

Draft Genome Sequence of *Pseudarthrobacter phenanthrenivorans* Strain MHSD1, a Bacterial Endophyte Isolated From the Medicinal Plant *Pellaea calomelanos*

Evolutionary Bioinformatics
Volume 16: 1–5
© The Author(s) 2020
DOI: 10.1177/1176934320913257



Khuthadzo Tshishonga and Mahloro Hope Serepa-Dlamini 

Department of Biotechnology and Food Technology, Faculty of Science, University of Johannesburg, Johannesburg, South Africa.

ABSTRACT: *Pseudarthrobacter phenanthrenivorans* strain MHSD1 is a bacterial endophyte isolated from sterilized leaves of *Pellaea calomelanos*, a medicinal plant capable of growing in arid environments. Here, we report the draft genome sequence and annotation of this bacterial endophyte. The draft genome sequence of *P. phenanthrenivorans* strain MHSD1 has 4 450 468 bp with a G + C content of 65.30%. The National Center for Biotechnology Information Prokaryotic Genome Annotation Pipeline identified a total of 4004 protein-coding genes, 56 genes coding for RNAs, and 82 pseudogenes. Biosynthesis pathways for various phytohormones such as auxin, salicylic acid, ethylene, cytokinin, jasmonic acid, abscisic acid, and gibberellins were identified. Putative genes involved in various characteristics of bacterial endophyte lifestyle such as transport, motility, adhesion, membrane proteins, secretion and delivery systems, plant cell wall modification, and detoxification were identified. Phylogenomic analysis showed *P. phenanthrenivorans* strain MHSD1 to be a subspecies of *P. phenanthrenivorans* Sphe3.

KEYWORDS: *Pseudarthrobacter phenanthrenivorans*, bacterial endophyte, genome sequencing, *Pellaea calomelanos*, phylogenomic analysis

RECEIVED: December 02, 2019. **ACCEPTED:** February 18, 2020.

TYPE: Short Report

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Research Foundation of South Africa (Thuthuka Grant No. TTK170405225920).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Mahloro Hope Serepa-Dlamini, Department of Biotechnology and Food Technology, Faculty of Science, University of Johannesburg, Doornfontein Campus, PO Box 17011 Doornfontein 2028, Johannesburg, South Africa. Email: hopes@uj.ac.za

Introduction

Endophytes are microorganisms, often bacteria or fungi, that are associated with plant tissues without causing any harm.¹ These microorganisms can spend part or all of their life cycle within their plant hosts.² In their relations with plants, they display various interactions that involve mutualism and antagonism but rarely parasitism.³ All plants are probably associated with endophytes.^{2,4} Endophytes promote plant growth by enhancing plant's uptake of nutrients such as nitrogen, phosphate, and potassium; also, they biologically control plant pathogens as well as the production of secondary metabolites with pharmaceutical or biotechnological interest, and phytostimulation through the production of phytohormones^{5–8}

Endophytes produce bioactive secondary metabolites; moreover, endophytes associated with medicinal plants are known to produce similar secondary metabolites as their plant host, with increased therapeutic potential.^{9,10} Thus, endophytes are alternate sources of bioactive secondary metabolites, as it is better to scale up the microbial fermentation process to increase the production of biologically active compounds, than use high amounts of plant materials, which can result in deforestation, decreased biodiversity, and conservation.^{9,11} As such, the prospects of isolating and identifying new endophyte species from plants can be beneficial. In a recent study, we isolated and identified bacterial endophytes associated with *Pellaea calomelanos*,¹² a medicinal plant capable of growing in arid conditions.

Pellaea calomelanos is a fern that belongs to *Pteridaceae* family.¹³ The plant has healing properties for ailments such as

asthma, head colds, coughs, and chest colds.^{14,15} One of the isolated bacterial endophytes was identified as *Arthrobacter* sp. MHSD1 using 16S ribosomal RNA (rRNA) gene sequence and biochemical characterization.¹² The whole genome of this strain has been sequenced and the sequence data submitted to National Center for Biotechnology Information (NCBI). The draft genome sequence is described here.

Materials and Methods

Genomic DNA isolation, library preparation, and sequencing

Total genomic DNA was extracted from glycerol stock cultures, maintained on nutrient agar at 30°C for 48 hours using the Nucleospin Microbial DNA extraction kit as per the manufacturer's protocol. The concentration and quality of isolated DNA were determined using the NanoDrop ND-2000 UV-Vis spectrophotometer. The DNA was sent to a commercial service provider, Agricultural Research Council, Onderstepoort, South Africa, for sequencing with Illumina MiSeq platform. Briefly, the library was prepared using NEBNextUltra II DNA kit following the manufacturer's protocol with a paired-end sequencing strategy (300 bp insert size) using Illumina MiSeq instrument v3.

Pre-processing, genome assembly, and annotation

All pre-annotation analyses were performed on Galaxy web platform available at <https://usegalaxy.org>.¹⁶ FastQC v 0.69



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

was used to assess the quality of the raw reads.¹⁷ Using default parameters, the sequence reads were *de novo* assembled using Unicycler v 0.4.1.1¹⁸ and assessed with Quast v 4.6.3.¹⁹ The draft genome sequence was submitted to NCBI and annotated using Prokaryotic Genome Annotation Pipeline (PGAP)²⁰ and Rapid Annotations using Subsystems Technology (RAST) server.^{21–23}

Bioinformatics

The phylogenomic analysis was undertaken with the Type Strain Genome Server (TYGS) available at <https://tygs.dsmz.de/>²⁴ and OrthoANI (Orthologous Average Nucleotide Identity) with the Orthologous Average Nucleotide Identity Tool (OAT) software.²⁵ The genomic islands (GI) were identified by screening the PGAP annotation file generated from NCBI on the IslandViewer 4 Web site (<http://www.pathogenomics.sfu.ca/islandviewer/>).²⁶ The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) were predicted by CRISPRCas finder software.^{27–29} The RAST server was used to annotate and classify predicted genes according to function.^{21–23} The shared and unique genes of MHSD1 were analyzed by comparing it with 1 bacterial endophyte genome (*Arthrobacter* sp. PAMC 25486) as well as 3 closely related genomes (*Pseudarthrobacter chlorophenolicus* A6, *P. phenanthrenivorans* Sphe3, *P. sulfonivorans* Ar51) using EDGAR 2.0.³⁰ The genome was masked for repeats using RepeatMasker.³¹

Accession of the genome sequence

The data from this Whole Genome Shotgun project have been deposited at DDBJ/ENA/GenBank with BioProject number PRJNA549841 and BioSample number SAMN12098155 under the accession VJJD00000000. The version described here is VJJD01000000.

Interpretation of Data Set

The draft genome of strain MHSD1 had 56 contigs with a total length of 4450468 bp, G + C content of 65.30%, and an *N50* value of 363437 bp. Using PGAP, the predicted number of genes was 4142, of which 4004 of them were protein-coding genes (CDSs), 56 were RNAs, 82 were pseudogenes, and 3 were non-coding RNAs (ncRNAs). The predicted RNA coding genes include 50 transfer RNAs (tRNAs) and 3 rRNAs (5S, 16S, and 23S). A total of 8566 bp comprising 0.19% were masked. The genome features are in Table 1. The TYGS whole genome-based taxonomic analysis showed that MHSD1 was closely related to *Pseudarthrobacter phenanthrenivorans* strain Sphe3 as shown in Figures 1 and 2 (Supplementary Data). MHSD1 formed a sister clade with Sphe3 in both phylogenetic trees.

MHSD1 had a digital DNA-DNA hybridization (dDDH) of 85.3% and G + C% content difference of 0.03 with *P. phenanthrenivorans* Sphe3 (Table 1, Supplementary Data). The observed

Table 1. Genome features of *Pseudarthrobacter phenanthrenivorans* strain MHSD1.

ATTRIBUTE	VALUE
Genomic size (bp)	4,450,468
GC content	65.30%
Total number of genes	4142
Protein coding genes	4004
Number of RNAs	56
rRNA genes	3
tRNA gene	50
Protein coding genes with function prediction	3553
CRISPR repeats	1

Abbreviations: CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; rRNA, ribosomal RNA; tRNA, transfer RNA.

dDDH of 85.3% was greater than the species boundary value of dDDH >70% for delineating bacterial species as closely related species³²; in addition, this value exceeded the dDDH >79%–80% for delineating subspecies.²⁴ Based on the TYGS phylogenomic classification, MHSD1 is delineated as a subspecies of *P. phenanthrenivorans* Sphe3; initial identification of MHSD1 using the 16S rRNA gene was *Arthrobacter* sp. strain MHSD1. MHSD1 showed lower OrthoANI values (<70%) (Figure 3, Supplementary Data) than the species boundary of >95%–96%.²⁵ Although the OrthoANI values were lower, the delineation of MHSD1 as a subspecies of *P. phenanthrenivorans* Sphe3 was based on the dDDH value in the TYGS because it was an enhanced method for species delineation and facilitates classification and identification of species as well as subspecies by comparison with published and described type strains.²⁴

Pseudarthrobacter phenanthrenivorans strain MHSD1 shared 682 common genes with other closely related *Pseudarthrobacter* species, and only 14 genes were shared with bacterial endophyte *Arthrobacter* sp. PAMC 25486 (Figure 1). A total of 1765 genes were common among MHSD1 and all the selected comparison species (Figure 1). The 14 exclusive common genes between MHSD1 and PAMC 25486 (results not shown) encode transport proteins and transcriptional regulators, which are essential bacterial endophyte genes that have been previously identified in other bacterial endophyte species.^{33,34} *Pseudarthrobacter phenanthrenivorans* strain MHSD1 had 367 unique genes, whereas *P. phenanthrenivorans* strain Sphe3 had 231 unique genes; this can be attributed to MHSD1 having a larger genome length than Sphe3.

Pseudarthrobacter phenanthrenivorans strain MHSD1 was found to consist of several sets of genes acquired through horizontal gene transfer. As such, 24 GI (Figure 2) were identified in *P. phenanthrenivorans* strain MHSD1 genome when aligned to reference genome *P. phenanthrenivorans* Sphe3.³⁵ The details of the genes clustered on the genomic islands are shown in Table 2

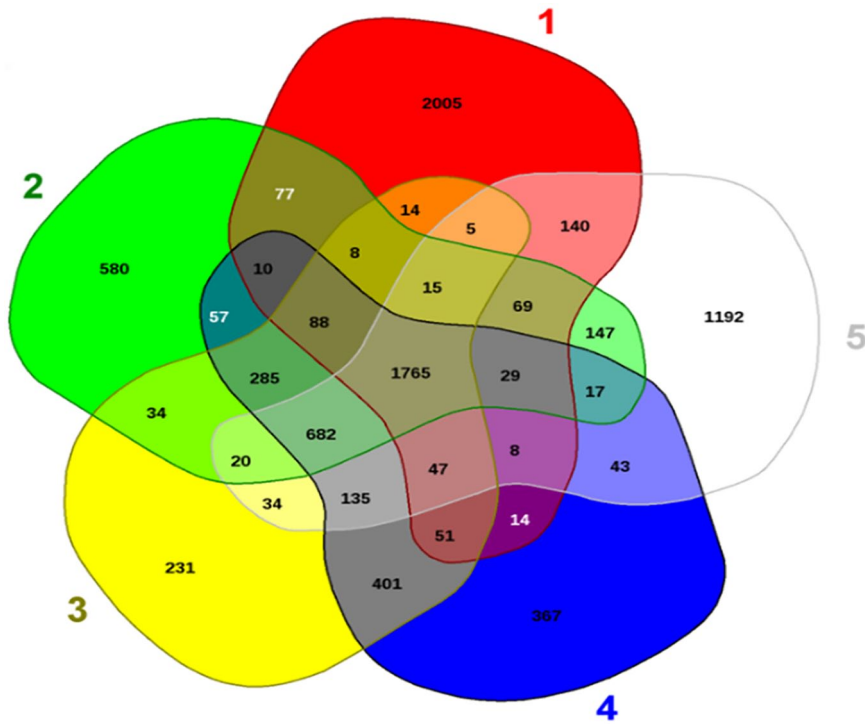


Figure 1. Venn diagram of shared and unique genes of *Pseudarthrobacter phenanthrenivorans* MHSD1 and selected comparison species. 1: *Arthrobacter* sp. PAMC 25486; 2: *Pseudarthrobacter chlorophenolicus* A6; 3: *P. phenanthrenivorans* Sphe3; 4: *P. phenanthrenivorans* MHSD1; 5: *Pseudarthrobacter sulfonivorans* Ar51.

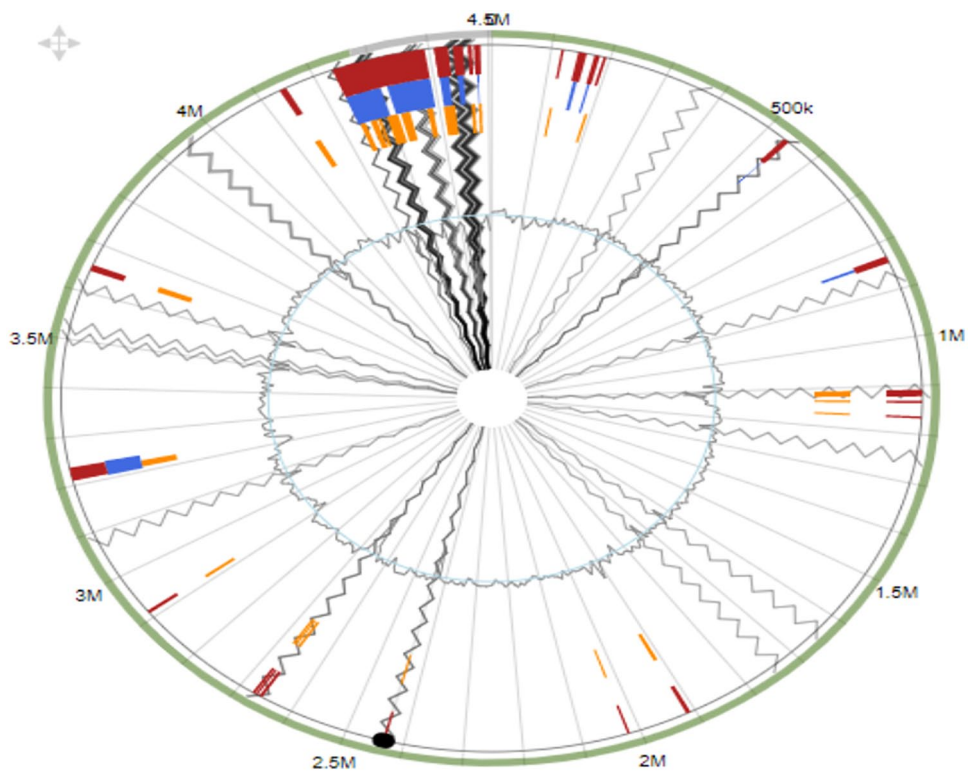


Figure 2. Genetic islands of *Pseudarthrobacter phenanthrenivorans* MHSD1 aligned with reference genome of *P. phenanthrenivorans* Sphe3. A total of 24 genetic islands were predicted using IslandViewer 4. The green outer circle represents the scale line of the genome in Mb, and the obtained genomic islands are represented by the following colors: IslandPath-DIMOB (blue), SIGI-HMM (orange), and integrated detection (red). Gray lines indicate contig boundaries.

Table 2. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) sequences present within *Pseudarthrobacter phenanthrenivorans* strain MHSD1 identified using CRISPRCasFinder.

ELEMENT	CRISPR ID/CAS TYPE	START	END	GENE SPACER	REPEAT CONSENSUS/CAS GENE	DIRECTION	EVIDENCE LEVEL
CRISPR	Contig_4_1	84903	85015	1	ACCGGCTGCGGGACGCGG ACCACCGGGACGGGGACC ACCGGCACCGGCTGCGGG ACGCGGACCACCGGGACG GTCGCCGTGCGGTGCGGCTT CCACCCGGACGGGGCATG CCCTGGGACGTGGCGAAC GGGTTGTTGCCGGGACGG GGAGCACGCTCGCCGTCC CCGCCGCGGCCGCGGGGC ATGCCCTGGGACGTGGCG AAGGGGTTGTTGCCCGGG CGGGGACCGCCGGGGCGG GGAGCTGAGCCACCCTGT GATCCGCCGGAGCGGGTG GGTGCTGCAGGGGTTTCA GCTTTGGGGGCCGCGCGT GCGCCG	ND	1

Abbreviation: ND, not determined.

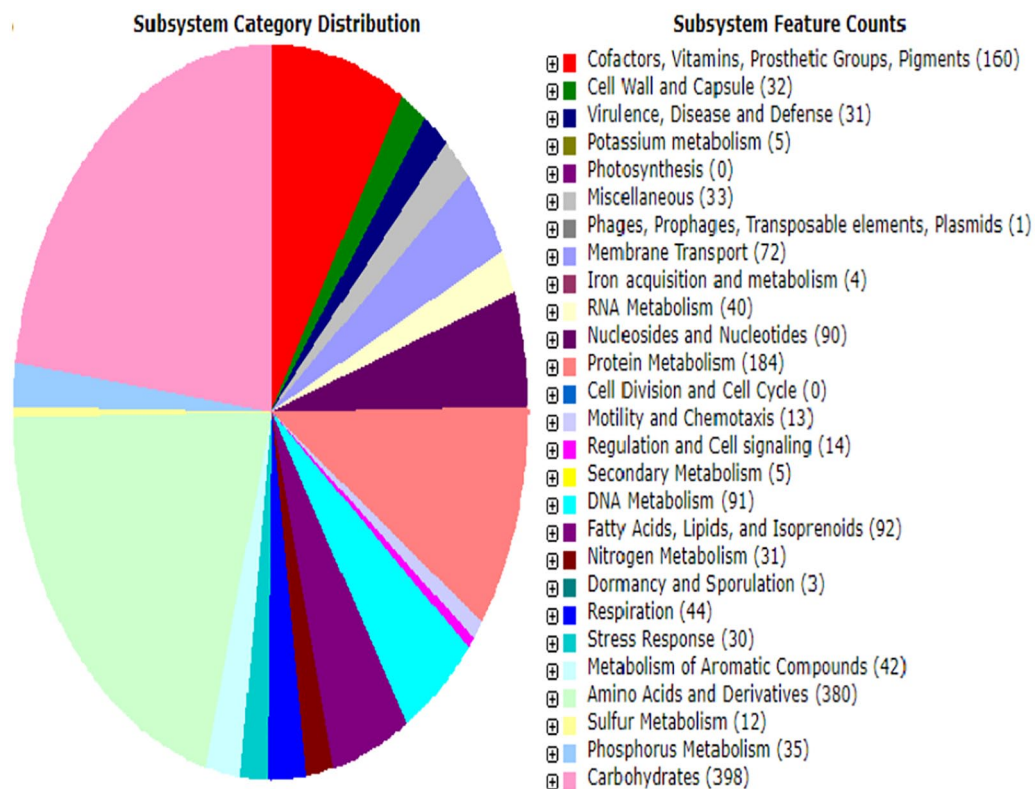


Figure 3. Functional classification of predicted genes of *Pseudarthrobacter phenanthrenivorans* MHSD1 genome based on RAST annotation server. RAST indicates Rapid Annotations using Subsystems Technology.

(Supplementary Data). We identified only one CRISPR system with 1 spacer and 11 repeats (Table 2). Functional classification of the genes in *P. phenanthrenivorans* strain MHSD1 based on RAST annotation (Figure 3) showed that most of the predicted genes are involved in carbohydrate metabolism, which is consistent with the bacterial endophyte lifestyle within which acquisition and mobilization of nutrients such as phosphate, nitrogen, and iron are important for symbiotic plant-bacteria interaction.³⁶

In this study, several putative genes involved in bacterial endophyte behavior or lifestyle were predicted and compared with bacterial endophyte *Enterobacter* sp. 638 as well as nonendophyte *P. phenanthrenivorans* Sphe3 (Table 3, Supplementary Data). Genes putatively involved in transport, motility, adhesion, membrane proteins, secretion and delivery systems, plant cell wall modification, detoxification, substrate utilization, stress protection, and transcriptional regulators were identified. In

addition, genes important in bacterial endophyte life style, such as those involved in nitrogen fixation and siderophore production, were identified in MHSD1. Although most of the genes present in bacterial endophytes, *Enterobacter* sp. 638 and *P. phenanthrenivorans* MHSD1, were also present in *P. phenanthrenivorans* Sphe3, distinctness of the bacterial endophytes was due to the presence of genes encoding transport proteins and transcriptional regulators important in endophytic behavior or lifestyle, which were not present in the latter. More work is currently underway to describe MHSD1 as a subspecies of *P. phenanthrenivorans* Sphe3.

Biosynthesis pathways of various phytohormones of *P. phenanthrenivorans* MHSD1 consist of various important plant hormones such as auxin, salicylic acid, ethylene, cytokinin, jasmonic acid, abscisic acid, and gibberellins (Figure 4, Supplementary Data). The phytohormones are essential for the development and growth of plants through various mechanisms such as cell elongation, division and differentiation, access to nutrients, stress tolerance, and defense against phytopathogens.³⁷⁻³⁹

Author Contributions

Work was planned by MHS-D and executed by KT.

ORCID iD

Mahloro Hope Serepa-Dlamini  <https://orcid.org/0000-0002-0573-181X>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Singh R, Dubey AK. Endophytic actinomycetes as emerging source for therapeutic compounds. *Indo Global J Pharm Sci.* 2015;5:106-116.
- Gouda S, Das G, Sen SK, Shin H, Patra JK. Endophytes: a treasure house of bioactive compounds of medicinal importance. *Front Microbiol.* 2016;7:1538. doi:10.3389/fmicb.2016.01538.
- Nair DN, Padmavathy S. Impact of endophytic microorganisms on plants, environment and humans. *Sci World J.* 2014;2014:250693.
- Dudeja SS, Giri R. Beneficial properties, colonisation, establishment and molecular diversity of endophytic bacteria in legumes and non-legumes. *Afr J Microbiol Res.* 2014;8:1562-1572. doi:10.5897/AJMR2013.6541.
- Hardoim PR, van Overbeek LS, Berg G, et al. The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol Mol Biol Rev.* 2015;79:293-320. doi:10.1128/MMBR.00050-14.
- Ma Y, Rajkumar M, Zhang C, Freitas H. Beneficial role of bacterial endophytes in heavy metal phytoremediation. *J Environ Manage.* 2016;174:14-25. doi:10.1016/j.jenvman.2016.02.047.
- Vejan P, Abdullah R, Khadiran T, Ismail S, Nasrullah Boyce A. Role of plant growth promoting rhizobacteria in agricultural sustainability—a review. *Molecules.* 2016;21:573. doi:10.3390/molecules21050573.
- Sharma IP, Chandra S, Kumar N, Chandra D. PGPR: heart of soil and their role in soil fertility. In: Meena S, Ishra PK, Bisht JK, Pattanayak A eds. *Agriculturally Important Microbes for Sustainable Agriculture*, Vol. 1. Singapore: Springer; 2017:51-67. doi:10.1007/978-981.
- Strobel GA. Endophytes as sources of bioactive products. *Microbes Infect.* 2003;5:535-544.
- Subbulakshmi GK, Thalavaipandian A, Bagyalakshmi RV, Rajendran A. Bioactive endophytic fungal isolates of *Biota orientalis* (L) Endl., *Pinus excelsa* wall and *Thuja occidentalis* L. *Int J Adv Life Sci.* 2012;4:9-15.
- Alvin A, Miller KI, Neilan BA. Exploring the potential of endophytes from medicinal plants as sources of antimicrobial compounds. *Microbiol Res.* 2014;169:483-495.
- Mahlangu SG, Serepa-Dlamini MH. First report of bacterial endophytes from the leaves of *Pellaea calomelanos* in South Africa. *S Afr J Sci.* 2018;114:55-63.
- Schuettpetz E, Schneider H, Huiet L, Windham MD, Pryer KM. A molecular phylogeny of the fern family Pteridaceae: assessing overall relationships and the affinities of previously unsampled genera. *Molec Phylogenet Evol.* 2007;44:1172-1185.
- Hutchings A, Scott AH, Lewis G, Cunningham AB. *Zulu Medicinal Plants: An Inventory*. Pietermaritzburg, South Africa: University of Natal Press; 1996.
- South African National Biodiversity Institute (SANBI). *Pellaea calomelanos*. <http://pza.sanbi.org/pellaea-calomelanos>. Updated 2019. Accessed November 21, 2019.
- Afgan E, Baker D, van den Beek M, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses. *Nucleic Acids Res.* 2016;44:3-10.
- Andrews S. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Updated 2010. Accessed September 2, 2019.
- Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13:e1005595. doi:10.1371/journal.pcbi.1005595.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072-1075.
- Tatousova T, DiCuccio M, Badretin A, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44:6614-6624. doi:10.1093/nar/gkw569.
- Aziz RK, Bartels D, Best AA, et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics.* 2010;9:75.
- Overbeek R, Olson R, Pusch GD, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 2014;42:D206-D214.
- Brettin T, Davis JJ, Disz T, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep.* 2015;5:8365.
- Meier-Kolthoff JP, Goker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun.* 2019;10:2182.
- Lee I, Kim YO, Park SC, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol.* 2015;66:1100-1103.
- Bertelli C, Laird MR, Williams KP, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 2017;45:W30-W35.
- Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 2007;35:W52-W57.
- Abby SS, Neron B, Menager H, Touchon M, Rocha EP. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-cas systems. *PLoS ONE.* 2014;9:e110726. doi:10.1371/journal.pone.0110726.
- Couvin D, Bernheim A, Toffano-Nioche C, et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for cas proteins. *Nucleic Acids Res.* 2018;46:W246-W251.
- Blom J, Kreis J, Spanig S, et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 2016;44:W22-W28. doi:10.1093/nar/gkw255.
- Smit AFA, Hubble R, Green P. RepeatMasker Open-4.0. Updated 2015. <http://www.repeatmasker.org/>.
- Auch AF, Klenk HP, Goker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci.* 2010;2:142-148.
- Taghavi S, van der Lelie D, Hoffman A, et al. Genome sequence of the plant growth promoting endophytic bacterium *Enterobacter* sp. 638. *PLoS Genet.* 2010;6:e1000943.
- Ali D, Duan J, Charles TC, Glick BR. A bioinformatics approach to the determination of genes involved in endophytic behaviour in *Burkholderia* spp. *J Theor Biol.* 2014;343:193-198. doi:10.1016/j.jtbi.2013.10.007.
- Kallimanis A, Labutti KM, Lapidus A, et al. Complete genome sequence of *Arthrobacter phenanthrenivorans* type strain (Sphe3). *Stand Genomic Sci.* 2011;4:123-130.
- Pinski A, Betekhtin A, Hupert-Kocurek K, Mur LAJ, Hasterok R. Defining the genetic basis of plant-endophytic bacteria interactions. *Int J Mol Sci.* 2018;20:1947. doi:10.3390/ijms20081947.
- Jasim B, Jimtha JC, Jyothis M, Radhakrishnan E. Plant growth promoting potential of endophytic bacteria isolated from *Piper nigrum*. *J Plant Growth Regul.* 2013;71:1-11.
- Taghavi S, Garafola C, Monchy S, et al. Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Environ Microbiol.* 2009;75:748-757.
- Hayat S, Ahmad A. *Salicylic Acid—A Plant Hormone*. Dordrecht, The Netherlands: Springer; 2007.