Bacterial Whole Genome Sequencing Report

Sample ID: W01022021_001

Received date: 01-02-2021

Reported date: 15-02-2021

Sample type: gDNA

Organism name: *Lactobacillus salivarius* strain Porcinocin

Reference genome: *L. salivarius* str. Ren (NP_CP011403.1)

Estimated genome size: 1.75 Mbp

NGS report details

☑ Sample information

☑ Data quality report

☑ Genome assembly

☑ Genome annotation

☑ Circular map

☑ Species identification and Phylogenetic tree analysis

☑ Bacterial variant calling (please provide reference accessions number)

Optional

☑ Functional analysis

☑ Specialty genes and Antimicrobial resistance gene analysis

☑ Secondary metabolite prediction

☑ Phage sequence identification

☑ Comparative genomic

Bacterial Whole Genome Sequencing Report

(ตัวอย่างผลการวิเคราะห์ WGS เบื้องต้น)

1. Sample information

The sample information table shows your sample ID and the name of fastq files obtained from illumina Miseq500 platform (**Table 1**).

Table 1. Sample information

Customer label	Sample ID	Raw seq file (fastq)	Filtered seq file (fastq)
SW1_1	SW_1_01	SW1_1_R1_001.fastq, SW1_1_R2_001.fastq	SW1_1_R1_001_filt.fastq, SW1_1_R2_001_filt.fastq

Raw seq files:

SW1_1_R1_001.fastq

SW1_1_R2_001.fastq

SW1_1_R1_001_filt.fastq

SW1_1_R2_001_filt.fastq

2. Quality profiles

The paired-end sequencing reads obtained from illumina Miseq500 platform were **699,603** reads for Read 1 (R1, SW1_1_R1_001.fastq) and **699,603** reads for Read 2 (R2, SW1_1_R1_002.fastq). The quality and adapter trimming of sequencing datasets are possessed by Trim Galore! with default parameters (Q score cutoff = 20). The summary of QC report is provided in **Table 2**. The quality of raw sequence datasets were analyzed using FastQC (1). The remaining good quality reads were **659,519** reads for R1 (SW1_1_R1_001_filt.fastq) and **659,519** reads for R2 (SW1_1_R2_001_filt.fastq). Next, the short reads were mapped against the provided reference genome and the result illustrated that the percent coverage compared to reference genome was 98.51% (**Table 2**).

Table 2. Summary of data quality report

Dataset information	
Total raw read (read)	
Read 1 (R1)	699,603
Read 2 (R2)	699,603
Filtered (read)	
Read 1 (R1)	659,519
Read 2 (R2)	659,519
Percent coverage to reference genome, *L. salivarius* str. Ren (NP_CP011403.1)	
Ref. length (bp)	1,751,565
Covered bases (bp)	1,725,381
Genome coverage	98.51X

File;

Constats.txt

3. Genome assembly

The datasets are submitted to an assembly pipeline for bacterial genomes, Unicycler, to produce complete and accurate assemblies (2). Briefly, the paired-end inputs are assembled with default setting. Unicycler output file is provided in fasta format (XX_final_assembly.fasta).

Next, the assembled genome is evaluated by quality assessment tool, QUAST (3, 4). The result showed that the assembled genome had 83 contigs, with estimated genome length of 1,936,708 bp and 32.74% of average GC content. The shortest sequence length at 50% of the genome, is 73,083 bp (N50). The L50 count, defined as the smallest number of contigs whose length sum produces N50, is 9. Summary of the assembly details are provided in **Table 3**.

Table 3. Assembly details

Dataset information	
Contigs	83
GC content	32.74
Largest contig	689,168
Contig L50	9
Contig N50	73,083
Genome length	1,936,708 bp
Chromosomes	0

File:

XX_final_assembly.fasta

QUAST_report.pdf

QUAST_report.html

Identification of prokaryotic genome contaminations

The ContEst16S is used to identify potential contaminations of prokaryotic genomes using 16S rRNA gene sequence from genome assemblies (5).

Suggestions by Porcinotec;

ConEst16S result shows that this project has a 16S rRNA gene fragment, so it cannot be checked for possible contamination.

4. Genome annotation

L. salivarius strain Porcinocin genome is annotated using rapid prokaryotic genome annotation (Prokka) by minimizing contig size to 200 bp (6). This genome is in the kingdom of bacteria, which is annotated using genetic code 11. In **Table 4**, this genome contains 1,855 protein coding sequences (CDS), 52 transfer RNA genes (tRNA), and, 4 ribosomal RNA genes (rRNA).

Table 4. Summary of annotated genome features

CDS	1,855
tRNA	52
rRNA	4

File;

XX_annotation.gff

XX_annotation.gbk

XX_annotation.feature_dna.fasta

XX_annotation.feature_protein.fasta

XX_annotation_summary.txt

XX_annotation_feature.txt

XX_annotation_feature.tsv

5. Circular graphical

A circular graphical display of bacterial DNA features was done using Cgview comparison tool (7, 8). Tracks from the outermost are as follows: CDS on the forward strand, CDS on the reverse strand, GC content, the contigs, and GC skew. The colors of the CDS on the forward and reverse strands are generated by the database of Clusters of Orthologous Groups of proteins (COGs) (9). GC content is shown in black ring and GC skews are shown in green-pink rings, respectively (**Figure 1**).

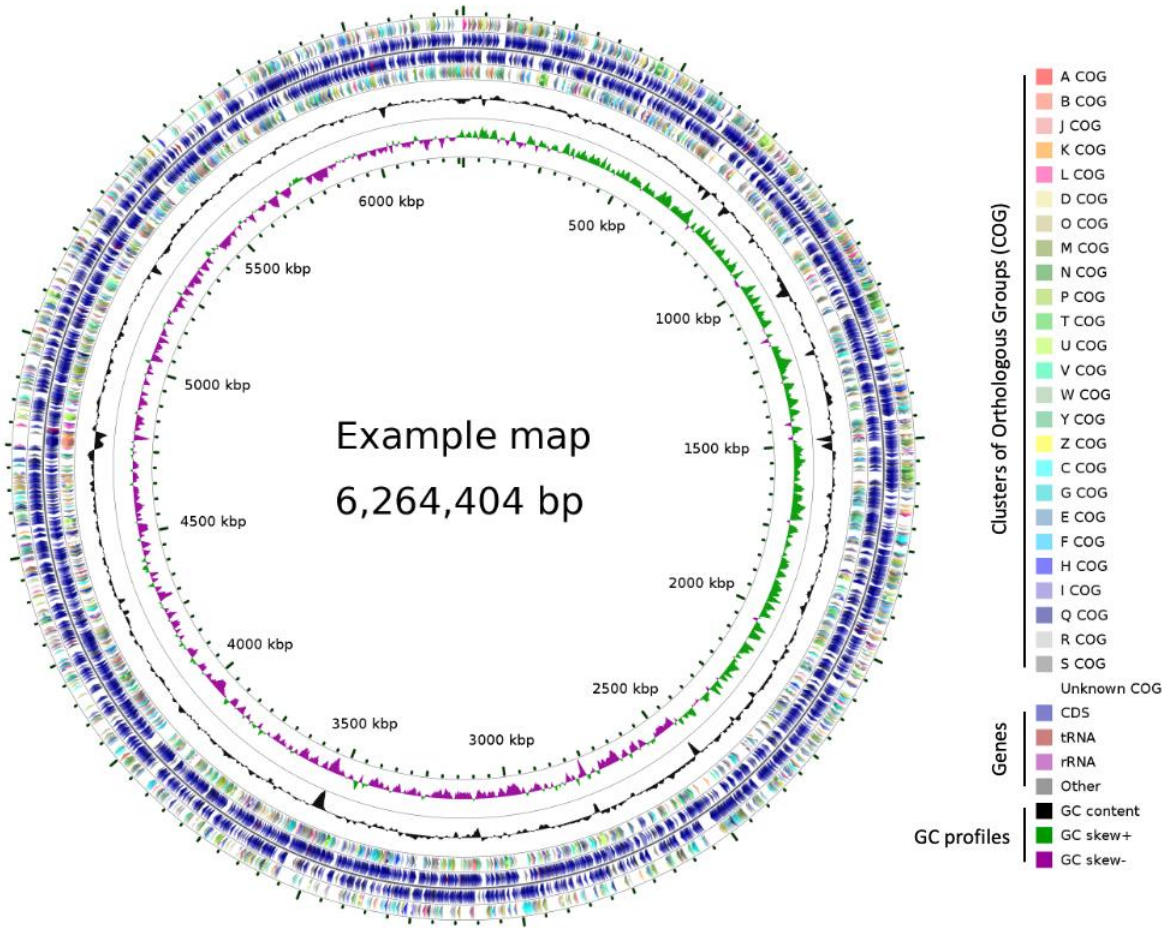


Figure 1. Circular representation of XX bacteria. The circular map is generated with the Cgview comparison tool.

File:

Circular_map.png

XX_annotation.gbk_cds_cogs.gff

6. Species identification and Phylogenetic tree analysis

JSpeciesWS is software tool for average nucleotide identity (ANI) calculation based on a BLAST algorithm and Tetra Correlation Search (TCS) function with default parameters (10). The ANI is a similarity index between a given pair of genomes that can be applicable to prokaryotic organisms independently of their G+C content, and a cutoff score of >95% indicates that they belong to the same species (11, 12).

Table 5. Comparison of ANI between each genome of the 11 selected *L. salivarius* strains

	***L. salivarius*_strain_porcinocin**	*L. salivarius* GJ-24	*L. salivarius* SMXD51	*L. salivarius* ATCC 11741	*L. salivarius* ATCC 11741 DSM 20555	*L. salivarius* UCC118	*L. salivarius* ACS-116-V-Col5a	*L. salivarius* CECT 5713	*L. salivarius* cp400	*L. salivarius* NIAS840	*Ligilactobacillus salivarius* CICC 23174
***L. salivarius*_strain_porcinocin**	*	96.98	97.12	97.05	97.09	96.4	96.64	96.32	96.74	97.22	97.26
L. salivarius GJ-24	97.14	*	97.62	97.12	97.13	96.75	96.74	96.57	96.94	97.74	97.26
L. salivarius SMXD51	97.47	97.7	*	97.38	97.37	96.94	96.89	96.77	97.23	97.78	97.57
L. salivarius ATCC 11741	97.02	96.9	97.08	*	99.97	97.16	97.23	96.96	97.42	96.94	96.97
L. salivarius ATCC 11741 DSM 20555	97.02	96.81	97.08	100	*	97.19	97.3	96.96	97.32	97.01	96.94
L. salivarius UCC118	96.5	96.49	96.8	97.37	97.4	*	97.97	98.4	97.23	96.81	96.58
L. salivarius ACS-116-V-Col5a	96.45	96.57	96.88	97.37	97.39	97.83	*	97.52	97.11	96.68	96.47
L. salivarius CECT 5713	96.41	96.54	96.7	97.07	97.07	98.42	97.69	*	96.86	96.69	96.38
L. salivarius cp400	96.9	96.71	97.17	97.52	97.47	97.08	97.09	96.88	*	96.87	97.1
L. salivarius NIAS840	97.37	97.71	97.86	97.07	97.08	97.01	96.97	96.82	97.26	*	97.66
Ligilactobacillus salivarius CICC 23174	97.28	97	97.25	96.99	97	96.55	96.42	96.42	96.86	97.49	*

File:

ANIb.cvs

In silico genome-to-genome comparison for microbial species discrimination is performed using DNA-DNA hybridization (DDH) which is calculated by the Genome-to-Genome Distance Calculator 2.1 (GGDC), using formular 2 (13). In silico DDH methods are based on the comparison of completely sequenced genomes using BLAST to determine high-scoring segment pairs (HSPs) and maximally unique matches (MUMs) between genome sequences after cutting them into small 1000 bp-long pieces to emulate the DDH procedure (14). In **Table 6**, the DDH (%) result was generated using formular 2 as recommended. Moreover, the GGDC reports the difference in G+C content, which can also be reliably used for species delineation (see explanation below).

Explanation:

Distances are inferred using three distinct formulas from the set of HSPs and MUMs obtained by comparing each pair of genomes with the chosen software. These distances are transformed to values analogous to DDH using a generalized linear model (GLM) inferred from an empirical reference dataset comprising real DDH values and genome sequences. Model-based confidence intervals are specified in square brackets but can also be obtained via bootstrapping. Logistic regression (a special type of GLM) is used for reporting the probabilities that DDH is ≥ 70% and ≥ 79%. Percent G+C content cannot differ by > 1 within a single species but by ≤ 1 between distinct species.

Table 6. In silico DDH percentages

	DDH (%)	Distance	Prob. DDH ≥ 70%	G+C difference
*L. salivarius*_strain_porcinocin	-	-	-	-
L. salivarius GJ-24	78.4	0.025	89.3	0.270
L. salivarius SMXD51	81.4	0.022	91.6	0.200
L. salivarius ATCC 11741	76.7	0.027	87.7	0.190
L. salivarius ATCC 11741 DSM 20555	76.7	0.027	87.6	0.250
L. salivarius UCC118	75.4	0.029	86.2	0.200
L. salivarius ACS-116-V-Col5a	74.6	0.030	85.3	0.020
L. salivarius CECT 5713	74.1	0.031	84.7	0.190
L. salivarius cp400	77.7	0.026	88.6	0.060
L. salivarius NIAS840	80.5	0.023	91.0	0.280
Ligilactobacillus salivarius CICC 23174	80.1	0.023	90.6	0.100

File:

GGDC_results.cvs

High-resolution of phylogenetic tree is constructed using the Automated Multi-Locus Species Tree (autoMLST). It uses Multi-Locus Sequence Analysis (MLSA) method with automatic selection of reference genomes. The out-group organisms are based on one or more query genomes with ultrafast Bootstrap analysis (15).

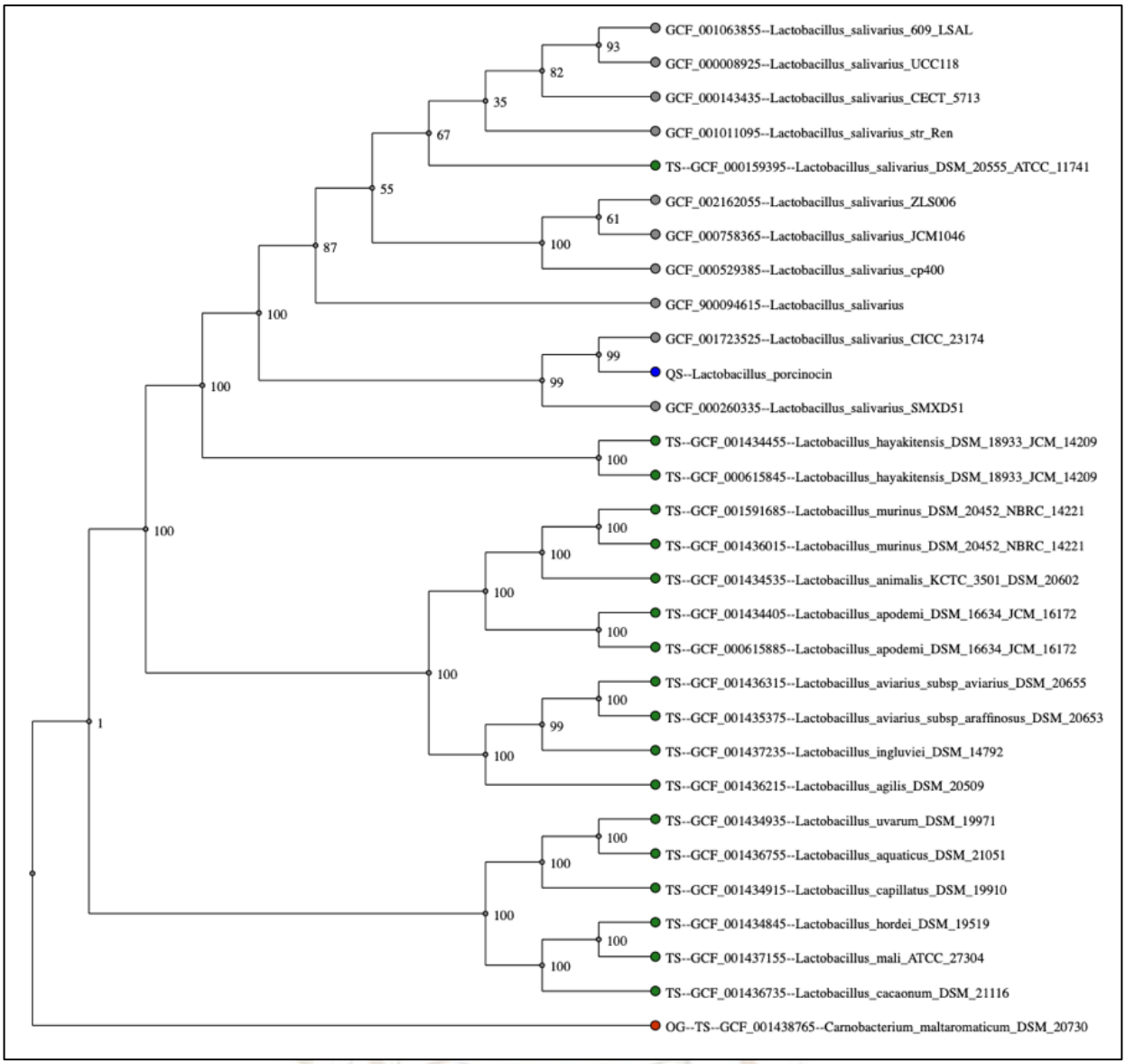


Figure 2. Phylogenetic tree of *L. salivarius* strain Porcinocin.

File:

Phylogenetic tree.svg

Phylogenetic tree.tree

Phylogenetic tree.png

Suggestions by Porcinotec;

In this study, the ANI values between the newly sequenced *L. salivarius* strain Porcinocin genome and the representative genomes of *Lactobacillus* spp. were calculated. As shown in Table 6, the ANI values between this genome and other reference strains were 96–98% which were considerably in threshold value of the boundary for species circumscription (Table 5). The DDH% values of *L. salivarius* strain Porcinocin against all reference genomes ranged from 84.65 to 91.58% (Table 5). The phylogenetic tree based on multi-locus alignment revealed that *L. salivarius* strain Porcinocin is closed to *Lactobacillus* genus and grouped with *L. salivarius* strain CICC23174 and *L. salivarius* strain SMXD51 (Figure 2). Therefore, the combination of ANI values, DDH values, and the phylogenetic tree demonstrated that *L. salivarius* strain Porcinocin belonged to genera of *Lactobacillus* which is closely related to *L. salivarius* strain CICC23174 and *L. salivarius* strain SMXD51.

7. Bacterial variant calling

Reference-based mapping for identifying single-nucleotide polymorphisms (SNPs) from bacterial sequencing data uses a known reference genome to guide this process, which is essential for monitoring outbreaks and predicting phenotypes, such as antimicrobial resistance. Snippy finds SNPs between a haploid reference genome and your NGS sequence reads. It will find both substitutions (snps) and insertions/deletions (indels) (16). Larger structural variation such as inversions, duplications and large deletions are not typically covered by this method.

The total variant of the query is 5,215, including 267 of deletion (DEL), 357 of insertion (INS), 928 of multiple nucleotide polymorphism (MNP), 2,895 of single nucleotide polymorphism (SNP), and 678 of combination of SNP/MNP (COMPLEX). Summary of the assembly details are provided in **Table 7**.

Table 7. Summary of bacterial variant calling

Software	**snippy v3.2**
Reference accession number	NZ_CP034551.1
Reference genome size	1,853,059
Variant-COMPLEX	678
Variant-DEL	267
Variant-INS	357
Variant-MNP	928
Variant-SNP	2,895
Variant Total	5,125

Example of Snippy result is shown in **Table 8**. The Snippy result provides the identification of positions where the sequenced sample is different from the reference sequence. It also annotates and predicts the effects of variants on genes (such as amino acid changes).

Table 8. Variant calling using Snippy

Chrom	Pos	Type	Ref	Alt	Evidence	Ftype	Strand	Nt_Pos	Aa_Pos	Effect	Locus_Tag	Gene	Product
LP_XX	9	snp	T	C	C:65 T:0	CDS	+	9/1341	3/446	synonymous_variant c.9T>C p.Asn3Asn	EJ379_RS00005	*dnaA*	DnaA
LP_XX	292	snp	C	T	T:648 C:0	CDS	+	292/1341	98/446	missense_variant c.292C>T p.Pro98Ser	EJ379_RS00005	*dnaA*	DnaA
LP_XX	1556	snp	G	A	A:959 G:0	CDS	+	37/1146	13/381	missense_variant c.37G>A p.Gly13Ser	EJ379_RS00010	*dnaN*	DNA polymerase III subunit beta
LP_XX	1684	complex	GTTT	TTTC	TTTC:809 GTTT:0	CDS	+	165/1146	55/381	synonymous_variant c.165_168delGTTTinsTTTC p.57	EJ379_RS00010	*dnaN*	DNA polymerase III subunit beta
LP_XX	4141	complex	AATAAC	CGTAAT	CGTAAT:938 AATAAC:1	CDS	+	1124/1128	375/375	stop_retained_variant&splice_region_variant c.1124_*1delAATAACinsCGTAAT p.Glu375Ala	EJ379_RS00020	*recF*	DNA replication/repair protein RecF
LP_XX	6824	snp	G	A	A:771 G:0	CDS	+	630/2472	210/823	synonymous_variant c.630G>A p.Leu210Leu	EJ379_RS00030	*gyrA*	DNA gyrase subunit A
LP_XX	6830	complex	ATAT	GTGC	GTGC:693 ATAT:1	CDS	+	636/2472	212/823	missense_variant c.636_639delATATinsGTGC p.Tyr213Cys	EJ379_RS00030	*gyrA*	DNA gyrase subunit A
LP_XX	2225828	snp	A	T	T:245 A:0	CDS	+	271/1146	91/381	missense_variant c.271A>T p.Ile91Leu	EJ379_RS11680	MFS transporter	MFS transporter
LP_XX	2225833	mnp	AT	GA	GA:230 AT:0	CDS	+	276/1146	92/381	missense_variant c.276_277delATinsGA p.Ser93Thr	EJ379_RS11680	MFS transporter	MFS transporter
LP_XX	2867904	snp	G	A	A:176 G:0	CDS	-	800/849	267/282	missense_variant c.800C>T p.Ala267Val	EJ379_RS14910	ABC transporter permease	ABC transporter permease

* snp; single nucleotide polymorphism, mnp; multiple nucleotide polymorphism, ins; insertion, del; deletion, complex; combination of snp/mnp

File:

Snippy.xlsx

Snps_summary.tubular

Specialist tools

8. Functional characterization of genome

The functional analysis is a method to identify genes or proteins that are presented in genome. BlastKOALA V2.2 is an automatic annotation server for genome sequence, which performs KO (KEGG Orthology) assignments to characterize individual gene functions and reconstruct Kyoto Encyclopedia of Genes and Genomes, KEGG pathways, BRITE hierarchies and KEGG modules to infer high-level functions of the organism using KOALA algorithm (17). The example of BlastKOALA functional category is shown in **Figure 3**. The 1,526 entries have been identified and characterized into functional processing pathways including cellular metabolism, genetic information processing, environmental information processing cellular processes, and human diseases.

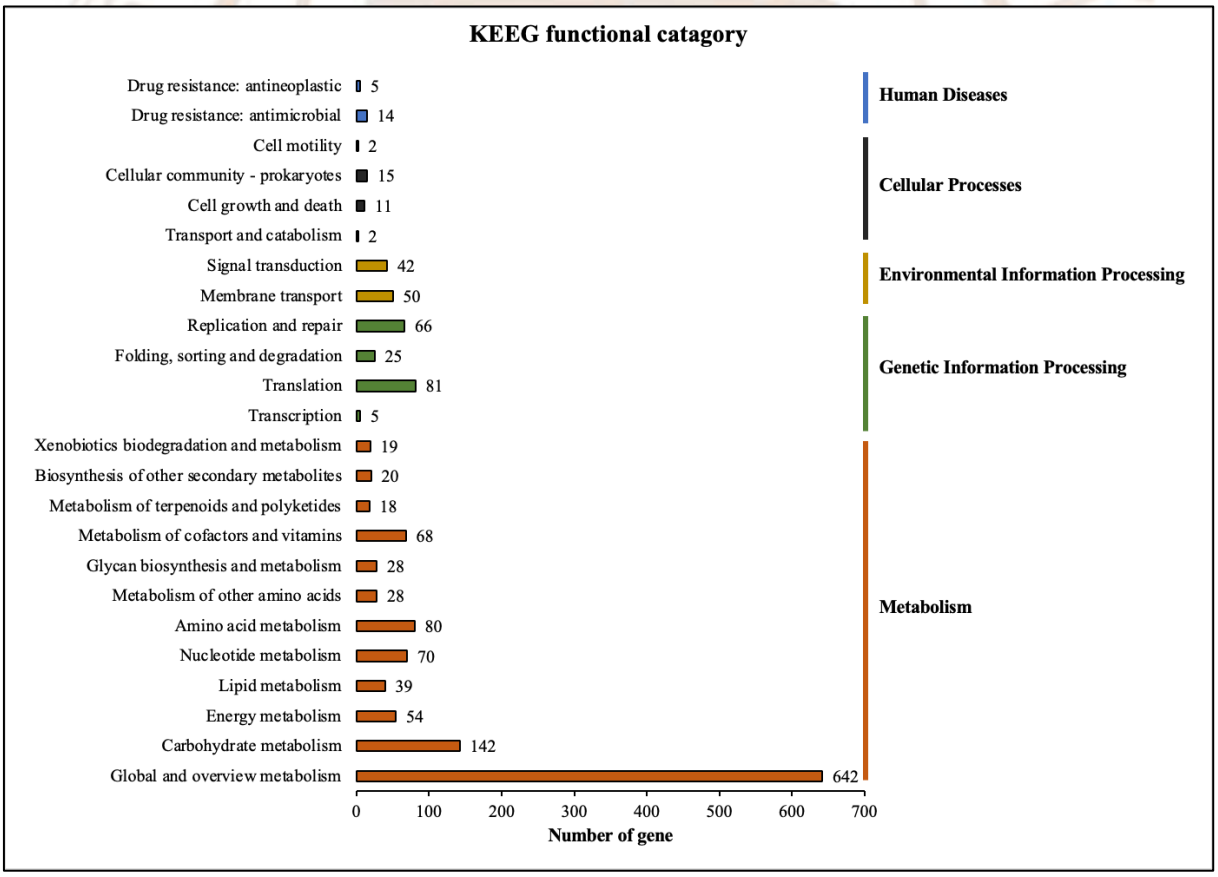


Figure 3. Functional characterization report of *L. Salivarius* strain Porcinocin genome.

File;

KEGG functional category.png

Functional analysis data; KO_definition.txt

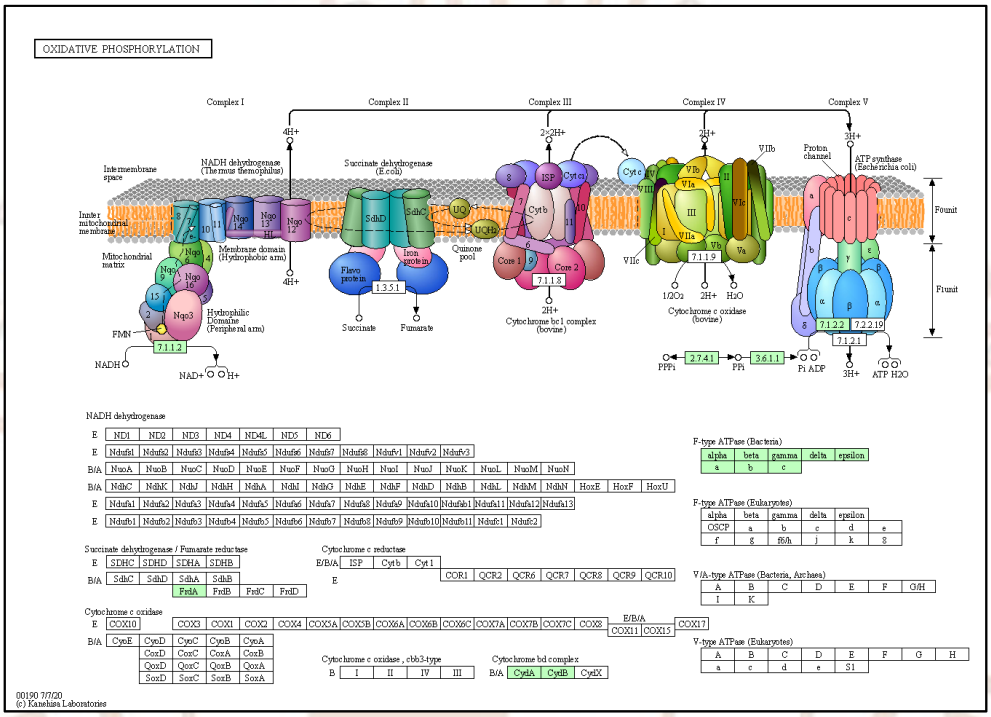
Note:

The KO functional analysis file (KO_definition.txt) will be provided for further visualization of KEGG pathway;

- Go to the KEGG Mapper tool link: https://www.genome.jp/kegg/tool/map_pathway.html

- Upload KO_definition.text

- KEGG will then give you a similar listing to the one you had in your initial result files, with the pathways listed and the number of hits per pathway. If you click a pathway, you can get:



Example of KEGG pathway. The green box is a subset of genes found in your input.

9. Specialty genes and Antimicrobial resistance gene analysis

WGS-based antimicrobial resistance analysis provides *in silico* antibiograms which assigns to each AMR gene functional annotation and specific antibiotic resistance. The number of annotated genes in this genome is homologous to known antibiotic resistance genes on The Comprehensive Antibiotic Resistance Database with default parameters (18, 19) and/or Resfinder 4.0 with 80% minimum DNA identity and DNA coverages (20) (**Table 9**).

The bacterial virulence factors are predicted against The virulence factor database (VFDB) with default parameters (21). The output data from the prediction of virulence genes are shown in (**Table 10**), including class and sub-class of virulence factor, related genes, and orf prediction of the input genome.

Table 9. Specialty antimicrobial resistance gene

Seq	**S**tart	End	Strand	Gene	% Coverage	% Identity	Database	Product	Resistance
1	170735	170778	+	aac(6')-laa_1	100.00	100.00	Resfinder	Acc(6')-laa	Amikacin Tobramycin
33	9430	11355	+	tet(M)_13	100.00	98.13	Resfinder	Tet(M)	Doxycycline Tetracycline Minocycline
35	3653	4381	+	erm©_12	98.18	99.45	Resfinder	Erm©	Erythromycin Lincomycin Clindamycin Quinupristin Pristinamycin Virginiamycin
44	1755	2240	-	lnu(a)_1	100.00	98.97	Resfinder	Lnu(A)	Lincomycin

File;

antibiotic resistance gene.xlsx

Table 10. Specialty virulence factor genes

Virulence factor class	Virulence factors	Related gene	Prediction
Secretion system	Type VII secretion system	*esxA*	orf00794
Toxin	Non-hemolytic enterotoxin (Nhe)	*nheC*	orf04578
	Cytolysin	*cylR2*	orf05041
Magnesium uptake	Mg2+ transport	*mgtB*	orf00391
Regulation	CheA/CheY	*cheA*	orf02477
	LisR/LisK	*lisR*	orf04897
Immune evasion	Polysaccharide capsule	Undetermined	orf01376; orf01386; orf01387; orf01388; orf01390; orf01391; orf01392; orf01393
Adherence	Hemorrhagic *E. coli* pilus (HCP)	*hcpA*	orf03267
		hcpB	orf03266
		hcpC	orf03265

File;

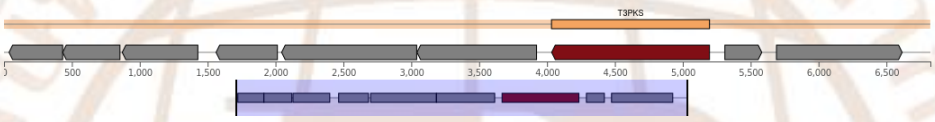
Virulence_factor.xlsx

10. Secondary metabolite prediction

The rapid identification, annotation, and analysis of secondary metabolite biosynthesis genome mining in bacterial genome are predicted using antibiotics & Secondary Metabolite Analysis Shell, antiSMASH, version 5.0 (22). antiSMASH is the most widely used tool for identifying and analyzing biosynthetic gene clusters (BGCs) in bacterial and fungal genome sequences.

For example, the genome of *L. Salivarius* strain Porcinocin was submitted to antiSMASH 5.0 and the results illustrated a type III polyketide synthases (T3PKS) gene cluster (**Figure 4A**). In which, this gene cluster shows the similarity around 18% to T3PKS from other *L. Salivarius* stains (**Figure 4B**).

A



B

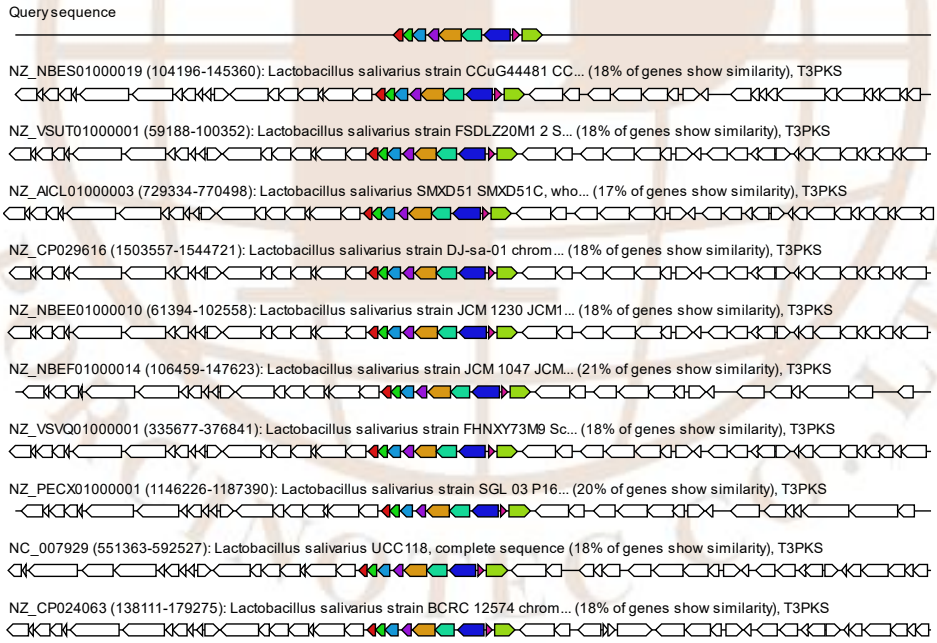


Figure 4. Graphical overview of the location of the identified regions on the chromosome. (A) The BGCs organization is displayed. (B) The ClusterBlast of BGCs are shown.

File;

AntiSMASH.html

11. Phage sequence identification

Phage search tool (PHAST) is an integrated search and annotation tool designed to rapidly and accurately identify, annotate and graphically display prophage sequences within bacterial genomes or plasmids (23). The PHAST results in different graphical mappings are shown in Figure 5.

Figure 5. A screenshot montage of some of PHAST's different graphical and tabular views including its linear and circular genome renderings as well as PHAST's corresponding prophage annotation (23).

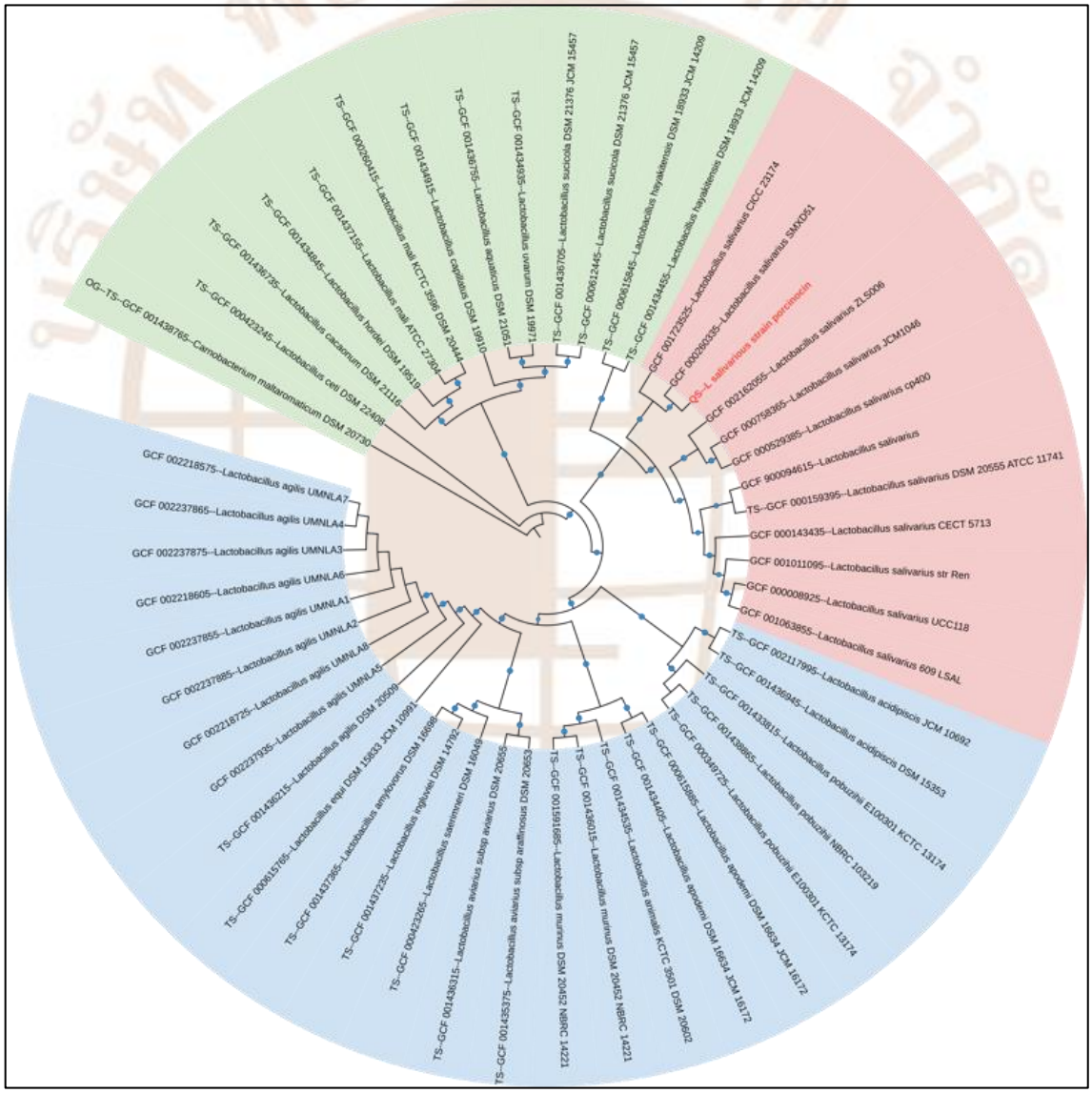
File:

Summary result.text

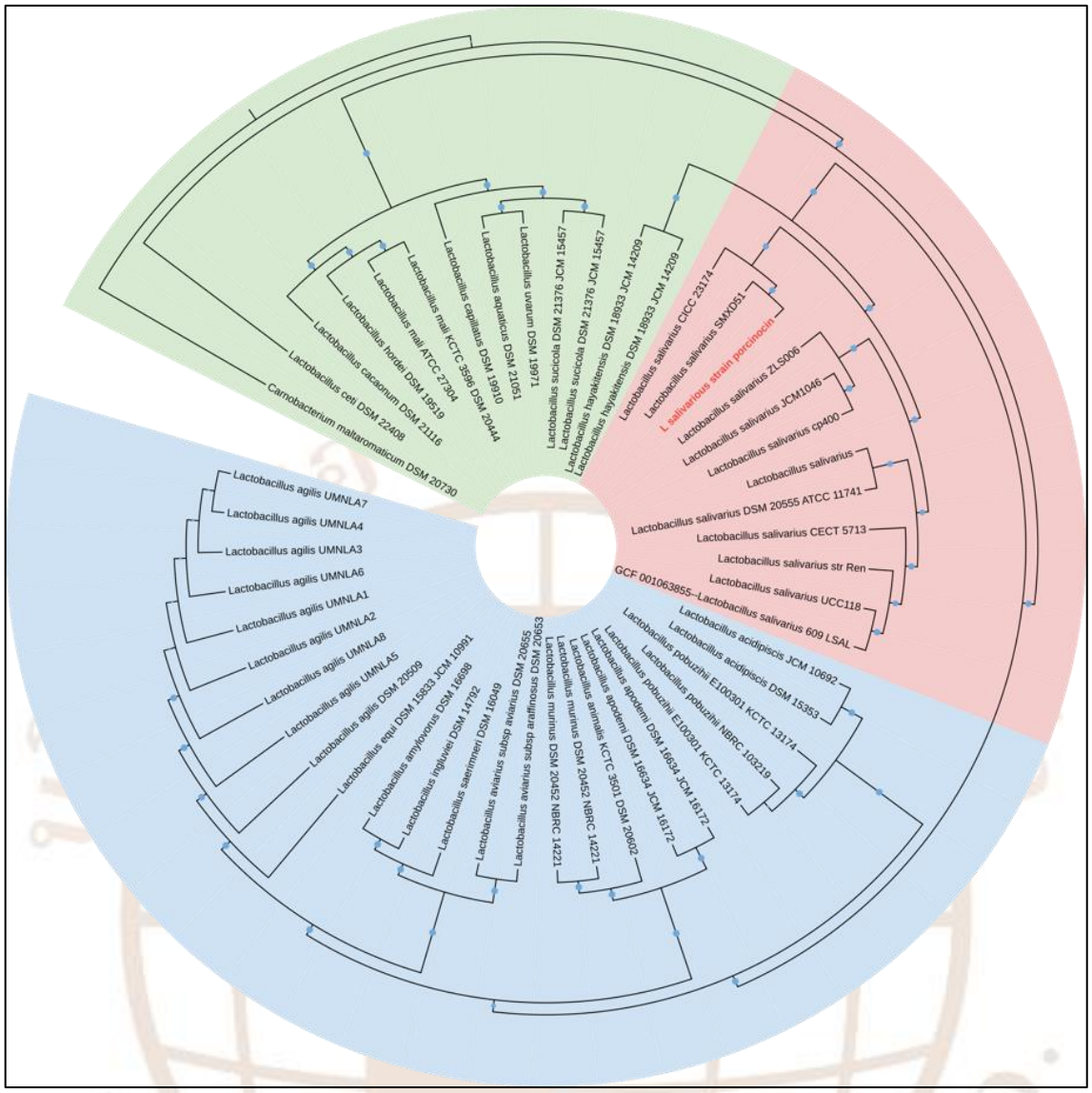
Detailed file.text

12. Phylogeny analysis

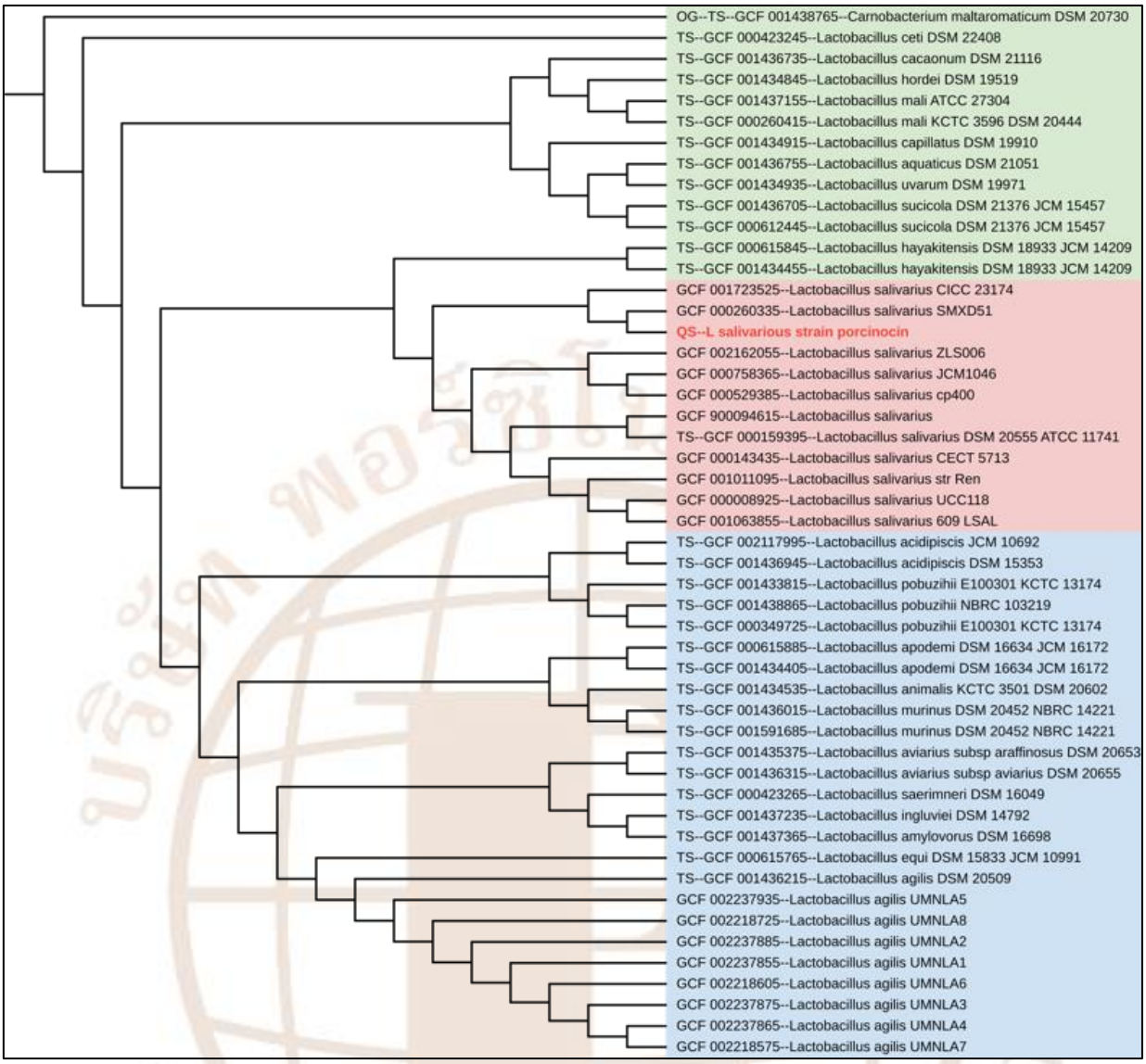
A phylogenetic tree is a branching simply diagram that represents the evolutionary relationships among species, organisms, or genes from a common ancestor (24). Phylogenetic trees are wildly used in a variety of biological and other scientific study. Here, we provide services for the display, manipulation, and annotation of phylogenetic tree using an Interactive Tree Of Life (iTOL) v5 (25). The phylogenetic analysis data can be visualized in various display modes including unrooted, circular, and regular phylograms. For example,



Example of tree of circular phylogram with adjustable colors and levels between various clades and with bootstrap values.



Example of tree of inverted circular phylogram with adjustable colors and levels between various clades.



Example of tree of rectangular phylogram with adjustable colors and levels between various clades.

File:

Phylogenetic tree.png

Phylogenetic tree.tree

13. Bacterial comparative genomics

Graphical genome maps

Comparative genomics is a comparison of biological information derived from WGS. Whole gene sets are compared to elucidate the common and different genomic features among two or more target organisms. Cgview Comparison Tool (CCT) generates maps displaying the result of sequence similarity comparisons between a bacterial genome of interest and other genomes (8). CCT generates several maps automatically, differing in terms of size and level of detail, as well as in terms of how the BLAST comparisons are done (at the nucleotide level or at the level of translated coding sequences). The maps depicting translated coding sequence comparisons also, by default, display COG (Cluster of Orthologous Groups) classifications, generated through the use of a COG sequence database.

For example, the map comparing *Pseudomonas aeruginosa* PAO1 with 11 additional strains of *P. aeruginosa* genome sequences are shown in **Figure 6**.

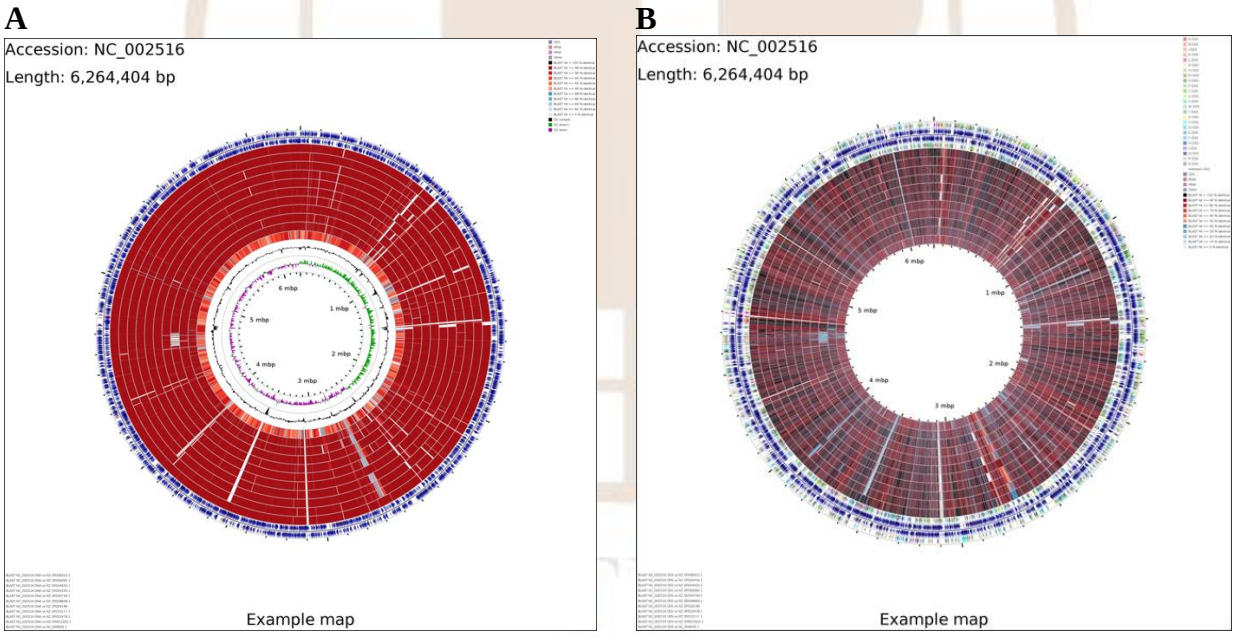


Figure 6. Circular graphical of the genome of *P. aeruginosa* PAO1 and 11 additional *P. aeruginosa* genome sequences generated using Cgview Comparison Tool. (A) BLAST comparing 11 complete genomes against *P. aeruginosa* PAO1 genome ordered from outer to inner ring following forward and reverse sequence features respectively. The remaining seven rings show the regions of sequence similarity detected by BLAST comparisons conducted between

nucleotide sequences from the PAO1 genome and 11 other *Pseudomonas* genomes. (B) Circles (from outside) represent the followings: 1. COG functional categories for forward coding sequence; 2. Forward sequence features; 3. Reverse sequence features; 4. COG functional categories for reverse coding sequence; 5. GC content; 6. GC skew.

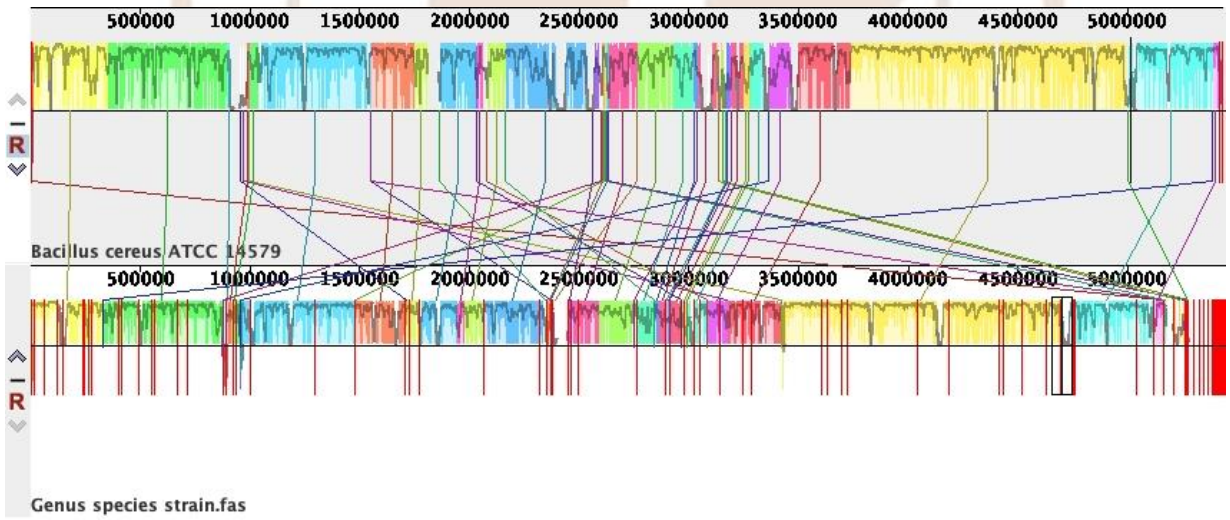
File;

Circular map.png

Multiple genome alignment using Mauve

Multiple alignment of conserved genomic sequence with rearrangements provides a basis for research into comparative genomics and the study of evolutionary dynamics using the Progressive Mauve algorithm with default parameters.

For example, the progressive Mauve is used to analyze virus genomes. ASF strain Benin 97/1 is the reference for alignments and comparisons for the two other strains (OURT 88/3 and E75) (**Figure 7**).



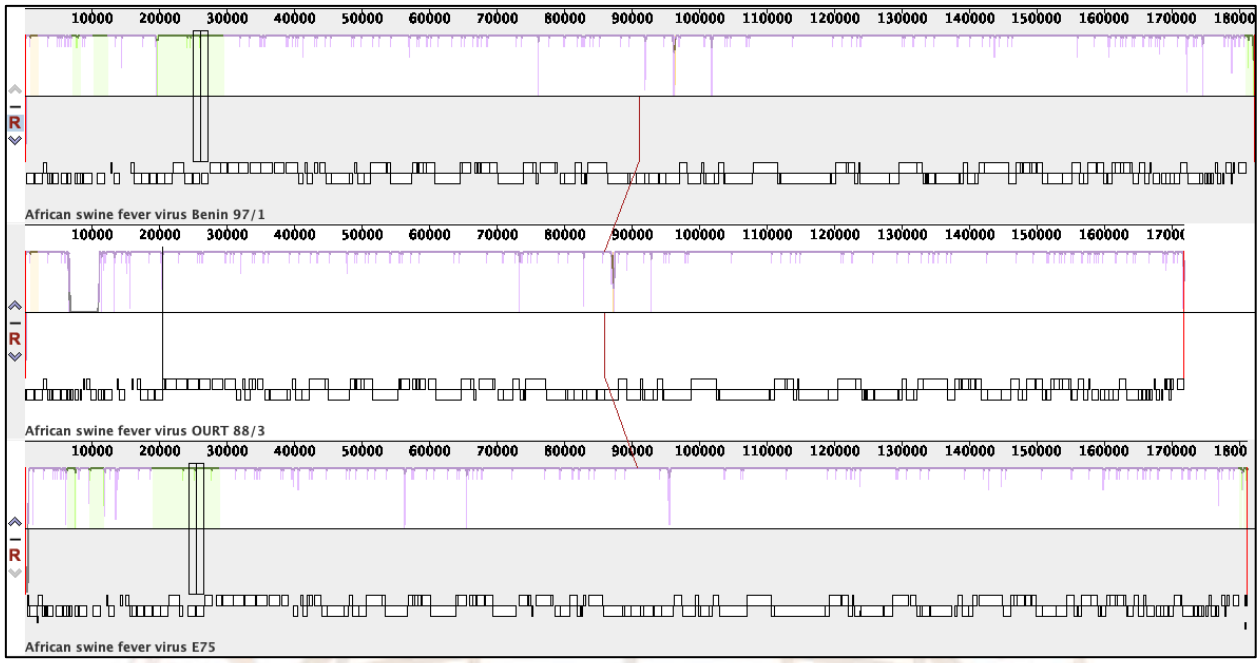


Figure 7. Multiple genome alignment using Mauve software comparing the ASF virus genomes. Boxes with identical colors represent local colinear blocks (LCB), indicating homologous DNA regions shared by two or more chromosomes without sequence rearrangements. LCBs indicated below the horizontal black line represent reverse complements of the reference LCB.

Genome browser and annotation tool using Artemis

Artemis is a widely used tool for a genome browser and annotation tool that allows visualization of sequence features, next generation data and the results of analyses within the context of the sequence (26).

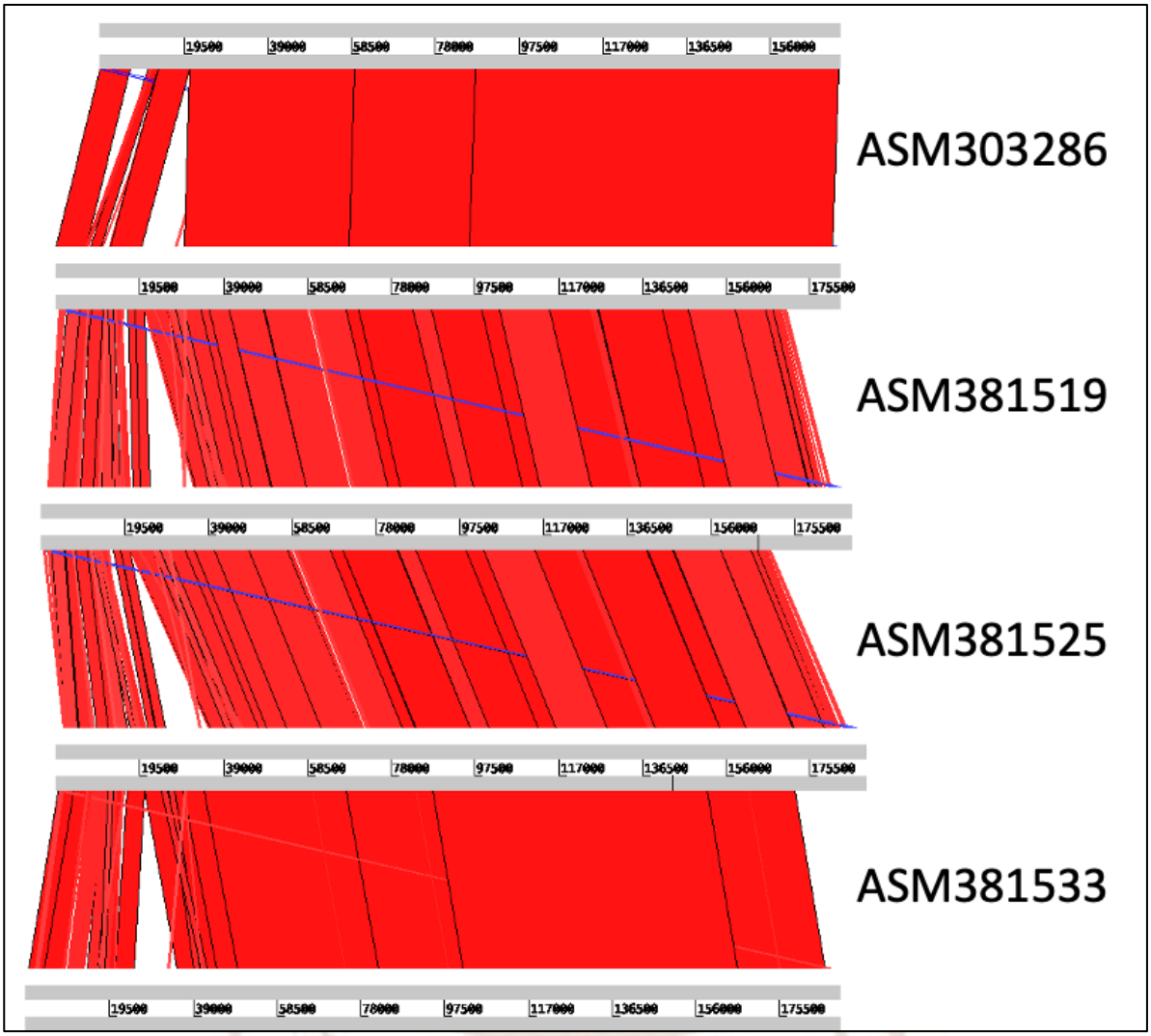


Figure 8. BLASTN genome alignments between ASM303288 genome against four other strains displayed using Artemis comparison tool (ACT). Genome sequences were aligned from the predicted KP86R and visualized in ACT with a cut-off set to blast scores >500. Red and blue bars indicate regions of similarity in the same orientation (red) and inverted (blue).

File:

Genome-to-genome alignment.crunch

Genome alignment.png

14. Material and methods

Genomic DNA extraction

Genomic DNA for prokaryotes was isolated using GF-1 Bacterial DNA Extraction Kit (Vivantis, Malaysia) according to the manufacturer's protocols. Briefly, bacteria cell pellet was extracted. The quality of the extracted DNA was determined via DeNovix QFX Fluorometer.

Whole genome library preparation and sequencing

The library preparation of genomic DNA was performed using the Qiagen QIAseq FX DNA Library kit (Qiagen, Hilden, Germany). The DNA fragments were labeled with different sequencing adaptors (Qiagen, Hilden, Germany). The quality and quantity of DNA libraries were evaluated using DeNovix QFX Fluorometer and QIAxcel Advanced (Qiagen, Hilden, Germany), respectively. DNA libraries were sequenced using an illumina Miseq500 platform (Illumina, San Diego, CA, USA).

15. References

1. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2015.

2. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.

3. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34(13):i142-i50.

4. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.

5. Lee I, Chalita M, Ha SM, Na SI, Yoon SH, Chun J. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. Int J Syst Evol Microbiol. 2017;67(6):2053-7.

6. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.

7. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. Bioinformatics. 2005;21(4):537-9.

8. Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using the CGView family of tools. Brief Bioinform. 2019;20(4):1576-82.

9. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 2021;49(D1):D274-D81.

10. Richter M, Rossello-Mora R, Oliver Glockner F, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 2016;32(6):929-31.

11. Richter M, Rossello-Mora R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106(45):19126-31.

12. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57(Pt 1):81-91.

13. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60.

14. Auch AF, von Jan M, Klenk HP, Goker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. 2010;2(1):117-34.

15. Alanjary M, Steinke K, Ziemert N. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. Nucleic Acids Res. 2019;47(W1):W276-W82.

16. Kwong JC, Mercoulia K, Tomita T, Easton M, Li HY, Bulach DM, et al. Prospective Whole-Genome Sequencing Enhances National Surveillance of *Listeria monocytogenes*. J Clin Microbiol. 2016;54(2):333-42.

17. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428(4):726-31.

18. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48(D1):D517-D25.

19. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother. 2013;57(7):3348-57.

20. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75(12):3491-500.

21. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res. 2016;44(D1):D694-7.

22. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47(W1):W81-W7.

23. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39(Web Server issue):W347-52.

24. McLennan DA. How to Read a Phylogenetic Tree. Evolution: Education and Outreach. 2010;3(4):506-19.

25. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021.

26. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 2012;28(4):464-9.