 99.78% completeness and 0.846% contamination, with a low GC content of 39.7%. Only 0.76% of this genome was occupied by repeat elements. Interspersed repeats were the predominant type of repeat region, which accounted for 0.55% of the whole genome.

The genomic bin has a single circular chromosome, and no plasmids were found. Blast searching genomic sequences against the NT database showed that the bin was classified as *Streptococcus* but could not be identified at the species level. Metagenomic taxonomy results showed that *Streptococcus_mitis_oralis_pneumoniae* were the most dominant species, accounting for 97.98%. Because *S. mitis*, *S. pneumoniae*, *S. pseudopneumoniae*, and *S. oralis* are closely related viridans group streptococcus species, they are difficult to discriminate at the species level (Scholz et al. 2012) (Ikryannikova et al. 2013). Therefore, there was a need to carry out genome comparisons of the *Streptococcus* MAG against related *Streptococcus* spp.

3.4 Comparative genome analysis

We used ANI and digital DDH for genome comparison analyses. The ANI is regarded as the most relevant comparative parameter to determine bacterial species. A whole-genome pairwise ANI > 95% indicates the same species (Jain et al. 2018). DDH is generally used to determine the genomic similarity among strains (Goris et al. 2007), and 70% similarity is considered the gold-standard threshold of DDH values for

species boundaries (Meier-Kolthoff et al. 2014). Table 1 contains the digital ANI and DDH values between this *Streptococcus* MAG and 5 type strains of related species. The digital DDH and ANI values of the *Streptococcus* MAG against the 5 type strains of related species were below 95% and 70%, respectively, which showed that the *Streptococcus* MAG represented a new species. This new *Streptococcus* strain was named *Streptococcus* sp. v1. nov. Consistent with the metagenomic taxonomic classification results, both the DDH and ANI values showed that *Streptococcus* sp. v1. nov. was highly similar to *Streptococcus gwangjuense* ChDC B345, *Streptococcus mitis* NCTC 12261, *Streptococcus pseudopneumoniae* CCUG 49455, *Streptococcus pneumoniae* NCTC 7465, and *Streptococcus oralis* subsp. dentisani CECT 7747. *Streptococcus gwangjuense* isolated from human pericoronitis is closely related to the mitis group of the genus *Streptococcus* (Park et al. 2019).

Table 1
The digital DDH and ANI values of comparisons between *Streptococcus* sp. v1. nov. and 5 type strains of related species.

Strains	DDH (%)	ANI (%)
<i>Streptococcus gwangjuense</i> ChDC B345 ^T	53.1	93.48
<i>Streptococcus mitis</i> NCTC 12261 ^T	54.7	92.84
<i>Streptococcus pseudopneumoniae</i> CCUG 49455 ^T	48.8	92.3
<i>Streptococcus pneumoniae</i> NCTC 7465 ^T	46.6	91.86
<i>Streptococcus oralis</i> subsp. dentisani CECT 7747 ^T	31.7	86.98
T: Type strain.		

The TCS function of JSpeciesWS can rapidly compare selected genomes against a continuously updated reference database (<ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank>) (Richter et al. 2016). TCS analysis showed that the draft genome of *Streptococcus* sp. v1. nov. was closest to the genome of *Streptococcus mitis* SK321, with a Z-score of 0.99859. To identify genetic differences between the two strains, we performed variant calling.

Comparing *Streptococcus* sp. v1. nov. with the reference *Streptococcus mitis* SK321, a total of 39979 high-quality SNPs were identified, including 28914 transitions (Ti) and 11065 transversions (Tv), with an average Ti/Tv ratio of 2.61. A >G|T > C type (14497) and G > A|C > T (14417) type accounted for the majority of all SNPs. The gain and loss of mutated genes are perceived as one of the most important contributors to functional changes (Nei and Rooney 2005). A total of 1588 InDels were identified, consisting of 809 insertions and 779 deletions. Moreover, 29 SVs containing 17 deletions, 1 inversion, and 11 translocations were uncovered in *Streptococcus* sp. v1. nov. genome. For CNV, 2 duplications and 43 deletions were identified (Fig. 2). Genomic variation is an important evolutionary driving force (Zhou et al. 2019).

3.5 Pan-core genome analysis

To find different characteristics between *Streptococcus* sp. v1. nov. and *Streptococcus mitis*, a comparative genome analysis for *Streptococcus* sp. v1. nov. and TCS top 25 hit *Streptococcus* strains (Additional Table 1) was performed by the BPGA pipeline. The core–pan plot showed that as the number of given genomes increased, the number of core genes decreased gradually (Fig. 3a). The core genome curve set plateaued, while the pan-genome trend curve grew continuously, indicating an open pan-genome and a conserved core genome. With the addition of new genomes, the number of genes in the core genome decreased from 1915 to 824, and the number of pan genome genes increased from 1915 to 5120. The pan-genome of all the strains had 718 core genes, 2591 accessory genes and 1810 unique genes. When compared with these related *Streptococcus* strains, *Streptococcus* sp. v1. nov. had the most accessory genes. A dendrogram of core genes showed that *Streptococcus* sp. v1. nov. had the closest phylogenetic relationship with *Streptococcus mitis* SK321 (Fig. 3b).

Functional analyses of COGs in the 26 *Streptococcus* genomes revealed that the highest proportion of core genome genes were related to “metabolism”, and the majority of unique gene families were mostly associated with “information storage and processing”. Previous research suggests that information storage and processing categories are linked to intracellular survival (Yang et al. 2018). The core genes were mainly enriched in [J] translation, ribosomal structure & biogenesis, [R] general function prediction only, and [E] amino acid transport & metabolism. In the categories of [K] transcription, [L] replication, recombination & repair, [M] cell wall/membrane/envelope biogenesis, and [V] defense mechanisms, the accessory and unique genes accounted for a greater proportion than the core genes (Fig. 4a). The KEGG analysis revealed that genes associated with “metabolism” accounted for the largest proportion of the core, accessory, and unique genomes. Among these genes, most were related to “carbohydrate metabolism”, “membrane transport”, “overview”, “amino acid metabolism”, “replication and repair”, and “nucleotide metabolism” (Fig. 4b).

In our new strain *Streptococcus* sp. v1. nov., 86 genes were unique. KEGG pathway annotations of these unique genes indicated that most KEGG Orthologies (KOs) were associated with the ABC transporter system, including ATP-binding cassette, subfamily B, bacteria (K06147); putative ABC transport system permease protein (K02004); putative ABC transport system ATP-binding protein (K02003); raffinose/stachyose/melibiose transport system permease protein (K101190); ABC-2 type transport system ATP-binding protein (K01990); and raffinose/stachyose/melibiose transport system permease protein (K10118) (Table 2). Pathogens acquire essential nutrients from the host by select ABC transporters, which enables rapid adaptation to changing host microenvironments and mediates toxicity (Tanaka et al. 2018)(Vitreschak et al. 2002)(Gutiérrez-Preciado et al. 2015). A total of 9 virulence factors were identified in the *Streptococcus* sp. v1. nov. genome using the virulence factor database (VFDB), including capsule, PavA, hyaluronic acid capsule, CBPs, PfbA, autolysin, and PsaA. Following the functional classification scheme in the VFDB, 71% of the virulence factors are related to “offensive function”, which contributes to successful infection of the host cells and tissues by colonization and

toxicity (Heermann and Fuchs 2008). This result also explains to a certain extent why the patient's condition progressed rapidly.

Table 2
KO annotation of the unique genes of *Streptococcus* sp. v1. nov.

KO	Definition	Score
K07407	alpha-galactosidase [EC:3.2.1.22]	369
K06147	ATP-binding cassette, subfamily B, bacterial	295
K09458	3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179]	294
K00652	8-amino-7-oxononanoate synthase [EC:2.3.1.47]	263
K02004	putative ABC transport system permease protein	198
K02003	putative ABC transport system ATP-binding protein	184
K10119	raffinose/stachyose/melibiose transport system permease protein	168
K01990	ABC-2 type transport system ATP-binding protein	139
K10118	raffinose/stachyose/melibiose transport system permease protein	114

4. Conclusions

The present study describes the microbial community and ARG distribution of the vitreous of an endophthalmitis patient in Wenzhou. This study also uncovered the genomic characteristics and comparative genomics of *Streptococcus* sp. v1. nov. Genomics-based analysis of patient specimens can promote faster and more accurate clinical diagnosis, which can help guide the clinical diagnosis of infectious diseases.

Declarations

Acknowledgments

The authors thank the Key Discipline of Zhejiang Province in Medical Technology (First Class, Category A). The authors also thank Jinyu Wu of the Institute of Genomic Medicine, Wenzhou Medical University for his support and help with data analysis.

Funding

This work was supported by the National Science and Technology Major Project (grant number 2018ZX10201001) and Science and Technology Bureau of Wenzhou (Y20190171). The funding source(s) had no involvement.

Competing interests

The authors declare that they have no competing interests.

Ethics approval

Approval was obtained from the ethics committee of the Eye Hospital of Wenzhou Medical University (approval number KYK [2015]34). The tenets of the Declaration of Helsinki were followed in this study.

Consent to participate

Written informed consent of this subject was obtained before sample collection.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

Y.C. contributed to conception, design, and interpretation of the data. M.Z. and Y.K. contributed to data analysis and drafted the article. Y.X. and F.X. contributed to sample collection and DNA extraction. All authors read and approved the manuscript.

Figures

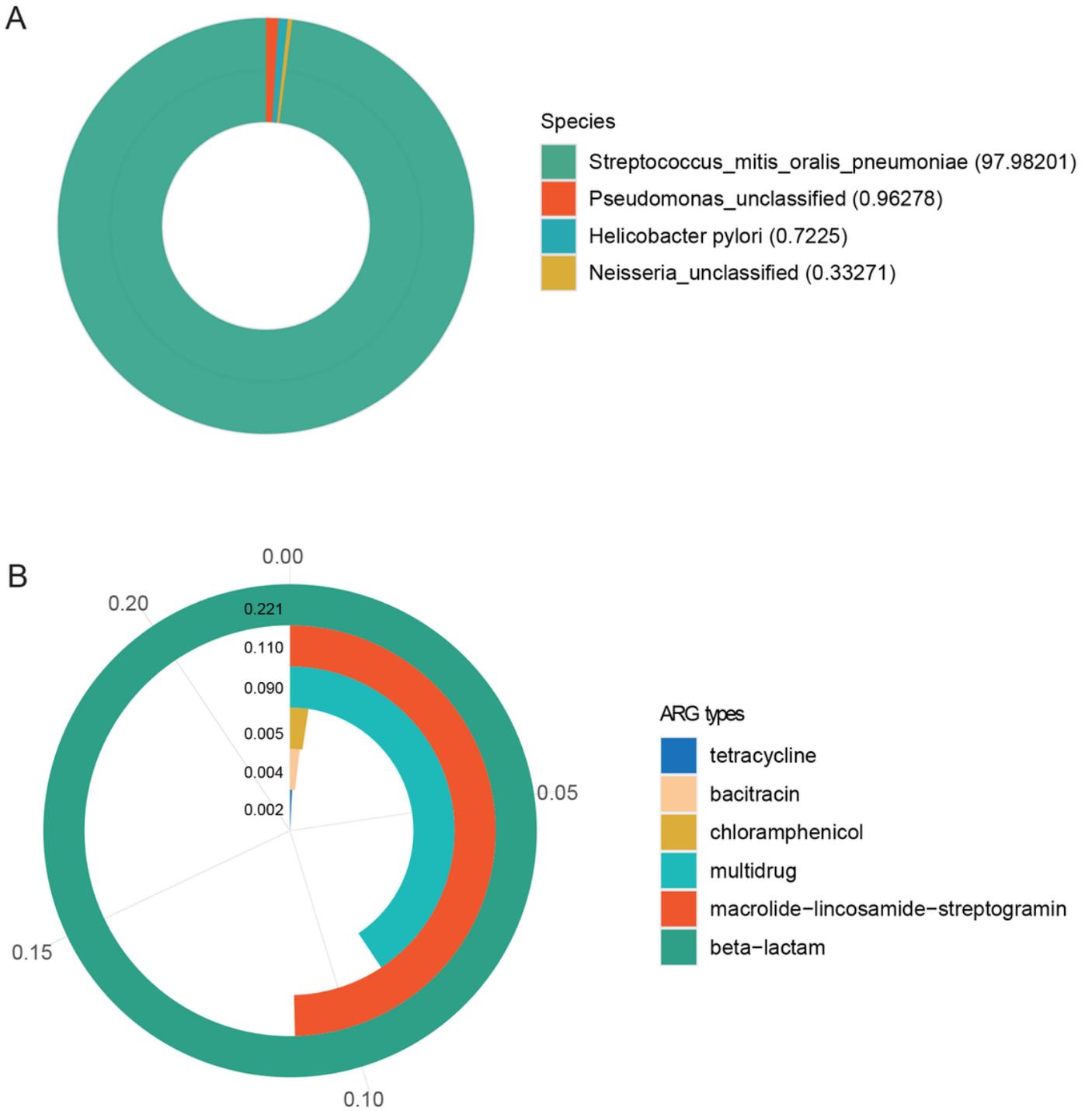


Figure 1

Metagenomic analysis of species classification and antibiotic resistance. a Sunburst plot showing the taxonomic composition at the species level. b Radial bar chart showing the antibiotic resistance gene (ARG) composition.

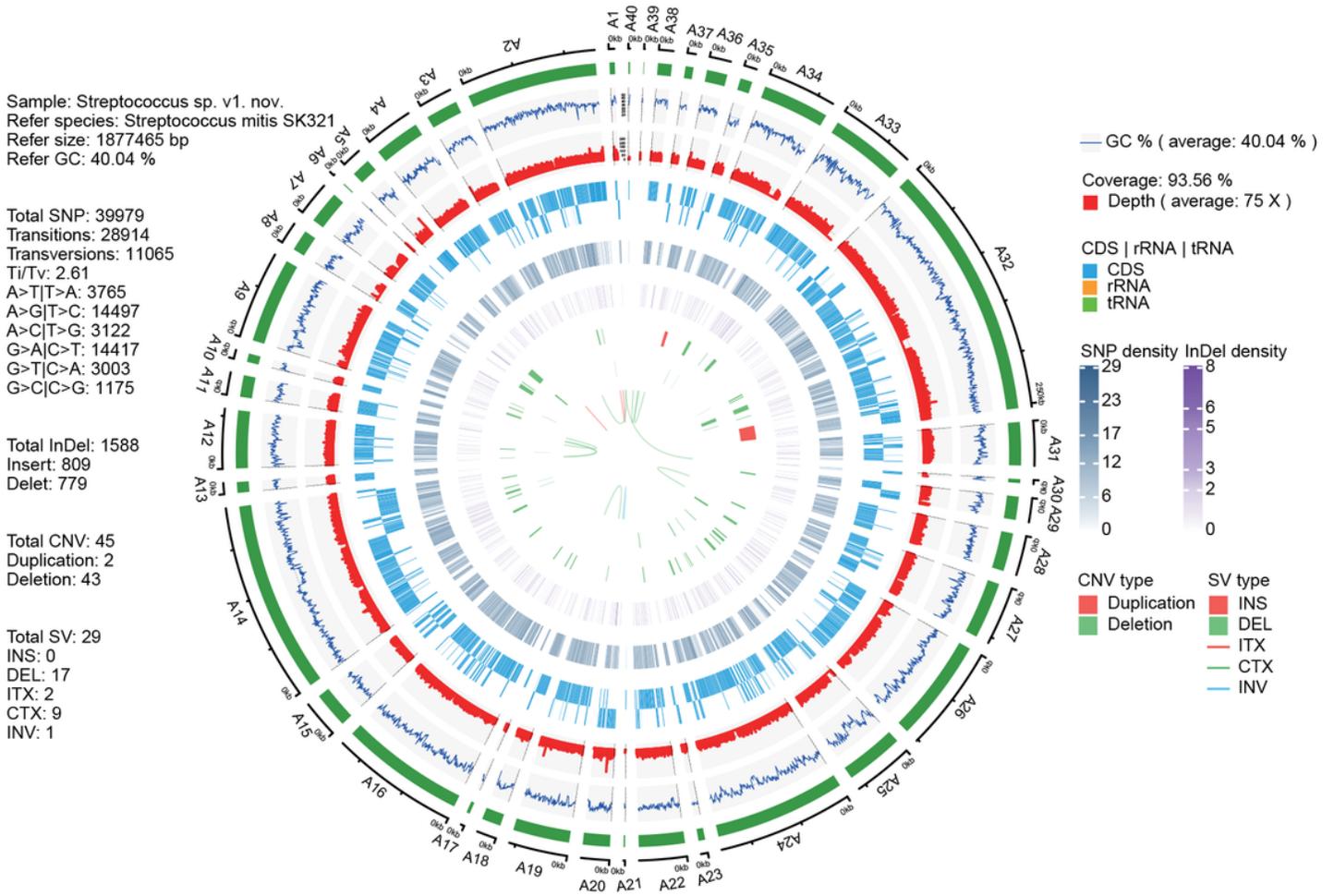


Figure 2

Genome-wide landscape of genetic variation. Circles 1 and 2 (outer to inner) represent sequence information and the GC content curve of the reference genome, respectively. Circle 3 represents sequencing depth and coverage. Circle 4 represents the gene coding region (CDS) and noncoding RNA region (rRNA, tRNA) in the reference genome (the outer circle represents the positive strand, and the inner circle represents the negative strand). Circles 5, 6, 7, and 8 represent the density of single-nucleotide polymorphisms (SNPs), small fragment insertion and deletion (InDel), copy number variation (CNV), and structural variation (SV), respectively.

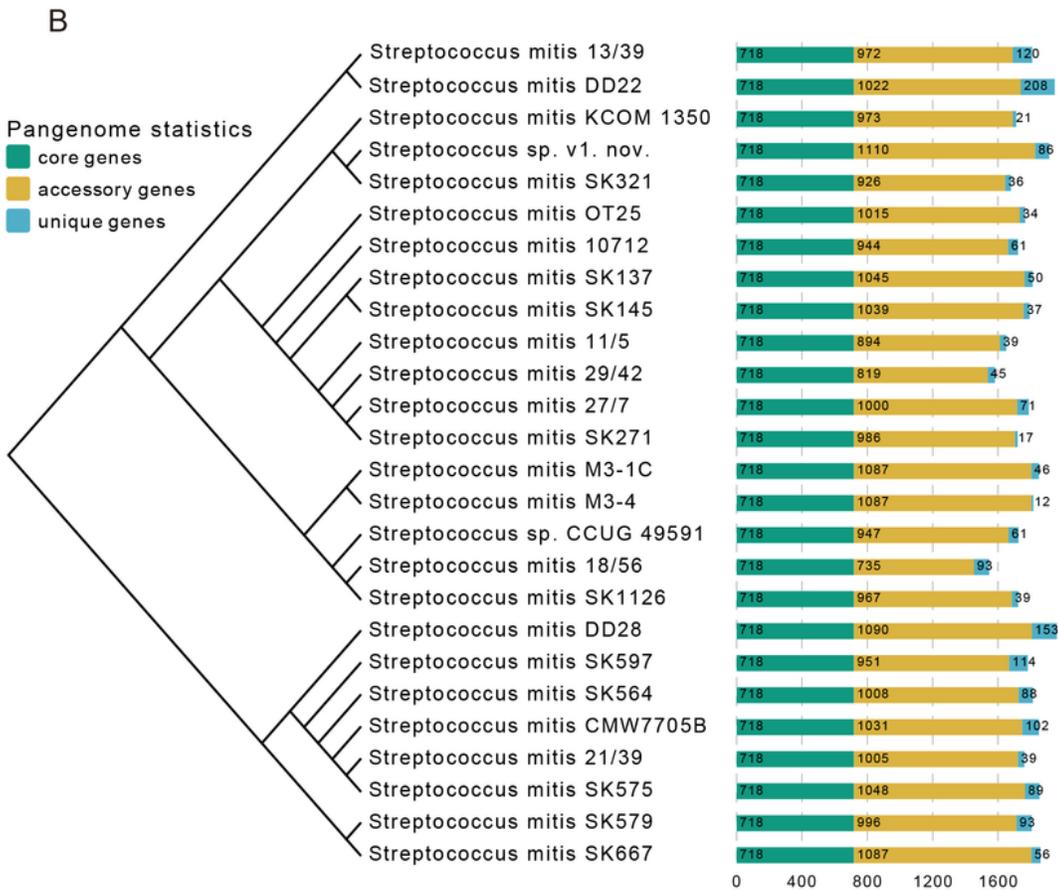
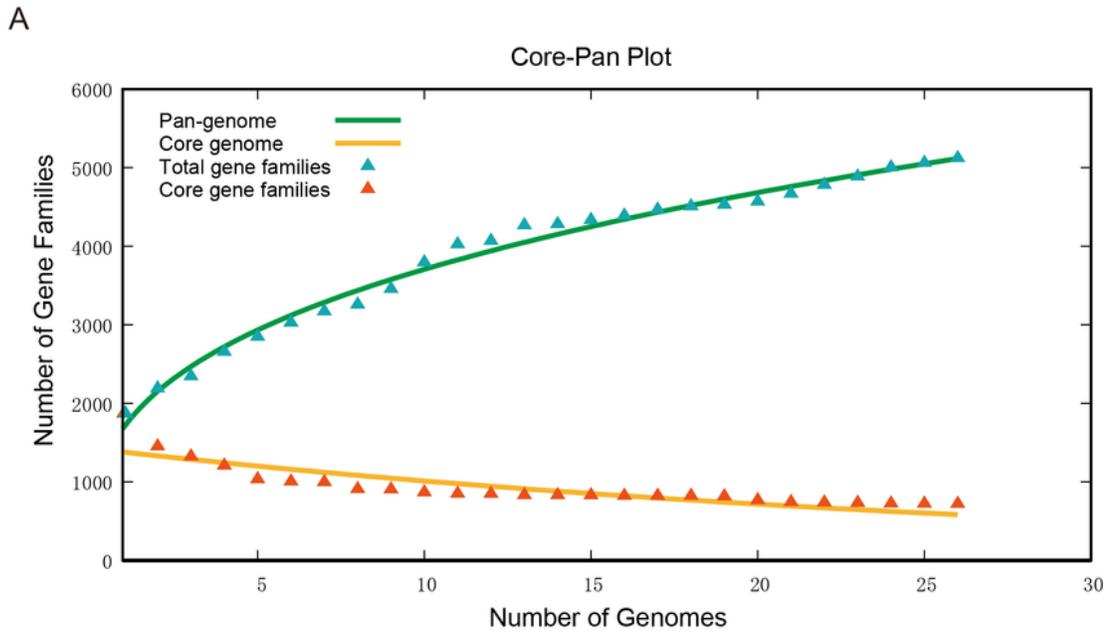


Figure 3

Pan-genome analysis of *Streptococcus* sp. v1. nov. and TCS top 25 hit *Streptococcus* strains. a Core-pan plot of *Streptococcus* sp. v1. nov. and TCS top 25 hit *Streptococcus* strains. The green and yellow lines indicate the numbers of pan-genomes and core genomes, respectively. b A UPGMA clustering tree based on core gene sequences of *Streptococcus* sp. v1. nov. and TCS top 25 hit *Streptococcus* strains

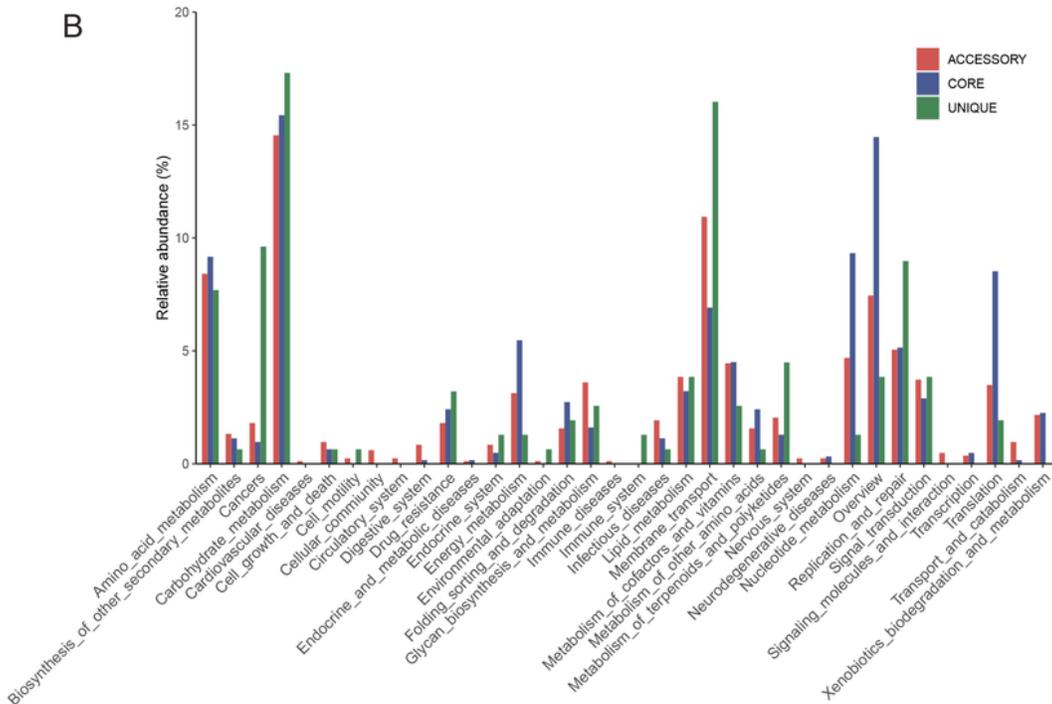
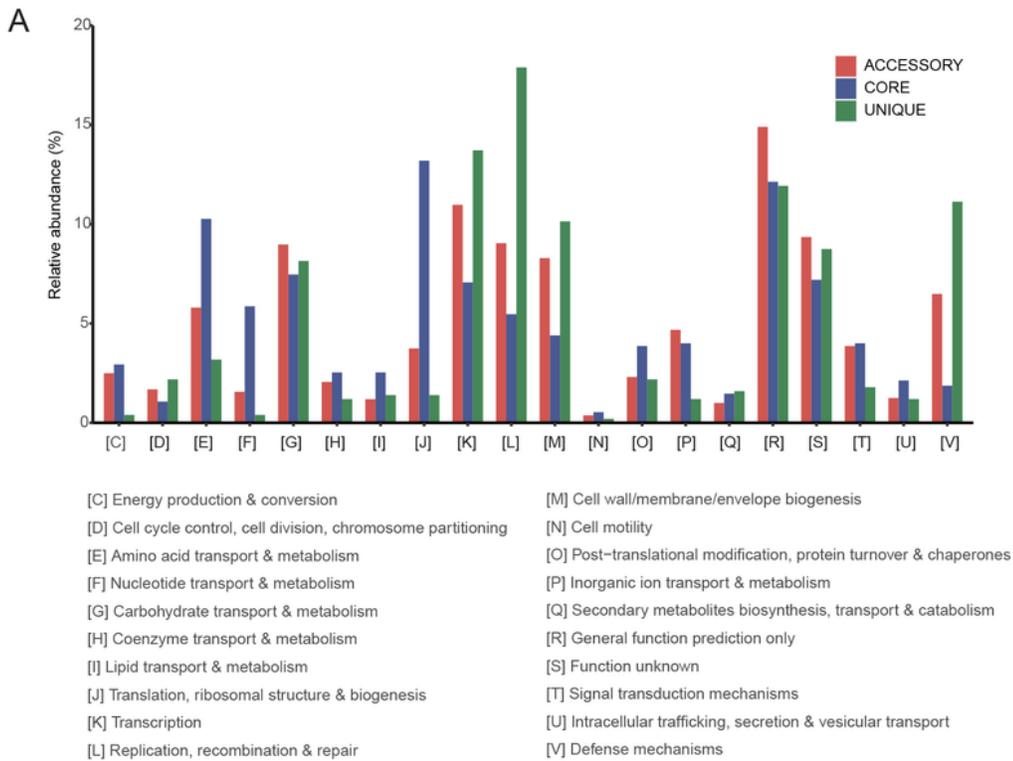


Figure 4

Detailed COG and KEGG pathway classification in the core, accessory, and unique genomes. a COG distribution. b KEGG distribution.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfigure1.pdf](#)
- [Additionalfigure2.pdf](#)
- [Additionaltable1.xls](#)