Complete genome sequence of endophytic Streptomyces strains with biocontrol and growth-promoting properties.

Ichrak Hamdene[1*], Rania Hammami[1], Dorra Souai[1], Najla-Sadfi-Zouaoui[1*]

[1]Laboratoire de Mycologie, Pathologies et Biomarqueurs (LR16ES05), Département de Biologie, Université de Tunis-El Manar, 2092 Tunis, Tunisia.

*Corresponding author: Ichrak Hamdene, Laboratoire de Mycologie, Pathologies et Biomarqueurs (LR16ES05), Département de Biologie, Université de Tunis-El Manar, 2092 Tunis, Tunisia.

Tel: 0021626092639

E-mail address: Ichrak.hamdene@gmail.com, najla.sadfi@fst.utm.tn

Abstract

This study reports the genome assemblies of three endophytic *Streptomyces* strains isolated from the halophytic plants *Arnebia decumbens* and *Ephedra alata*, collected from desert regions of southern Tunisia. Whole-genome sequencing of *Streptomyces* sp. Ju416(a), Av1, and Act26 were performed using the Illumina platform, generating 250bp pair-end reads that were assembled into 72, 47, and 68 large contigs (≥1000 bp), respectively. The genomic characterization of environmentally adapted and understudied *Streptomyces* expands our understanding of microbial diversity and adaptation to extreme environments. These genomes represent a valuable resource for identifying genetic determinants of key functional traits and establishing a foundation for exploring their potential in biotechnological applications, including biocontrol and plant growth promotion.

1. Introduction

Beneficial microorganisms, including bacteria and fungi used as biocontrol agents, seem to be the most promising methods for maintaining plant health, ensuring environmental biosafety, and improving crop production quality [1]. In changing climate scenarios, intervention with microbes is

considered a new sustainable strategy in agricultural production and mitigation of the resilient impacts of stress [2]. Over the past few decades, extensive research has focused on evaluating the effectiveness of selected microorganisms, particularly endophytes, in alleviating the negative impacts of climatic stresses on plant health, physiology, and biochemical aspects. Literature indicates that endophytic associations drive the evolution of biosynthetic gene clusters (BGCs) responsible for secondary metabolite production, making them significant sources for drug discovery and development [3]. Among endophytes, actinobacteria associated with arid plants remain relatively understudied, despite their potential to produce novel bioactive compounds [4]. In particular, endophytic *Streptomyces* represent a valuable source of bioactive secondary metabolites with diverse biological activities, including anticancer, antifungal, antibacterial, insecticidal, and other antimicrobial properties [5]. The diversity and metabolic potential of endophytic Streptomyces are influenced by several factors, such as plant species, developmental stage, environmental conditions, and geographical origin. These factors can also regulate the expression of BGCs through biotic and abiotic stresses in the natural habitat [6]. Therefore, the isolation and characterization of Streptomyces species from diverse and unexplored ecosystems have become important strategies for bioprospecting. Exploring such niches not only facilitates the discovery of novel species but also increases the likelihood of identifying new bioactive compounds [7]. Recent advances in whole-genome sequencing have provided powerful tools for exploring the genetic potential of microorganisms [8]. Genome analysis enables the identification of biosynthetic gene clusters, functional genes involved in plant growth promotion, and mechanisms underlying biocontrol activities [9]. Such approaches offer deeper insights into the metabolic capabilities of Streptomyces and facilitate the discovery of novel bioactive compounds [10]. In this context, the present study reports the whole-genome sequencing of three endophytic Streptomyces strains isolated from halophytic plants in southern Tunisia. These strains were selected based on their biostimulant, antibacterial, and antifeedant activities.

2. Materials and Methods

2.1 Bacterial isolates

Streptomyces strains Ju416(a) and Av1 were isolated from the aerial parts and roots of *Arnebia decumbens*, respectively, while strain Act26 was obtained from the aerial parts of *Ephedra alata*. All strains were isolated as endophytes from halophytic plants in southern Tunisia.

2.2 Sequencing of bacterial genomic DNA

A single colony of pure culture from each strain is picked from a fresh plate and inoculated into 20 ml of GLM broth. After 7 days of incubation, cells are pelleted, washed with nuclease-free water, and repelleted. Around 50 mg of the cell pellet was resuspended in DNA/RNA shield buffer (Zymo Research, USA) and sent to MicrobesNG (Birmingham, UK) for sequencing.

2.2.1 DNA Processing

Five to forty microlitres of the cell suspension were lysed with 120 µL of TE buffer containing lysozyme (final concentration 0.1 mg/mL) and RNase A (ITW Reagents, Barcelona, Spain) (final concentration 0.1 mg/mL), incubated for 25 min at 37°C. Proteinase K (VWR Chemicals, Ohio, USA) (final concentration 0.1mg/mL) and SDS (Sigma-Aldrich, Missouri, USA) (final concentration 0.5% v/v) were added and incubated for 5 min at 65°C. Genomic DNA was purified using an equal volume of SPRI beads and resuspended in EB buffer (10mM Tris-HCl, pH 8.0). The extracted DNA was then quantified using the Quant-iT dsDNA HS kit (ThermoFisher Scientific) assay in an Eppendorf AF2200 plate reader (Eppendorf UK Ltd, UK) and diluted as appropriate.

2.2.2 Illumina Sequencing

Genomic DNA libraries were prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: input DNA was increased 2-fold, and PCR elongation time was increased to 45 seconds. DNA quantification and

library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). Libraries were sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, USA) using a 250 bp paired-end protocol.

Reads are adapter-trimmed using Trimmomatic version 0.30 [11] with a sliding window quality cutoff of Q15. De novo assembly is performed on samples using SPAdes version 3.7 [12], and contigs are annotated using Prokka 1.11 [13].

2.2.4 Genome assembly and comparisons

Genome assemblies were evaluated for completeness using BUSCO v6, with the *Streptomyces_odb12* dataset [15]. Digital DNA–DNA hybridization (dDDH) values were calculated using the Genome-to-Genome Distance Calculator (GGDC) 3.0, applying Formula 2, via the GGDC web server (https://ggdc.dsmz.de/), accessed on 14 August 2025) [16]. Average Nucleotide Identity (ANI) was calculated using FastANI [17]. The heatmap was generated using the *pheatmap* package in R (version 4.4.1). Average Amino Acid Identity (AAI) analysis was performed using EzAAI version 1.2.4, released in July 2025[18].The resulting interaction network was visualized and analyzed using Cytoscape version 3.10.4. Reference genome sequences were retrieved from the NCBI database for comparative analyses.

3. Results

3.1 Genomic Features

A summary of the genomic features for *Streptomyces* strains is listed in Table 1. The genomes of *Streptomyces* sp. Ju416(a), Av1, and Act26 were assembled, and outcomes were 72, 47, and 68 large contigs (≥1000 bp). The Ju416(a) genome features an average GC-content of 71,51%, while the N50 value for contigs was 218220 bp with an average contig size of 169,638 bp. The Av1 genome features an average GC-content of 72,35%, while the N50 value for contigs was 286057 bp with an average

contig size of 170,978 bp. The Act26 genome features an average GC-content of 71,63%, while the N50 value for contigs was 308817 bp with an average contig size of 98,0424 bp.

3.2 Nucleotide sequence accession numbers

The complete genome sequences of the three *Streptomyces* strains Ju416(a), Av1, and Act26 have been deposited in the NCBI GenBank under the accession numbers listed in **Table 1**.

Table 1: Genomic features of *Streptomyces* strains.

Bacterial species	*Streptomyces sp.*	*Streptomyces tunisiensis*	*Streptomyces anulatus*
Strain	Ju416(a)	Av1	Act26
Total size of the Genomes (bp)	8733275	7544725	8581043
Number of contigs (>=1000 bp)	72	47	68
Largest contig (bp)	854394	860190	710853
Number of reads	3101243	2692211	1800390
Contigs *N*50	218220	286057	308817
Mean coverage (X)	167	171	98
GC content (%)	71.51	72.35	71.63
BioProject	PRJNA1284273		
BioSample	SAMN49717262	SAMN49717263	SAMN49717264
Assemblies	JBPSDK000000000	JBPSDJ000000000	JBPSDI000000000
SRA	SRR35279849	SRR35279848	SRR35279847

3.3 Whole genome taxonomy-based analyses

3.3.1 Genome quality assessment (BUSCO)

Genome completeness was assessed using BUSCO v6.0.0 with the Streptomyces_odb12 dataset (**Table 2**). All assemblies showed high completeness (>99%), with very low proportions of fragmented and missing BUSCOs, indicating high-quality genome assemblies suitable for downstream taxonomic analyses.

Table 2: BUSCO assessment of Streptomyces genome assemblies.

	Ju416(a)	Av1	Act26
BUSCO	C:99.4%[S:99.1%,D:0.3%],F:0.4%,M:0.2%,n:1438	C:99.7%[S:99.4%,D:0.3%],F:0.1%,M:0.2%,n:1438	C:99.4%[S:99.2%,D:0.2%],F:0.3%,M:0.2%,n:1438
Complete BUSCOs (C)	1429	1433	1430
Complete and single-copy BUSCOs (S)	1425	1429	1427
Complete and duplicated BUSCOs (D)	4	4	3
Fragmented BUSCOs (F)	6	2	5
Missing BUSCOs (M)	3	3	3
Total BUSCO groups searched (n)	1438	1438	1438

3.3.2 Genome-based taxonomic analyses (dDDH, ANI, AAI):

Digital DNA–DNA hybridization (dDDH) analyses were performed using the Genome-to-Genome Distance Calculator (GGDC) server based on Formula 2. The strain Ju416(a) exhibited a dDDH value of 79.81% with *Streptomyces parvus*, strain Av1 showed a high dDDH value of 96.74% with *Streptomyces tunisiensis*, and strain Act26 displayed a dDDH value of 79.46% with *Streptomyces anulatus*. All these values exceed the 70% species delineation threshold, supporting the assignment of each strain to the corresponding species.

Table 3: Closest Reference Genomes and Probability of Species and Subspecies Delineation Based on Digital DNA-DNA Hybridization (DDH).

Query strain	**Closest reference genome**	**Species name**	**Probability that DDH > 70% (i.e., same species)**	**Probability that DDH > 79% (i.e., same subspecies)**
Ju416(a)	GCF_014648875.1	*Streptomyces parvus*	79.81%	30.31%
Av1	GCF_039538125.1	*Streptomyces tunisiensis*	96.74%	70.09%
Act26	GCF_001434355.1	*Streptomyces anulatus*	79.46%	29.93%

According to ANI analysis (fig.1), Ju416(a) and Act26 showed ANI values of 96.87% with *Streptomyces parvus* and 97.08% with *Streptomyces anulatus*, respectively. Moreover, the results of ANI comparisons between *Streptomyces sp*. Av1 and *Streptomyces tunisiensis* yielded values of 99.25% identity with 93.6% genome coverage.

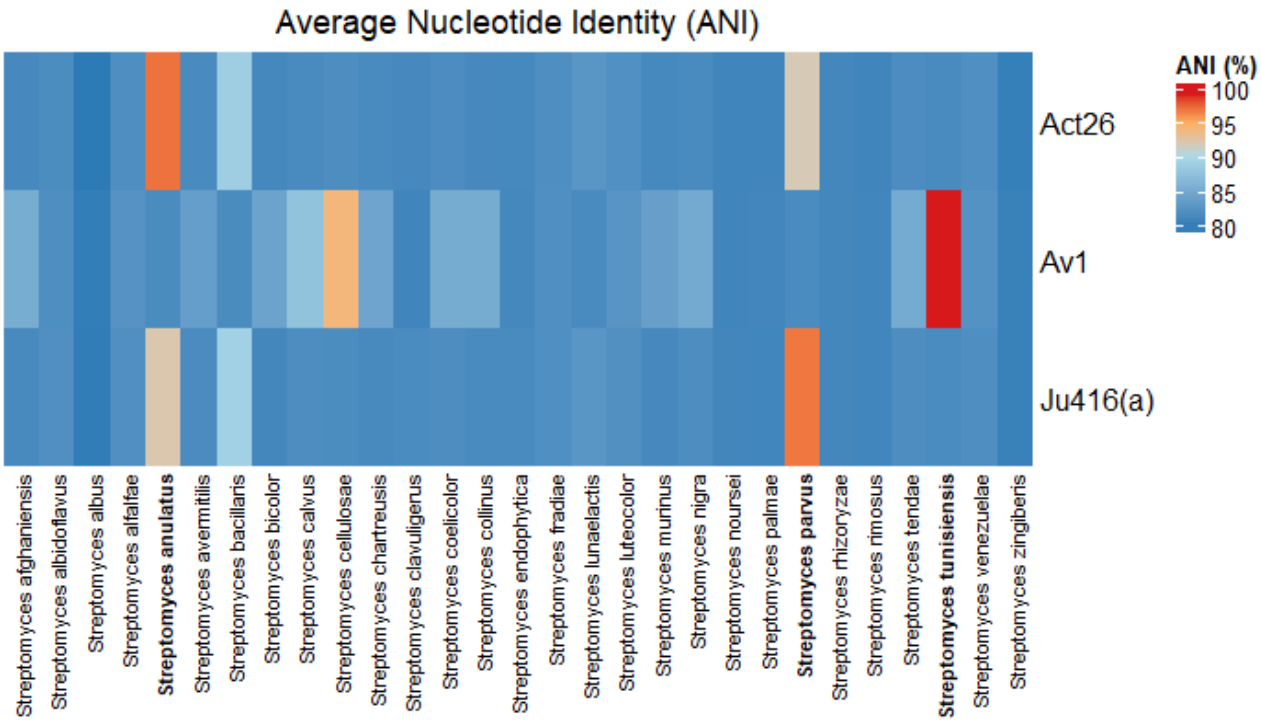


Figure 1: Heatmap of FastANI values of *Streptomyces* strains Ju416(a), Av1, and Act26. Higher values correspond to greater genetic homology between strains (red). Pairwise ANI values were calculated using FastANI version 1.34, which provides a very fast and relatively accurate approximation of ANI.

The Average Amino Acid Identity (AAI) analysis (**fig. 2**) revealed distinct levels of genomic relatedness between the query genome and the reference genome. The AAI identities of the different Streptomyces range from 70.04–99.10%, indicating that the strains share at least two-thirds of the presented protein-coding genes. The strain Av1 showed the highest AAI value with *Streptomyces tunisiensis* (99.10%), followed by *Streptomyces cellulosae* (93.75%), indicating a very close phylogenetic relationship. For the strain Act26, the highest AAI values were observed with

Streptomyces anulatus (96.83%), *Streptomyces griseus* (91.93%), and *Streptomyces parvus* (90.91%), supporting its close affiliation with these taxa. Similarly, the strain Ju416(a) exhibited high similarity with *Streptomyces parvus* (97.66%), *Streptomyces griseus* (91.89%), and *Streptomyces anulatus* (90.77%).

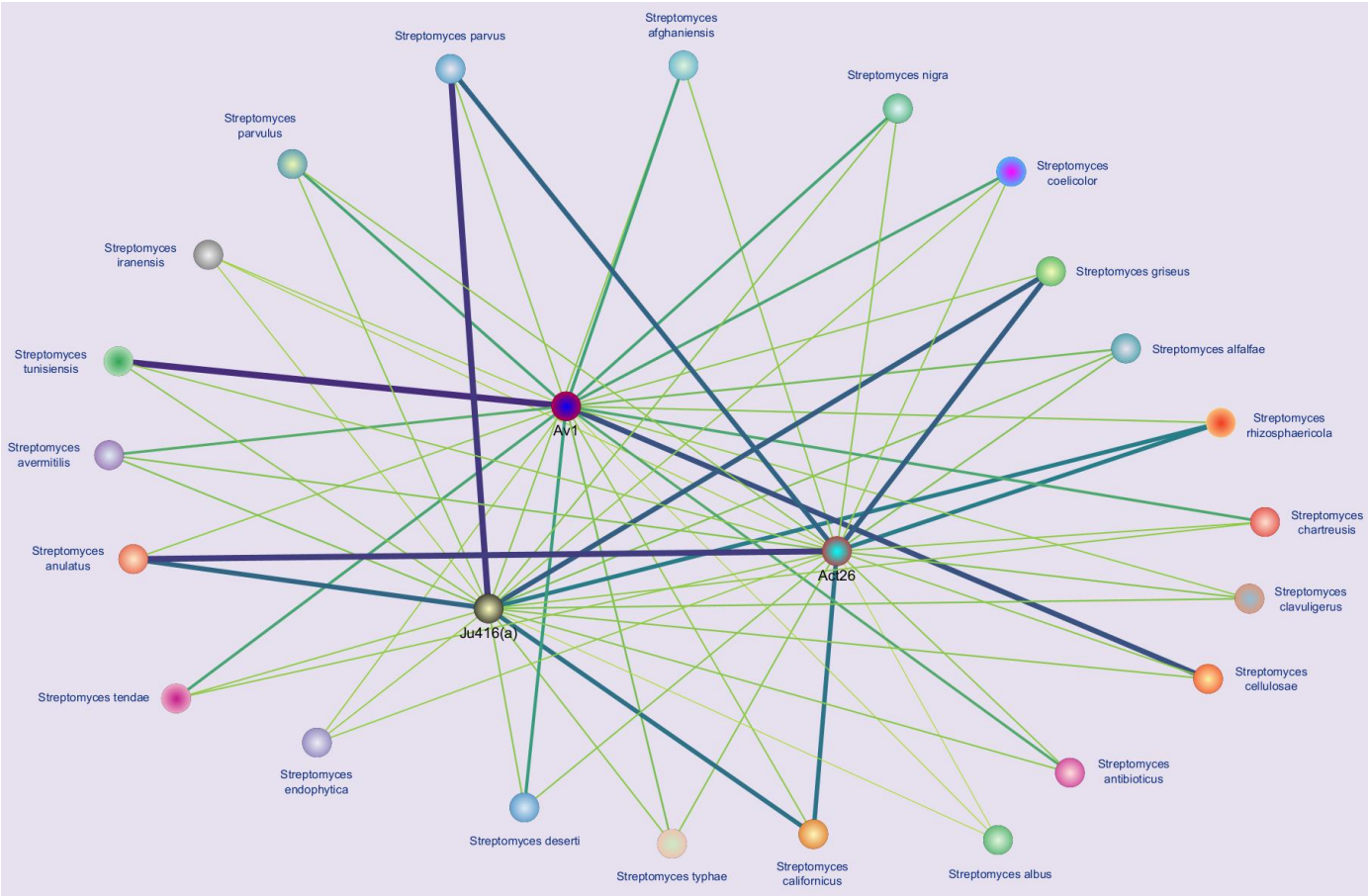


Figure 2: Average Amino Acid Identity (AAI) network constructed using Cytoscape based on 22 reference strains and *Streptomyces* strains Ju416(a), Av1, and Act26. Nodes represent the studied and reference strains and are labeled by their names. Edges indicate pairwise relationships between strains; thicker and darker edges correspond to higher AAI values, reflecting greater genomic similarity.

The genomic relatedness of the studied strains was further assessed using ANI, AAI, and digital DNA–DNA hybridization (dDDH) analyses (**Table 4**). Strain Av1 showed very high similarity with *Streptomyces tunisiensis* (ANI = 99.25%, AAI = 99.10%, dDDH probability = 96.74%), clearly indicating that it belongs to the same species.

Similarly, strain Ju416(a) exhibited high similarity with *Streptomyces parvus* (ANI = 96.87%, AAI = 97.66%, dDDH probability = 79.81%), supporting its assignment to this species. Strain Act26 also

showed strong genomic relatedness to *Streptomyces anulatus* (ANI = 97.08%, AAI = 96.83%, dDDH probability = 79.46%).

Overall, all three strains exceeded the accepted species delineation thresholds (ANI ≥ 95% and dDDH ≥ 70%), confirming their affiliation with known *Streptomyces* species, although the lower probabilities for subspecies delineation suggest potential intraspecific diversity.

Table 4: ANI, dDDH, and AAI among strains queries and other closely related *Streptomyces* members.

Query strain	**Closest reference genome**	**Species name**	**Probability that DDH > 70% (i.e., same species)**	**Probability that DDH > 79% (i.e., same subspecies)**	**ANI values**	**AAI values**
Ju416(a)	GCF_014648875.1	*Streptomyces parvus*	79.81%	30.31%	96.87%	97.66%
Av1	GCF_039538125.1	*Streptomyces tunisiensis*	96.74%	70.09%	99.25%	99.10%
Act26	GCF_001434355.1	*Streptomyces anulatus*	79.46%	29.93%	97.08%	96.83%

Statement on continuing work

The datasets are available before their formal publication and should be considered preliminary. We welcome feedback from the community. For further information, please contact ichrak.hamdene@gmail.com, najla.sadfi@fst.utm.tn.

Acknowledgements

The sequencing of the bacterial strains was supported by the GetGenome initiative of The Sainsbury Laboratory, Norwich, UK, with support from the Gatsby Charitable Foundation and the Biotechnology and Biological Sciences Research Council (BBBSRC). We would like to thank Prof. Sophien Kamoun and Dr. James Canham for their valuable assistance.

References

1. Bakr, R., et al., *The Potential of Beneficial Microbes for Sustainable Alternative Approaches to Control Phytopathogenic Diseases.* Microbiology Research, 2025. **16**(5).
2. Jalal, A., et al., *Beneficial Microorganisms Improve Agricultural Sustainability under Climatic Extremes.* Life (Basel), 2023. **13**(5).
3. Mohamed, R.M., M.E. El Awady, A.M. Fahim, and A.E. Abo-Amer, *Antibacterial potential of endophytic Streptomyces spp. isolated from peanut (Arachis hypogaea) roots: bioactiveprofiling and molecular docking studies.* Sci Rep, 2026. **16**(1): p. 6351.
4. Singh, R. and A.K. Dubey, *Diversity and Applications of Endophytic Actinobacteria of Plants in Special and Other Ecological Niches.* Front Microbiol, 2018. **9**: p. 1767.
5. Yadav, R.P., et al., *Antibacterial, Antifungal, and Cytotoxic Effects of Endophytic Streptomyces Species Isolated from the Himalayan Regions of Nepal and Their Metabolite Study.* Biomedicines, 2024. **12**(10).
6. Sivalingam, P., et al., *Endophytic Streptomyces: an underexplored source with potential for novel natural drug discovery and development.* Arch Microbiol, 2024. **206**(11): p. 442.
7. Veilumuthu, P., et al., *Genomic insights into an endophytic Streptomyces sp. VITGV156 for antimicrobial compounds.* Front Microbiol, 2024. **15**: p. 1407289.
8. Mohite, O.S., et al., *Pangenome mining of the Streptomyces genus redefines species' biosynthetic potential.* Genome Biol, 2025. **26**(1): p. 9.
9. Heonjun Jeong, Y.C., Jiyoon Nam, Yeon Hee Ban, *A guide to genome mining and genetic manipulation of biosynthetic gene clusters in Streptomyces.* Journal of Microbiology, 2025. **63**.
10. Du, Y.H., W.; Hao, P.; Hu, and T.Z. Y.; Hu, Y, *A Genomics-Based Discovery of Secondary Metabolite Biosynthetic Gene Clusters in the Potential Novel Strain Streptomyces sp. 21So2-11 Isolated from Antarctic Soil.* Microorganisms, 2024. **12:1228**.
11. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.
12. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.
13. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
15. Tegenfeldt, F., et al., *OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes.* Nucleic Acids Res, 2025. **53**(D1): p. D516-D522.
16. Meier-Kolthoff, J.P., J.S. Carbasse, R.L. Peinado-Olarte, and M. Goker, *TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes.* Nucleic Acids Res, 2022. **50**(D1): p. D801-D807.
17. Jain, C.R.-R., L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S, *High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.* nature communications, 2018. **9:5114**.
18. Kim, D., S. Park, and J. Chun, *Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity.* Journal of Microbiology, 2021. **59**(5): p. 476-480.