

# Phylo-taxonomgenomic reconciliations of genus *Streptococcus* reveals the existence of sixteen novel genomospecies

Sanjeet Kumar (✉ [ksanjeet.ibab@gmail.com](mailto:ksanjeet.ibab@gmail.com))

Gangadhar Meher University <https://orcid.org/0000-0001-5143-2557>

**Kanika Bansal**

CSIR- Institute of Microbial Technology

**Santosh Kumar Sethi**

Gangadhar Meher University

---

## Research Article

**Keywords:** Streptococcus, phylogeny, taxonomgenomic, species group, genomospecies

**Posted Date:** April 12th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1416559/v1>

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

[Read Full License](#)

---

# Abstract

Over two centuries, genus *Streptococcus* has been extensively explored in clinical and industrial settings. However, its taxonomy is based on low-resolution methods such as biochemical, housekeeping genes, or conserved sequence identity of species groups resulting in various amendments. In the current study, we have used genome similarities and phylogeny to provide a robust taxonomic framework of 115 type strains and 23 unclassified species. Phylogeny based on 16S rRNA suggests no integrity of the earlier classified species groups, which cannot be used as a taxonomic classification. Whole genome phylogeny counter supports the non-existence of the earlier defined species groups. Genome similarity assessment suggest the existence of sixteen novel genomospecies including some of the amendments in the earlier classified species and subspecies. Highly stringent antibiotic resistance genes search resulted in widespread of fluoroquinolones resistance. Interestingly, no well-curated antibiotic resistance genes were not found in 42 genomes, including strains of probiotic importance like *S. thermophilus* NCTC 12958 (T) and *S. salivarius* NCTC 8618 (T), etc. However, presence of partial hits for resistance genes in such strains requires further experimental validation before their downstream application.

## Introduction

*Streptococcus* was first observed in 1868 by the German surgeon Theodor Billroth as chain-forming cocci from wounds (Wilson 1987). In 1884, *Streptococcus* was first used in a generic sense by Rosenbach for *Streptococcus pyogenes* which was isolated from suppurative lesions (Rosenbach 1884). The genus *Streptococcus* belongs to phylum *Firmicutes*, class *Bacilli*, order *Lactobacillales*, and family *Streptococcaceae* (Whiley and Hardie 2015). It constitutes a broad range of Gram-positive, facultatively anaerobic, spherical, or ovoid cells arranged in chains or pairs. The nomenclature of the genus has lacked international consensus (Facklam 1984). Due to which the genus had witnessed major revisions in the past with the implementation of novel techniques to understand its taxonomy. *Streptococcus* is misspelled with more than 64 different names similar to its true taxonomic name (<https://lpsn.dsmz.de/genus/streptococcus>). Genus *Streptococcus* was divided into three separate genera *Streptococcus (sensu stricto)*, *Enterococcus*, and *Lactococcus* (Schleifer and Kilpper-Bälz 1984; Schleifer and Kilpper-Bälz 1987). Early taxonomy of the genus was based on biochemical, physiological, and serological characterization (Jones 1978). Lancefield serological grouping scheme (1933) is based on cell-wall polysaccharides where strains are designated as A, B, C, E, F, G, etc. (Lancefield 1933). This is extremely useful for identifying  $\beta$ -haemolytic streptococci (pyogenic group) from human and animal infections. However, phenotypic characteristics in earlier studies like haemolytic reaction, Lancefield grouping, or lactose fermentation cannot resolve the species of genus *Streptococcus* (Whiley et al. 1990). Since then, with the advancement of chemotaxonomic, whole genome DNA hybridization, and 16S rRNA gene-based taxonomy, the genus has undergone considerable taxonomic revisions (Schleifer and Kilpper-Bälz 1987; Bentley et al. 1991; Kawamura et al. 1995).

According to current standing in nomenclature, the genus *Streptococcus* encompasses 111 species as of October 2021 (Parte et al. 2020) (<https://lpsn.dsmz.de/genus/streptococcus>). The genus *Streptococcus*

is isolated from a wide range of habitats like humans, animals, and plants. Some species such as *S. thermophilus* are beneficial or harmless to humans (Doron and Snyderman 2015), while most streptococcal species are obligate or opportunistic pathogens like *S. pyogenes*, *S. pneumoniae*, *S. equi*, *S. afalactiae* causing pneumonia, meningitis, or endocarditis (Mitchell 2003). On the basis of 16S rRNA gene sequence and reassociation data, the constituent species of the genus *Streptococcus* were divided into seven “species groups” along with several ungrouped species. These seven groups are “Pyogenic”, “Bovis”, “Mutans”, “Mitis”, “Anginosus”, “Salivarius” and “Hyovaginalis” group (Bentley et al. 1991; Kawamura et al. 1995; Devriese et al. 1997; Whiley et al. 1997; Täpp et al. 2003). However, the relationship amongst these species groups is not clear and most of the species of the genus *Streptococcus* remain ungrouped (Richards et al. 2014; Pontigo et al. 2015) (<https://lpsn.dsmz.de/genus/streptococcus>). A plethora of other studies based on single genes or multilocus genes like: 16S rRNA, *rpoB*, *gyrB*, *dnaJ*, *mpB*, *recN*, *sodA*, *groEL*, *tuf* and, *rpnB* have been conducted to understand its taxonomy (Kawamura et al. 1995; Täpp et al. 2003; Itoh et al. 2006; Glazunova et al. 2009; Glazunova et al. 2010). Further, genome-based investigations of the genus *Streptococcus* or some species groups revealed highly dynamic pan-genome (Gao et al. 2014; Richards et al. 2014; Delorme et al. 2015; Jans et al. 2015; Jensen et al. 2016; Salvadori et al. 2019). In 2018, the genus *Streptococcus* was reported to constitute 103 species with only 70 type strain genomes available. A study based on a limited number of species using phylogenomic and comparative genomics based on conserved sequence indels (CSI) of the genus delineated its constituent species in 2 major clades (“Mitis-Suis” and “Pyogenic-Equinus-Mutans” clades). These two major clades were further delineated into 14 other subclades (Patel and Gupta 2018). However, the analysis to redefine the species groups of the genus *Streptococcus* were not described in taxonomic terms. Taxonomic investigations using deep phylo-taxonomogenomics (DEEPT) including genomic resources of all the representatives could assign major reconciliations in the genus *Mycobacterium* and *Xanthomonas* (Kumar et al. 2019; Bansal et al. 2021; Meehan et al. 2021). Such DEEPT investigation is required in the genus *Streptococcus* for a robust taxonomic framework.

Over the past two centuries, genus *Streptococcus* has significantly expanded with 27 novel species in past five years (<https://lpsn.dsmz.de/genus/streptococcus>). Currently, genus *Streptococcus* includes around 170 taxa (including validly published, synonyms, and not validly published species). With this unprecedented level of diversity of the species, its taxonomy becomes cumbersome. Previously attempts to understand the taxonomy of the genus *Streptococcus* were primarily aimed to identify the species groups and not the species. Such investigation was often based on the low resolution or limited sequence information to designate species groups based on certain physicochemical features such as the ability to haemolytic process (Brown 1920) etc. Correct taxonomic assessment can be helpful in the screening and characterization of potential probiotic strains from their pathogenic counterparts. Owing to the importance of strains of *Streptococcus* in dairy, healthcare, and industry, understanding their true taxonomy based on the latest robust methods is a must.

The present classification of the genus is confusing with several species, subspecies, and synonyms, which are in utmost need of a robust taxonomy. The current advances in the next-generation sequencing technologies have enabled us with generation of several genomes of the type or representative strains.

Aim of the current study is to provide a robust phylo-taxono-genomic framework to the genus *Streptococcus* based on the gold standards for the modern taxonomy, i.e., average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) and core genome-based phylogeny (Chun et al. 2018). In addition to the type strains of the species, we have also included 23 strains with high-quality genomes which were not classified at the species level. Such large-scale genomic investigation can be helpful in biosafety assessment of the strains of genus *Streptococcus*. Moreover, the present study also suggests that species groups in the genus cannot be used as an absolute parameter for its taxonomy and phylogeny. Moreover, species groups cannot be used for identification and characterization of any new species of the genus. Presence of several species such as *S. pneumoniae*, *S. pyogenes*, *S. agalactiae* raises concern over the beneficial effect of the strain of genus *Streptococcus* such as *S. thermophilus* and *S. salivarius* for human (Kaci et al. 2014; Spellerberg and Brandt 2015; Alexandraki et al. 2019). Antibiotic resistance profile estimation will be crucial to understand the resistance profile of the constituent species.

## Methods

### Procurement of genomes from the public repository and their quality check

All the available genomes of the type species, undefined strains of genus *Streptococcus*, and representative of the species were procured from the NCBI genome (<https://www.ncbi.nlm.nih.gov/genome/?term=Streptococcus>) and EZBioCloud (<https://www.ezbiocloud.net/taxonomy?tn=Streptococcus&depth=2>). 16S rRNA integrity was assessed using ContEst16S (Lee et al. 2017) (<https://www.ezbiocloud.net/tools/contest16s>). Further, CheckM was performed for quality assessment of the genomes. Genomes showing more than 95% completeness and less than 5% contamination were considered for the analysis in the present study. Altogether, 138 high-quality genomes were used in the study from genus *Streptococcus*. *Lactococcus lactis* ATCC-19435 was used as an outgroup in the study. All the genomes were annotated using prokka v1.12 (Seemann 2014).

### Phylogenomics of the genus *Streptococcus*

16S rRNA gene sequences were obtained for all the type strains (including species and subspecies) of genus *Streptococcus* (<https://lpsn.dsmz.de/genus/streptococcus>). A maximum likelihood (ML) based phylogenetic was constructed using MEGA v7 (Kumar et al. 2016) with a bootstrap replication of 500. This phylogenetic tree was rooted with *Lactococcus lactis* ATCC 19435 (T). To construct an accurate core genome phylogeny of genus *Streptococcus*, PEPPAN was implemented with default settings. PEPPAN implements single representative gene identification and alignment for construction of phylogeny (Zhou et al. 2020). Phylogenetic tree was annotated using iTOL (<https://itol.embl.de/>) for better visualization (Letunic and Bork 2019).

### Taxonogenomic analysis of the genus *Streptococcus*

Several genome similarity assessment tools such as OrthoANI, GGDC fastANI was implemented for robust inference. OrthoANI v1.2 (Yoon et al. 2017) was implemented with a cut-off of 96% for species

delineation. Digital DNA-DNA hybridization (dDDH) was estimated with a webserver of the genome-to-genome distance calculator 2.1 GGDC (<http://ggdc.dsmz.de/ggdc.php>) with a species delineation of 70% (Auch et al. 2010). dDDH was calculated against all the identified genomospecies across all the genome (n = 139). Heatmap for orthoANI, dDDH and was generated using GENE-E (<https://software.broadinstitute.org/GENE-E/>).

### **Resistome estimation using RGI module of CARD**

RGI module (Alcock et al. 2020) (<https://github.com/arpcard/rgi>) of the Comprehensive Antibiotic Resistance Database (CARD, v5.2.0) was implemented for identification of antibiotic resistance gene among all the genomes in the study (n = 138). A heatmap was generated using the heatmap module of the RGI including a perfect and strict hit. Also, to obtain distant homologs of antibiotic resistance genes, a loose algorithm was used with a cut-off of > 30% and > 70% for best identity and length, respectively.

## **Result**

### **Genomic features of strains used in the study**

A total of 138 high-quality genomes were used in the present study with an average genome size ranging from 1.5 mb to 2.8 mb and 33–46% GC. A descriptive table of all the species, representative and undefined species of genus *Streptococcus* with their genomic attributes including numbers of CDS, genome size, GC content, isolation source etc. is summarized in supplementary table 1. Genomes of species such as *S. danieliae*, *S. symci*, *S. devriesei* were removed from the study because of the non-permissible range of contamination and completeness or the absence of 16S rRNA gene in their genome.

### **Phylogenomics resolved intermingled clades and species groups within the genus *Streptococcus***

16S rRNA-based phylogeny of all the type strains, representative and undefined species of genus *Streptococcus* suggested several intermingling phylogenetic positioning of earlier defined species groups (Fig. 1). We could identify several phylogenetic positioning such as, some of the type strains of Mitis-Suis clade such as *S. suis*, *S. azizii*, *S. marmotae*, *S. himanayensis*, *S. cuniculi*, *S. varani*, *S. minor*, *S. ovis* and *S. merionis* shares phylogeny with Pyogenes-Equinus-Mutans clades. Likewise, the *S. downii* of the Pyogenes-Equinus-Mutans clade is within the Mitis-Suis clade. These evidences clearly indicate that 16S rRNA-based phylogeny could not provide sufficient evidence for supporting the species groups among genus *Streptococcus*.

Moreover, the current phylogeny of the genus *Streptococcus* is based on MLST or CSI, including a maximum of 70 species. In the present analysis, inclusion of 115 type strains along with 23 unclassified strains at whole genome level provides us robust phylogenetic grouping of the genus *Streptococcus* (Fig. 2). We obtained two phylogroups in accordance with the earlier clades of genus *Streptococcus* i.e., Mitis-Suis and Pyogenes-Equinus-Mutans. However, there were some exceptions such as: *S. acidominimus*, *S. porci* and *S. plurextorum* belong to Mitis-Suis clade and *S. pharyngis* belong to Pyogenes-Equinus-Mutans clade contrary to their previous phylogeny. Additionally, we could also designate 23 unclassified strains

of genus *Streptococcus* to their respective clades (14 in Mitis-Suis clade and 9 in Pyogenes-Equinus-Mutans clade).

Overall, several ambiguities in subclade level grouping could not be resolved by 16S rRNA or whole genome-based phylogeny. For instance, members of subclade Halotolerans (*S. acidominimus*, *S. hyovaginalis*, *S. pluranimalium*, *S. halotolerans* and *S. thoralensis*) and subclade Entericus (*S. entericus*, *S. marimammalium*) of Pyogenes-Equinus-Mutans clade had intermingled positions at both the levels of phylogeny. Interestingly, out of all these seven members of Pyogenes-Equinus-Mutans clade, *S. acidominimus* now belongs to Mitis-Suis clade. To further examine these ambiguities in the genus *Streptococcus*, we performed genome similarity assessment.

### **Taxonomgenomics suggest the presence of several genomospecies and subspecies reclassification**

The current taxonomy of the genus *Streptococcus* is majorly based on biochemical and physiological methods. In the study, implementation of several genome similarity assessment methods helped in identification of several novel genomospecies and in truly demarcating the species and subspecies of genus *Streptococcus* (Fig. 3, Data sheet 1). Genomotaxonomy revealed twelve novel genomospecies amongst the unclassified species: *S. sp* HKU75 (GS1), *S. sp*. zq-86 (GS2), *S. sp*. KS 6 (GS3), *S. sp*. LPB0220 (GS4), *S. sp* A12 (GS5), *S. sp* oral taxon 431 F0610 5-114 (GS6), *S. sp* oral taxon 061 F0704 (GS7), *S. sp*. 116-D4 (GS8), *S. sp* oral taxon 064 W10853 (GS9), *S. sp* NPS 308 (GS11) and *S. sp*. 1643 (GS12) and *S. sp* CNU G3, *S. sp* CNU G2, *S. sp* CNU 77 – 61 (GS15). Further, three of the subspecies of *S. oralis* (*S. oralis* subsp *oralis* NCTC 11427 (T), *S. oralis* subsp *dentisani* CECT 7747 (T) and *S. oralis* subsp *tigurinus* AZ 3a (T)) and *S. equi* subsp *ruminatorum* CECT 5772 (T) were found to be novel species and designated as GS10, GS13, GS14, and GS16.

Further, the species status of some of the species of genus *Streptococcus* needs to be investigated in details like: *S. ilei* I-G2 (T) and *S. koreensis* JS71 (T) (ANI: 96.32; dDDH: 68.8); *S. ratti* FA-1 (T) and *S. ursoris* DSM 22768 (T) (ANI: 98.6; dDDH: 88) and *S. bovis* ATCC-33317 (T) with *S. equinus* NCTC 12969 (T) (ANI: 96.89; dDDH: 72.7). In the present study, we could also assign nine unclassified species to already known species of the genus *Streptococcus*. Like, *S. sp* DAT741 to *S. ruminantium* GUT-187 (T) (ANI:99; dDDH: 90.4), *S. sp*. NSJ-72 to *S. constellatus* subsp *pharyngis* CCUG-46377 (T) (ANI:97.73; dDDH: 78.3), *S. sp* ZB199 to *S. parasanguinis* ATCC 15912 (T) (ANI: 96.78; dDDH: 70.8), *S. sp* FDAARGOS 192 to *S. salivarius* NCTC 8618 (T) (ANI: 96.08; dDDH: 66.5); *S. sp* NCTC 11567 to *S. dysgalactiae* subsp *dysgalactiae* NCTC 13731 (T) (ANI:96.02; dDDH: 67.4); *S. sp* FDAARGOS 522, *S. sp* group B FDAARGOS 229, *S. sp* FDAARGOS 521, *S. sp* FDAARGOS 520 belong to *S. agalactiae* NCTC 8181 (T) (ANI value in the range of 98.7% – 99.8% and dDDH value in the range of 89.6% – 98.6%).

In the present study, we have obtained 16 novel genomospecies among the genus *Streptococcus*. These genomospecies were classified earlier as either subspecies of already reported species or as unclassified species. Further, merger of a few already defined species such as *S. ilei* and *S. koreensis*, *S. bovis*, *S. equinus* and *S. equinus*, *S. ratti* and *S. ursoris*.

### **Genus-wide resistome analysis**

Infection caused by several species of genus *Streptococcus*, such as *S. pneumoniae*, *S. pyogenes* etc. are the major cause of community-acquired respiratory infections worldwide (Appelbaum 2002). Antimicrobial resistance in the bacterial population raises concern over the application of microbes in industry and agriculture. Resistome analysis of the genus *Streptococcus* resulted in several classes of drug resistance genes using high stringent cut-offs (see methods) (Fig. 2). Resistome gene profile included: AAC(6')-Ie-APH(2'')-Ia, ANT(6)-Ia, APH(3')-IIIa and *aad(6)* for aminoglycosides; *dfrF* for diaminopyrimidines; *patA*, *patB* and *pmrA* for fluoroquinolones; *ErmB*, *RlmA(II)*, *cfr(D)*, *ImrP*, *InuA*, *InuB*, *InuC*, *IsaC*, *IsaE* and *mel* for lincosamides; *ErmB*, *RlmA(II)*, *ImrP*, *IsaC*, *IsaE*, *mel* for macrolides; SAT-4 for nucleosides; *cfr(D)*, *IsaC*, *IsaE* and *mel* for oxazolidinones; *S. agalactiae mprF* for peptidic antibiotic; *Enterococcus faecalis* chloramphenicol acetyltransferase; *Lactobacillus reuteri cat-TC*, *catQ*, *cfr(D)*, *IsaC*, *IsaE* and *mel* for phenicols; *IsaC*, *IsaE* and *mel* for pleuromutilin antibiotic; *ErmB*, *cfr(D)*, *ImrP*, *IsaC*, *IsaE* and *mel* for streptogramin antibiotic and *ImrP*, *IsaC*, *IsaE*, *mel*, *tet(L)*, *tet(W/N/W)*, *tetA(46)*, *tetM* and *tetO* for tetracycline antibiotic (Data sheet 2). Most of these resistance genes were present in *S. porci* DSM 23759, *S. gallolyticus subsp pasteurianus* NCTC 13784, *S. hyovaginalis* DSM 12219, *S. parasuis* H35, *S. plurextorum* DSM 22810, *S. orisratti* DSM 15617, *S. pluranimalium* TH11417, *S. sp* FDAARGOS 522 and *S. vaginalis* P1L01 (Fig. 2). Further, 75 out of 138 genomes were harbouring fluoroquinolone resistance genes (*patA* and *patB*). Interestingly, well-curated antibiotic resistance genes were not found in 42 genomes in the present study. These genomes were majorly present in Pyogenes-Equinus-Mutans clade in phylogroup containing strains of probiotic importance like: *S. thermophilus* NCTC 12958 (T) and *S. salivarius* NCTC 8618 (T) etc. Further, Mitis-Suis clade also had 14 genomes not having perfect hits for resistance genes (Fig. 2).

However, low-stringent cut-offs (see methods) resulted in a highly enriched resistome profile for all the species of the genus *Streptococcus* (Data sheet 3). A total of 61,274 antibiotic resistance-related genes could be identified using the standalone RGI module of CARD. Out of which, 15,976 genes were efflux pump related genes. Antibiotic resistance genes related to  $\beta$ -lactam were detected in all the strains in large numbers (up to 178 copies). Aminoglycoside and glycopeptide antibiotics were also present in large numbers among all the strains. Few strains of genus harbour gene related to ethionamide, fluoroquinolone antibiotic; aminoglycoside antibiotic, isoniazid; triclosan, mupirocin, polyamine antibiotic, prothionamide, pyrazinamide, and sulfone antibiotic.

## Conclusion

Present study is an attempt to understand the taxonomy and status of species group previously described based on the whole genome information. We have curated all the type strain information of the species of genus *Streptococcus* to perform a systematic phylo-taxonomogenomic investigation. We found several taxonomic ambiguities and existence of genomospecies within genus *Streptococcus*. Such a genome wide investigation will be very important in understanding the nature of constituent species including clinical, commensal and industrial important strains. Resistome profile for species of probiotic or industrial importance like *S. thermophilus* NCTC 12958 (T) or *S. salivarius* NCTC 8618 (T) etc was very diverse or partially matched. These antibiotic resistance gene hits at low stringent cutoffs needs to be

further experimentally verified. This will enable the research and industrial community to truly demarcate strains for their downstream applications.

## Abbreviations

CSI: conserved sequence indels; ANI: Average Nucleotide Identity; dDDH: digital DNA-DNA hybridization; ML: maximum likelihood; CARD: Comprehensive Antibiotic Resistance Database; RGI: Resistance gene identifier; GS: genomospecies

## Declarations

### Funding Information

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

### Conflict of Interest

All the authors declare no conflict of interest.

### Author Contributions

SK and KB have performed all the investigation and drafted the manuscript. SK, KB, SKS and SK have conceived the manuscript.

### Ethics approval

This is an observational study which does not require any ethical approval.

### Acknowledgement

All the authors convey their regards for support from Ms. Anu Singh, research scholar at CSIR-Institute of Microbial Technology, Chandigarh for her valuable discussion and support. We also convey a special thanks to Dr. Prabhu Patil for their immense support and guidance through the whole study.

### Data Availability

All the data sheets generated in the manuscript has been provided with the figshare link: ([https://figshare.com/articles/dataset/Data\\_sheets\\_for\\_the\\_study/19312352](https://figshare.com/articles/dataset/Data_sheets_for_the_study/19312352)). Description of each data sheets can be found as fellow:

**Data sheet 1:** Digital DNA-DNA hybridization (dDDH) values of the strains with taxonomic ambiguities in the genus *Streptococcus*.

**Data sheet 2:** Antibiotic resistance drug classes identified by RGI module of CARD with high stringent cutoffs.

**Data sheet 3:** All the identified antibiotic resistance gene with identity and coverage cutoff greater than 30% and 70% respectively.

## References

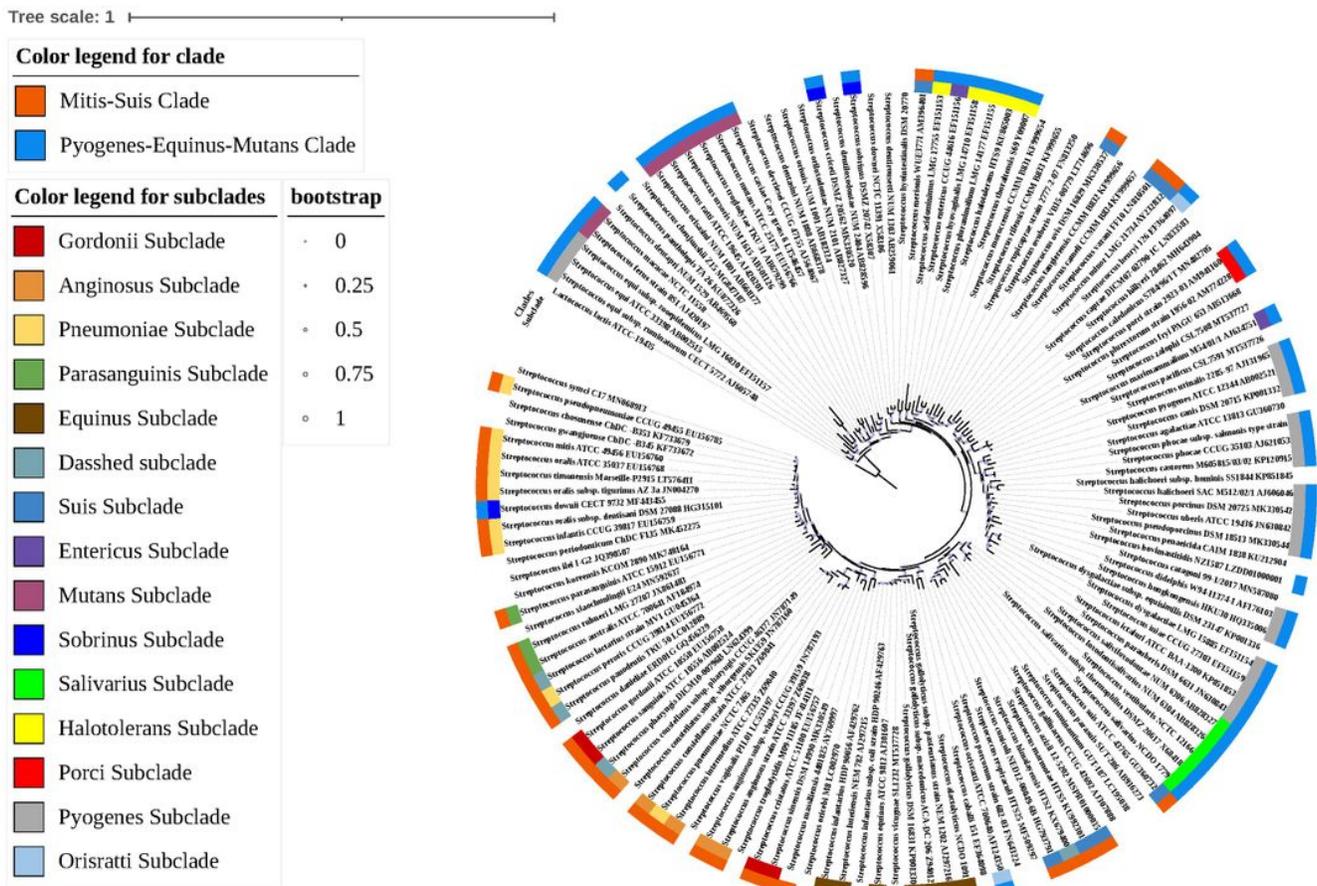
1. Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research* **48**: D517-D525.
2. Alexandraki V, Kazou M, Blom J, Pot B, Papadimitriou K, Tsakalidou E. 2019. Comparative genomics of *Streptococcus thermophilus* support important traits concerning the evolution, biology and technological properties of the species. *Frontiers in microbiology*: 2916.
3. Appelbaum PC. 2002. Resistance among *Streptococcus pneumoniae*: implications for drug selection. *Clinical infectious diseases* **34**: 1613-1620.
4. Auch AF, Klenk H-P, Göker M. 2010. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Standards in genomic sciences* **2**: 142-148.
5. Bansal K, Kumar S, Kaur A, Singh A, Patil PB. 2021. Deep phylo-taxono genomics reveals *Xylella* as a variant lineage of plant associated *Xanthomonas* and supports their taxonomic reunification along with *Stenotrophomonas* and *Pseudoxanthomonas*. *Genomics* **113**: 3989-4003.
6. Bentley RW, Leigh JA, Collins MD. 1991. Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. *International Journal of Systematic and Evolutionary Microbiology* **41**: 487-494.
7. Brown JH. 1920. The cultural differentiation of beta hemolytic streptococci of human and bovine origin. *The Journal of Experimental Medicine* **31**: 35.
8. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu X-W, De Meyer S. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *International journal of systematic and evolutionary microbiology* **68**: 461-466.
9. Delorme C, Abraham A-L, Renault P, Guédon E. 2015. Genomics of *Streptococcus salivarius*, a major human commensal. *Infection, Genetics and Evolution* **33**: 381-392.
10. Devriese L, Pot B, Vandamme P, Kersters K, Collins M, Alvarez N, Haesebrouck F, Hommez J. 1997. *Streptococcus hyovaginalis* sp. nov. and *Streptococcus thoralensis* sp. nov., from the genital tract of sows. *International Journal of Systematic and Evolutionary Microbiology* **47**: 1073-1077.
11. Doron S, Snyderman DR. 2015. Risk and safety of probiotics. *Clinical Infectious Diseases* **60**: S129-S134.
12. Facklam R. 1984. The major differences in the American and British *Streptococcus* taxonomy schemes with special reference to *Streptococcus milleri*. Vol 3, pp. 91-93. Springer.

13. Gao X-Y, Zhi X-Y, Li H-W, Klenk H-P, Li W-J. 2014. Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. *PLoS one* **9**: e101229.
14. Glazunova OO, Raoult D, Roux V. 2009. Partial sequence comparison of the *rpoB*, *sodA*, *groEL* and *gyrB* genes within the genus *Streptococcus*. *International journal of systematic and evolutionary microbiology* **59**: 2317-2322.
15. Glazunova OO, Raoult D, Roux V. 2010. Partial *recN* gene sequencing: a new tool for identification and phylogeny within the genus *Streptococcus*. *International journal of systematic and evolutionary microbiology* **60**: 2140-2148.
16. Itoh Y, Kawamura Y, Kasai H, Shah MM, Nhung PH, Yamada M, Sun X, Koyana T, Hayashi M, Ohkusu K. 2006. *dnaJ* and *gyrB* gene sequence relationship among species and strains of genus *Streptococcus*. *Systematic and applied microbiology* **29**: 368-374.
17. Jans C, Meile L, Lacroix C, Stevens MJ. 2015. Genomics, evolution, and molecular epidemiology of the *Streptococcus bovis*/*Streptococcus equinus* complex (SBSEC). *Infection, Genetics and Evolution* **33**: 419-436.
18. Jensen A, Scholz CF, Kilian M. 2016. Re-evaluation of the *Mitis* group of the genus *Streptococcus* based on whole genome phylogenetic analyses, and proposed reclassification of *Streptococcus dentisani* as *Streptococcus oralis* subsp. *dentisani* comb. nov., *Streptococcus tigurinus* as *Streptococcus oralis* subsp. *tigurinus* comb. nov., and *Streptococcus oligofermentans* as a later synonym of *Streptococcus cristatus*. *International journal of systematic and evolutionary microbiology* **66**: 4803-4820.
19. Jones D. 1978. Composition and differentiation of the genus *Streptococcus*. In *Society for Applied Bacteriology Symposium Series*, Vol 7, pp. 1-49.
20. Kaci G, Goudercourt D, Dennin V, Pot B, Doré J, Ehrlich SD, Renault P, Blottière HM, Daniel C, Delorme C. 2014. Anti-inflammatory properties of *Streptococcus salivarius*, a commensal bacterium of the oral cavity and digestive tract. *Applied and environmental microbiology* **80**: 928-934.
21. Kawamura Y, Hou X-G, Sultana F, Miura H, Ezaki T. 1995. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *International Journal of Systematic and Evolutionary Microbiology* **45**: 406-408.
22. Kumar S, Bansal K, Patil PP, Patil PB. 2019. Phylogenomics insights into order and families of *Lysobacterales*. *Access microbiology* **1**.
23. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* **33**: 1870-1874.
24. Lancefield RC. 1933. A serological differentiation of human and other groups of hemolytic streptococci. *The Journal of experimental medicine* **57**: 571-595.
25. Lee I, Chalita M, Ha S-M, Na S-I, Yoon S-H, Chun J. 2017. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. *International journal of systematic and evolutionary microbiology* **67**: 2053-2057.

26. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* **47**: W256-W259.
27. Meehan CJ, Barco RA, Loh Y-HE, Cogneau S, Rigouts L. 2021. Reconstituting the genus *Mycobacterium*. *International Journal of Systematic and Evolutionary Microbiology* **71**.
28. Mitchell TJ. 2003. The pathogenesis of streptococcal infections: from tooth decay to meningitis. *Nature Reviews Microbiology* **1**: 219-230.
29. Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M. 2020. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology* **70**: 5607.
30. Patel S, Gupta RS. 2018. Robust demarcation of fourteen different species groups within the genus *Streptococcus* based on genome-based phylogenies and molecular signatures. *Infection, Genetics and Evolution* **66**: 130-151.
31. Pontigo F, Moraga M, Flores S. 2015. Molecular phylogeny and a taxonomic proposal for the genus *Streptococcus*. *Genet Mol Res* **14**: 10905-10918.
32. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. 2014. Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome biology and evolution* **6**: 741-753.
33. Rosenbach AJF. 1884. *Mikro-organismen bei den Wund-infections-krankheiten des Menschen*. JF Bergmann.
34. Salvadori G, Junges R, Morrison DA, Petersen FC. 2019. Competence in *Streptococcus pneumoniae* and close commensal relatives: mechanisms and implications. *Frontiers in cellular and infection microbiology* **9**: 94.
35. Schleifer K, Kilpper-Bälz R. 1987. Molecular and chemotaxonomic approaches to the classification of streptococci, enterococci and lactococci: a review. *Systematic and Applied Microbiology* **10**: 1-19.
36. Schleifer KH, Kilpper-Bälz R. 1984. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* **34**: 31-34.
37. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.
38. Spellerberg B, Brandt C. 2015. *Streptococcus*. *Manual of clinical microbiology*. 383-402.
39. Täpp J, Thollesson M, Herrmann B. 2003. Phylogenetic relationships and genotyping of the genus *Streptococcus* by sequence determination of the RNase P RNA gene, *rnpB*. *International journal of systematic and evolutionary microbiology* **53**: 1861-1871.
40. Whiley R, Fraser H, Hardie J, Beighton D. 1990. Phenotypic differentiation of *Streptococcus intermedius*, *Streptococcus constellatus*, and *Streptococcus anginosus* strains within the "*Streptococcus milleri* group". *Journal of clinical microbiology* **28**: 1497-1501.
41. Whiley R, Hall L, Hardie J, Beighton D. 1997. Genotypic and phenotypic diversity within *Streptococcus anginosus*. *International Journal of Systematic and Evolutionary Microbiology* **47**: 645-650.

42. Whiley R, Hardie J. 2015. Streptococcus. Bergey's Manual of Systematics of Archaea and Bacteria. Chichester, UK: John Wiley & Sons, Ltd.
43. Wilson LG. 1987. The early recognition of streptococci as causes of disease. *Medical history* **31**: 403-414.
44. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**: 1281-1286.
45. Zhou Z, Charlesworth J, Achtman M. 2020. Accurate reconstruction of bacterial pan-and core genomes with PEPPAN. *Genome research* **30**: 1667-1679.

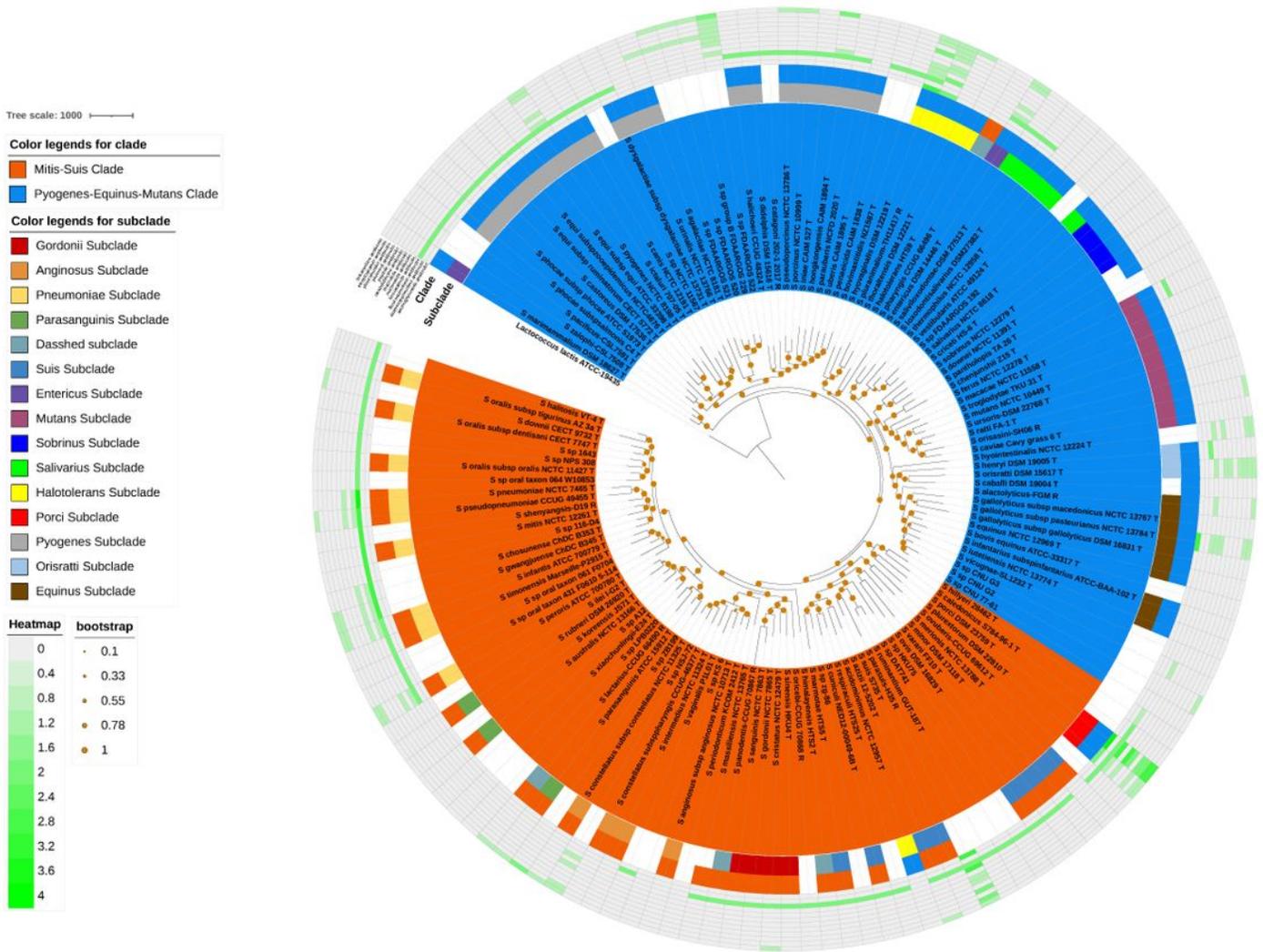
## Figures



**Figure 1**

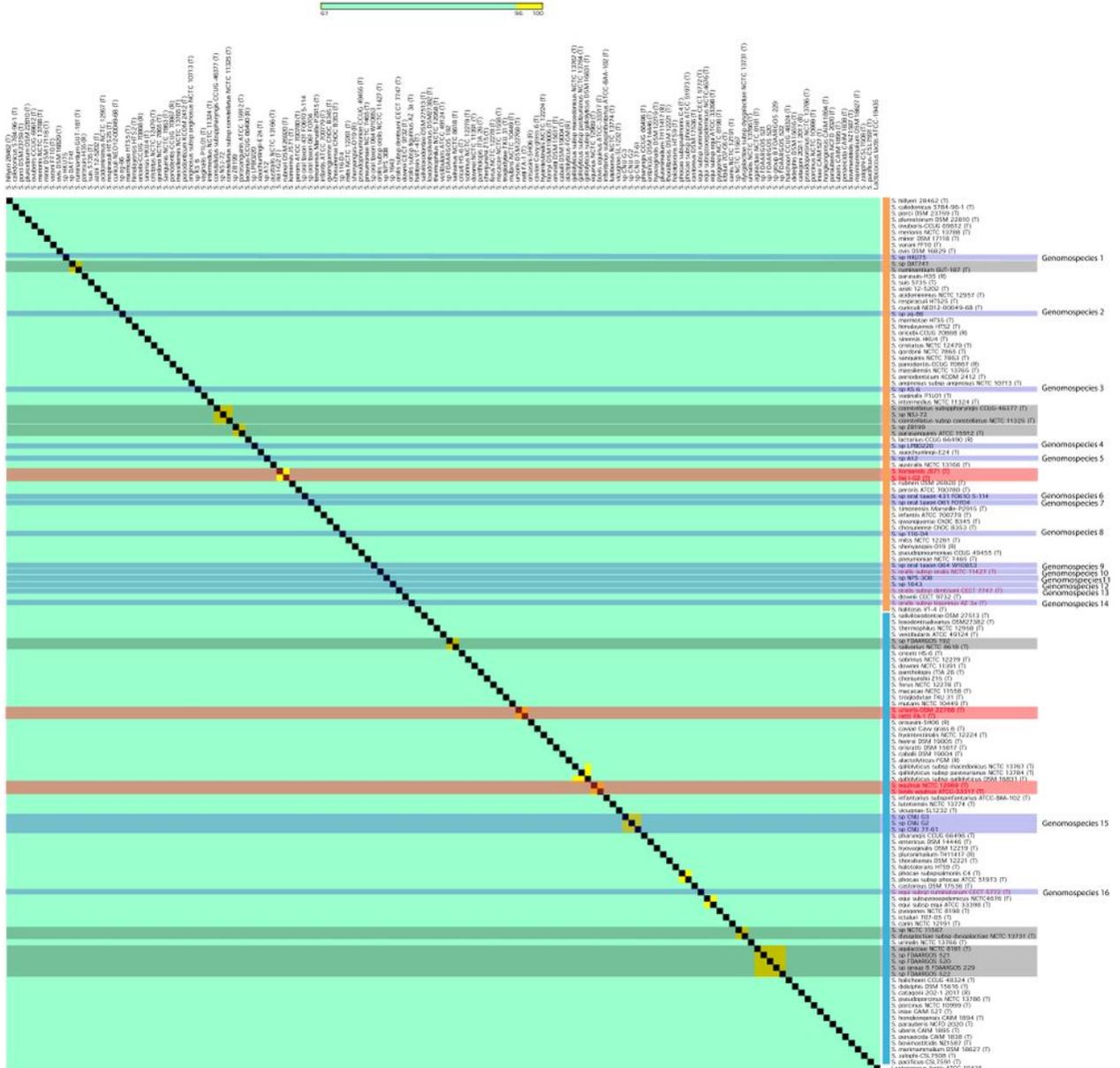
ML based phylogenetic tree construction of all the 16S rRNA from the type strains of genus available in LPSN. The phylogenetic tree was constructed with a bootstrap replication of 500. *Lactococcus lactis* ATCC-19435 was used as an outgroup. All the species classified in several clades and subclades are marked

on the circumference of the phylogenetic tree. Bootstrap values of all the earlier classified species group are represented by spherical dots on the branch.



**Figure 2**

Whole genome based phylogenetic tree constructed using PEPPAN (Phylogeny enhanced pipeline for pangenome). *Lactococcus lactis* ATCC-19435 was used as an outgroup. All the species classified in several clades and subclades are marked on the circumference of the phylogenetic tree. Bootstrap values of all the earlier classified species group are represented by spherical dots on the branch. Clades and subclades earlier defined are marked with the color strips at circumference from inside to outside. Mitis-Suis and Pyogenes-Equines-Mutans clades identified in the present study are marked by orange and blue color ranges respectively in the phylogenetic tree. Antibiotic resistance drug classes identified by RGI module of CARD is represented as heatmap on the outer circumference. Type strains are suffixed with (T) and representative strains marked by (R).



**Figure 3**

Heatmap of average nucleotide identity of the genus *Streptococcus*. Here, species identified with taxonomic ambiguities are highlighted as red box or red font and novel genomospecies are highlighted as purple boxes. Type strains are designated as (T). Unclassified *Streptococcus* sp. could be designated to the species highlighted by grey box.

## Supplementary Files

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

- [supplementarytable1.docx](#)