## Article Title

Draft Genome Sequence of *Sanguibacter* species strain 25GB23B1, Cultivated from Arctic Surface Water


## Authors

Kiana Mitchell, Grace Busch, Indu Sharma[#]

## Affiliations

Department of Biological Sciences, Hampton University, Hampton, Virginia, USA


## Running Title

Draft Genome Sequence of *Sanguibacter* species strain 25GB23B1


## Corresponding Author's Email Address

[#]Address correspondence to Indu Sharma, indu.sharma@hamptonu.edu



## Abstract

We report a draft genome sequence for *Sanguibacter* species strain 25GB23B1, isolated from arctic surface water off the coast of Alaska. The whole genome sequence will provide knowledge of the bacteria's relationship to its environment and possibly a new species of *Sanguibacter*.


## Announcement

*Sanguibacter* is a genus of gram-positive, coryneform bacteria under the family of *Cellulomonadaceae*. Members of this genus have been found in marine environments; specifically in coastal sediment from the Eastern China Sea (1) and in ice in Antarctica (2). The *Sanguibacter* strain of this study was isolated from arctic seawater accessed one meter below the water's surface through drilling holes with an ice auger near North Slope Borough, Alaska on May 22, 2023 at the coordinates 70° 27' 30.0" N, 148° 25' 06.2" W and named 25GB23B1.

An enrichment was performed with a medium modified from Unfried et al. (3)(342.2 mM NaCl, 14.8 mM $MgCl_2 \cdot 6H_2O$, 1.0 mM $CaCl_2 \cdot 2H_2O$, 6.71 mM KCl, 100 mg casamino acids, 100 mg peptone, 20 mM MOPS [pH 8.0], 10 mM $NH_4Cl$, 10 mM $KH_2PO_4$, 1 mM $Na_2SO_4$, 1 mM trace metals, 0.5% Laminarin). The culture was incubated for 7 days at 30°C on orbital shaker (100 rpm shaking). From this enrichment, 100 µL was plated on an artificial seawater complete medium (342.2 mM NaCl, 14.8 mM $MgCl_2 \cdot 6H_2O$, 1.0 mM $CaCl_2 \cdot 2H_2O$, 6.71 mM KCl, 5 g Bacto-tryptone, 1 g yeast extract, 20 mM MOPS [pH 8.0], 0.3% glycerol) and agar at a final concentration of 1.5%. The colony was selected for its bright yellow color and streaked for isolation on the same artificial seawater media.

For whole genome sequencing, DNA was extracted from 2ml culture of microbial isolate *Sanguibacter* sp. 25GB23B1 using the Zymo Quick-DNA Fungal/ Bacterial Miniprep Kit (Zymo). The sample was received and quantified using an Invitrogen Qubit 4 Fluorometer and 1x dsDNA High Sensitivity Assay Kit (ThermoFisher Scientific). The genomic library was prepared using DNA extracts and the Nextera XT DNA Library Preparation kit (Illumina) according to the manufacturer's protocol. The library was quality checked using an Agilent 2100 Bioanalyzer and DNA High Sensitivity kit and then pooled in an equimolar ratio. The pool was gel purified using a 2% agarose gel and the Qiagen QIAquick gel extraction kit (Qiagen). Following purification, the pool was sequenced on an Illumina NextSeq 550 instrument using a Mid-Output v2.5 chemistry 300-cycle kit to produce 2x150 bp reads. Amplification of the 16S rRNA gene using the primers 533F and 1100R and sequencing confirmed that the isolate is a member of the genus *Sanguibacter*.

The raw genome data was analyzed using FastQC (4) and found to have a total number of 7,418,186 raw sequences with a sequence length of 151 bp. Trimmomatic v0.39 (5) was used for quality filtering and trimming the raw Illumina reads using the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:60. A combined 3,513,311 individual forward and reverse reads were retained which were used to generate the genome assembly. For all the following programs, default settings were used. Assembly was performed from the trimmed reads using SPAdes v3.13.0 (6), and assembly summary statistics were generated with QUAST v5.0.2 (7). The draft genome's total length is 4,303,413 bp, composed of 66 contigs ($N_{50}$: 96,636 bp, $L_{50}$: 14), with an average GC content of 70.68% and an average genome coverage of ~123-fold. The genome was annotated by Rapid Prokaryotic Genome Annotation (PROKKA) (8) and found to contain 3,972 protein coding sequences, three complete rRNA genes, 54 tRNA genes, and one tmRNA gene. A phylogenomic tree was created using the Interactive Tree of Life (iTOL) (9) and GToTree program (10) with *Cellulomonas fimi* as an outlier for comparison (fig. 1). The average nucleotide identity, calculated using pyANI v0.1.2 (11) default settings, between 25GB23B1 and *S. anarcticus* (GCF_002564005.1) was 84.57%, with ~68% of the two genomes aligning. As a genome needs to share 97% or more DNA with another genome to be considered the same species, the *Sanguibacter* strain 25GB23B1 ascribes to one of the criteria of being a new species of *Sanguibacter* (12).

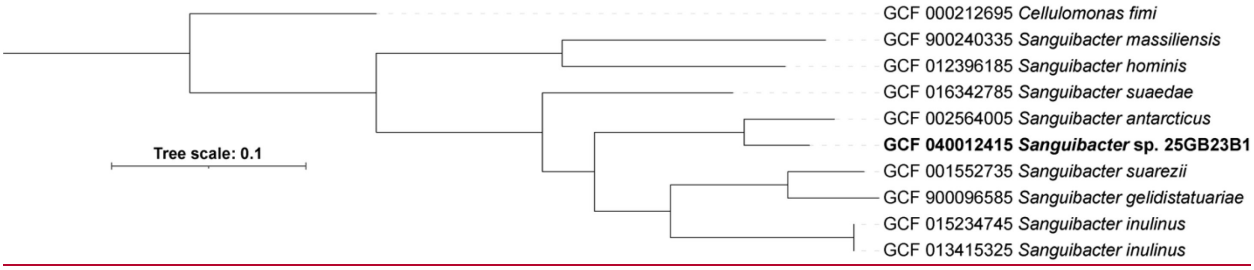


**Figure 1. Phylogenomic tree** for *Sanguibacter* sp. isolate 25GB23B1 based on 74 unique single-copy core genes (SCG) depicting the phylogenomic position relative to other *Sanguibacter* sp. *Cellulomonas fimi* was used as an outlier.

## Data Availability Statement

This whole-genome shotgun genome assembly has been deposited at NCBI GenBank under accession number JBDYKL000000000.1. The BioProject accession number is PRJNA1113969 and the BioSample accession umber is SAMN41472747. Raw sequencing reads were deposited in NCBI's Sequence Read Archive under accession number SRX25616898.

## Acknowledgments

This work was supported by National Science Foundation award #2205569. We would like to thank Dr. Krista Longnecker from Woods Hole Oceanographic Institution for providing water samples and Dr. John McNabb, Research Assistant Professor, Department of Atmospheric and Planetary Sciences, Hampton University, Hampton, VA, for computational help in establishing and maintaining a genome assembly pipeline at the Hampton University pirate cluster. We thank Wright Labs for the whole genome sequencing services and to Hampton University in Hampton, Virginia for providing the laboratory space and equipment to conduct this research.

## References

1. Huang Y, Dai X, He L, Wang Y-N, Wang B-J, Liu Z, Liu S-J. 2005. Sanguibacter marinus sp. nov., isolated from coastal sediment. International journal of systematic and evolutionary microbiology 55:1755-1758.
2. Pikuta EV, Lyu Z, Williams MD, Patel NB, Liu Y, Hoover RB, Busse H-J, Lawson PA, Whitman WB. 2017. Sanguibacter gelidistatuariae sp. nov., a novel psychrotolerant anaerobe from an ice sculpture in Antarctica, and emendation of descriptions of the family Sanguibacteraceae, the genus Sanguibacter and species S. antarcticus, S. inulinus, S. kedieii, S. marinus, S. soli and S. suarezii. International Journal of Systematic and Evolutionary Microbiology 67:1442-1450.
3. Unfried F, Becker S, Robb CS, Hehemann J-H, Markert S, Heiden SE, Hinzke T, Becher D, Reintjes G, Krüger K. 2018. Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. The ISME journal 12:2894-2906.
4. Andrews S. 2017. FastQC: a quality control tool for high throughput sequence data. 2010.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.
6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology 19:455-477.

7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.
8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.
9. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research 49:W293-W296.
10. Lee MD. 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics 35:4162-4164.
11. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Analytical Methods 8:12-24.
12. Bauer M, Kube M, Teeling H, Richter M, Lombardot T, Allers E, Würdemann CA, Quast C, Kuhl H, Knaust F. 2006. Whole genome analysis of the marine Bacteroidetes 'Gramella forsetii'reveals adaptations to degradation of polymeric organic matter. Environmental Microbiology 8:2201-2213.