CODEN [USA]: IAJPBB ISSN : 2349-7750

INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES

SJIF Impact Factor: 7.187

https://doi.org/10.5281/zenodo.14251363

https://www.iajps.com/volumes/volume11-november-2024/73-issue-11-november-24/

Available online at: http://www.iajps.com *Research Article*

CHARACTERIZATION AND PAN-GENOME ANALYSIS OF *STREPTOCOCCUS SALIVARIUS*

Anas Tariq[1] , Maged Al mezgagi[1] , Irum Saleem[2] , Wang Qianqian[1] , Xing Jiangwa*[1]

[1]Research Center of Basic Science, Medical College, Qinghai University, Xining 810016, anastariq@hotmail.co.uk

[1]Research Center of Basic Science, Medical College, Qinghai University, Xining 810016 1902244017@qq.com

[2]Physics Department of zakariya University, Multan , Pakistan imschims@gmail.com

[1]Research Center of Basic Science, Medical College, Qinghai University, Xining 810016 1498619249@qq.com

[1]Research Center of Basic Science, Medical College, Qinghai University, Xining 810016 xingjiangwa@qhu.edu.cn

Abstract:

Background: *Streptococcus salivarius plays a pivotal role in maintaining oral and upper respiratory health. It serves as a natural barrier against pathogens and contributes to the equilibrium of the oral microbiome. Their pro-biotic attributes are garnering increasing recognition for their potential to support both oral and respiratory health.* ***Aim:*** *To perform comprehensive analysis of Streptococcus Salivarius genome data, investigating its secondary metabolites, antibiotic resistance genes, and virulence gene variants.* ***Materials and Methods:*** *The genomic sequences of 20 S. Salivarius strains were obtained from the NCBI database. Genome analysis tools, secondary metabolite detection software, and databases for antibiotic resistance and virulence genes were used to examine genomic information.* ***Result:*** *This analysis elucidated the distinctive genomic attributes of FDAARGOS_1045, encompassing its varied secondary metabolism gene clusters, singular phylogenetic placement, and substantial number of unique genes. The distinctive characteristics exhibited by this strain suggest the potential presence of novel bio synthetic pathways and potentially valuable metabolites. These findings merit further investigation to explore potential biotechnological applications and to enhance our understanding of bacterial diversity within this species.* ***Conclusion:*** *It demonstrates the relationship between pan-genome, core genome, and gene count. This research enhances our understanding of S. Salivarius and provides a theoretical foundation for its medical applications. Further investigation of FDAARGOS_1045 is necessary to gain a more comprehensive understanding of S. Salivarius.*

Keywords: *Streptococcus salivarius, pan-genome, antimicrobial resistance genes, virulence genes, genome annotation, pro-biotic.*

Corresponding author:

Xing Jiangwa,

Research Center of Basic Science,
Medical College, Qinghai University, Xining 810016
xingjiangwa@qhu.edu.cn

QR code

Please cite this article in press ***Xing Jiangwa*** *et al.,* ***Characterization And Pan-Genome Analysis Of Streptococcus Salivarius..,*** *Indo Am. J. P. Sci, 2024; 11 (11).*

INRTODUCTION:

Streptococcus salivarius is a bacterium prevalent in the oral cavity of healthy individuals shortly after birth. It is a gram-positive bacterium and *S. salivarius* is spherical, non-motile, non-sporing, and catalyze-negative. *S. salivarius* is also present in the stomach and jejunum, and plays a significant role in oral and digestive tract ecology [5]. Research has demonstrated that *S. salivarius* can secrete antimicrobial peptides that can be utilized as probiotic to enhance oral health [1]. *S. salivarius K12* is the first strain employed commercially to develop an oral pro-biotic that aids in improving oral health and reducing halitosis [2].

Concurrently, *S. salivarius* is an opportunistic pathogen capable of infecting individuals with a compromised immune system. *S. salivarius* infections are infrequent under normal circumstances, but are more prevalent in institutionalized individuals, including those receiving care in hospitals or long-term care facilities, patients with compromised immune systems, and those with periodontal disease or poor dental hygiene. Biofilm formation by *S. salivarius* may contribute to dental plaque formation and oral biofilm development, resulting in dental caries and periodontal diseases. Infections caused by *S. salivarius* include bacteremia, meningitis, endocarditis, pneumonia, urinary tract infections, and cholecystitis [3,4]. *S. salivarius* is a rare cause of sepsis in newborns [5].

S. *salivarius* has been observed to exhibits drug resistance, particularly macrolides and tetracyclines [6]. This gene resistance indicated that *S. salivarius* can function as a reservoir of antibiotic resistance genes in the oral microbiota. Consequently, multiple drug resistance (MDR) has been detected in oral streptococcal isolates, with a combination of erythromycin, tetracycline, and ofloxacin resistance being prevalent [6,7]. These findings underscore the importance of monitoring the susceptibility of *S. salivarius* to antimicrobial agents as it can potentially mediate the transfer of resistance determinants to more pathogenic microorganisms [8]. The objective of this study was to examine the genomic characteristics of *S. salivarius* and explore its secondary metabolites, antibiotic resistance genes, and virulence factor types. To enhance the clinical application of *S. salivarius* and antibiotics by providing valuable insights.

MATERIAL AND METHODS:

Material

20 strains of *S. salivarius* with complete genome sequences were selected from the NCBI database, in which 12 strains had clear isolation sources, and their complete genome sequences were downloaded in both fasta and GenBank (full) formats.

METHODS:

Average nucleotide identity (ANI) analysis and genome annotation

ANI analysis assesses genetic relationships at the whole-genome level. ANI/AAI-Matrix software (http://enve-omics.ce.gatech.edu/g-matrix/) was used to investigate the genomic evolutionary distances among different strains [9,10]. An ANI value exceeding 95% indicated conspecificities. Genome sequences were annotated using Prokka v1.13, obtained from Conda (v. 4.14.0) [11].

Genome phylogenetic analysis

The 20 genome sequences in FASTA format were submitted to REALPHY (https://realphy.unibas.ch/realphy/) for whole-genome comparison, using the ICDC1 strain as the reference and the other 19 strains for comparison [12]. The resultant tree file was imported into FigTree (v1.4.4, http://tree.bio.ed.ac.uk/software/figtree) to construct and refine the evolutionary tree.

Pan-genome and core genome analysis

The GenBank genome sequences of 20 bacterial strains were submitted to PGAweb (http://pgaweb.vlcc.cn/) [13] for PGAP analysis using the GeneFamily Method (GF). Results in the Orthologs_Clusters.txt file were fitted using PanGP (https://pangp.zhaopage.com/).

Analysis of secondary metabolites, drug resistance genes, and virulence genes

Genome sequences in FASTA format were submitted to anti SMASH (https://antismash.secondarymetabolites.org/) [14,15] to analyze the secondary metabolites of all 20 S. salivarius strains using the relaxed prediction method. The Resistance Gene Identifier(; RGIhttps://card.mcmaster.ca/) was employed to analyze resistomes from protein, genome, or genome assembly data in FASTA format. Comprehensive documentation and a command-line version of RGI are available at https://github.com/arpcard/rgi for selecting the Comprehensive Antibiotic Research Database (CARD) [16]. The Virulence Factors Database (VFDB) [17] was utilized via BLASTN/BLASTP searches at http://www.mgc.ac.cn/VFs/ to identify the sequence homologs.

RESULTS:

Average Nucleotide Identity (ANI)

Average nucleotide identity (ANI) analysis was used to determine the connection between species in S. salivarius, with a 95% ANI value indicating the same species. This suggests an ANI identification criterion of 65%–90% for species and 90%–96% for subspecies [18]. Among all 20 strains of *S. salivarius*,this study discovered that the ANI values in 12 strains of *S. Salivarius* were 94% (Fig. 1), which is comparable to previously reported strain information. These 12 strains were ATCC25975, ATCC29745, DB-B5, FDAARGOS_259, FDAARGOS_771, FDAARGOS_1045, NCTC_7366, ICDC_3, LAB_813, ICDC_2, SALI_10, and 57. I. ATCC25975 and ATCC29745, which were obtained from saliva, whereas SALI_10 and DB-B5 were collected from the oral cavity. FDAARGOS_259 was derived from the blood. FDAARGOS_771 was obtained from normal skin on the right arm. The ICDC_1 and ICDC_2 were collected from children aged 4-6 years. The ANI value between ICDC_1and ICDC_3 was 100%, and the ANI values across LMG 11489, JF, NCTC_8618, and NCTC861 were all 100%. All strains that shared an ANI value of 100% originated from the mouth cavity.

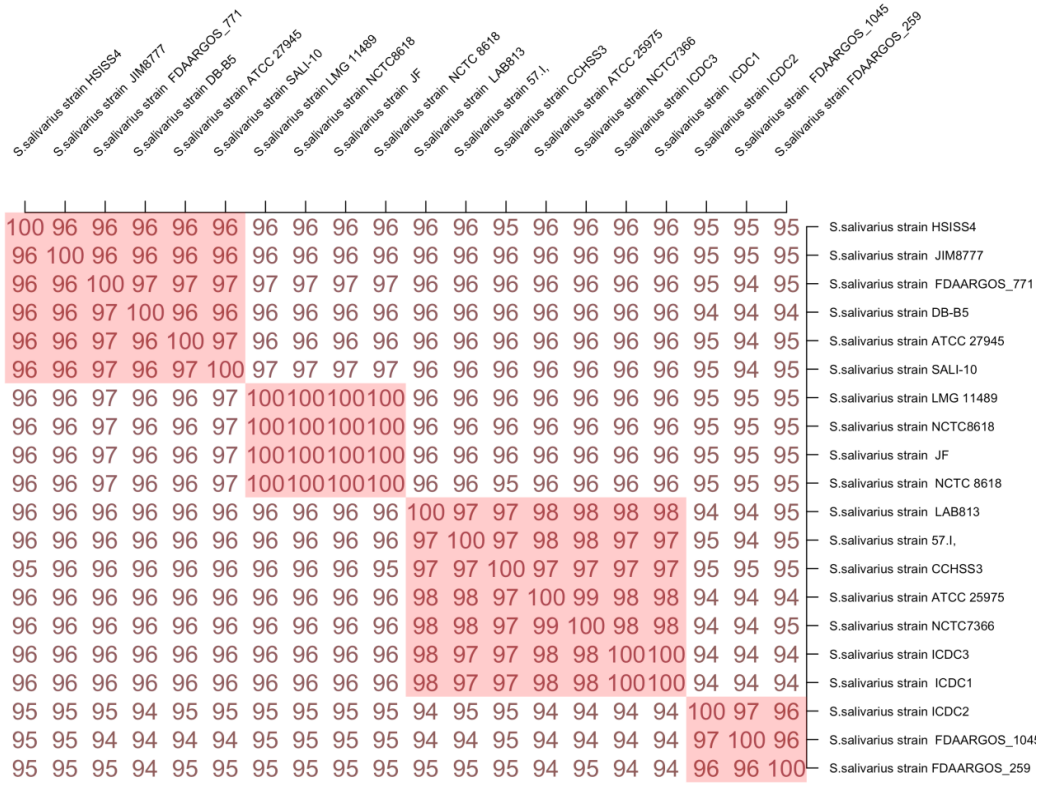

Fig. 1. Average Nucleotide Identity (ANI)

Annotation of the genome sequence

The number of rRNA of 20 strains was in the range of 15–23, and the number of tRNA was in the range of 67–68. Except for strains, FDAARGOS_1045 had a range of 58, the least among all other strains, LAB813 had a range of 74, which is above the range of other strains, and JF had the highest range of 91 among all strains. All the strains contained one transfer-messenger RNA in their genomes. The RNA, referred to as transport-messenger, or 10SaRNA, was first identified in *E. coli* and has the ability to break down aberrant messengers more quickly and recycle ribosomes that have stalled [19]. The genome size of all strains ranged from 2.0 to 2.4 MB, with the largest being the ICDC_2 strain from a 4-6-year-old child with healthy oral and the smallest being the FDAARGOS_771 strain from the normal skin of the right arm (Table 1). The whole-genome GC% content of all strains ranged from 39.5% to 40.5%.

Table 1. Genomic annotation information of 20 strains of *S. salivarius*

Strain	rRNA	CDS	tRNA	tmRNA	Genome size (MB)	GC content (%)	Isolation source	NCBI accession number
57.I	18	1947	68	1	2.1	40	Not available	NC_017594.1
ATCC_25975	18	1981	68	1	2.3	40	Saliva	NZ_CP015283.1
ATCC_29745	18	1885	68	1	2.1	40	Saliva	NZ_CP015282.1
CCHSS3	18	2004	68	1	2.2	40	Not available	NC_015760.1
DB-B5	18	1904	68	1	2.3	40	Oral	NZ_CP054153.
FDAARGOS_259	18	2051	68	1	2.3	40	Oral	NZ_CP054153.1
FDAARGOS_771	18	1853	68	1	2	40	Normal skin of the right arm	NZ_CP053998.1
FDAARGOS_1045	15	2123	58	1	2.3	39.5	Not available	NZ_CP066093.1
HSISS4	18	1876	68	1	2.1	40	Ileostomy effluent eveningsample obtained from a 79-year-old male ileostomy plated on Mitis-salivarius agar	NZ_CP013216.1
ICDC_1	18	1974	68	1	2.2	40	4-6-year-old child with healthy oral	NZ_CP018186.1
ICDC_2	18	2028	68	1	2.4	40	4-6-year-old child with healthy oral	NZ_CP018187.1
ICDC_3	18	1976	68	1	2.2	40	4-6-year-old child with healthy oral	NZ_CP018189.1
JF	22	1987	91	1	2.2	40	Oral	NZ_CP014144.1
JIM8777	18	1971	68	1	2.2	40	Not available	NC_017595.1
LAB_813	23	2027	74	1	2.4	40	Not available	NZ_CP040804.1
LMG11489	15	1993	67	1	2.2	40	Not available	NZ_CP133476.1
NCTC_8618	18	1992	68	1	2.2	40	Oral cavity	NZ_CP009913.1
NCTC_7366	18	1986	68	1	2.2	40	Not available: to be reported later	NZ_LS483366.1
NCTC8618	18	2001	68	1	2.1	40	Not available	NZ_LR134274.1
SALI-10	18	1849	68	1	2.3	40.5	Human oral cavity	NZ_CP090007.1

Analysis of genome-based on phylogenetic tree

The whole-genome phylogenetic tree **(Fig. 2)** showed that the 20 *S. Salivarius* strains were divided into three main branches. The first branch is the *S. Salivarius* strain ICDC_2 source, which is a 4–6-year-old child with a healthy oral cavity.. The second branch indicates *S. Salivarius* strain FDAARGOS_1045, the source of which is unknown, whereas the third branch contains the remaining strains.

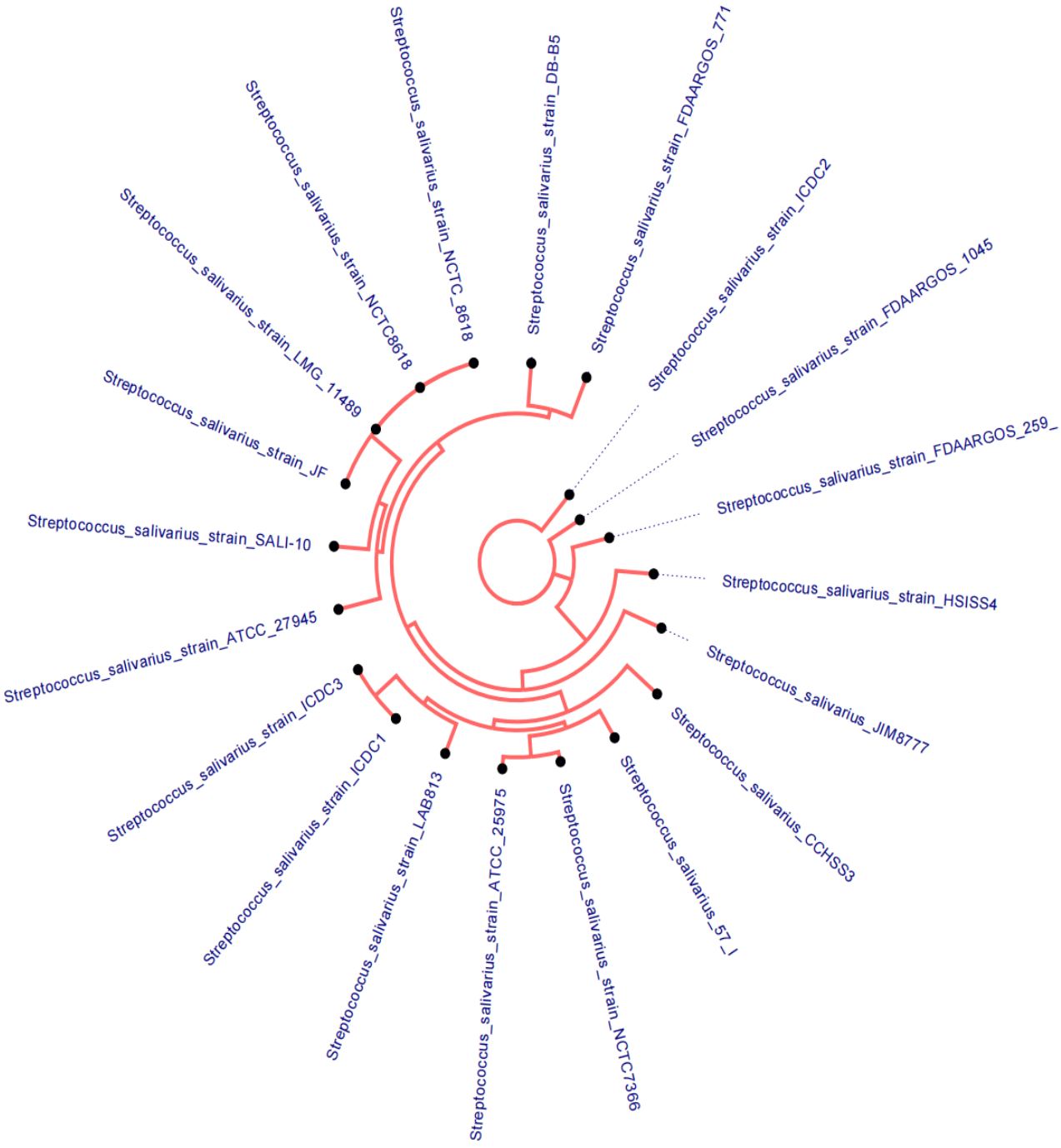

Fig. 2 Whole genome phylogenetic tree of 20 *S. Salivarius* strain.

Pan-genome and core genome analysis

20 representative *S. salivarius* isolates from different sources were subjected to pan-genome and core-genome analyses **(Fig. 3)**. A total of 38,370 genes were identified, of which 27,690 genes (72.16%) were shared by all genomes of the *S. Salivarius* strains, 950 genes were specific genes (2.47%), and 9,730 genes were non-essential genes (25.35%). Of the strains analyzed, strain FDAARGOS_1045 had the highest number of unique genes (179), and strain LAB813 had the most core genes (1396).

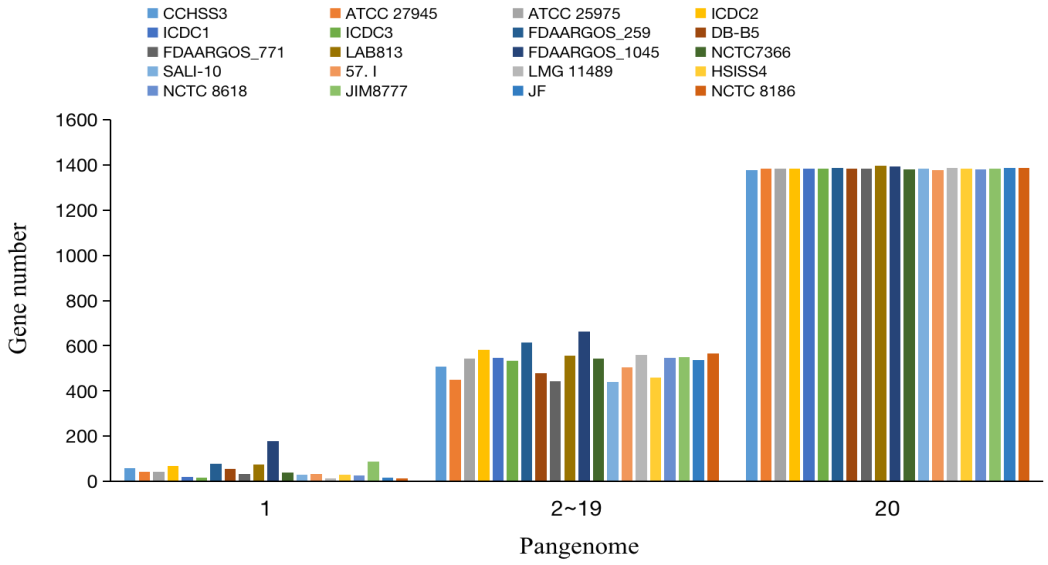

Fig. 3. Core and nonessential genes in the genome.

To infer the relationship between pan-genome, core genome, and bacterial count, PanGP was used to fit the data and obtain box plots of pan-genome and core genome, with blue indicating pan-genome size and green indicating core-genome size. As shown in **(Fig .4)**, the size of the pan-genome increases with the number of genomes, and the fitting equation between the pan-genome (y) and genome size (x) is $y=668.48x^{0.39}+1186.9$; the size of the core genome decreases with the increase in the number of genomes. The fitting equation of the relationship between the core genome (M) and genome size (N) is $M=577.16e^{-0.31x}+1376$, which indicates that *S. Salivarius* has an open pan-genome.

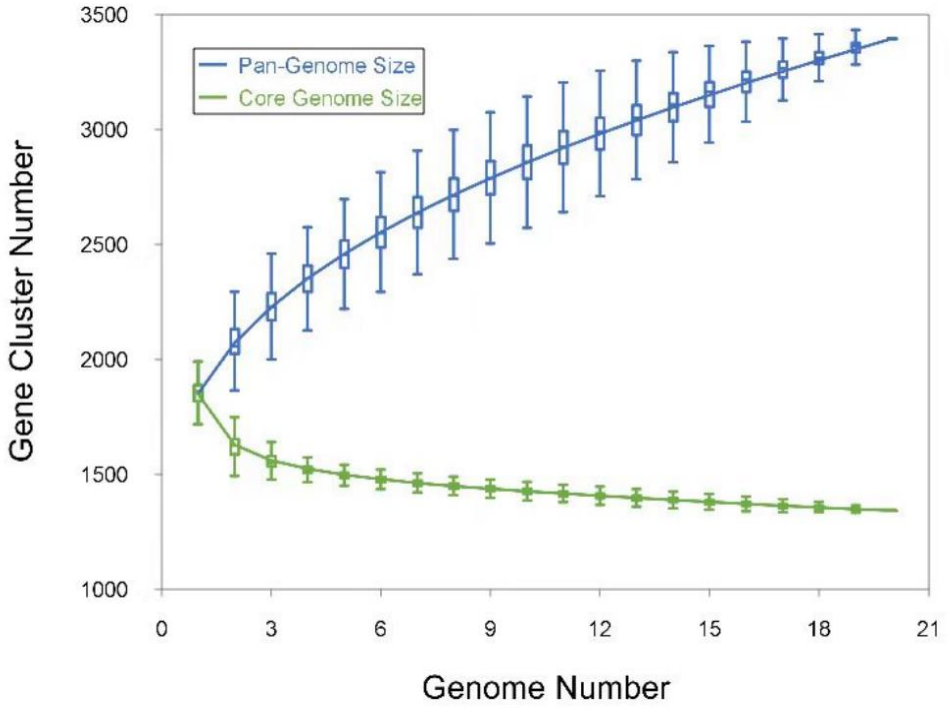

Fig. 4. Pangenome and core genome characteristic curves

The relationship between the number of new genes (T) and the number of genomes (P) was calculated using PanGP, and the fitting equation was $T=316.085P^{-0.65}$
As shown in **(Fig .5)**, it can be seen from the figure that as the number of genomes increased, the number of new genes gradually decreased until it leveled off and ceased to increase (P≥7).

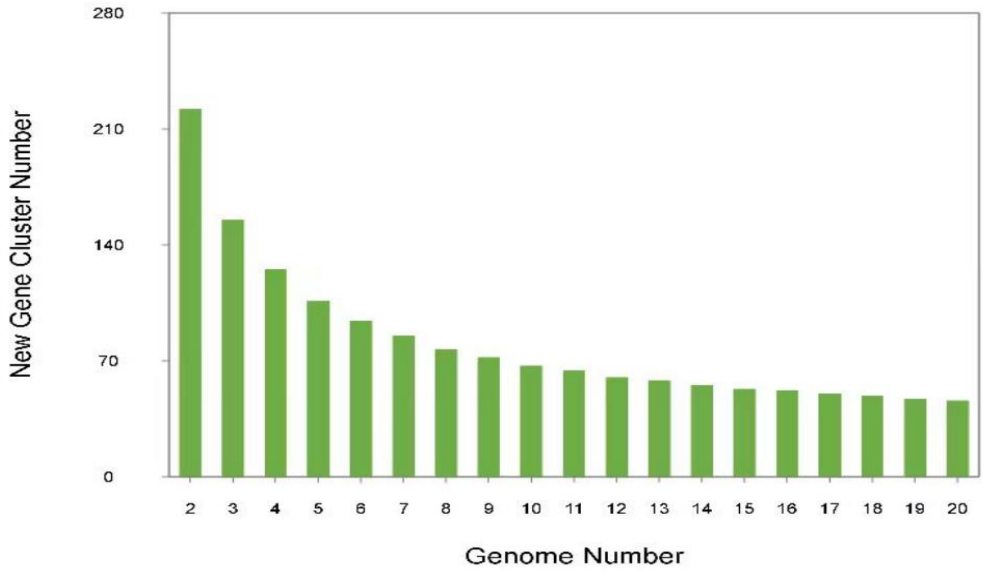

Fig. 5. Number of new genes produced by *S. Salivarius*

Secondary metabolite analysis
Using the antiSMASH online software, a total of eight8 categories, 81 gene clusters related to secondary metabolism, and 82 secondary metabolites were predicted. The (Table 4.2) included 47 *RiPP-like*, 20 *T3PKS*, 5 *Ras_Ripp*,4 *Lanthipeptide class-i*, three *Lanthipeptide class-ii*, two *RRE-containing*, one *Lassopeptide*, one *LAP*, *thiopeptide*. FDAARGOS_1045 strain had more abundant secondary metabolism gene cluster types than the other strains.

Table 2. Secondary metabolite gene cluster and types of 20 strains of *S. salivarius*

Strain	Cluster	Type	From	To
57. I	Cluster 1	*RiPP-like*	311,902	330,735
	Cluster 2	*T3PKS*	1,533,945	1,5575,111
	Cluster 3	*RiPP-like*	1,863,631	1,875,727
CCHSS3	Cluster 4	*RiPP-like*	264,373	276,469
	Cluster 5	*lanthipeptide-class-ii,RaS-Ripp*	742,810	776,926
	Cluster 6	*T3PKS*	1,579,285	1,620,451
	Cluster 7	*RiPP-like*	1,782,054	1,792,350
	Cluster 8	*RiPP-like*	1,875,773	1,894,610
JIM8777	Cluster 9	*RiPP-like*	254,276	266,372
	Cluster 10	*lasso peptide*	519,883	542,318
	Cluster 11	*T3PKS*	574,836	616,005
	Cluster 12	*RiPP-like*	1,878,141	1,900,930
ATCC 25975	Cluster 13	*RiPP-like*	202,310	221,670
	Cluster 14	*RiPP-like*	800,229	812,325

	Cluster 15	*RiPP-like*	836,160	848,316
	Cluster 16	*RaS-RiPP*	1,264,148	1,286,572
	Cluster 17	*T3PKS*	2,108,816	2,149,982
ATCC 27945	Cluster 18	*T3PKS*	950,866	992,032
	Cluster 19	*RiPP-like*	1,258,856	1,270,952
	Cluster 20	*RiPP-like*	1,832,650	1,842,883
DB-B5	Cluster 21	*RiPP-like*	242,921	255,017
	Cluster 22	*T3PKS*	512,915	554,081
	Cluster 23	*RaS-RiPP*	1,331,210	1,353,501
	Cluster 24	*RiPP-like*	1,781,088	1,799,925
FDAARGOS_259	Cluster 25	*T3PKS*	352,775	393,941
	Cluster 26	*RiPP-like*	682,836	694,932
	Cluster 27	*RiPP-like*	1,330,010	1,348,795
FDAARGOS_771	Cluster 28	*RiPP-like*	804,636	823,458
	Cluster 29	*RiPP-like*	1,376,253	1,388,349
	Cluster 30	*RiPP-like*	1,415,590	1,427,746
	Cluster 31	*T3PKS*	1,669,477	1,710,643
FDAARGOS_1045	Cluster 32	*RiPP-like*	204,141	222,964
	Cluster 33	*T3PKS*	1,437,501	1,478,667
	Cluster 34	*lanthipeptideclass-ii*	1,534,986	1,557,820
	Cluster 35	*lanthipeptide-class-ii*	1,600,707	1,623,487
	Cluster 36	*RRE-containing*	1,624,114	1,646,975
	Cluster 37	*RRE-containing*	1,650,517	1,670,777
	Cluster 38	*RiPP-like*	1,958,733	1,970,829
HSISS4	Cluster 39	*RiPP-like*	244,219	256,315
	Cluster 40	*T3PKS*	533,072	574,238
	Cluster 41	*RiPP-like*	1,775,315	1,794,142
	Cluster 42	*RiPP-like*	1,858,797	1,868,997
ICDC1	Cluster 43	*RiPP-like*	260,291	272,387
	Cluster 44	*T3PKS*	1,559,574	1,600,740
	Cluster 45	*RiPP-like*	1,766,982	1,777,254
	Cluster 46	*RiPP-like*	1,861,023	1,888,089
ICDC2	Cluster 47	*RiPP-like*	274,734	286,830
	Cluster 48	*T3PKS*	577,443	618,609
	Cluster 49	*RiPP-like*	1,910,107	1,928,941
ICDC3	Cluster 50	*RiPP-like*	260,291	272,387
	Cluster 51	*T3PKS*	1,559,549	1,600,715
	Cluster 52	*RiPP-like*	1,766,977	1,777,249
	Cluster 53	*RiPP-like*	1,861,018	1,888,084
JF	Cluster 54	*lanthipeptide-class-i*	81,639	107,618
	Cluster 55	*RiPP-like*	256,222	268,318
	Cluster 56	*T3PKS*	524,033	565,199
	Cluster 57	*RiPP-like*	1,710,927	1,733,713
LAB813	Cluster 58	*RiPP-like*	507,752	526,586

	Cluster 59	*RiPP-like*	597,368	607,640
	Cluster 60	*T3PKS*	769,534	810,700
	Cluster 61	*RiPP-like*	2,076,684	2,088,780
LMG 11489	Cluster 62	*lanthipeptideclass-i*	324,444	350,423
	Cluster 63	*RiPP-like*	499,045	511,141
	Cluster 64	*T3PKS*	779,635	820,801
	Cluster 65	*RiPP-like*	2,057,995	2,080,781
NCTC 8618	Cluster 66	*lanthipeptide-class-i*	135,668	161,647
	Cluster 67	*RiPP-like*	316,085	328,181
	Cluster 68	*T3PKS*	596,691	637,857
	Cluster 69	*RiPP-like*	1,863,368	1,886,155
NCTC7366	Cluster 70	*RiPP-like*	264,416	276,512
	Cluster 71	*RaS-RiPP*	732,792	755,216
	Cluster 72	*LAP,thiopeptide*	866,465	891,138
	Cluster 73	*T3PKS*	1,553,690	1,594,856
	Cluster 74	*RiPP-like*	1,843,989 -	1,863,349
NCTC 8618	Cluster 75	*lanthipeptide-class-i*	135,931	161,910
	Cluster 76	*RiPP-like*	315,757	327,853
	Cluster 77	*T3PKS*	596,347	637,513
	Cluster 78	*RiPP-like*	1,874,707	1,897,493
SALI-10	Cluster 79	*RiPP-like*	93,809	812,648
	Cluster 80	*RiPP-like*	1,370,670	1,382,766
	Cluster 81	*T3PKS*	1,658,885	1,700,051

Drug resistance genes and virulence genes

Analysis of resistance genes

Eight resistance genes are available from the CARD resistance gene database, namely *VanT in the VanG* cluster, *VanY in the VanG* cluster, *VanY in VanB* cluster, *VanY in VanM* cluster, *qacJ*, *tet(M)*, *ErmB*, and *catQ* (Table. 3). In strain 57.1, *VanT in the VanG* cluster and *VanY in the VanG* cluster shared genes, whereas CHSS3 contained a separate gene *VanY in the VanB* cluster. *VanY* is located in the *VanM* cluster-specific gene of the strain JIM8777. *VanY* was found in the *VanB* cluster-specific gene of the strain ATCC 25975. FDAARGOS_259 contains *qacJ*, *tet(M)*, and *ErmB*-specific genes. FDAARGOS_771 contains a *tet (M)*-specific *ErmB* gene. The ICDC-1 strain contained an *ErmB*-specific gene. The ICDC-2 strain contained a catQ-specific gene. The ICDC-3 strain contained an *ErmB*-specific gene. The *VanB* cluster-specific genes in the LAB813 strain included *qacJ*, *tet(M)*, and *VanY*. Finally, the NCTC7366 strain contains *VanY in the VanB* cluster-specific gene.All genes share glycopeptide antibiotic as a drug class and antibiotic target alteration as a resistance mechanism, except for *tet(M)* which involves tetracycline antibiotic and antibiotic target protection. *qacJ* is associated with disinfecting agents and antiseptics as the drug class, and antibiotic efflux as the resistance mechanism. *ErmB* is a macrolide antibiotic, lincosamide antibiotic, streptogramin antibiotic, streptogramin A antibiotic, and streptogramin B antibiotic as drug classes, with antibiotic target alteration as the resistance mechanism. *catQ* is linked to phenicol antibiotics as a drug class and antibiotic inactivation as a resistance mechanism. The CARD resistance gene database was used as the protein homolog model. The RGI criteria were Perfect, Strict, or complete.

Virulence Genes Resistance

To determine the virulence gene resistance, we used VFDB and performed BLASTN (nucleotide sequences from the VFDB core dataset {setA}). This procedure was performed using 20 bacterial strains. We obtained seven different nucleotide genes from the database with a score equal to or greater than 80. All strains shared five genes, including *hasC* (UDP-

glucose pyrophosphorylase), responsible for a hyaluronic capsule, and four others belonging to the protein family: *psaA* (manganese-binding adhesion lipoprotein), *lap* (Listeria adhesion protein Lap), *fbp54* (Fibronectin-binding proteins also known as fbpA), and *fbp54* (fbp from the family Streptococcus pyogenes M1 GAS). However, only a few strains have the specific genes *clpC* (endopeptidase Clp ATP-binding chain C) and *pavB/pfbB*. *clpC* was detected in strains 57, I, CCHSS3, ATCC 25975, ATCC 27945, FDAARGOS_259, ICDC1, ICDC3, JF, LAB 813, LMG 11489, NCTC 8618, NCTC7366, NCTC8618, and SALI-10. In contrast, *pavB/pfbB*, a Plasminogen- and Fibronectin-binding protein B, was only present in three strains: JIM8777, FDAARGOS_771, and SALI-10.

Table 3 Drug resistance genes and virulence genes

Strain	Secondary metabolites	Resistant genes		Virulence genes	
		Shared genes	Specific genes	Shared genes	Specific genes
57.I	RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG	*VanY* in VanB	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
CHSS3	RiPP-like lanthipeptide-class-ii,RaS-Ripp T3PKS RiPP-like RiPP-like	*VanT* in VanG *VanY* in VanG	*VanY* in VanB	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
JIM8777	RiPP-like lasso peptide T3PKS RiPP-like	*VanT* in VanG*VanY* in VanG	*VanY* in VanM	*hasC*;*psaA*; f*bp54*;*lap*	*pavB*;*pfbB*
ATCC 25975	RiPP-like RiPP-like RiPP-like RaS-RiPP T3PKS	*VanT* in VanG *VanY* in VanG	*VanY* in VanB	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
ATCC 27945	T3PKS RiPP-like RiPP-like	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
DB-B5	RiPP-like T3PKS RaS-RiPP RiPP-like	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	
FDAARGOS_259	T3PKS RiPP-like RiPP-like	*VanT* in VanG *VanY* in VanG	*qacJ*;*tet(M)*; *ErmB*	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
FDAARGOS_771	RiPP-like RiPP-like RiPP-like T3PKS	*VanT* in VanG *VanY* in VanG	*ErmB*	*hasC*;*psaA*; f*bp54*;*lap*	*pavB*;*pfbB*
FDAARGOS_1045	RiPP-like T3PKS	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	

	lanthipeptide-class-ii lanthipeptide-class-ii RRE-containing RRE-containing RiPP-like				
HSISS4	RiPP-like T3PKS RiPP-like RiPP-like	*VanT* in VanG*VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	
ICDC1	RiPP-like T3PKS RiPP-like RiPP-like	*VanT* in VanG*VanY* in VanG	*ErmB*	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
ICDC2	RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG	*catQ*	*hasC*;*psaA*; f*bp54*;*lap*	
ICDC3	RiPP-like T3PKS RiPP-like RiPP-like	*VanT* in VanG *VanY* in VanG	*ErmB*	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
JF	lanthipeptide-class-i RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
LAB813	RiPP-like RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG	*VanY* in VanB *qacJ*;*tet(M)*	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
LMG 11489	lanthipeptide-class-i RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
NCTC 8618	lanthipeptide-class-i RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
NCTC7366	RiPP-like RaS-RiPP LAP,thiopeptide T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG	*VanY* in VanB	*hasC*;*psaA*; f*bp54*;*lap*	*clpC*
NCTC8618	lanthipeptide-class-i RiPP-like T3PKS RiPP-like	*VanT* in VanG *VanY* in VanG		*hasC*;*psaA*; f*bp54*;*lap*	*clpC*

SALI-10	RiPP-like RiPP-like T3PKS	*VanT* in VanG *VanY* in VanG	*hasC*;*psaA*; f*bp54*;*lap*	*pavB*;*bfbB*; *clpC*

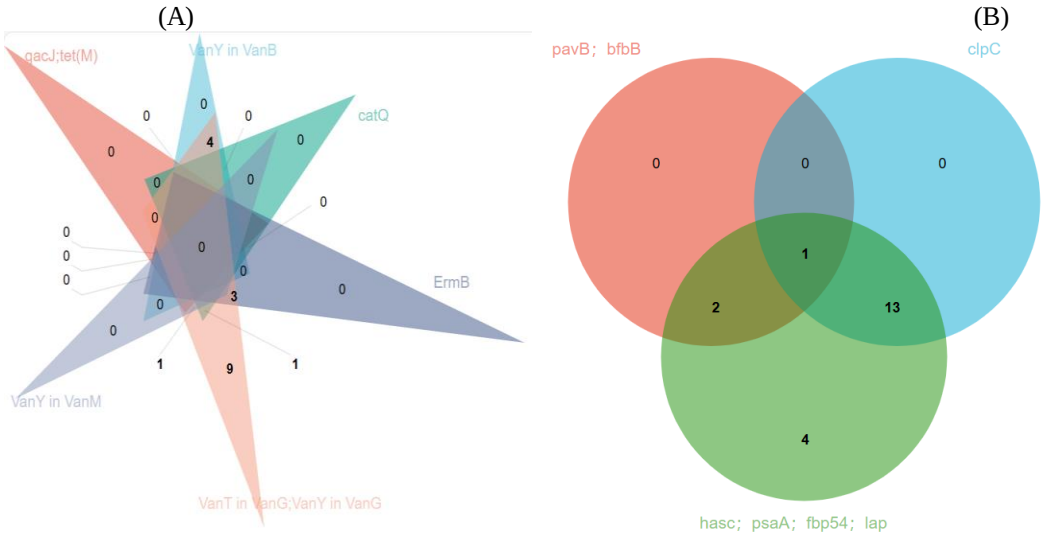

Fig. 6. Venn diagram of shared genes for resistance genes (A) and virulence genes (B)

DISCUSSION:

Twelve *S. Salivarius* strains had an average nucleotide identity (ANI) of 94%, including ATCC25975, ATCC29745, DB-B5, FDAARGOS_259, FDAARGOS_771, FDAARGOS_1045, NCTC_7366, ICDC_3, LAB_813, ICDC_2, SALI_10, and 57. I. This supports the existing knowledge. All the strains were isolated from the oral cavity. FDAARGOS_1045 had the lowest range (58), whereas LAB813 had the highest (74). All the strains contained 10SaRNA, which degrades aberrant messengers. Genome sizes ranged from 2.0 to 2.4 MB, with ICDC_2 being the largest and FDAARGOS_771 the smallest. The phylogenetic tree showed three main branches: ICDC_2, FDAARGOS_1045, and the remaining strains, each from different sources. Analysis of the 20 isolates identified 38,370 genes, of which 72.16% were shared. A total of 950 were specific genes and 9,730 were non-essential. FDAARGOS_1045 had the most unique genes (179), whereas LAB813 had the most core genes (1396). The PanGP data indicated an open pan-genome in S. Salivarius. The pan-genome size increased, and the core genome size decreased with more genomes. The relationship between the core genome and genome size was M=577.16e-0.31x+1376. Box plots indicate the relationship between the pan-genome, core genome, and bacterial count. The new gene count and genome relationship used PanGP with T=316.085P-0.65. AntiSMASH predicted 81 gene clusters for secondary metabolism and 82 secondary metabolites, with FDAARGOS_1045 having the most abundant gene cluster types. The CARD resistance gene database identified eight resistance genes, including *VanT*, *VanY*, *qacJ*, *tet(M)*, *ErmB*, and *catQ*, sharing glycopeptide antibiotics as a drug class, with the exception *tet(M)* (tetracycline antibiotics) and *qacJ* (disinfecting agents and antiseptics). *ErmB* is associated with multiple antibiotics, with target alterations as a resistance mechanism. *CatQ* involves phenicol antibiotics and their inactivation. The detection criteria used the protein homolog model, with the RGI criteria being perfect, strict, and complete genes only. VFDB and BLASTN identified virulence genes in 20 strains, with seven different nucleotide genes scoring 80 or higher. All strains shared five genes, including *hasC* for a hyaluronic capsule, and a few strains possessed *clpC* and *pavB/pfbB*.

CONCLUSION:

In this study, we conducted a comprehensive analysis of *Streptococcus salivarius*, focusing on its genetic composition and potential clinical implications. This study examined the presence and distribution of drug resistance and virulence genes in *S. salivarius* strains, providing valuable insights into the pathogenic

potential of the organism and its ability to withstand antimicrobial treatments. Additionally, this study delves into the genomic architecture of *S. salivarius*, exploring its core genome (genes shared by all strains) and pan-genome (the total gene pool of the species). This genomic analysis not only sheds light on the genetic diversity within the species but also helps in understanding the evolutionary adaptations of *S. salivarius*. Furthermore, this study predicted secondary metabolites produced by *S. salivarius*, which could have implications for its interactions with the host and other microorganisms in its environment. Elucidation of the relationships between gene numbers, core genome, and pan-genome provides a deeper understanding of the species' genetic flexibility and adaptability. These findings collectively contribute to the theoretical foundation for potential medical treatments targeting *S. salivarius* and enhance our understanding of this important oral commensal. However, the study acknowledges that further research, particularly focusing on strain FDAARGOS_1045, is necessary to gain a more comprehensive understanding of *S. salivarius* and its potential clinical significance.

REFERENCES:

1. Hakalehto E, Vilpponen-Salmela T, Kinnunen K, von Wright A. 2011. Lactic acid bacteria enriched from human gastric biopsies. ISRN Gastroenterol. 2011:109183. 10.5402/2011/109183.
2. Masaaki, Minami., Shunsuke, Akahori., Michio, Ohta. (2023). Amylase from Streptococcus pyogenes inhibits biofilm formation in Streptococcus salivarius. GSC Advanced Research and Reviews, doi: 10.30574/gscarr.2023.14.2.0050.
3. Burton JP, et al. 2006. A preliminary study of the effect of probiotic Streptococcus salivarius K12 on oral malodour parameters. Journal of Applied Microbiology, 100(4): 754-764.
4. Ferretti J, et al. 2016. Streptococcus pyogenes: Basic Biology to Clinical Manifestations [Internet]. Oklahoma City (OK): University of Oklahoma Health Sciences Center; 2016.
5. Franklin, W., Yates., Christopher, Bruno. (2014). Streptococcus salivarius Meningitis as a Presenting Manifestation of a Rare Underlying Basicranial Defect. Infectious Diseases in Clinical Practice, doi: 10.1097/IPC.0000000000000135.
6. Nataliia, Valerievna, Davidovich., A, S, Galieva., N, G, Davydova., O, G, Malygina., N, N, Kukalevskaya., G, V, Simonova., T., A., Bazhukova. (2020). Spectrum and resistance determinants of oral streptococci clinical isolates.. Klinicheskaia laboratornaia diagnostika, doi: 10.18821/0869-2084-2020-65-10-632-637.
7. Fanny, Chaffanel., Florence, Charron-Bourgoin., Virginie, Libante., Nathalie, Leblond-Bourget., Sophie, Payot., Sophie, Payot. (2015). Resistance Genes and Genetic Elements Associated with Antibiotic Resistance in Clinical and Commensal Isolates of Streptococcus salivarius.. Applied and Environmental Microbiology, doi: 10.1128/AEM.00415-15.
8. Thais, Palma, Erika, N., Harth-Chu., Jodie, C., Scott., Rafael, Nobrega, Stipp., Heike, Boisvert., Mariana, F., Salomao., Jéssica, Dias, Theobaldo., Rosana, de, Fátima, Possobon, Leandro, Costa, do, Nascimento, Jonathan, W., McCafferty., Lina, L., Faller., Margaret, J., Duncan., Renata, O., Mattos-Graner. (2016). The oral cavities of healthy infants harbor high proportions of Streptococcus salivarius strains with phenotypic and genotypic resistance to multiple classes of antibiotics. Journal of Medical Microbiology, doi: 10.1099/JMM.0.000377.
9. Rodriguezr, L.M. and K.T. Konstantinidis. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes [J]. PeerJ Preprints, 2016 4:e1900v1.
10. Luis, M., Rodriguez-R., Roth, E., Conrad., Tomeu, Viver., Dorian, J., Feistel., B., Lindner., Fanus, Venter., Luis, H., Orellana., Rudolf, Amann., Ramon, Rosselló-Móra., Konstantinos, T., Konstantinidis. (2023). An ANI gap within bacterial species that advances the definitions of intra-species units. bioRxiv, doi: 10.1101/2022.06.27.497766.
11. Frederic Bertels, Olin K. Silander, Mikhail Pachkov. Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads [J]. Molecular biology and evolution, 2014. 31(5): 1077-1088.
12. Frederic Bertels, Olin K. Silander, Mikhail Pachkov. Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads [J]. Molecular biology and evolution, 2014. 31(5): 1077-1088.
13. Xinyu, Chen., Yadong, Zhang., Zhewen, Zhang., Yongbing, Zhao., Chen, Sun., Ming, Yang., Jinyue, Wang., Jinyue, Wang., Qian, Liu., Baohua, Zhang., Meili, Chen., Jun, Yu., Jiayan, Wu., Zhong, Jin., Jingfa, Xiao., Jingfa, Xiao. (2018). PGAweb: A Web Server for Bacterial Pan-Genome Analysis.. Frontiers in

Microbiology, doi: 10.3389/FMICB.2018.01910

14. Kai, Blin., Simon, Shaw., Hannah, E., Augustijn., Zachary, L., Reitz., Friederike, Biermann., Mohammad, Alanjary., Artem, Fetter., Barbara, R., Terlouw., William, W., Metcalf., Eric, J., N., Helfrich., Gilles, P., van, Wezel., Marnix, H., Medema., Tilmann, Weber. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures, and visualization. Nucleic Acids Research, doi: 10.1093/nar/gkad344.
15. Kai, Blin., Simon, Shaw., Alexander, M., Kloosterman., Zach, Charlop-Powers., Gilles, P., van, Wezel., Marnix, H., Medema., Marnix, H., Medema., Tilmann, Weber. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Research, doi: 10.1093/NAR/GKAB335
16. Mcarthur, A G, et al. The comprehensive antibiotic resistance database [J]. Antimicrobial Agents & Chemotherapy, 2013. 57(7):3348-3357.
17. Liu B, Zheng DD, Zhou SY, Chen LH, and Yang J, 2022. VFDB 2022: a general classification scheme for bacterial virulence factors.
18. Bo, L. I. U., H. U. Gui-ping, and T. A. N. G. Wei-qi. "Characteristic of average nucleotide identity (ANI) based on the whole genomes from Bacillus species in Bacillus-like genus." 福建农业学报 28.9 (2013): 833-843.
19. KoMINE YU, Kitabatake M, Yokogawa T, Nishikawa K, Inokuchi H. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from Escherichia coli. Proceedings of the National Academy of Sciences. 1994 Sep 27;91(20):9223-7.