



Data Article

Draft genome sequencing data of multidrug-resistant *Staphylococcus haemolyticus* from Bangladeshi hospitals

Jarin Tabassum, Afia Anjum, Sohidul Islam, Abdul Khaleque, Ishrat Jabeen, Sabbir R. Shuvo*

Department of Biochemistry & Microbiology, North South University, Dhaka, Bangladesh

ARTICLE INFO

Article history:

Received 24 July 2024

Revised 8 August 2024

Accepted 2 September 2024

Available online 7 September 2024

Dataset link: [SAMN35731443 \(Original data\)](#)

Dataset link: [SAMN35731649 \(Original data\)](#)

Keywords:

S. haemolyticus

WGS

Antibiotic resistance

Phylogenomic analysis

ABSTRACT

Here, we report the draft genome sequence data of two multidrug-resistant (MDR) *Staphylococcus haemolyticus* strains, SAC2 and SAC7, isolated from clinical samples from Dhaka, Bangladesh. The sequence raw read files were generated using Ion Torrent Sequencing Technology using the genomic DNA from the pure culture of the strains. These two Bangladeshi *S. haemolyticus* strains had an average genome size of 2.49 million base pairs with a GC content of 32.6 % and an average of 1783 coding sequences. We conducted genomic studies using bioinformatics tools focusing on resistance genes, virulence factors, and toxin-antitoxin systems. A phylogenomic study with *S. haemolyticus* strains isolated worldwide revealed that these two Bangladeshi strains are in different nodes but clustered together. The data can be used as a starting point for understanding the genomic content, epidemiology, and evolution of *S. haemolyticus* in Bangladesh. The genome sequence data of SAC2 and SAC7 strains have been deposited in the NCBI database under BioSample accession numbers SAMN35731443 and SAMN35731649, respectively.

© 2024 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

* Corresponding author.

E-mail address: sabbir.shuvo@northsouth.edu (S.R. Shuvo).

Specifications Table

Subject	Biological Sciences
Specific subject area	Omics: Genomics
Type of data	Raw, table, figure, filtered, analyzed
Data collection	The strains were isolated from clinical samples from Bangladesh, and the genomic DNA was extracted. The raw read files were generated using Ion Torrent Sequencing Technology, followed by assembly and annotation through services provided by BV-BRC. A genomic map was generated by Proksee. The resistome was predicted using the RGI tool offered by the CARD database. The Victors and VFDB databases were used to predict the virulome. TA systems were predicted using TASmania. MLST-2.0 web server was used to determine the MLST of the isolates. TYGS and CSIPhylogeny v1.4 were used to generate phylogenomic trees based on whole genome and SNP, respectively, and iTOL was used for their visualization.
Data source location	Institution: • Dhaka Medical College Hospital, (SAC2, SAMN35731443) • LABAID Diagnostic Mirpur (SAC7, SAMN35731649) City/Town/Region: Dhaka Country: Bangladesh Latitude: 23.8041° N, Longitude: 90.4152° E
Data accessibility	Repository name: NCBI BioSample Data identification number: SAMN35731443 Direct URL to data: https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN35731443 Data Identification number: SAMN35731649 Direct URL to data: https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN35731649
Related research article	None

1. Value of the Data

- This study represents the first WGS of *S. haemolyticus* isolated from Bangladesh. The data could serve in investigating genomic mechanisms for MDR and the epidemiology of the species, possibly aiding in developing better public health strategies, particularly through targeted antimicrobial stewardship efforts to combat infections.
- The genome sequences of *S. haemolyticus* strains SAC2 and SAC7 could provide information for future comparative inter-species and intra-species genomic research.
- The data would help to understand the genotype responsible for the increased virulence and phylogenomic studies of *S. haemolyticus* in Bangladesh.

2. Background

In recent years, *S. haemolyticus* has exhibited increased resistance to a wide range of drugs [1]. However, the factors linked to their spread, pathogenicity, and virulence are yet to undergo detailed characterization. There is inadequate information available about the frequency of *S. haemolyticus* infections and the genomic content of the MDR Bangladeshi *S. haemolyticus* [2]. The objective of the present study was the genomic characterization of the two MDR clinical isolates of *S. haemolyticus* from Dhaka, Bangladesh, assessing their antimicrobial resistance, virulence profile, toxin-antitoxin systems, and phylogenomic relationships.

3. Data Description

3.1. Antibiotic resistance profile of the Bangladeshi *S. haemolyticus* isolates

The two studied strains were confirmed to be MDR. Notably, SAC2 and SAC7 were resistant to 78.94 % and 63.15 %, respectively, of all antibiotics tested (Table 1).

Table 1Antibiotic sensitivity pattern of *S. haemolyticus* isolates.

Antibiotic class	Antibiotics	Strains	
		SAC2	SAC7
Penicillin	AMP	R	R
Penicillin + beta-lactamase inhibitors	AMC	I	S
	TZP	R	I
Cephalosporins	FOX	R	R
	CAZ	R	R
	CFM	R	R
	CRO	R	R
	CTX	R	R
	FEP	R	R
Monobactams	ATM	R	R
Carbapenems	MEM	R	I
Aminoglycosides	AMK	R	I
	GEN	R	S
Fluoroquinolones	CIP	R	R
	LEV	S	R
Macrolides	ERY	R	R
Tetracyclines	TET	S	S
	TGC	S	S

R = Resistant, S=Sensitive, I=Intermediate.

Table 2Genomic characteristics of Bangladeshi *S. haemolyticus* strains.

Attribute	SAC2	SAC7
Place of isolation	Dhaka, Bangladesh	Dhaka, Bangladesh
Isolation source	Wound	Blood
Sequencing coverage	56.34	49.97
Genome size (Mbp)	2.48	2.5
Number of contigs	97	135
^a N50	70,474	52,525
^b L50	12	16
GC (%)	32.58 %	32.61 %
Coding genes	1773	1793
tRNA	25	56
rRNA	4	3
MLST	40	42
Bio-sample ID/reference	SAMN35731443	SAMN35731649

^aN50 = Half of the genome assembly is contained in contigs equal to or larger than this value; ^bL50= smallest number of contigs (each with its length) in the genome assembly needed to cover approximately half of the total genome size.

3.2. Genomic features of the isolated *S. haemolyticus* strains

Two Bangladeshi strains exhibited an average genome size of 2.49 million base pairs (Mbp) with a GC content of 32.6 % and an average of 1783 coding sequences (CDS). To our knowledge, the multilocus sequence types (MLST) of the isolated strains had not been previously documented in Bangladesh (Table 2).

The local alignment of the Bangladeshi *S. haemolyticus* strains against the reference genome of *S. haemolyticus* S167 is visualized in Fig. 1.

3.3. Antibiotic resistance and virulence genes prediction of the *S. haemolyticus* strains

Using the Comprehensive Antibiotic Resistance Database (CARD), we found antibiotic resistance genes (ARG) in the studied strains and compared them with a reference strain S167 (Fig. 2A).

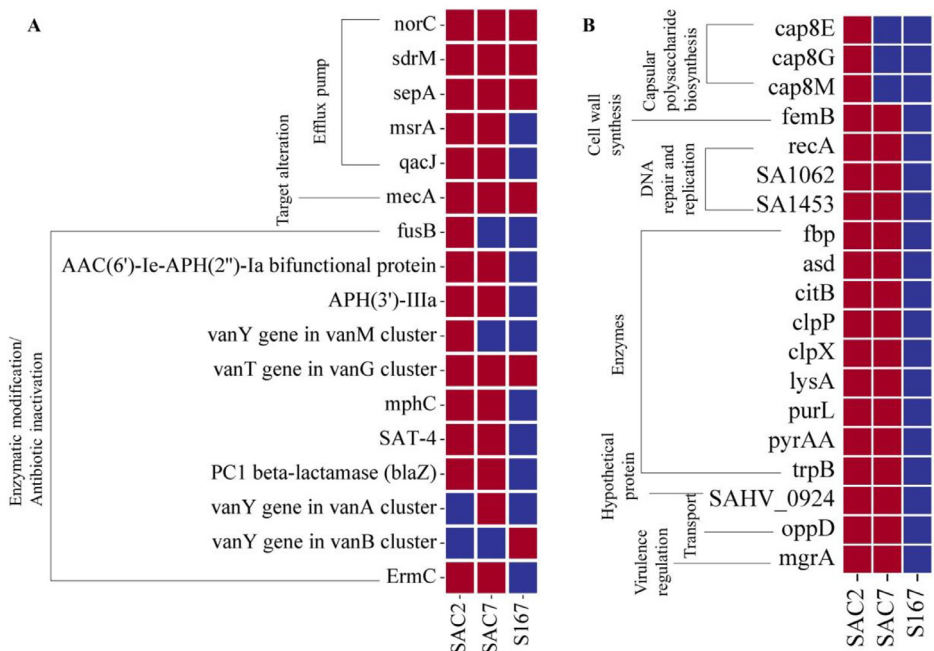


Fig. 2. Antibiotic resistance genes (A) and Virulence Factors (B) of the two isolated Bangladeshi *S. haemolyticus* strains and the reference strain *S. haemolyticus* S167. The red squares denote the presence of the genes, and the blue squares denote the absence of the genes listed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

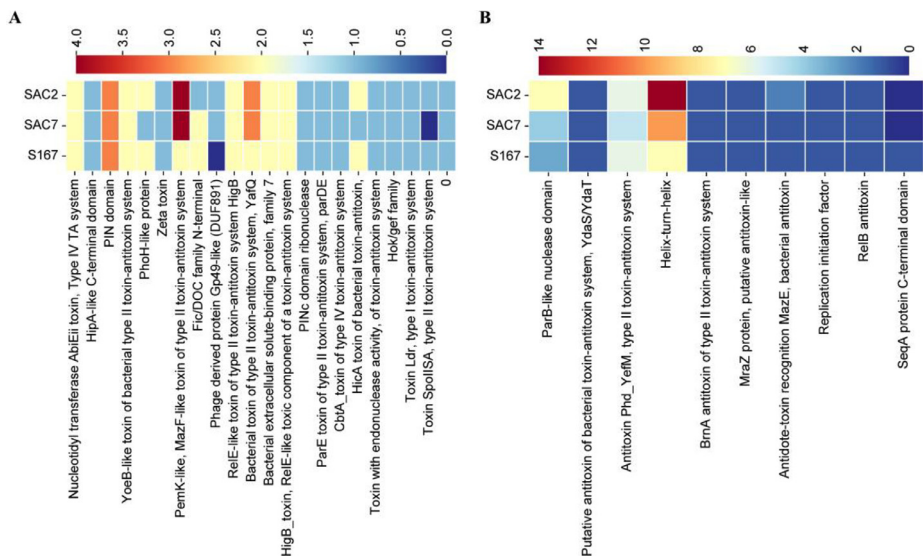


Fig. 3. Toxin-antitoxin system of the isolated strains and the reference strain *S. haemolyticus* S167. Varying colors represent different numbers of toxin genes (A), and antitoxin genes (B). The identified toxins and antitoxins are shown at the Y axis. The names of the strains are mentioned at X axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

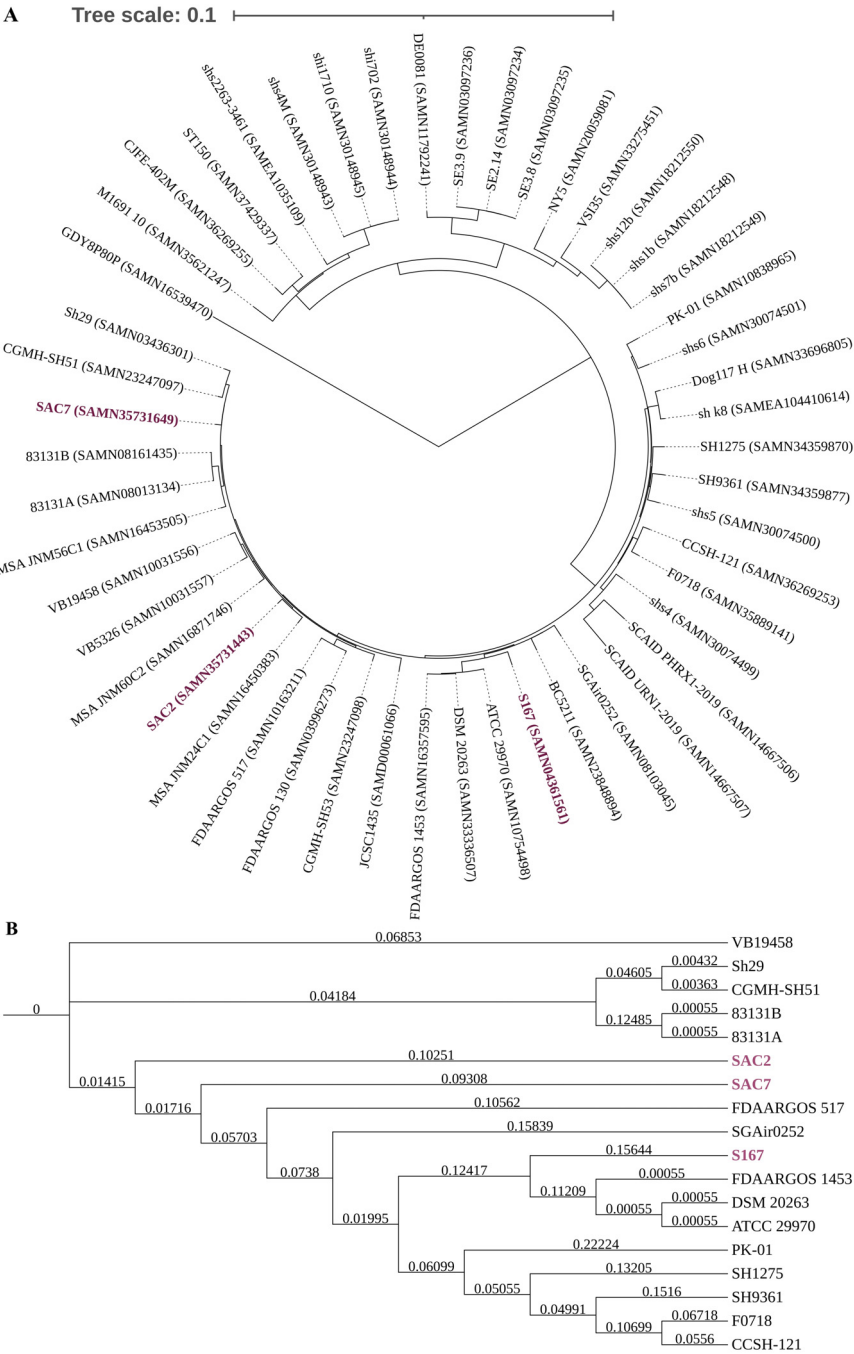


Fig. 4. Phylogenomic tree based on (A)WGS of 50 *S. haemolyticus* strains, (B) SNP of 18 *S. haemolyticus* strains including 2 Bangladeshi strains. In Figure A, the BioSample Numbers are given inside the bracket.

4. Experimental Design, Materials and Methods

4.1. Isolation and phenotypic characterization of *S. haemolyticus* strains

The *S. haemolyticus* strains from clinical samples were isolated on Mannitol Salt Agar (MSA; Condalab, Spain), followed by the subculture on MSA to observe their distinctive characteristics. The confirmed *S. haemolyticus* isolates were tested to identify their antimicrobial resistance pattern by using the Kirby-Bauer disk diffusion technique [3] on Mueller-Hinton agar (HiMedia, India) according to the Clinical and Laboratory Standards Institute (CLSI) (CLSI 2022) guideline (Table S1). The antibiotics tested were ampicillin (AMP 30 µg), amoxicillin with clavulanic acid (AMC 30 µg), tazobactam with piperacillin (TZP 110 µg), ceftazidime (CAZ 30 µg), cefixime (CFM 5 µg), ceftriaxone (CRO 30 µg), cefotaxime (CTX 30 µg), cefepime (FEP 5 µg), aztreonam (ATM 30 µg), meropenem (MEM 10 µg), amikacin (AMK 30 µg), gentamicin (GEN 10 µg), ciprofloxacin (CIP 5 µg), levofloxacin (LEV 5 µg), erythromycin (ERY 15 µg), tetracycline (TET 30 µg), tigecycline (TGC 15 µg), and colistin (COL 10 µg) (Bioanalyse, Turkey).

4.2. Genome assembly and annotation

WGS was performed on two MDR Bangladeshi *S. haemolyticus*. The genomic DNA was extracted using Wizard® Genomic DNA Purification Kit following the manufacturer's protocol (Promega, USA). The quantity and quality of the extracted DNA were assessed with a NanoDrop™ 2000 spectrophotometer (Thermo Scientific, USA). Raw sequence read files for *S. haemolyticus* strains were obtained using Ion Torrent sequencing technology on an Ion GeneStudio™ S5 System (ThermoFisher Scientific, USA) following the manufacturer's instructions at DNA Solution Ltd., Dhaka, Bangladesh. Each sample underwent multiple read generation, with quality control and adapter trimming executed using the integrated Torrent Suite™ Software version 5.10.0. The assembly of the sequences was performed using Unicycler version v0.4.8.0 [4], Quast version 5.0.2 [5], and Samtools version 1.11 [6]. The annotations were done by using the Rapid Annotation using Subsystem Technology (RAST) tool kit (RASTtk) [7]. The assembly and annotation services were provided by the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) [8,9]. Proksee was used to generate a circular map reflecting the local alignment of the strains (Fig. 1) [10]. Assembled draft genomes were stored at the National Center for Biotechnology Information (NCBI) (BioSample IDs: SAMN35731443; SAMN35731649).

4.3. Genomic characterization, prediction of antibiotic resistance genes, virulome, and toxin-antitoxin systems

Assembled draft genomes were used in the Resistance Gene Identifier (RGI) tool offered by the Comprehensive Antibiotic Resistance Database (CARD) to predict the resistomes of the strains. The presence of virulence factors for the strains was predicted using the Virulence Factor Database (VFDB) (<http://www.mgc.ac.cn/VFs/>) and the Victors Database. Multilocus sequence typing (MLST) of *S. haemolyticus* isolates was determined by the MLST-2.0 web server. MLST was performed by identifying different variants located in seven housekeeping genes, carbamate kinase(arcC), D-ribose ABC transporter substrate-binding protein (Ribose_ABC), cell surface elastin binding protein (SH1431), D-3-phosphoglycerate dehydrogenase (SH_1200), ribulose-phosphate 3-epimerase (cfxE), ferrochelatase (hemH), 3-isopropylmalate dehydrogenase (leuB). All the analysis was performed following the method described in [11,12]. Toxin-Antitoxin Systems were predicted using Toxin-Antitoxin systems mania (TASmania) [13]. Default parameters were applied for all analyses using the mentioned tools.

4.4. Comparative analysis

Average Nucleotide Identity (ANI) analysis was performed using Kostas Lab [14] (Fig. S1). A Single Nucleotide Polymorphism (SNP) tree was constructed from closely related 18 *S. haemolyticus* strains (with average nucleotide identity >99 %) isolated worldwide based on ANI (Fig. S1 and Table 2). Phylogenomic analysis was carried out in the Type Strain Genome Server (TYGS) (<https://tygs.dsmz.de>) for a whole genome-based taxonomic analysis [15]. Later, CSIPhylogeny version 1.4 was used for the generation of a phylogenomic tree against the reference genome *S. haemolyticus* strain S167 (Accession: SAMN04361561) based on SNP [16]. Both the phylogenomic trees were visualized by the Interactive Tree of Life (iTOL) [17]. The heatmaps were generated using Science and Research (SR) online Plot (www.bioinformatics.com.cn).

Limitations

Not applicable.

Ethics Statement

All procedures were approved by North South University Research Ethics Committee (IRB: 2022/OR-NSU/IRB/0703).

Data Availability

[SAMN35731443 \(Original data\)](#) (NCBI BioSample).
[SAMN35731649 \(Original data\)](#) (NCBI BioSample).

CRediT Author Statement

Jarin Tabassum: Writing – original draft, Visualization; **Afia Anjum:** Writing – original draft, Visualization; **Sohidul Islam:** Writing – review & editing; **Abdul Khaleque:** Writing – review & editing; **Ishrat Jabeen:** Writing – review & editing; **Sabbir R. Shuvo:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition.

Acknowledgments

We want to thank North South University (grant number NSU: CTGRC-21-SHLS-05 to SRS) for the financial support of this work. We would further like to thank Ms. Munmun Faria Iqbal, Department of Biochemistry & Microbiology, North South University, for assistance with the isolation of the strains and Dr. Laila Siddika, Department of Biochemistry & Microbiology, North South University, for helping with the initial study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.dib.2024.110918](https://doi.org/10.1016/j.dib.2024.110918).

References

- [1] K. Becker, C. Heilmann, G. Peters, Coagulase-negative staphylococci, *Clin. Microbiol. Rev.* 27 (2014) 870–926.
- [2] M.A. Yusuf, S. Islam, A. Shamsuzzaman, et al., Burden of infection caused by methicillin-resistant staphylococcus aureus in bangladesh: a systematic review, *Glob. Adv. Res. J. Microbiol.* 2 (2013) 213–223.
- [3] J.H. Jorgensen, J.D. Turnidge, *Susceptibility Test Methods: Dilution and Disk Diffusion Methods*, Man Clin Microbiol (2015) 1253–1273.
- [4] R.R. Wick, L.M. Judd, C.L. Gorrie, et al., Unicycler: resolving bacterial genome assemblies from short and long sequencing reads, *PLoS Comput. Biol.* 13 (2017) e1005595.
- [5] A. Gurevich, V. Saveliev, N. Vyahhi, et al., QUAST: quality assessment tool for genome assemblies, *Bioinformatics* 29 (2013) 1072–1075.
- [6] H. Li, B. Handsaker, A. Wysoker, et al., The Sequence Alignment/Map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079.
- [7] T. Brettin, J.J. Davis, T. Disz, et al., RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes, *Sci. Rep.* 5 (2015) 8365.
- [8] R.D. Olson, R. Assaf, T. Brettin, et al., Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR, *Nucleic Acids Res.* 51 (2023) D678–D689.
- [9] A.R. Wattam, J.J. Davis, R. Assaf, et al., Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center, *Nucleic Acids Res.* (2017) 45.
- [10] J.R. Grant, E. Enns, E. Marinier, et al., Proksee: in-depth characterization and visualization of bacterial genomes, *Nucleic Acids Res* 51 (2023) W484–W492.
- [11] N. Ullah, T. Raza, H.A. Dar, et al., Whole-genome sequencing of a new sequence type (ST5352) strain of community-acquired methicillin-resistant *Staphylococcus aureus* from a hospital in Pakistan, *J. Glob. Antimicrob. Resist.* 19 (2019) 161–163.
- [12] S. Sayers, L. Li, E. Ong, et al., Victors: a web-based knowledge base of virulence factors in human and animal pathogens, *Nucleic Acids Res.* 47 (2019) D693–D700.
- [13] H. Akarsu, P. Bordes, M. Mansour, et al., TASmania: a bacterial toxin-antitoxin systems database, *PLOS Comput. Biol.* 15 (2019) e1006946.
- [14] L.M. Rodriguez-R, K.T. Konstantinidis, The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes, *PeerJ Preprints* 4 (2016) e1900v1.
- [15] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, *Nat. Commun.* 10 (2019) 2182.
- [16] R.S. Kaas, P. Leekitcharoenphon, F.M. Aarestrup, et al., Solving the problem of comparing whole bacterial genomes across different sequencing platforms, *PLoS ONE* 9 (2014) e104984.
- [17] I. Letunic, P. Bork, Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool, *Nucleic Acids Res.* 2024 (2024) 1–5.