République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Ghardaia

Faculté des Sciences de la Nature et de Vie et Sciences de la Terre

Département de Biologie

Mémoire présenté en vue de l'obtention du diplôme de

MASTER

Filière : Sciences biologiques

Spécialité : Microbiologie appliquée

Par : BENGAID Soumia & OULAD KOUIDER Daouia

Thème

In silico genomic analysis of *Saccharothrix algeriensis* DSM 44581

Soutenu publiquement, le 03/ 07 / 2024, devant le jury composé de :

M. MAHAMEDI A E	Maitre de Conférences B	Univ. Ghardaia	Président
M. DIF G	Maitre de Conférences A	ENS Laghouat	Directeur de mémoire
M^{elle} DJEMOUAI N	Maitre de Conférences A	Univ. Ghardaia	Co-directeur de mémoire
M. BAKELI A	Maitre de Conférences B	Univ. Ghardaia	Examinateur

Année universitaire : 2023 - 2024

DEDICATIONS

I have made this success and I confer my profound gratitude in memorial of my late beloved parents and to my happy family.

I am grateful and give special thanks to my dear husband, who helped, and supported me. He receives on this occasion, the expression of my affectionate feelings as well as my gratitude for his patience throughout the period of the completion of this dissertation. Many thanks to my precious children Hiba Tallah, El Mountassir Billah and El Moutassim Billah.

May Allah bless them for supporting me.

My appreciation and thanks go to my family, especially to my dearest sister Souhila and brother Faycel for pushing and supported me.

I respectfully provide my improvement and consecrate this research

To the remarkable collectives who supported me in making and achieving our dreams.

Soumia

DEDICATIONS

First and foremost, praise be to God alone, and may blessings and peace be upon His Messenger And worship our Prophet Muhammad and his family and all his companions.

I present my success to my parents. They were my support since I was young until I reached this level, and they were with me in all my circumstances. I want to offer special thanks to my dear mother, Nabaa Hanan.

Thanks to her upbringing and support for me, I have reached this level without forgetting my dear father, for his grace upon me, which cannot be valued at any price. Likewise, I extend my greetings to my sisters, each one by the name of my religion, with whom I grew up and raised, and the large family, my aunts, uncles, and my grandmother and my mother's mother. I would like to thank them very much.

I also thank my husband, Aziz Ahmed Ibrahimi, who was my friend and companion in this life and supported me during the final period of my studies. I wish him success.

I would like to thank my husband Uzair's family, especially my husband's mother. I ask God Almighty to have mercy on her and make her abode in Paradise. Her death was a tragedy for us and our family. I ask God to forgive her and have mercy on her, Amen.

We are grateful to my dear friends for their invaluable support and encouragements throughout my studies. Their presence and close friendship have made this journey more fulfilling, especially my beloved friend, mother, sister, and colleague Soumia BENGAID. There are no words to express my thanks and gratitude to her, who was my second support after my family, and I wish success, and I hope that God Almighty fulfills all her aspirations.

We would also like to thank our research advisors and mentors for their invaluable guidance, experience and insights throughout this research journey. Their mentor-ship and commitment to our growth have made us the researchers we are today.

Daouia

Acknowledgements

Firstly, we praise ALLAH for his blessings in our work and for giving us the courage to complete this work.

We want to extend our warmest thanks to our supervisor Mr DIF G., Associate professor A from the Department of Natural Sciences, Higher Normal School of Laghouat and Miss DJEMOUAI N., Associate professor A at the Faculty of Nature and Life Sciences and Earth Sciences from the University of Ghardaia, for their encouragement, patience, guidance, help, and sound scientific advice.

We are honored to have Mr MAHAMEDI A. E, Associate professor B at the Faculty of Nature and Life Sciences and Earth Sciences from the University of Ghardaia, to be a president of this jury and to Mr BAKELLI A., Associate professor B at the same university for accepting to examine our work.

A big thank you to everyone who participated directly or indirectly in the accomplishment of this modest work.

We would like to extend our heartfelt appreciation to Miss DJEMOUAI N. and Miss BAALI F. for helping us and supporting our work, inspiring us and making us love researching in the field of biology.

Our sincere thanks to our colleague's with whom we made such good memories

Special thanks to our university of Ghardaia and all staff and members, from the headmaster to the security grade.

Thankyou

ملخص

تهدف هذه الدراسة إلى إجراء تحليل جينومي للجنيوم الكلي لسلالة من السلالة الشعاعية الاكتنوبكتيرية معزولة من منطقة صحراوية في الجزائر .والغرض من ذلك هو تحديد جينات محددة تؤدي دورا في تعزيز نمو النباتات وتوفير الحماية من العوامل المسببة للأمراض .في هذه الدراسة، تم معزولة سلالة تدعى *Saccharothrix algeriensis* DSM 44581 من عينة التربة التي جمعت في بستان النخيل في أدرار، الجزائر .وقد أجري تحليل تصنيف الجينوم على المنصة الرقمية TYGS، وتحققنا من التشابه الوراثي بين هذه السلالة والنوع *Saccharothrix australiensis* DSM 43800 في التحليل الفوليولجيني والتحليل الفولوجينتيك .وتُجرى التحليلات العامة للسمات ، DFAST بما في ذلك مجموع طول الجينوم والنسبة المئوية ل GCوCODs .وركز تحليل الجينومات لهذا الإجهاد على دراسة الجينات المرتبطة بفئات الجينات الوظيفية(COGs) ، وشمل هذا التحليل شرحاً جينياً باستخدام خادم RAST و BV-BRC للنظر إلى الجينات خريطة دائرية، وتلاه التنبؤ بالجينات باستخدام برنامج Prokka في GALAXY لتحديد الجينات الداخلة في الآليات التي تعزز نمو النباتات .وبالإضافة إلى ذلك، فإنantiSMASH ينطوي على جينات ذات صلة بالحد من الإجهاد، مثل ACC-deaminaseومختلف أشكال حماية النبات مثل الغلوتاثيون، فضلاً عن العديد من سيدروفور (YclQوYecD) وتشترك جينات مثل DhbC و ResXفي التركيب الأحيائي لمواد نمو النبات .ويُظهر هذا الضغط تكيفاً استثنائياً مع الظروف المالحة للغاية .وبالإضافة إلى ذلك، حددنا العديد من الجينات التي قد تكون مرتبطة بعمليات متعددة من اعمال مضادات الأكسدة الأنزيمية .ويمكن لهذه الآليات أن تقلل من التكوّن المفرط لأنواع الأكسجين التفاعلية الناجمة عن الإجهاد الملحي .وأُبلغ عن وجود جينات تشارك في إنتاج/تدهور مركبات الكربون الهيدروفلورية) مثل ف (IlvB1 وAcuC، واستشعارات القُصَر (SsgB و KDAو Soj)وإنتاج/تعديل الهرمونات النباتية مثل (PotB و PuuPو GabT)وأكدت الدراسات المقارنة للإحصاءات الموجزة في برنامج (GALAXY) Roary نتائج الدراسات الفيزيائية والفيلوجينومية . وتبين النتائج الواسعة النطاق أن سلالة *S. Algeriensis* DSM 44581 لها أهمية كبيرة بوصفها خيارا واعدا للغاية لتعزيز تنمية النباتات وتطبيقات التكنولوجيا الأحيائية.

الكلمات الرئيسية : *Saccharothrix algeriensis* , أدوات المعلوماتية البيولوجية، تحليلات الجينوم، البكتيريا المعززة لنمو النباتات.

Résumé

Cette étude vise à effectuer une analyse génomique d'une souche actinobactérienne isolée à partir une région saharienne de l'Algérie. L'objectif est d'identifier des gènes spécifiques qui jouent un rôle dans l'amélioration de la croissance des plantes et dans la protection contre les agents pathogènes. Dans cette étude, une souche actinobactérienne appelée *Saccharothrix algeriensis* DSM 44581 isolée à partir d'un échantillon de sol collecté dans une palmeraie à Adrar, en Algérie. L'analyse de la classification génomique a été réalisée sur la plateforme numérique TYGS et nous avons vérifié la similarité génétique entre cette souche et l'espèce actinobactérienne *Saccharothrix australiensis* DSM 43800 dans l'analyse phylogénétique et phylogénomique. Les analyses des caractéristiques générales sont effectuées dans DFAST, y compris la longueur totale du génome, le GC% et les COD. L'analyse génomique de cette souche s'est concentrée sur l'étude des gènes associés aux catégories de gènes fonctionnels (COG), cette analyse a impliqué l'annotation génétique en utilisant le serveur RAST et BV-BRC pour la vue circulaire du génome, suivie de la prédiction des gènes en utilisant le programme Prokka dans GALAXY pour identifier les gènes impliqués dans les mécanismes qui favorisent la croissance des plantes. En outre, les gènes impliqués dans l'antiSMASH et liés à la réduction du stress, tels que l'ACC-désaminase et diverses formes d'osmoprotecteurs comme le glutathion, ainsi que plusieurs sidérophores (YecD et YclQ), ont également été mis en évidence. Des gènes tels que DhbC et RecX sont impliqués dans la biosynthèse de substances de croissance végétale. La souche présente une adaptation exceptionnelle à des environnements extrêmement salés. En outre, nous avons identifié de nombreux gènes qui peuvent être liés à de multiples processus enzymatiques antioxydants. Ces mécanismes peuvent réduire la formation excessive d'espèces réactives de l'oxygène causée par le stress salin. Les gènes impliqués dans la production/dégradation des COV (par exemple IlvB1 et AcuC), le Quorum sensing (SsgB, kDa et Soj) et la production/modulation des phytohormones (par exemple PotB, PuuP et gabT) ont été trouvés. Les études comparatives des statistiques du programme Roary (GALAXY) ont confirmé les résultats des études phylogénétiques et phylogénomiques. Les résultats détaillés démontrent que la souche *S. algeriensis* DSM 44581 revêt une importance significative en tant que choix très prometteur pour améliorer le développement des plantes et pour des applications biotechnologiques.

Mots clés: *Saccharothrix algeriensis*, Outils bioinformatiques, Analyses génomiques, Bactéries favorisant la croissance des plantes.

Abstract

This study aims to perform a genomic analysis of the whole genome of an actinobacterial strain isolated from a Saharan region of Algeria. The study allow us identify specific genes that play an important role in enhancing plant growth and providing protection against disease-causing agents. In this study, an actinobacterial strain called *Saccharothrix algeriensis* DSM 44581 isolated from a soil sample collected from palm grove in Adrar province, Algeria. The genomic classification analysis was conducted on the TYGS digital platform (16S RNA) and phylogenomic (whole genome). The general features analyses are conducted in DFAST, including the total length genome, GC % and CODs. Genomics analysis of this strain focused on studying genes associated with functional gene categories (COGs), this analysis involved genetic annotation using the RAST server and BV-BRC for circular genome view, followed by gene prediction using the Prokka program in GALAXY to identify genes involved in mechanisms that promote plant growth. In addition, the antiSMASH involved genes related to the reduction of stress, such as ACC-deaminase and various forms of osmoprotection such as glutathione, as well as several siderophores (YecD and YclQ), were also emphasized. Genes such as DhbC and RecX are involved in the biosynthesis of plant growth substances. The strain exhibits exceptional adaptation to extremely salty settings. In addition, we have identified numerous genes that may be linked to multiple enzymatic antioxidant processes. These mechanisms can reduce the excessive formation of reactive oxygen species caused by salt stress. Genes involved in VOCs production/degradation (e.g IlvB1 and AcuC), Quorum sensing (SsgB, kDa and Soj) and phytohormones production/modulation (e.g PotB, PuuP and gabT) were reported. The comparison studies of summary statistics in the Roary program (GALAXY) confirmed the results of phylogenetic and phylogenomic studies. The extensive results demonstrate that strain *S. algeriensis* DSM 44581 holds significant importance as a highly promising choice for enhancing plant development and in biotechnological applications.

Keywords: *Saccharothrix algeriensis*, Bioinformatics tools, Genomic analyses, Plant Growth-Promoting Bacteria.

Glossary

Whole genome sequencing (**WGS**) : As full genome sequencing, is the process of determining the entirety, or nearly the entirety, of the DNA sequence of an organism's genome at a single time. This entails sequencing all of an organism's chromosomal.

Plant growth-promoting bacteria (**PGPB**)**:** Is bacteria that can enhance plant growth and protect plants from disease and abiotic stresses through a wide variety of mechanisms; those that establish close associations with plants, such as the endophytes, could be more successful in plant growth promotion. Several important bacterial characteristics, such as biological nitrogen fixation, antibiotic, phosphate solubilization, ACC deaminase activity, and production of siderophores and phytohormones, can be assessed as plant growth promotion (PGP) traits.

Core-genome: It is part of the pan-genome, which means similar genes are found in all genomes and are compared. It expresses the basic genes that must be present in the strains being compared and the pool of genes shared by all the strains of the same bacterial species.

Pan-genome (pangenome or supragenome): The whole set of genes from all strains within a clade in molecular biology and genetics. It's the sum of a clade's genomes in a broader perspective. The pan-genome is divided into three sections: a "core pangenome" that contains genes found in all individuals, a "shell pangenome" that contains genes found in two or more strains, and a "cloud pangenome" that only has genes seen in one strain.

DDH: DNA hybridization distance is a measure used to assess the genetic similarity between two bacterial species. The value of the hybrid DDH often compared to the value of the DDH obtained by hybridizing a reference genome with itself. DDH values less than or equal to 70% are considered an indication that the tested organism belongs to a species different from the type strain(s) used as a reference.

Bootstrap: The probability of getting the same tree or phylogenetic relationship if you redraw tree using the same data 1000 or 100 bootstrap value means if you draw the tree using the same data sets you get a particular number at the node that specifies the correctness of the tree.

Quorum sensing (QS): A cellular communication mechanism used by bacteria to regulate the expression of genes involved in virulence, competition, pathogenicity and resistance. It involves the production, detection and response to extracellular signaling molecules called antibiotic autoinducers.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Figures

ABBREVIATION LIST

Figures

Abbreviation list

16S rRNA	16s Ribosomal Ribonucleic Acid
ACC	Aminocycopropane-1-Carboxylic Acid
AntiSMASH	Antibiotics and Secondary Metabolite Analysis SHell
BLAST	Basic Local Alignment Search
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
C	Cytosine
COGs	Clusters of Orthologous Genes.
dDDH	Digital DNA-DNA Hybridise
DFAST	DDBJ Fast Annotation and Submission Tool
DNA	Deoxyribo Nucleic Acid
FAO	Food and Agriculture Organization of The United Nations
FASTA	Fast Alignment Search Tool
G	Guanine
Gb	Giga base
GBDP	Genome BLAST Distance Phylogeny
HGAP	Hierarchical Genome Assembly Process
HGP	Human Genome Project
IAA	Indole-3-Acetic Acid
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NGS	New Generation Sequencing
Pacbio	Pacific Biosciences
PGP	Plant Growth Promotion
PGPB	Plant Growth Promoting Bacteria
PPI	Protein-Protein Interaction
Prokka	Rapid Porkaryotic genome Annotation
RAST	Rapid Annotation using Subsystem Technology
rRNA	Ribosomal Ribonucleic Acid
tRNA	Transfer RNA
TYGS	Type Strain Genome Server
WGS	Whole Genome Sequencing

Introduction

Global food security is a major preoccupation as the future demand for food will increase by 60% (Serraj *et al*., 2018). Furthermore, advocating for encouraging the practices of sustainable agriculture with intensive farming these, include the cultivation of high-yield crops, and the application of fertilizers and pesticides, is a strategy used nowadays. Food production has been on the rise, but there are occasional undesirable consequences. To mitigate the risks associated with the use of fertilizers and pesticides, scientists have sought out methods that enable increased production while simultaneously safeguarding human health and the environment. Among all these methods, the biological solution, particularly the microbiological one that is used with bacterial strains, can be more advantageous for plants. These bacteria, referred to as Plant Growth-Promoting Bacteria (PGPB), can supply the essential nutrients required for plant growth and protect plants against phytopathogens (Barka *et al*., 2016).

Actinobacteria are a group of bacteria that have a high G+C DNA content and are classified as Gram-positive bacteria. They are found in both aquatic and terrestrial environments and are one of the biggest bacterial groups. A significant number of them possess a comprehensive secondary metabolism and generate approximately two-thirds of all naturally occurring antibiotics currently employed in therapeutics, along with several chemicals that exhibit anticancer, anthelmintic, antifungal and other properties. As a result, these bacteria play a significant role in the fields of biotechnology, medicine, and agriculture (Barka *et al*., 2016).

In our work, we selected a strain whose phylogenetic analysis, based on 16S rDNA sequences, as well as its morphological, chemotaxonomic, and physiological traits, confirms its classification within the *Saccharothrix* genus. Furthermore, it is identified as a distinct species that can be easily differentiated from all previously identified *Saccharothrix* species. The isolate is proposed to be named *Saccharothrix algeriensis* sp. nov., with the type strain designated as SA 233T (=NRRL B-24137T=DSM 44581T) (Zitouni *et al*., 2004). The genus *S. algeriensis* demonstrates antimicrobial properties against bacteria, filamentous fungi, and yeasts. The antifungal activity of *S. algeriensis* is significant in combating phytopathogenic filamentous fungi (Merrouche *et al*., 2017). The use of biological methods to manage plant diseases presents a more advantageous option compared to chemical control in order to achieve sustainable agriculture (Asif *et al*., 2023).

In this study, we performed a genomic analysis using high-accuracy, academically approved computer tools and programs for the genome of a strain named *Saccharothrix algeriensis* DSM 44581 that has not previously been subjected to an in-depth genomic study. Through this research, we target within the genome of this strain that have a positive role in stimulating and supporting plant growth, that is, to predict the importance of the strain as biological vaccines that enhance crop growth and protect them from diseases.

Our Manuscript consists of parts, the first chapter focuses on a bibliographical survey of Actinobacteria in general (ecology, taxonomy, and significance) and specifically on the genus *Saccharothrix* and its species *Saccharothrix algeriensis*. It also delves into genomic analyses, which encompasses bioinformatic science and, genome annotation and the PGPB. The second chapter comprises the materials and methods employed for genomic analysis and tools used for the NCBI website, which functions as an extensive repository of gene information from various sources. We practiced in many servers, such as TYGS, which is a web server that enables the efficient categorization of bacteria using genome sequences. DFAST for general features of genes. The BV-BRC offers computational tools for data analysis. The RAST server provides a fully annotated service for the identification and annotation of genes within bacterial genomes. The Prokka in GLAXSY is useful for identifying virulence, pathogenicity, and infectivity-related genes. Using antiSMASH server in order to identify sets of biosynthetic genes encoding secondary metabolites is a must. Also, the Roary program was used for comparative analyses. The third chapter focuses on presenting the results and outcomes of various bioinformatic tools and engaging in a discussion about them. The conclusion, highlights the results that were obtained and provides suggestions for potential future perspectives.

Bibliographic review

1. Actinobacteria

1.1. History

Ferdinand Cohn was the first to characterize an Actinomycete in 1875. In 1878, Harz described *Actinomyces bovis*, a parasitic fungus found in a bovine jaw infection (Garrity and Cole, 2007). Waksman categorized Actinomycetes' evolutionary history into four major groups. With the contributions of Kraisky, Cohn, Waksman, and Curtis, the second phase (1900-1919) was associated with the identification and investigation of soil Actinomycetes. Furthermore, Waksman, Lieske and Krassilnikov's studies helped advance our knowledge of actinobacterial strains between 1919 and 1940. Antibiotics were synthesized by Actinomycetes, with Waksman's name inextricably linked to it. The name Actinomycetes was changed to Actinobacteria (Ouargli, 2018).

1.2. General features of Actinobacteria

Actinobacteria is a prominent phylum among bacteria, characterized by Gram-positive staining that have a high G+C content in their DNA. This bacterial group exhibits a diverse array of forms and sizes, ranging from coccoid to fragmenting hyphae, along with distinct metabolic and physiological characteristics. In addition, actinobacterial strains have various lifestyles (Ventura *et al*., 2007). Actinobacteria are primarily aerobic, although there are some cases where this is not the case. In addition, they can exhibit heterotrophic or chemoautotrophic characteristics, while the majority are chemoheterotrophic and possess the ability to use a diverse range of nutritional sources, including complex polysaccharides. Actinobacteria are predominantly mesophilic, thriving best at temperatures ranging from 25 to 30°C. Thermophilic Actinobacteria can thrive in temperatures that fall between the range of 50 to 60°C. Actinobacteria mostly thrive in soil environments characterized by a pH level that is neither acidic nor alkaline (Barka *et al*., 2016).

1.3. Classification of Actinobacteria

The phylum *Actinobacteria* is currently classified into six subclasses: *Actinobacteria*, *Acidimicrobiia*, *Coriobacteriia*, *Nitriliruptoria*, *Rubrobacteria*, and *Thermoleophilia*. These subclasses are additionally separated into various orders and suborders (**Figure 1**) (Barka *et al*., 2016).

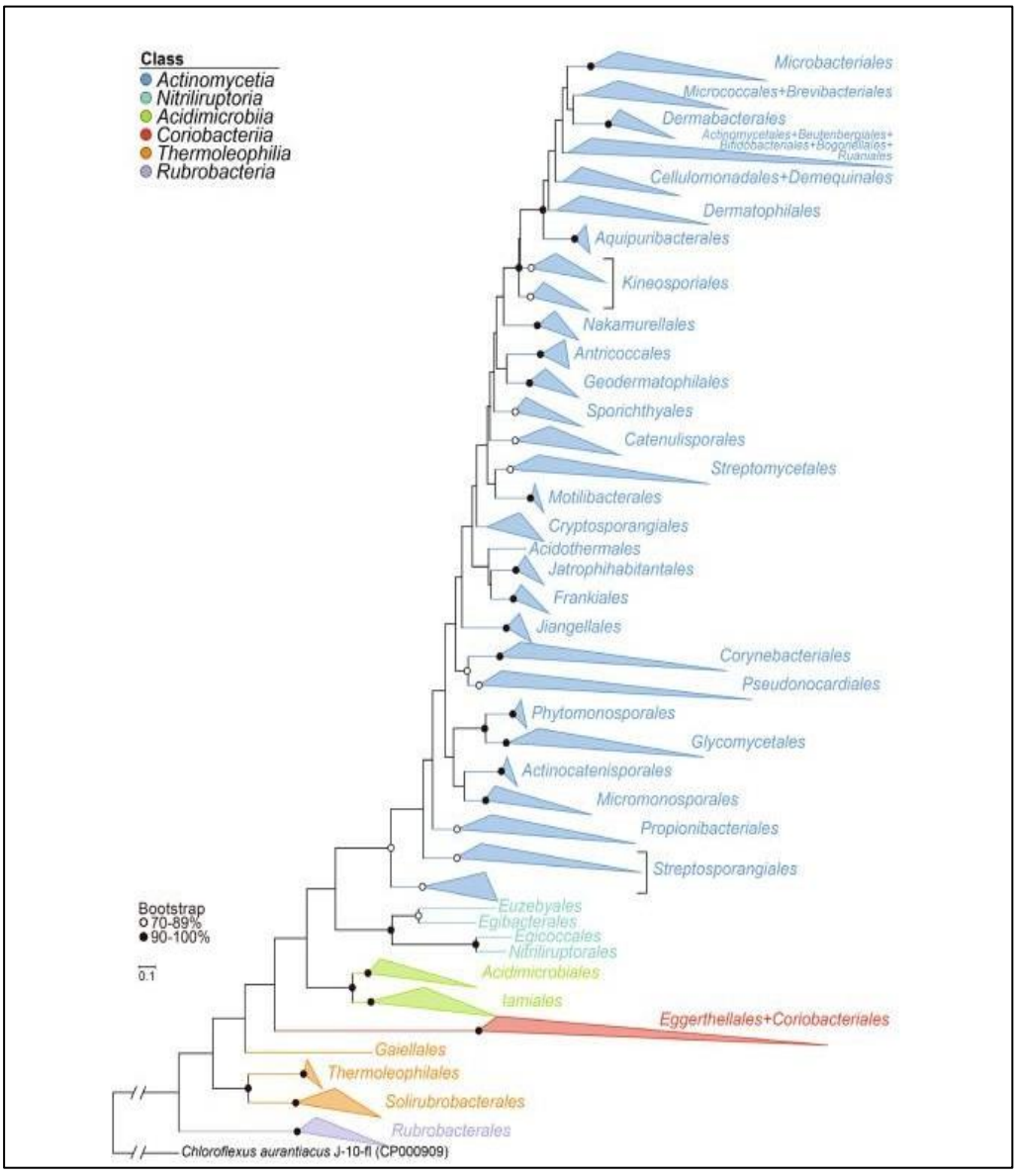


Figure 1. Phylogenetic tree for 98 actinobacterial species whose genomes have been sequenced (Salam *et al*., 2020).

1.4. Morphological characteristics

Mycelial fragmentation is viewed as a distinct type of asexual reproduction nevertheless, Actinobacteria, which predominantly exhibit mycelial lifestyles typically reproduce through the formation of asexual spores. Actinobacteria reveal a diverse range of morphologies, primarily characterized by the presence or absence of a substrate mycelium or aerial mycelium, the color of the mycelium, the synthesis of diffusible melanoid pigments, and the structure and appearance of their spores (Barka *et al*., 2016) (**Figure 2**).

.

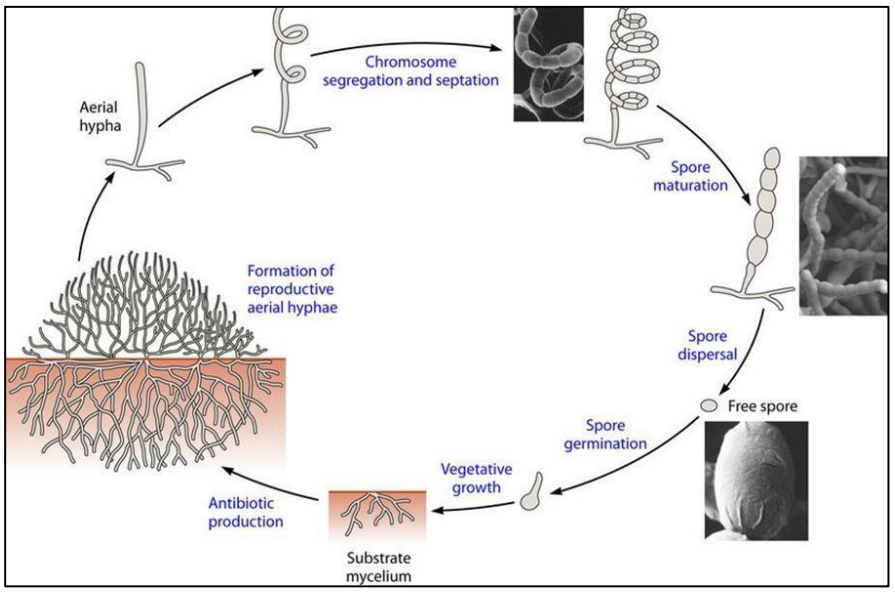


Figure 2. Schematic representation of the life cycle of sporulating Actinobacteria (Barka *et al*., 2016).

The substrate mycelium originates from the sprouting of a germinating spore. The substrate mycelium often exhibits monopodial branching, although there are occasional cases of dichotomous branching observed in Actinobacteria, such as *Thermoactinomyces*. In contrast, individuals belonging to the *Micromonosporaceae* family generate a substantial substrate mycelium, whereas their aerial mycelium is either nonexistent or underdeveloped (Barka *et al*., 2016).

1.5. Molecular characteristics

The morphological and chemical classification, which is a polyphasic approach to identifying Actinobacteria, has been called into question due to molecular taxonomic data, much of which were obtained thanks to the rapid advancement of genome sequencing. Remarkably, current molecular investigations have led to the reclassification of several species that were previously incorrectly assigned to specific taxonomic groupings (Zhi *et al*., 2009).

1.6. Distribution of Actinobacteria

Actinobacteria are now understood to be global microorganisms present in almost every habitat on Earth (Lewin *et al*., 2016). The natural ecosystem habitats of soil, rhizosphere soil, plant tissues, hypersaline soil, freshwater, marine, sponges, volcanic cave-hot spot, air, insect stomach, earthworm castings, goat feces are highly populated with Actinobacteria (Selim *et al*., 2021).

1.7. Importance of Actinobacteria

The *Actinobacteria* phylum, consists of bacteria that are both phylogenetically and physiologically varied. These bacteria play a crucial role in maintaining the functioning of Earth's ecosystems. Actinobacteria play a role in the global carbon cycle by breaking down plant materials, both as independent organisms and as symbionts of herbivorous animals. Actinobacteria, with their advanced cellulolytic capabilities and wide range of chemical properties, show great potential for the bioenergy sector. More precisely, the enzymes produced by these organisms play a crucial role in efficiently breaking down large amounts of cellulosic plant biomass into easily digestible sugars, which can subsequently be turned into biofuels on a large scale. Moreover, using their capacity to produce various types of small molecules holds promise for the production of specialized biofuels (Lewin *et al*., 2016). Actinobacteria have been found to be a valuable reservoir of bioactive secondary metabolites, including antibiotics, biopesticides, plant growth regulators, antitumor agents, antiviral compounds, pigments, enzymes, enzyme inhibitors, anti-inflammatory substances, and biosurfactants. In addition, the Actinobacteria found in marine animals and plants, macroalgae, cyanobacteria and lichens have a high capacity to create active metabolites that serve as chemical defenses. These metabolites help protect the host from predators and microbial infections and are well regarded for their wide range of expertise and demonstrated success in generating innovative bioactive chemicals (Chen *et al*., 2021). Actinobacteria are responsible for producing 70% of the over 22000 identified microbial secondary metabolites (Subramani and Aalbersberg, 2012). The Actinobacteria provide significant potential advantages for humans as reservoirs of innovative antibiotics, antifungals, and anticancer medicines. Due to the rapid advancements in genomics, synthetic biology, and ecology, as well as the urgent need for new antimicrobial compounds to address antimicrobial resistance, the study of Actinobacteria is an extremely dynamic research area (Barka *et al*., 2016).

Actinobacteria have great potential for pest management and promoting plant growth. The presence of Actinobacteria that exhibit Plant Growth-Promoting (PGP) traits, such as nitrogen fixation, phosphate solubilization, and production of phytohormones like auxins, gibberellins, and cytokinins, as well as siderophores production, enhances the importance of actinobacterial taxa in agriculture (Palaniyandi *et al*., 2013).

2. Genus *Saccharothrix*

2.1. Characteristics and taxonomic position

The genus *Saccharothrix* was first described and initially characterized by Labeda and coauthors in 1984, with the reference species *Saccharothrix australiansis* (Labeda *et al*., 1984). The members in this genus are Gram-positive, heterotrophic chemoorganotrophs that are saprophytic and strictly aerobic.

Saccharothrix is capable of inhabiting several environments, including wastewater, desert, salty and alkaline soils, as well as ocean sediments and mineral deposits (Schippers *et al*., 2002; Hozzein *et al*., 2004; Zitouni *et al*., 2005). This genus, although rare, is distinct from other Actinobacteria. The proportion of *Saccharothrix* isolates in relation to the total Actinobacteria is less than 0.5%, but it reaches 15% in Saharan soil samples (Sabaou *et al*., 1998). Mycelial filaments undergo random fragmentation, often exhibiting a "zig-zag" pattern, leading to the development of elongated chains of spores **(Figure 3)**. The spores are stationary and possess a morphology that is either cylindrical (1 to 2 µm x 0.7 to 1 µm) or egg-shaped. After the creation of this genus, various species of *Saccharothrix* were identified using advanced techniques in digital and molecular taxonomy.

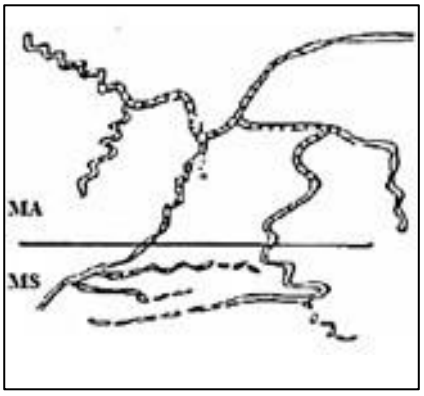


Figure 3. Micromorphology of *Saccharothrix* with MA as aerial mycelium and MS as substrate mycelium (Bergey, 1989).

2.2. Species belonging to the genus *Saccharothrix*

The 16S rDNA sequencing and DNA-DNA hybridization have helped assign *Saccharothrix* species. Currently, there are a total of 25 species within the genus *Saccharothrix* **(Table I)**.

Table I. Different species of the genus *Saccharothrix* in chronological order according to the National Center for Biotechnology Information and genome size (http://www.ncbi.nlm.nih.gov/Taxonomy)

Species		**Types strains**	**Reference**	**Genome size (Mb)**
Saccharothrix australiensis		NRRL 11239, NBRC 14444, JCM 3370, DSM 43800, ATCC 31497, IFO 14444, VKM Ac-894	Labeda *et al*. (1984)	7816
Saccharothrix coeruleofusca		NRRL B-16115, DSM 43679, NBRC 14520, JCM 3313, INA 1335, IFO 14520, ATCC 35108	Labeda and Lechevalier (1989)	7779
Saccharothrix espanaensis		NRRL 15764, ATCC 51144, DSM 44229	Labeda and Lechevalier (1989)	9361
Saccharothrix longispora		NRRL B-16116, DSM 43749, NBRC 14522, JCM 3314, ATCC 35109, IFO 14522, INA 10222, VKM Ac-907	Labeda and Lechevalier (1989)	8446
Saccharothrix mutabilis	**subsp. *capreolus***	NRRL 2773, DSM 40225, NBRC 12847, JCM 4630, JCM 4248, IFO 12847, ATCC 23892	Grund and Kroppenstedt (1989)	***
	subsp. *mutabilis*	NRRL B-16077, ATCC 31520, DSM 43853, IFO 14310, IFM 240	Labeda and Lechevalier (1989)	***
Saccharothrix syringae		NRRL B-16468, DSM 43886, NBRC 14523, JCM 6844, INA 2240, ATCC 51364, IFO 14523, VKM Ac-1858	Labeda and Lechevalier (1989)	1093
Saccharothrix texasensis		NRRL B-16134, DSM 44231, NBRC 14971, JCM 9113, IFO 14971, ATCC 51593, VKM Ac-1968	Labeda and Lechevalier (1989)	9178
Saccharothrix algeriensis		**NRRL B-24137, NBRC 101915, JCM 13242, DSM 44581**	**Zitouni *et al*. (2004)**	**6879**

Saccharothrix xinjiangensis	NBRC 101911, JCM 12329, AS 4.1731	Hu *et al.* (2004)	***
Saccharothrix violaceirubra	NBRC 102064, JCM 16955, KCTC 19326	Otoguro *et al.* (2009)	7354
Saccharothrix variisporea	NRRL B-16296, DSM 43911, NBRC 14104, JCM 3273, IFO 14104, ATCC 31203, KCC A-0274, D409-5	Kim *et al.* (2011)	9409
Saccharothrix yanglingensis	KCTC 19722, CGMCC 45627	Yan *et al.* (2012)	8286
Saccharothrix hoggarensis	DSM 45457, CCUG 60214	Boubetra *et al.* (2013)	****
Saccharothrix saharensis	DSM 45456, CCUG 60213	Boubetra *et al.* (2013)	8925
Saccharothrix carnea	DSM 45878, CGMCC 4.7097	Liu *et al.* (2014)	8918
Saccharothrix lopnurensis	YIM LPA2h, KCTC 39545, CGMCC 4.7246	Li *et al.* (2015)	***
Saccharothrix ecbatanensis	DSM 45486, UTMC 00537, CCUG 63021	Hamedi *et al.* (2015)	9689
Saccharothrix tamanrassetensis	DSM 45947, CECT 8640	Boubetra *et al.* (2015)	8054
Saccharothrix isguenensis	DSM 46885, CECT 9045	Bouznada *et al*. (2016)	***
Saccharothrix stipae	JCM 30560, ACCC 19714	Guo *et al.* (2016)	***
Saccharothrix ghardaiensis	DSM 46886, CECT 9046	Bouznada *et al*. (2017)	***
Saccharothrix tharensis	TD-093, KCTC 39724, MCC 2832	Ibeyaima *et al.* (2018)	***
Saccharothrix deserti	B8144, CGMCC 4.7490, KCTC 49001	Liu *et al.* (2020)	1083
Saccharothrix obliqua	SC076, TBRC 14540, NBRC 115117	Azad *et al.* (2022)	8040
Saccharothrix luteola	CCTCC: AA 2020037, JCM 34800, NEAU-S10	He *et al.* (2022)	1031

***: Genome not sequenced.

2.3. Biological functions and actions of *Saccharothrix*

Saccharothrix are uncommon Actinobacteria that can synthesize various antibiotics with distinct structures and biological properties. For instance, *Saccharothrix xinjiangensis* NRRL B-24321 produces the antibacterial compound Tianchimycins, while *Saccharothrix mutabilis* R516-16 produces the antiviral compound Fluvirucin. Studies have found that the dry soils of the Algerian Sahara contain a variety of species from the *Saccharothrix* genus, including discovered species such as *Saccharothrix algeriensis* NRRL B-24137T, *S. hoggarensis* DSM 45457T, *S. saharensis* DSM 45456T, *S. tamanrassetensis* DSM 45947T, *S. isguenensis* DSM 46885T, and *S. ghardaiensis* DSM 46886T. Furthermore, numerous studies have demonstrated that *Saccharothrix* strains found in these soils can synthesize both novel and established antibiotics, including Dithiolopyrrolones (Lamari *et al*., 2002; Bouras *et al*., 2008; Merrouche *et al*., 2011), Anthracyclines (Zitouni *et al*., 2004), and Chloramphenicol (Aouiche *et al*., 2012).

2.4. *Saccharothrix algeriensis*

The strain *S. algeriensis* was isolated from a soil sample collected from a palm grove in Adrar, Algeria. It was discovered and categorized as a member of the genus *Nocardiopsis* by Boudjella in 1994. It was later categorized within the genus *Saccharothrix*, which exhibits the same micromorphology. After conducting a phylogenetic study using 16S rDNA sequencing and DNA-DNA hybridization, followed by a numerical taxonomy study using 77 physiological tests. The results showed distinct differences between *S. algeriensis* and other known species of the genus. The microorganism was designated as *S. algeriensis* (Zitouni *et al*., 2004) and has been stored in two international collections: Leibniz-Institut DSMZ (Leibniz-Institut Deutsche Sammlung von Mikroorganismen und Zellkulturen) and ARC Culture (Agricultural Research Service Culture) with the respective identification numbers DSM 44581 and NRRL B-24137, respectively. The 16S RNA sequence has an accession number of AY054972 (Zitouni *et al.*, 2004). The taxonomic classification of *S. algeriensis* NRRL B-24137 is presented as follows :

Domain: Bacteria
Phylum: *Actinomycetota*
Class: Actinomycetes
Order: Pseudonocardiales
Family: Pseudonocardiaceae

Genus: *Saccharothrix*

Species: *Saccharothrix algeriensis* (Zitouni *et al*., 2004)

S. algeriensis exhibits a high production of yellow-orange aerial mycelium on Bennet, ISP2, and ISP5 media, which breaks apart into rod-shaped spores in terms of morphology. The substrate mycelium exhibits vibrant hues of yellow, orange-yellow, or yellowish brown. A vivid yellow soluble pigment is produced, but no melanoid pigments are formed (Zitouni *et al*., 2004). *S. algeriensis* is a bacterium that was first isolated in Algeria, hence its name. The microorganism is aerobic, Gram-positive, and catalase-positive. The hydrolysates of the cell wall contained meso-diaminopimelic acid but did not contain glycine. On the other hand, the hydrolysates of the entire cell contained galactose, mannose, and rhamnose, in addition to glucose and ribose. These components are characteristic of cell wall type III. The most abundant fatty acid is iso-C16 : 0, accounting for 31.26% of the total. It is followed by iso-H-C16: 0 at 14.00%, iso-2-hydroxy-C16 : 0 at 10.44%, and iso-C15 : 0 at 10.06%.

2.5. Biological activities of *Saccharothrix algeriensis*

This genus is being studied and discovered that it generates bioactive substances that are part of the dithiolopyrrolone category of antibiotics, which are distinguished by having a distinct pyrrolinonodithiole core. Dithiolopyrrolones are recognized for their capacity to exhibit bacteriostatic effects on both Gram-positive and Gram-negative bacteria, as well as has antifungal properties. The discovery of the anticancer characteristics of this class of antibiotics sparked significant attention (Bouras *et al*., 2008; Saker *et al*., 2014; Merrouche *et al*., 2019). *S. algeriensis* NRRL B-24137 has been identified as a promising biocontrol agent for combating *Botrytis cinerea* and other plant infections (Muzammil *et al*., 2013). This strain can be used as a biocontrol agent to combat *Fusarium* cotton wilt, which is caused by *Fusarium oxysporum*. This application of *Saccharithrix algeriansis* enhances cotton growth characteristics and increases production. Also, it can be regarded as an optimal and efficient biocontrol agent for combating soil-borne diseases and promoting the growth of cotton plants. Biological management of plant diseases is a favorable alternative to chemical control for sustainable agriculture (Asif *et al*., 2023).

3. Genomic analyses

3.1. Bioinformatics

Bioinformatics is a hybrid word which refers to biology and information technology. It consists of the application of informatic tools to analyze biological problems and find solutions (Sharma *et al*., 2021). Bioinformatics is an emerging area of biology that involves the use of computational methods for bettering traditional biological studies conducted in natural environments (*in situ*), living organisms (*in vivo*), and experimental settings (*in vitro*). Bioinformatics is widely acknowledged as the leading scientific field of the 21st century, with immense potential for uncovering complex biological systems through the examination and synthesis of multi-omic data. In the age of omics and the advancements in high-throughput sequencing technologies, a substantial volume of molecular data is regularly uploaded to international data repositories by groups of scientists worldwide (Pathak *et al*., 2022). Bioinformatics is a multidisciplinary field that combines biology with information technology, mathematics, statistics, physics, and chemistry. The inaugural endeavor in the field of bioinformatics sequencing and annotation consisted of the Human Genome Project (HGP) initiated in 1990 (Sharma *et al*., 2021).

3.2. Genomics

Genomics is a scientific topic that focuses on studying a organism's genomes. This involves understanding its physiological, chemical, and biological processes, as well as the structure of genes, including their sequences and functional annotations (Varshney *et al*., 2018).

The process entails examining the genome's structure, function, evolution, mapping, and changes through recent progress in molecular biology techniques, specifically high-throughput genome sequencing and gene analysis. Basically, genomics is a discipline of study that focuses on the genes that distinguish various species and determine their whole set of genetic materials, known as the genome. In recent years, there have been substantial advancements in this field, transitioning from a descriptive phase to a phase focused on functional experimentation. Genome analysis is an important step in the investigation of organisms and their biological processes (Charlotte *et al*., 2019).

3.3. Next Generation Sequencing (NGS)

Next-generation sequencing (NGS) techniques offer multiple sequencing of a substantial number of DNA molecules (Metzker, 2010; Innocenti *et al*., 2015). Currently, a solitary computer has the capability to sequence 1 Gigabase (Gb) over a short span of hours. Furthermore, the advancements in technology that have been reported are even more astonishing (Lacoste *et al*., 2017).

NGS platforms, such as Illumina, Qiagen, and ThermoFisher Scientific, are typically assessed by considering the size and number of reads, as well as the overall quantity of data generated (Kumar *et al*., 2019). Using commercial software, automatic analytical pipelines enable efficient data processing within a short timeframe of a few hours (Meinel *et al*., 2017) (**Table II**).

Table II. Description of the five main families of next generation sequencing platforms (Hodkinson and Grice, 2015; Pericard, 2017).

Technology	Reading length	Precision	Number of readings per run	Cost (US$/Mbp)
Sanger	400-900 bp	99.9%	N/A	2400
Illumina	MiSeq : 50-600 bp HiSeq : 50-500 bp	99.9%	MiSeq : 1-25M HiSeq : 0.3-2 G	0.05-0.15
Roche-454	700 bp	99.9%	1 M	10
Ion Torrent	Up to 400 bp	98%	Up to 80 M	1
Pacific Biosciences	10-15 kbp average	87%	500-1000 M	0.13-0.60
Nanopore	max > 40 kbp to 500 kbp	~92-97%	Variable	~0.5

3.4. Genomic analysis

Genomics is a contemporary branch of biology. Advancements in sequencing technologies have facilitated the sequencing and analysis of multiple genomes from different organisms individually. The genome refers to the complete set of genetic material contained within the DNA of an individual or species (Vannier, 2017).

3.4.1. Structural genomics

Structural genomics is efficiently determining and analyzing the three-dimensional structures of biological macromolecules. It has traditionally been categorized as either "curiosity-driven" or

"hypothesis-driven" research. A significant outcome of structural biology is the identification of novel connections between amino acid sequences and protein structures, as well as among various protein structures. This development signifies a fresh approach in structural biology, driven by the objective of determining the maximum number of structures (Goldsmith-Fischman *et al*., 2003). An early objective was to identify a representation for each protein fold, ensuring comprehensive coverage of "fold space" (Harrison *et al*., 2002). The various steps involved in genomics are as follows: development of precise genetic and physical maps with great precision; genome sequencing and proteome profiling of an organism (Pradeep *et al*., 2021).

3.4.2. Functional genomics

Functional genomics is a field that combines the study of molecular biology and cell biology to examine the complete structure, function, and regulation of a gene (Kaushik *et al*., 2018). Genomic research includes the comprehensive examination of DNA function, including genes. It utilizes genome data to investigate the expression and activities of genes and proteins across the entire genome. To do this, high-throughput techniques are employed to comprehend gene transcription, translation, and Protein-Protein Interaction (PPI). Functional genomics studies focus on understanding the specific roles and functions of genes within an organism's genome. The gene's role and its modification under a specific condition are essential for gaining a deeper grasp of an organism's biology (De Sousa *et al*., 2018).

3.4.3. Comparative genomics

Comparative genomics is one of the main uses of genomic analysis. The comparison of biological data obtained from whole genome sequences is the simplest definition of comparative genomics, and it has grown rapidly to be a crucial tool in the field (Setubal *et al*., 2018). This analysis can be used to investigate the roles and mechanisms of action of genes; comparative methods need genomes or sets of genes that can be compared, as well as matching and similarity-assessing algorithms; the lineage, niche, and phenotype of the species from which a protein originates, as well as all available data on its metabolism and the arrangement of its neighboring genes on the chromosome, are all taken into consideration when comparing a protein with all of its characterized homologs to learn more about its biological processes and molecular functions (Ulitsky, 2016).

3.4.4. Reconstruction of sequences (assembly)

Genome assembly involves the procedure of ordering nucleotide sequences in the appropriate sequence. Assembly is necessary due to the relatively small sequence read lengths, which are currently shorter than the majority of genomes and genes. The process of genome assembly is facilitated by the presence of publicly accessible databases, which may be found on the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov, Betsy, 2012). The fragments that have been aligned and linked are referred to as an assembly, which is derived from a longer DNA sequence in order to reconstitute the original sequence. These techniques can be classified into two groups based on the presence or absence of a reference genome. Genome assembly is crucial for predicting the behavior of organisms in terms of the structure and function of genes. Genome assembly is the procedure of arranging nucleotide sequences in an accurate sequence. The process of genome assembly is facilitated by the presence of publicly accessible datasets, which may be found on the website of the National Center for Biotechnology Information (Pallavi *et al*., 2022). Whenever we have access to the sequencing and assembly of nucleotide data, we can initiate the annotation step of our target genome (Ortet, 2018).

3.4.5. Annotation of genome

Genome annotation includes the prediction and identification of all physiologically functioning sequences (genes) in a genome. It also includes establishing the structure, function, and interactions between biological entities associated with the genome (Amador, 2017). The objective of gene function annotation is to compile a comprehensive inventory of an organism's molecular and/or metabolic capabilities **(Figure 4)**; (Mercier, 2017).

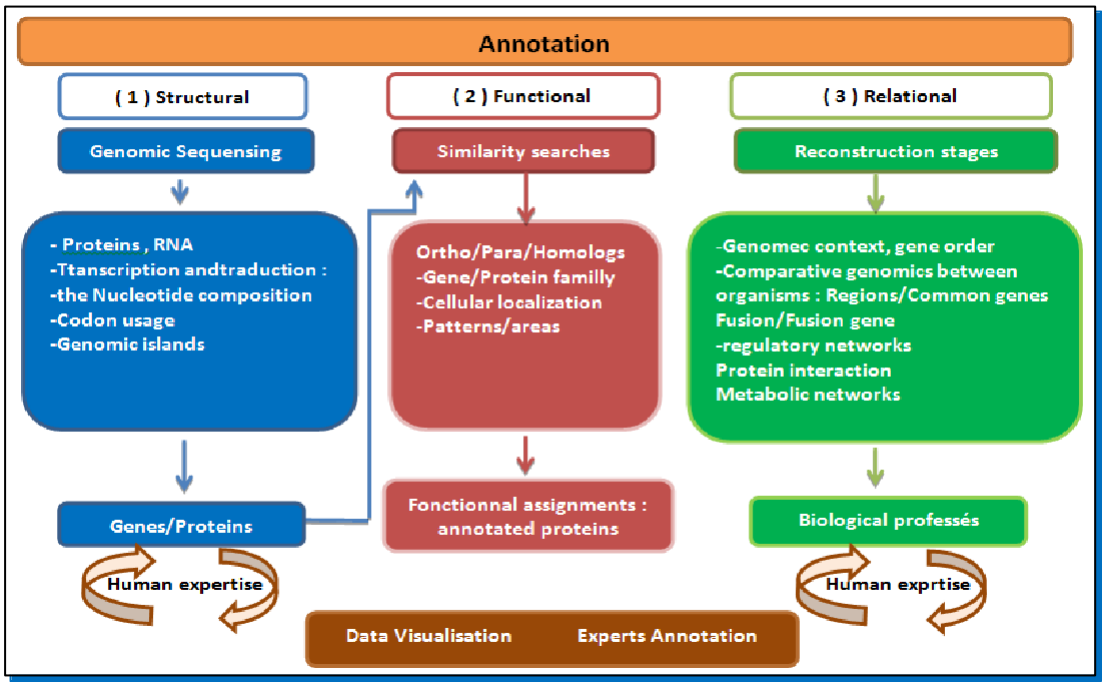


Figure 4. Presentation of the three levels of annotation (Vannier, 2017). 1) Structural: identification of objects of biological interest on the DNA sequence (genes of protein, RNA genes, signals, etc.), 2) Functional annotation: assignment of objectives biologically achieved by previously identified sequences (usually by homology with a known function sequence), 3) Relational annotation: different grouping relationships established between sequences to describe biological objectives or modules by example a metal track.

Materials and Methods

All analyses were conducted with DESKTOP type laptop DELL Precision 7550 processor Generation 10th, Intel(R) Core (TM) I7-10875H 16M CPU @ 2.30-up to 5.00GHz, with 32 GB RAM, storage 1 TB SSD, graphical card NVIDIA GeForce RTX 3000 6GB, system type: - 64 operating system, bit and a Windows 11 Pro.

1.1. Tools used (platform's)

1.1.1. NCBI (National Center for Biotechnology Information) website

The NCBI website (https://www.ncbi.nlm.nih.gov) serves as a comprehensive database containing gene information from multiple sources, accessible to molecular biology professionals and researchers (Brown *et al*., 2015).

1.1.2. TYGS (Type Strain Genome Server) platform

TYGS (http://tygs.dsmz.de) is a high-throughput web server facilitating the classification of bacteria based on genome sequences. It is linked to a vast database of genomic and taxonomic information, aiding in microbial classification and taxonomy (Göker and Meier-Kolthoff, 2019).

1.1.3. DDBJ Fast Annotation and Submission Tool (DFAST)

DDBJ Fast Annotation and Submission Tool (DFAST) (https://dfast.ddbj.nig.ac.jp/) is a bacterial genome annotation pipeline integrated with quality and taxonomy assessment methods. DFAST has been developed so that all the procedures required for submission can be done seamlessly online. Thus, it can be used as an online workspace to prepare submission files to the DDBJ Mass Submission System (MSS) (Tanizawa *et al*., 2018).

1.1.4. BV-BRC platform

The BV-BRC platform (https://www.bv-brc.org) provides access to bacterial and viral bioinformatic resources. It offers a comprehensive dataset of microorganisms, especially those related to pathogens and infectious diseases. It also offers computational tools for data analysis, including predictive analytics using artificial intelligence techniques (Olson *et al*., 2023).

1.1.5. RAST Server

The RAST server (https://rast.nmpdr.org) offers a fully annotated service for identifying and annotating genes within bacterial genomes. It classifies gene functions and groups into functional subsystems (COGs), providing insights into genomic functionality (Aziz *et al*., 2008).

1.1.6. Prokka program

The Prokka program, integrated within the Galaxy global bioinformatics platform (https://usegalaxy.eu), analyses genome sequences to accurately identify genes. It is particularly useful for identifying virulence, pathogenicity, and infectivity-related genes (Seemann *et al*., 2014).

1.1.7. AntiSMASH Server

The antiSMASH version 7.1 <**anti**biotics and **S**econdary Metabolites **A**nalysis **Sh**ell= server (https://antismash.secondarymetabolites.org) is a web-based tool used to identify sets of biosynthetic genes encoding secondary metabolites in the complete genome of the organism under study, aiding in the exploration of secondary metabolite production (Blin *et al*., 2023). The tool allows for rapid identification, annotation, and analysis of gene clusters that produce secondary metabolites in bacterial and fungal genomes on a large scale. This platform seamlessly combines and links with a diverse range of computational tools for in silico investigating secondary metabolites previously described. antiSMASH employs various open-source technologies such as NCBI BLAST+, HMMer 3, Muscle 3, Glimmer 3, FastTree, TreeGraph 2, Indigo-depict, PySVG, and JQuery SVG. Also, to identify groups of co-occurring biosynthesis genes in genomes, referred to as **B**iosynthetic **G**ene **C**lusters (BGCs). This later often encompasses the whole set of genes necessary for producing specific **N**atural **P**roducts (NPs), also referred to as specialized or secondary metabolites that exhibit intriguing biological properties and have been successfully transformed into vital drugs such as antibiotics (penicillin and streptomycin), anti-cancer treatments (bleomycin and doxorubicin), and cholesterol-lowering compounds (lovastatin). The NPs and their corresponding BGCs are highly significant commercially and scientifically (Blin *et al*., 2023) (**Figure 5**).

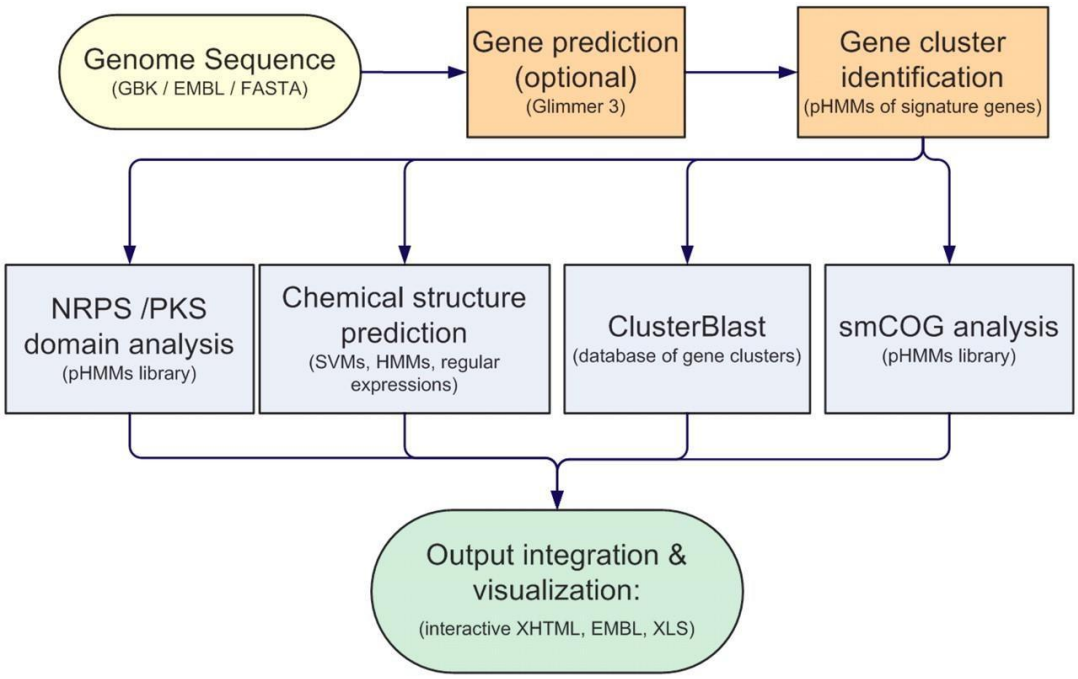


Figure 5. Outline of the pipeline for genomic analysis of secondary metabolites. Genes are extracted or predicted from the input nucleotide sequence, and gene clusters are identified with signature gene pHMMs. Subsequently, several downstream analyses can be performed: NRPS/PKS domain analysis and annotation, prediction of the core chemical structure of PKSs and NRPSs, ClusterBlast gene cluster comparative analysis, and smCOG secondary metabolism protein family analysis. The output is visualized in an interactive XHTML web page, and all details are stored in an EMBL file for additional analysis and editing in a genome browser. A Microsoft Excel file with an overview of all detected gene clusters and their details is also generated (Medema *et al*., 2011).

1.1.8. Roary program

The Roary program (Version 3.13.0) is designed to compare bacterial strain genomes, determine pan-genome, core genes, shell genes, and cloud genes and provide insights into genomic diversity and evolution (Page *et al*., 2015).

1.2. Process of selecting and loading the genome of *Saccharothrix algeriensis*

The complete genome sequence file of the actinobacterial strain under investigation, namely the *Saccharothrix algeriensis* DSM 44581 strain, was retrieved from the NCBI database in FNA format, with the genome's digital address being the accession number PRJNA695776. This strain

originating isolated from a soil sample collected from a palm grove in Adrar (Algeria) underwent Whole-Genome Sequencing (WGS) using PacBio Sequencing with HGAP v.smrtlnk/8.0.0.80529, HGAP 4 (1.0) assembly method, conducted by NCBI. A summary of the steps performed is provided in the following screenshots (**Figure 6**).

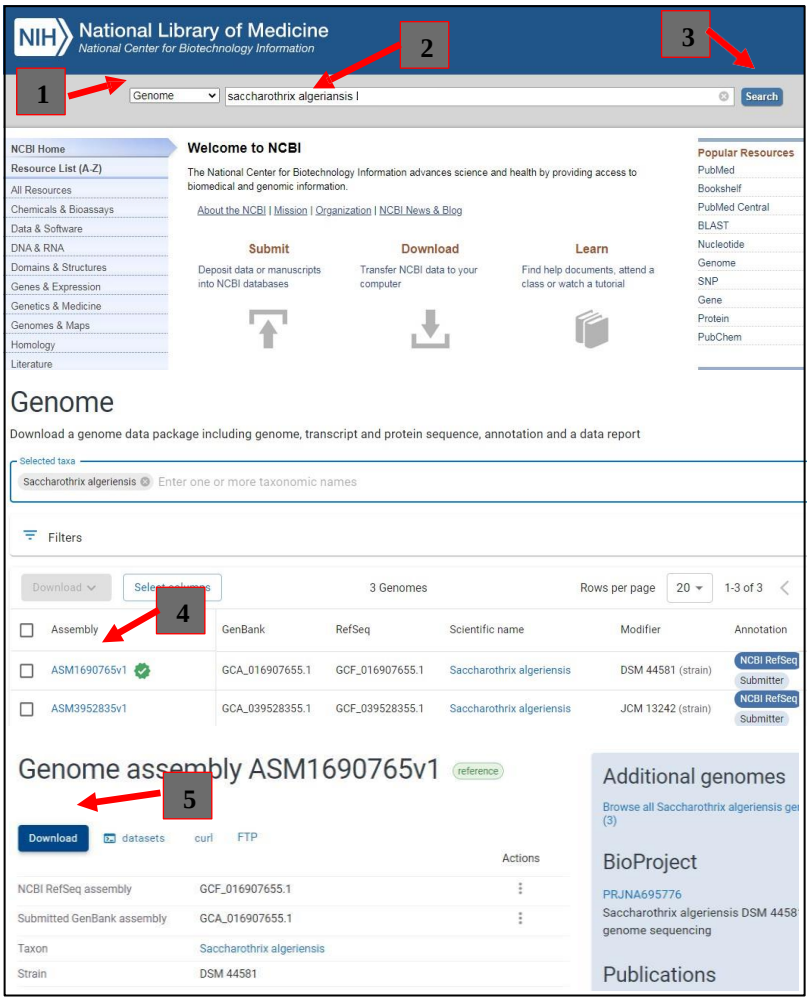


Figure 6. Steps performed to download files of *Saccharothrix algeriensis* DSM 44581sequences. (**1**) Choose genome, (**2**) Write the studied strain *Saccharothrix algeriensis*, (**3**) Click on search, (**4**) Click on assembly number of strain, (**5**) Download the file in FNA format.

1.3. Comparative genomic analysis of *Saccharothrix algeriensis*

Following the genome acquisition, a series of computational processes and analyses were conducted utilizing various academically endorsed programs and platforms (Dif and Zitouni, 2023). The methodological approach is outlined as follows:

1.3.1. Phylogenomic and phylogenetic analysis by TYGS (Type Strain Genome Server)

For the taxonomic study, the genome of the studied strain was uploaded to the TYGS platform to confirm its identity, leveraging the platform's algorithms and extensive taxonomic database for accurate species identification. A summary of the steps performed is provided in the following screenshots (**Figure 7**).

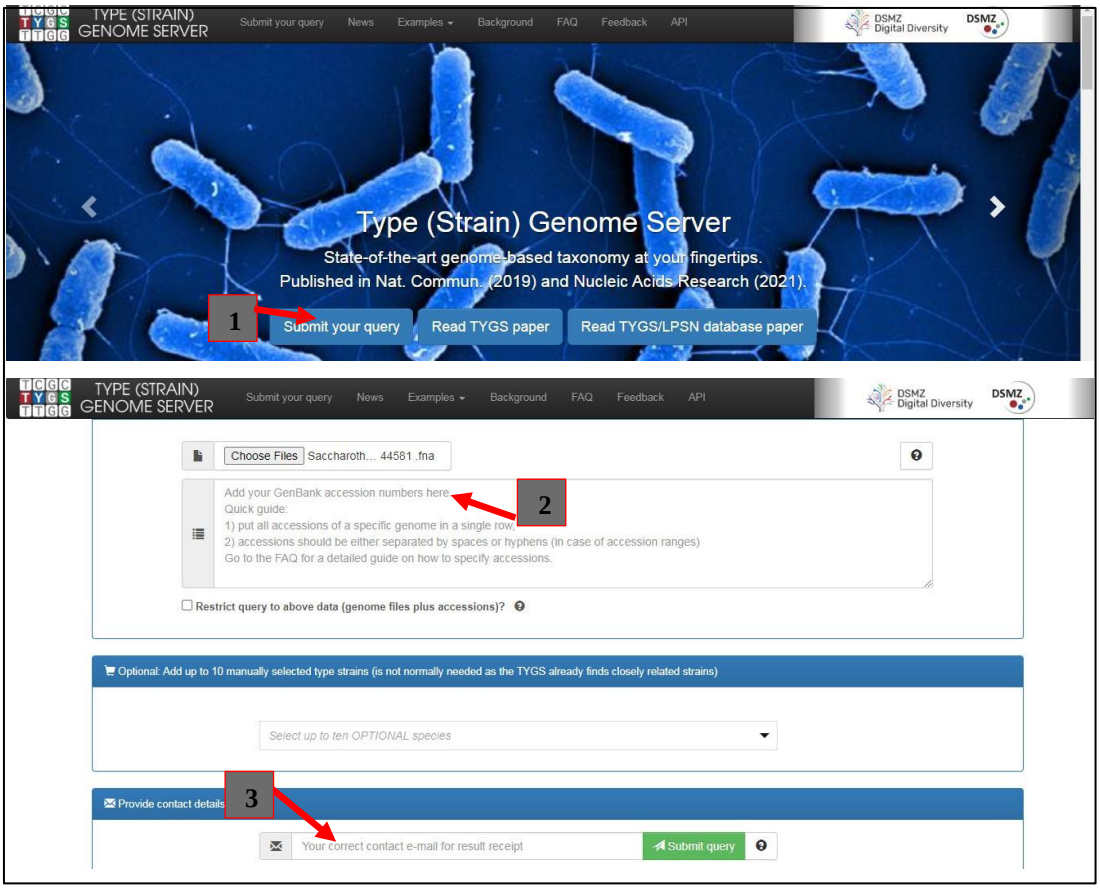


Figure 7. Steps performed to upload the FNA format of the strain *Saccharothrix algeriansis* DSM 44581. (**1**) Submit your query, (**2**) Choose files, (**3**) Put the email address for obtaining the results and submit a query.

1.3.2. Genome features with DFAST

The digital DNA-DNA hybridization (dDDH) values were calculated using a genome-to-genome distance calculator. The screenshot (**Figure 8**) provides a summary of the steps performed.

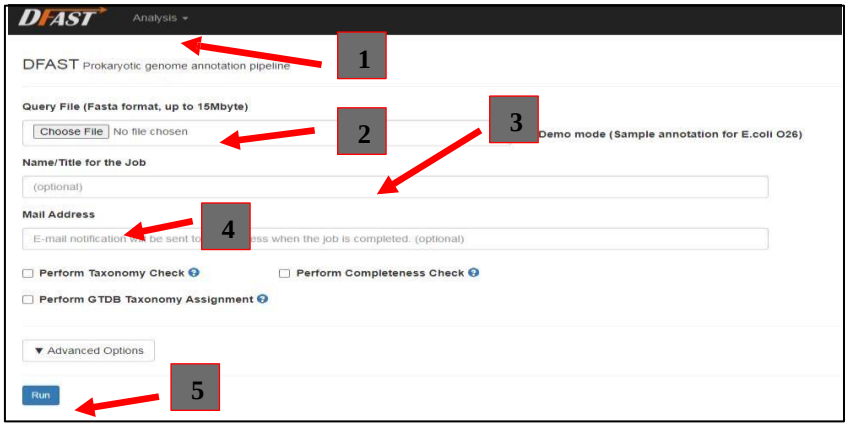


Figure 8. DFAST steps. (**1**) Choose genome annotation, (**2**) Choose files, (**3**) Give a name for the job, (**4**) Put the email address for obtaining the results, (**5**) Run.

1.3.3. Circular view of the genome

The genome was processed using the BV-BRC platform. Using the annotation service within this platform, we gained insights into the general features of the actinobacterial strain and its circular representation. Summary of the steps performed in the provided screenshots (**Figures 9 and 10**).

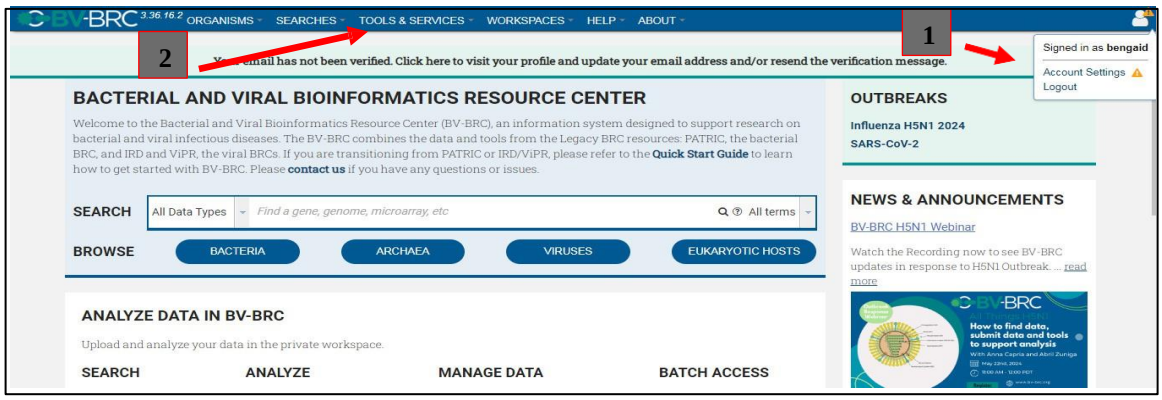


Figure 9. BV-BRC steps. (**1**) Open an account (to save data), (**2**) Choose tools and services: genome annotation.

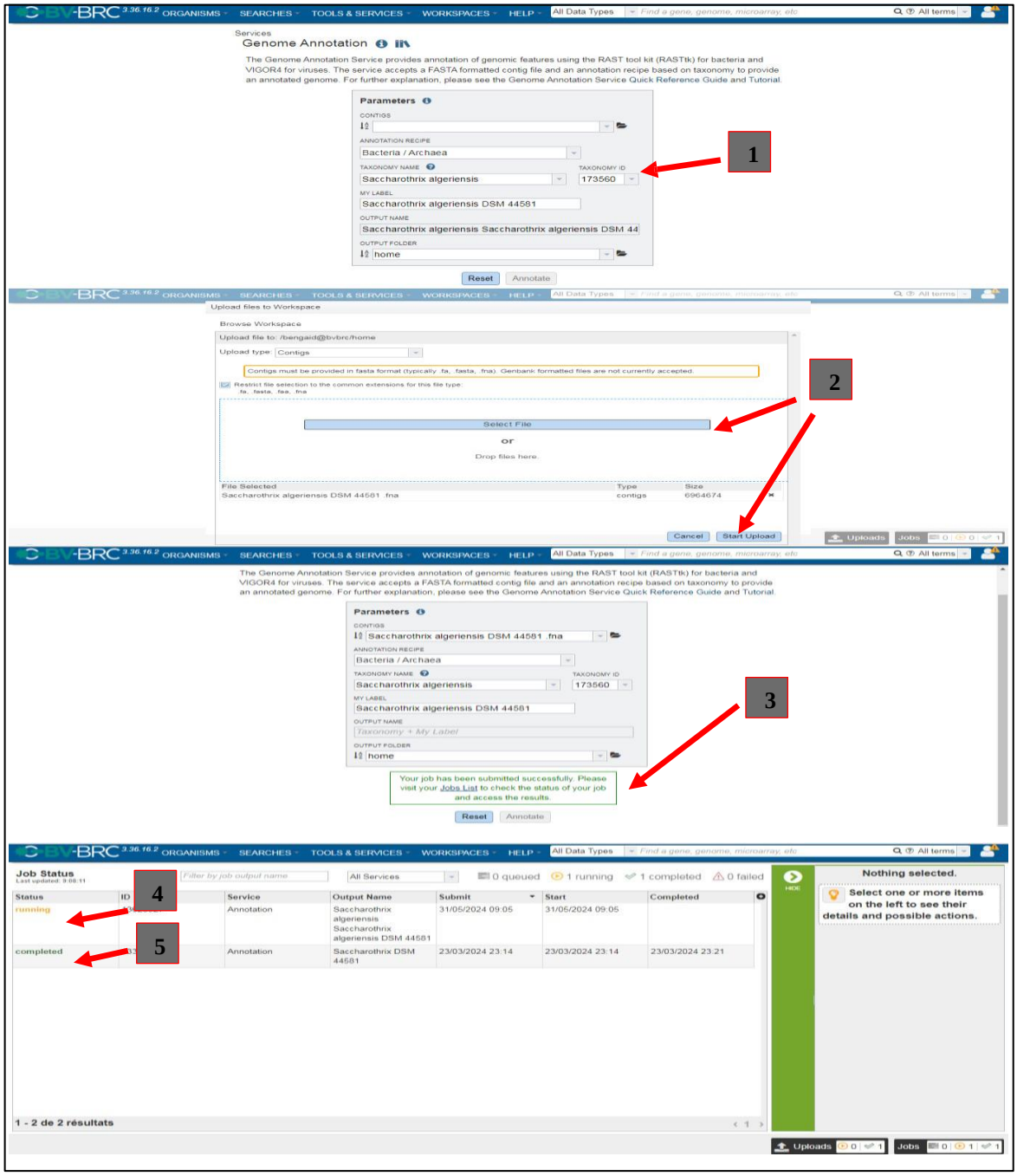


Figure 10. Continued BV-BRC steps. (**1**) Fill in the blanks and upload the folder, (**2**) Select the file and start upload, (**3**) Job has been submitted, (**4**) Job list not yet finished, (**5**) Job list finished (check the status and access the results).

1.3.4. Clusters of Orthologous Groups of proteins (COGs)

The processing was launched using the RAST server which facilitated the functional classification of genes within the actinobacterial strain's genome into functional subsystems (COGs), providing information on the frequency of genes in each functional category. **Figures 11, 12 and 13** provide a summary of the steps performed.

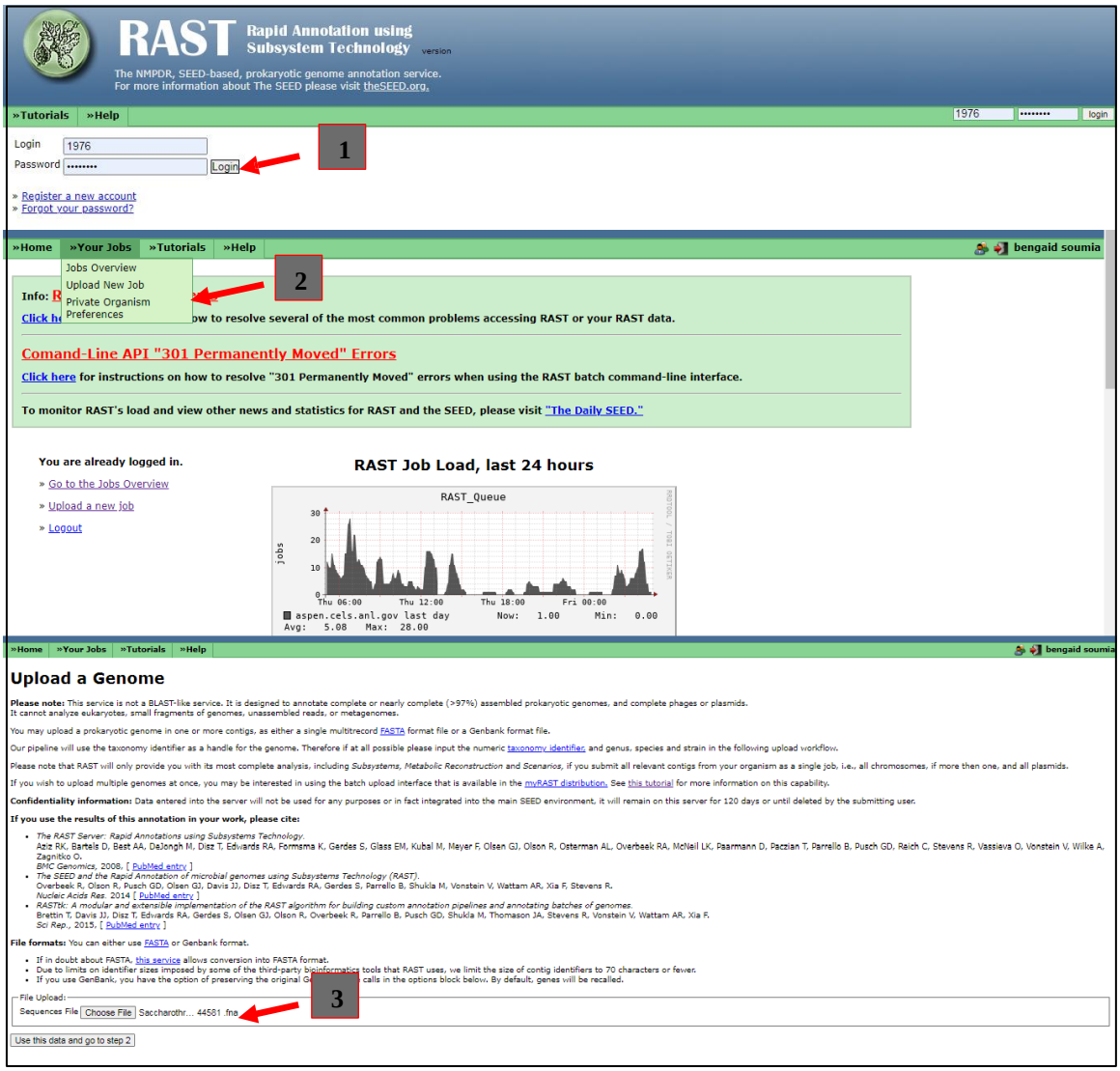


Figure 11. RAST server steps. (**1**) Open an account (to save data), (**2**) Select tools job: upload new job, (**3**) Select the sequence file and use this data.

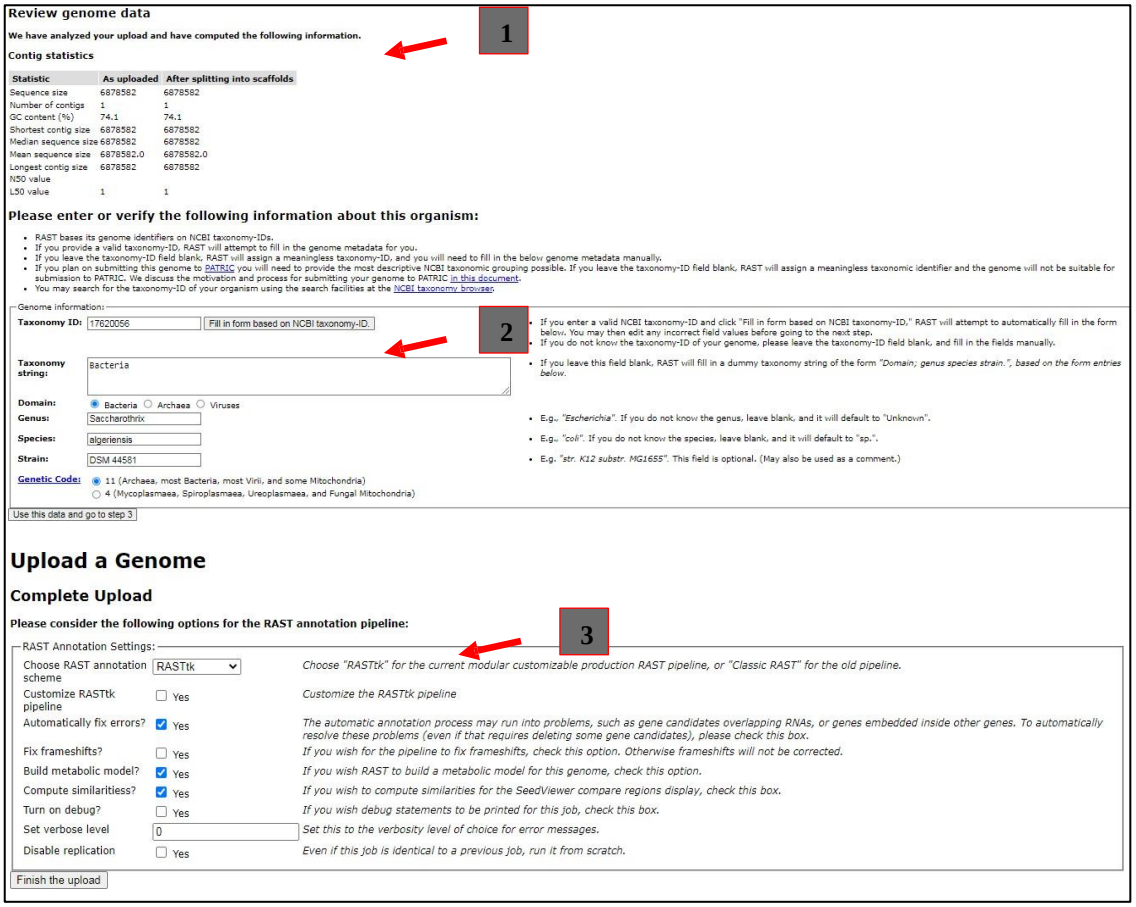


Figure 12. Continued RAST server steps. (**1**),(**2**),(**3**) With the write words, fill in the blanks and cross correct propositions and then finish the upload.

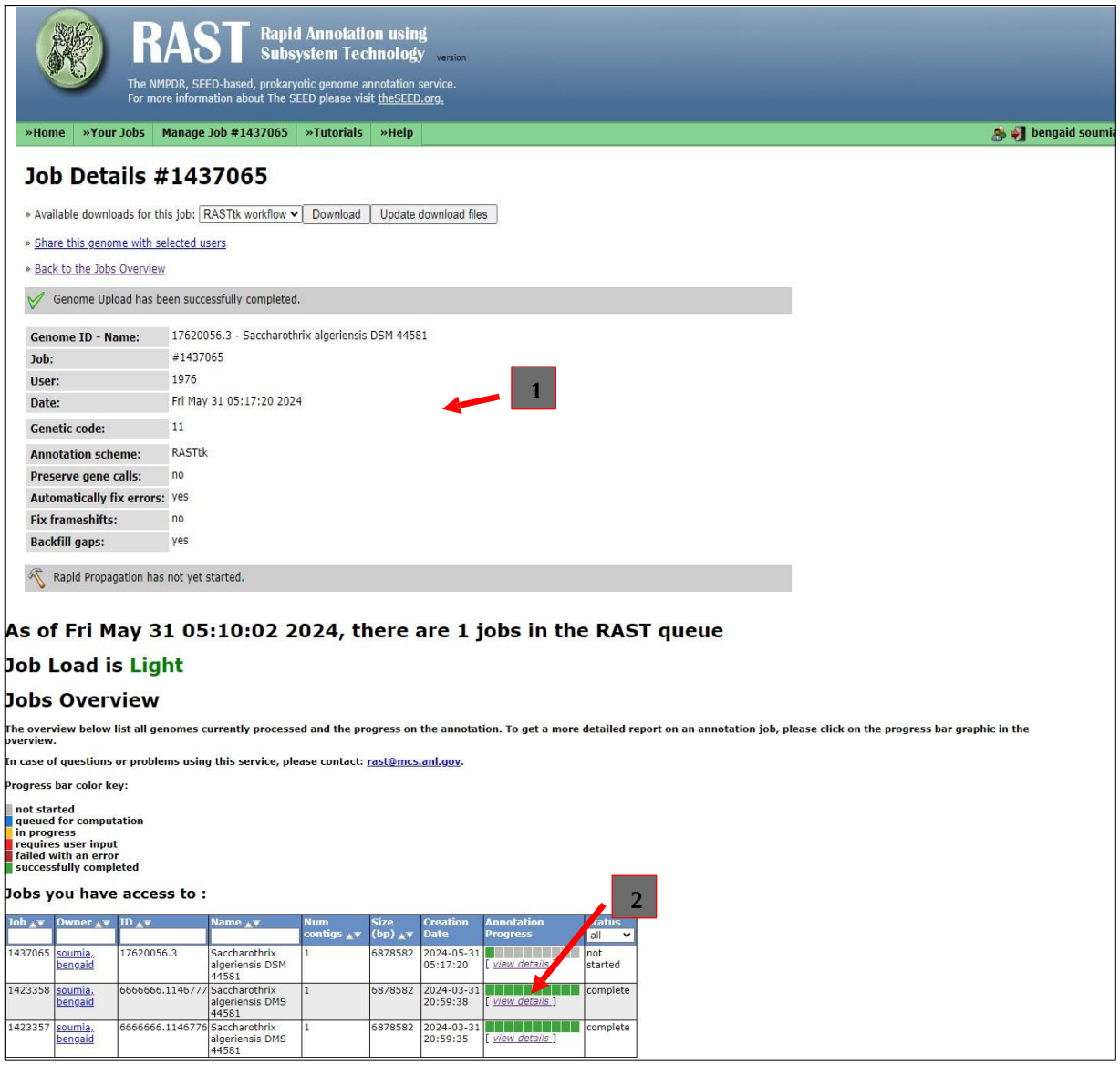


Figure 13. Continued RAST server steps. (**1**) Window of job details, (**2**) Job Overview consulting (When the view details light is green 100%, the job is completed.

1.3.5 Galaxy analyses

Using the Prokka annotation program enabled a more in-depth exploration of the genes within the genome, with a focus on identifying significant genes associated with virulence, pathogenicity, and infectivity. A summary of the steps performed is provided in the screenshots of **Figures 14 and 15**.

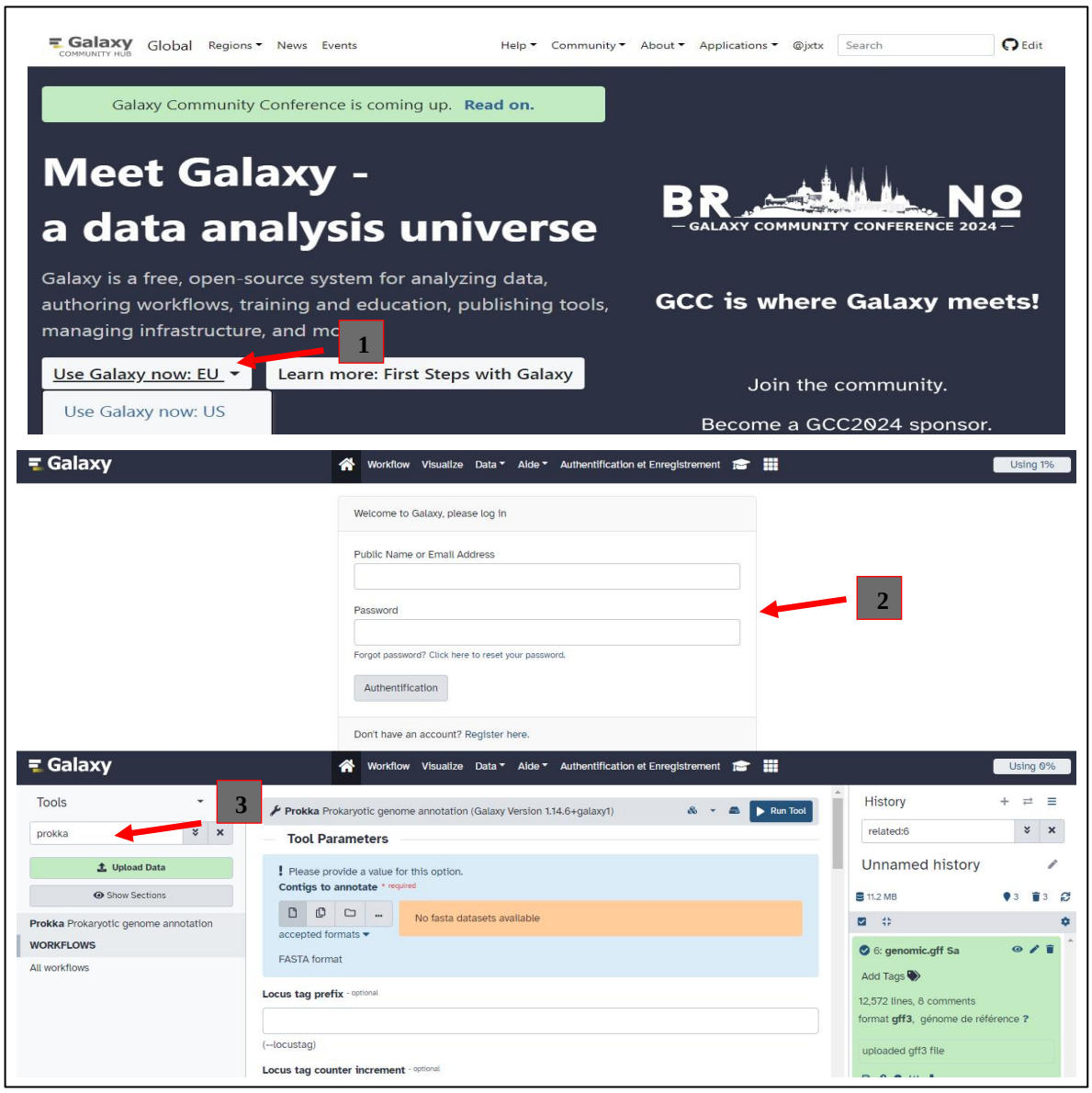


Figure 14. Galaxy steps. (**1**) Choose the European platform, (**2**) Create an account (to save data), (**3**) Pick Prokka annotation in the tools box.

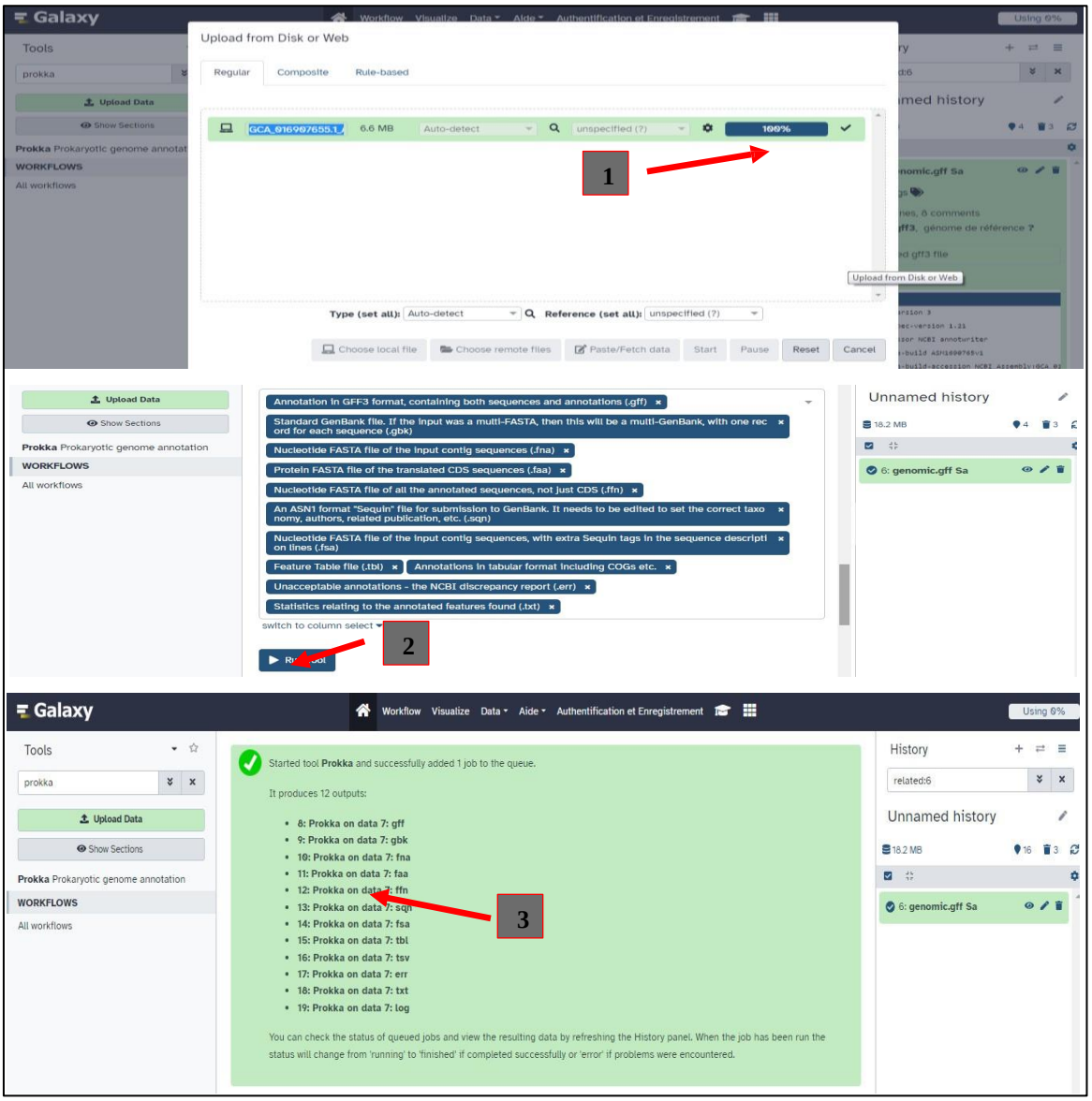


Figure 15. Prokka steps. (**1**) Upload file 100%, (**2**) Run tools, (**3**) Successful job (The FFN file).

1.3.6 Biosynthetic gene clusters and secondary metabolites

The outputs of the processing using antiSMASH identified genetic clusters within the genome responsible for secondary metabolites, shedding light on potential biochemical pathways and functions. **Figure 16** provides the steps performed.

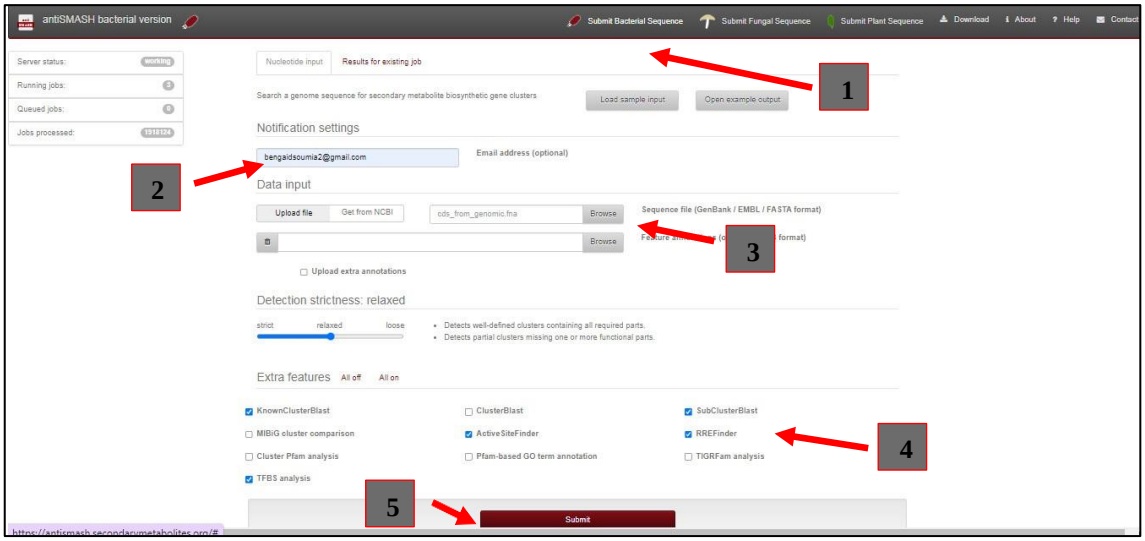


Figure 16. AntiSMASH steps. (**1**) Select bacteria, (**2**) Notification settings, (**3**) Upload the file from GenBank <NCBI3=, (**4**) Do not change the cross, (**5**) Submit (check and access the results).

1.3.7 Roary analyses

Roary is a high-speed standalone pan-genome pipeline that takes annotated assemblies in GFF3 format (produced by Prokka) and calculates the pan-genome. Using a standard desktop PC can analyze datasets with thousands of samples, something which is computationally infeasible with existing methods, without compromising the quality of the results. A number of 128 samples can be analyzed in under 1 hour using 1 GB of RAM and a single processor. To perform this analysis using existing methods, it would take weeks and hundreds of GB of RAM. The three strains of *S. algeriensis*, *S. saharensis*, and *S. tamanrassetensis* that were isolated from Algerian Saharan soils were selected, and the genome was annotated in the NCBI platform. A summary of the performed steps is shown in **Figure 17**.

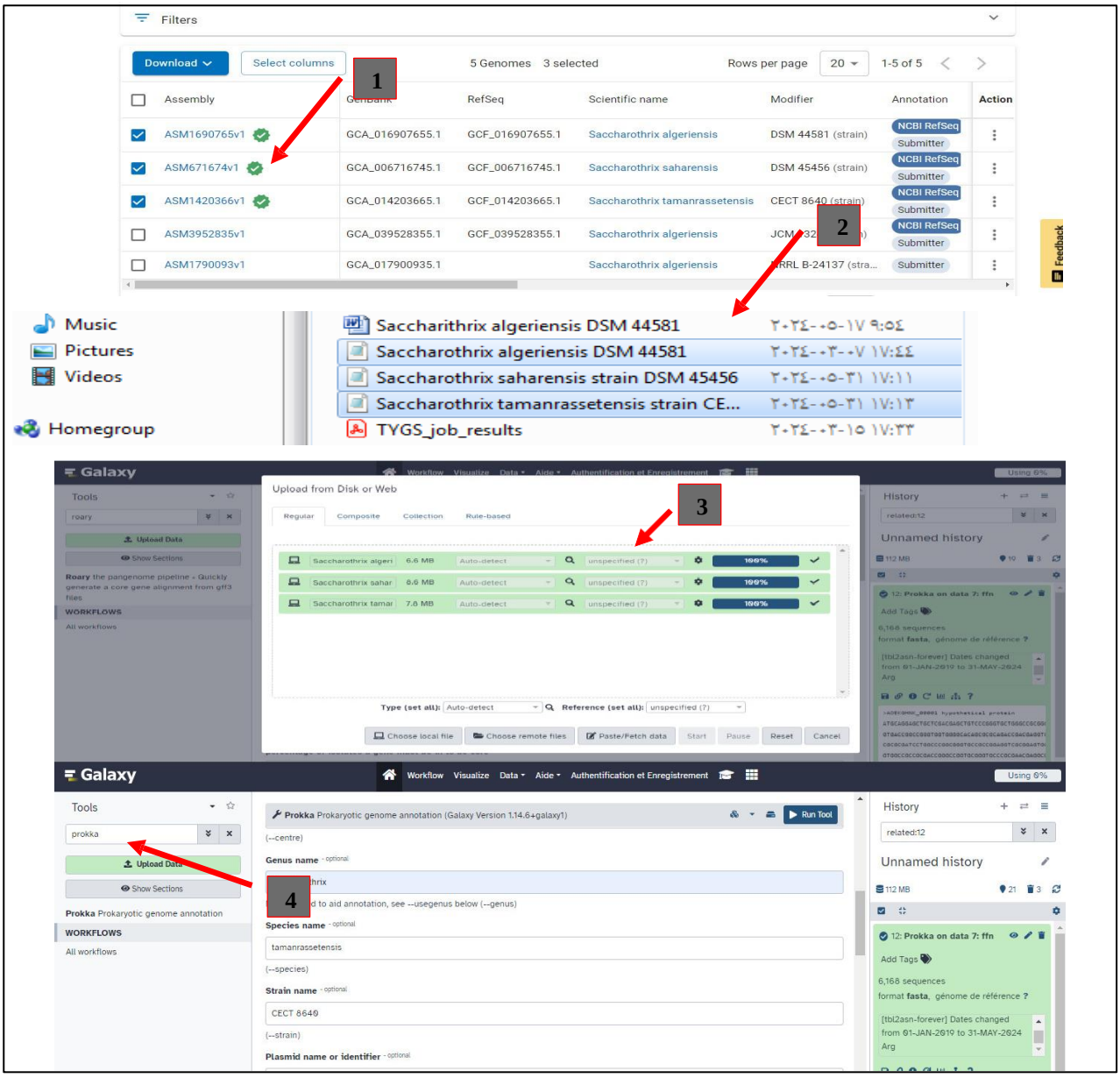


Figure 17. Roary analyses steps. (**1**) Upload the FNA file of *Saccharothrix algeriensis* DSM 44581 strain with the two other strains *Saccharothrix saharensis* strain DSM 45456 and *Saccharothrix tamanrassetensis* strain CECT 8640, (**2**) Select the file FNA (**3**) Place in the Prokka, (**4**) Choose Roary tool.

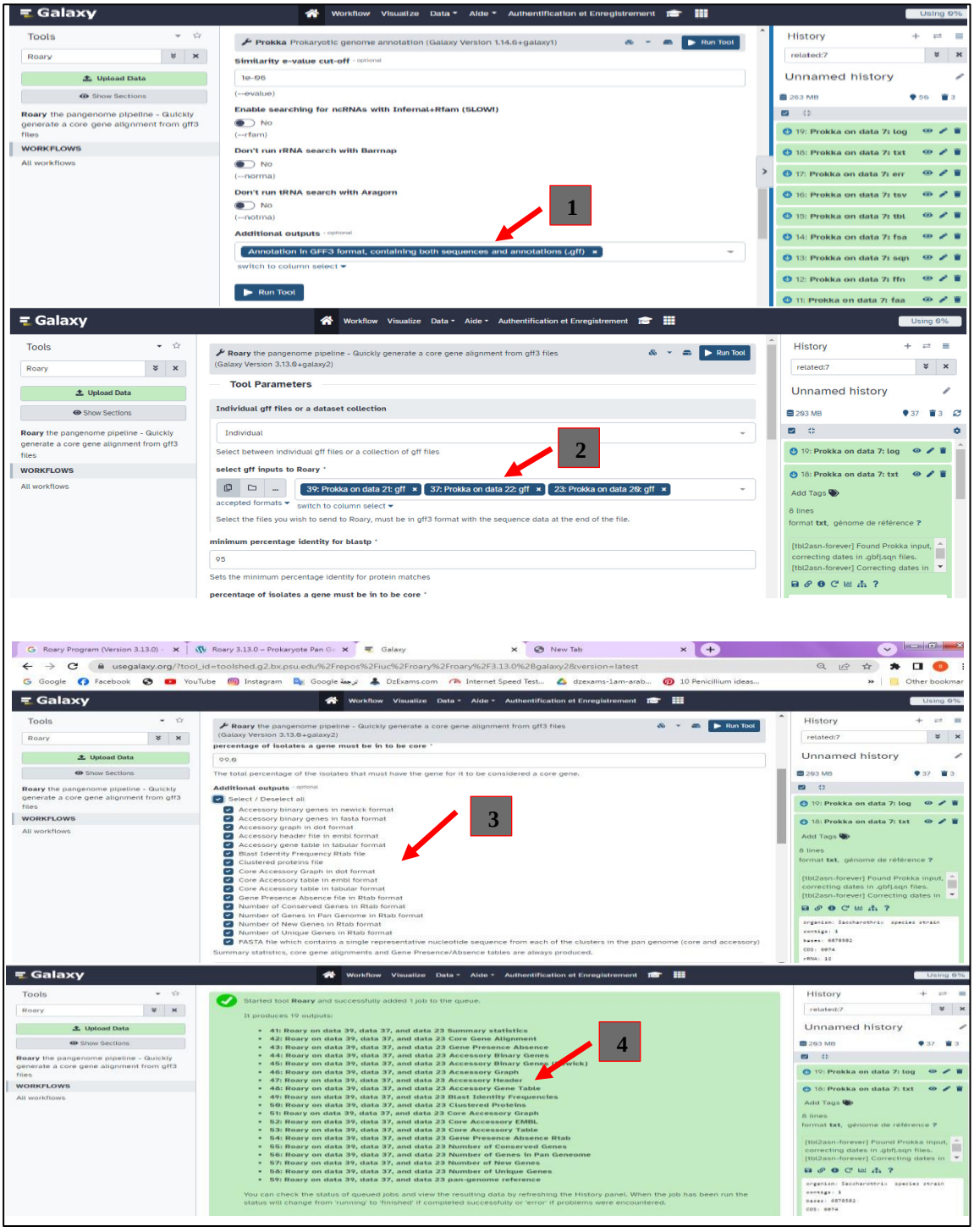


Figure 18. Continued Roary analyses steps. (**1**) Chose annotation Gff3 format only (to save time in the analysis), (**2**) Upload the files in Gff3 format for the three strains, (**3**) Select additional outputs, (**4**) Job is successful.

Results and Discussion

1. Results

1.1. Phylogenetic and phylogenomic analyses

Using TYGS for taxonomic classification, the studied strain was assigned to the species *Saccharothrix algeriensis* strain DSM 44581 (**Figures 19** and **20**). This interactive tree view displays information on the underlying genome data, the GBDP distance formula and the algorithm. Additional annotation is displayed to the right-hand side of the viewer, including species and subspecies clusters, genomic G+C content (in %), · values, total sequence length (in bp), and number of proteins. The alignments consist of 15 strains.

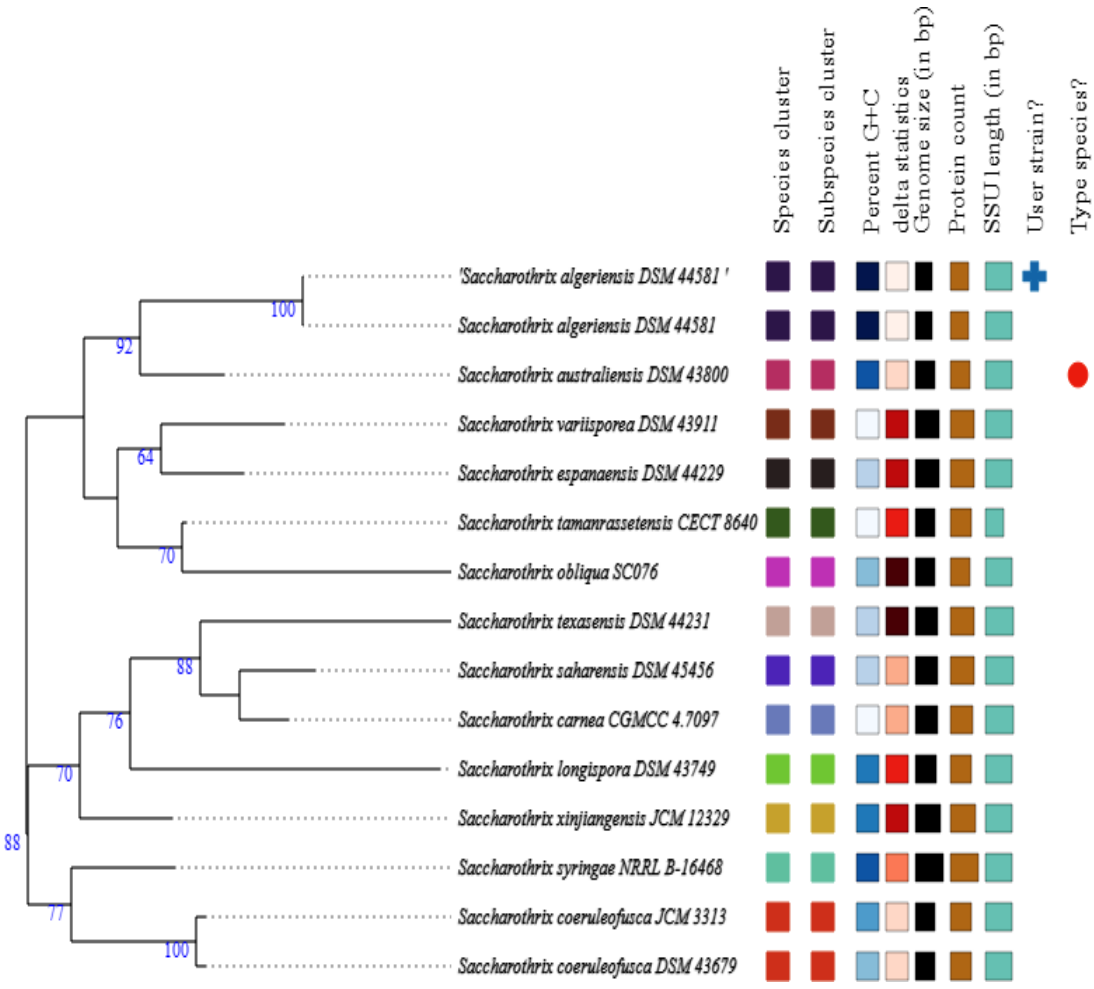


Figure 19. TYGS-generated 16S rDNA-based phylogenetic tree showing the relatedness of *Saccharothrixr algeriensis* strain DSM 44581 to the most closely related type strains, including those that have not yet been subjected to genome sequence analysis.

The tree inferred with FastME 2.1.6.1 from GBDP distances calculated from 16S rDNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 77.7 %. The tree was rooted at the midpoint, and with a percentage of Bootstrap 92%.

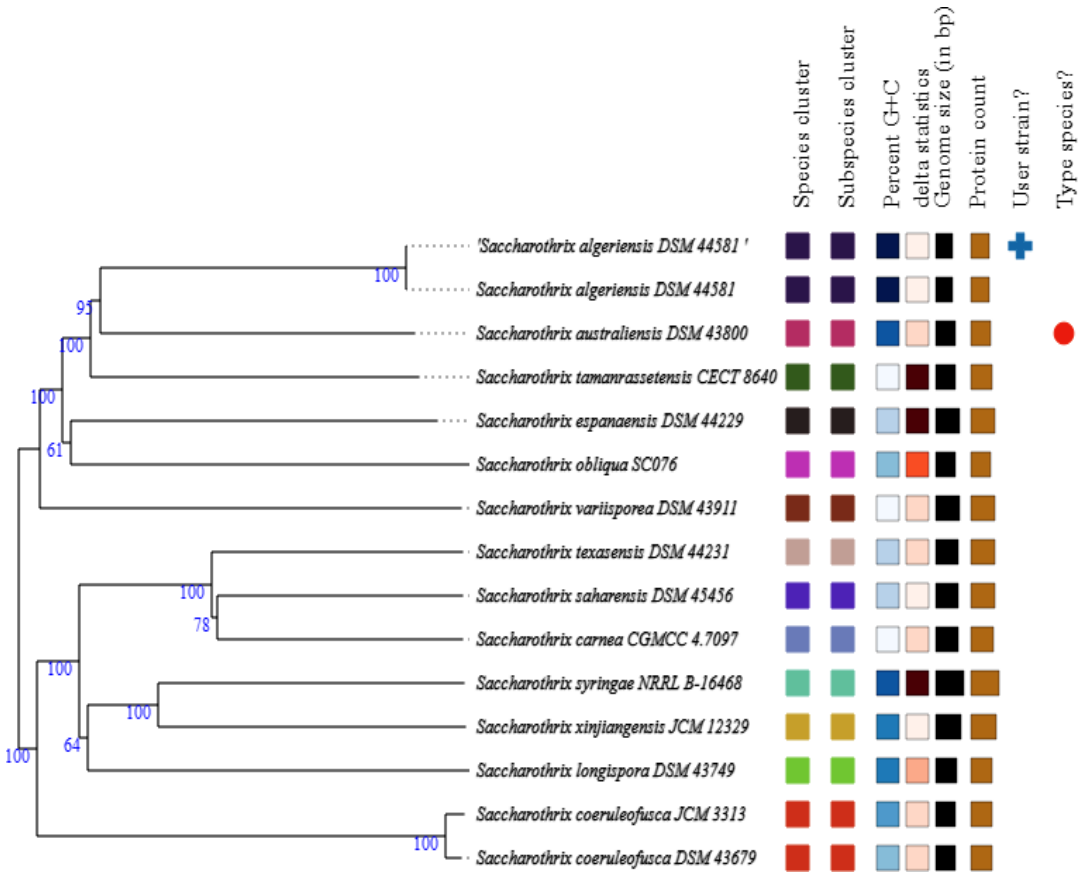


Figure 20. Phylogenomic tree based on genome sequences in the TYGS tree inferred with FastME 2.1.6.1 from the Genome BLAST Distance Phylogeny approach (GBDP) (Meier-Kolthof and Göker, 2019). Distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d5. The numbers above branches are GBDP pseudo-bootstrap support values>60% from 100 replications, with an average branch support of 93.7%. The tree was rooted at the midpoint (Meier-Kolthof *et al*., 2013).

The TYGS platform allowed us to obtain a JOB SUMMARY with all the necessary results information, including phylogeny, identification, pairwise comparisons, and strains in your dataset (Appendix 1).

1.2. General genome features

1.2.1. DFAST features

The genome of *Sccharothrixr algeriensis* strain DSM 44581 consisted of 6,878,582 bps with a GC content of 74.1%. The genome carries CDSs 6,229 coding genes, being 80 for tRNA, 12 for rRNA, five for ncRNAs, and 5 prophage regions (**Table III**).

Table III. Genome features of *Saccharotthrix algeriensis* strain DSM 44581 as a result of combined output from DFAST annotation.

Features	**Chromosome**
Total Length (bp)	6,878,582
No. of sequences	1
GC content (%)	74.1%
N50	6,878,582
Gap ratio (%)	0.0%
No. of CDSs	6,229
No. of rRNA	12
No. of tRNA	80
No. of CRISPRS	5
Coding ratio (%)	87.7%

1.2.2. Circular genome viewer

The circular genome viewer was used for *S. algeriensis* and produced a circular diagram (interactive circular viewer) that displays the genes, their orientation, homology, virulence factors, and GC content skew. A separate downloadable SVG or PNG of a circular graph image is available in the jobs list (**Figure 21**).

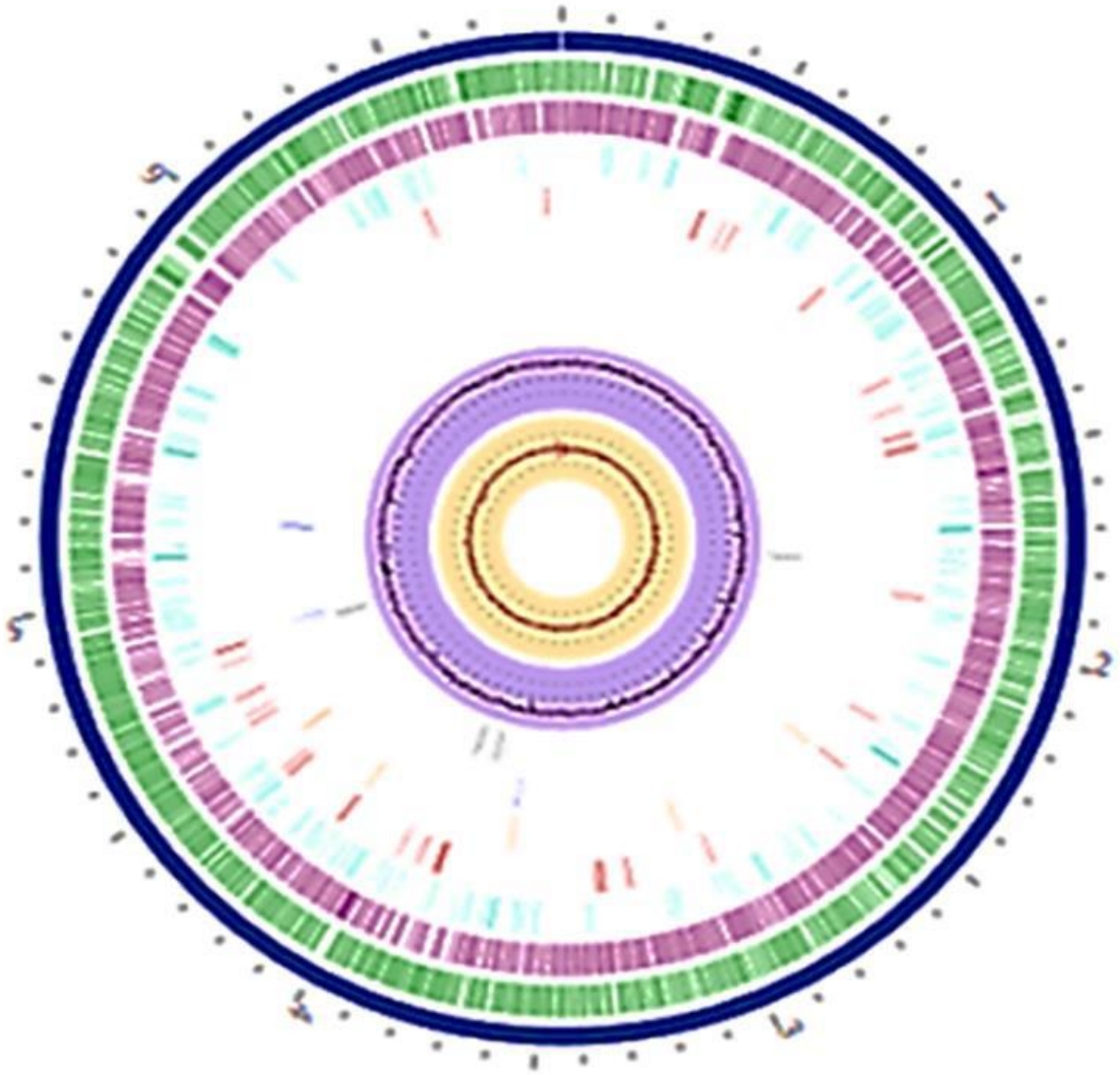


Figure 21. A circular interactive view of the genome of strain *Saccharothrix algeriensis* DSM 44581 with a circular interactive graphical representation of the alignment of genes. Tracks are displayed as concentric rings, from outermost to innermost: 1, reference position in the genome; 2, position and order of the 94 assembled contigs; 3, CDS-forward strands; 4, CDS-reverse strand; 5, non-coding features; 6, anti-microbial resistance genes; 7, virulence factors genes; 8, genes for transporters; 9, drug-targets; 10, GC content; and 11, GC skew.

1.3. Clusters of Orthologous Groups of proteins (COGs)

A complete genome annotation was conducted using the RAST server, and a total of 303 subsystems were classified with 82% of subsystems coverage (**Figure 22**). Subsystem features

belonged mostly to amino acids and derivatives (357 genes); followed by carbohydrates (258 genes); protein metabolism (181 genes); cofactors, vitamins, prosthetic groups, and pigments (168 genes); nucleosides and nucleotides (93 genes); respiration (82 genes); virulence, disease and defense (45 genes) and stress response (37 genes). Of note, this strain had 37 genes corresponding to the stress response subsystems, including oxidative stress (14 genes), stress response-no subcategory (18 genes), osmotic stress (2 genes), detoxification (2 genes) and periplasmic stress (1 gene) (**Figure 22**).

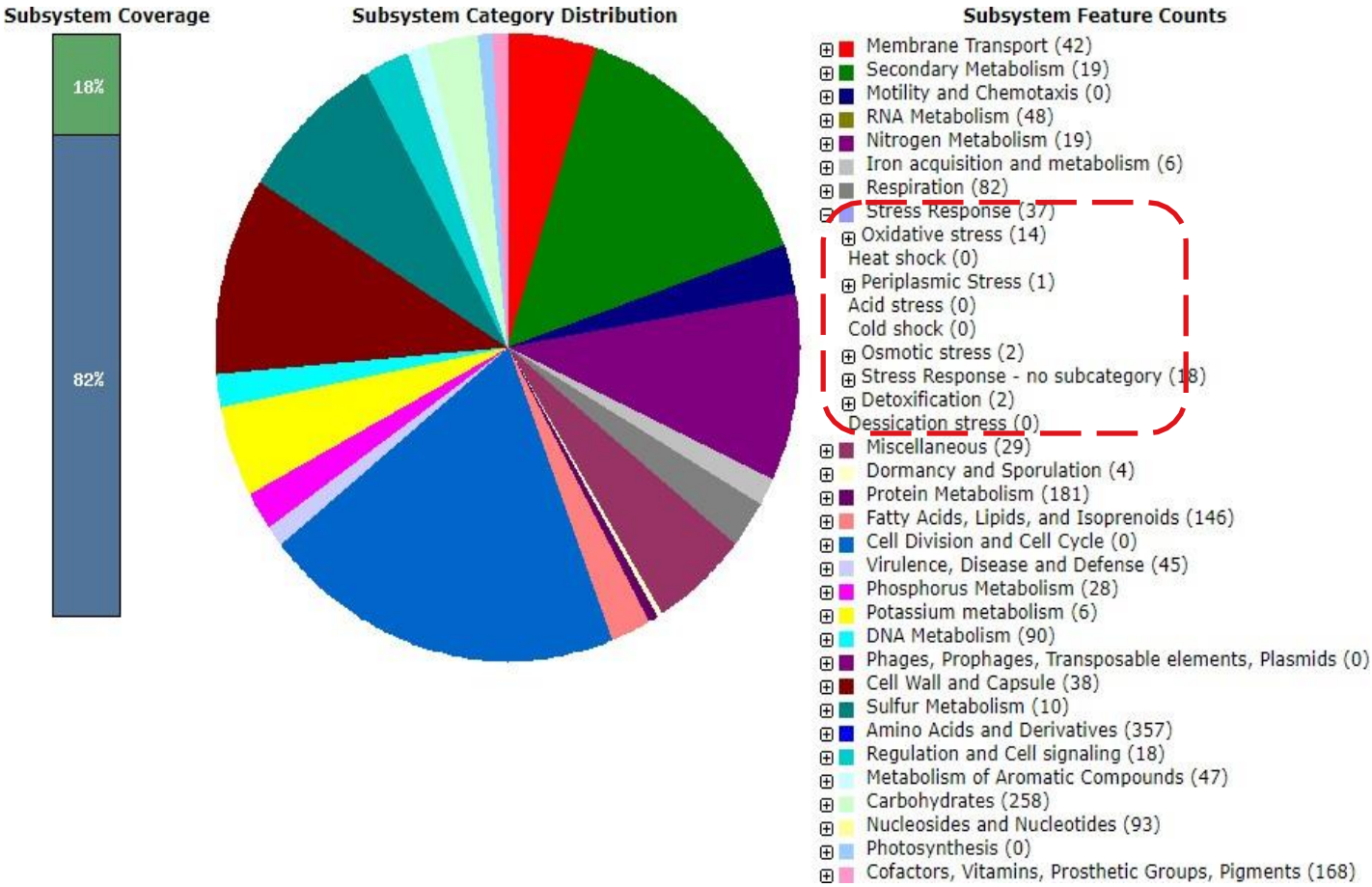


Figure 22. An overview of the subsystem categories of the *Saccharothrix algeriensis* strain DSM 44581. Genome annotation was conducted using the RAST server.

1.4. Genetic elements involved in *Saccharothrix algeriensis* DSM 44581

1.4.1. Osmotic stress tolerance

S. algeriensis strain DSM 44581 genome analysis revealed the presence of several genes that are involved in osmotic stress tolerance, including: sodium and potassium transporters, the osmoprotectant transport system, osmolyte biosynthesis and transport (e.g. proline, glutamate, glycine and glutamine) (**Table IV**). The abundance of genetic elements involved in osmotic stress tolerance in the strain DSM 44581 genome is in agreement with the functional data showing that strain DSM 44581 tolerates high salinity concentrations.

Table IV. Genes involved in osmotic stress tolerance.

Locus tag	Gene	Product	Pathway
FHGOJNDK_01817	ProA	Gamma-glutamyl phosphate reductase	**Proline biosynthesis**
FHGOJNDK_01821	ProB	Glutamate 5-kinase	
FHGOJNDK_03431	ProC	Pyrroline-5-carboxylate reductase	
FHGOJNDK_02149 FHGOJNDK_02153 FHGOJNDK_02164	GlnA	Glutamine synthetase	**Glutamine biosynthesis**
FHGOJNDK_03002	BetL	glycine betaine transporter	**Glycine betaine/proline transport system**
FHGOJNDK_04507 FHGOJNDK_05107	-	Glutamate synthase [NADPH] small chain	**Glutamate biosynthesis**
FHGOJNDK_01839	-	Folylpolyglutamate synthase	
FHGOJNDK_02168	-	Glutamine-dependent NAD(+) synthetase	**Glutamine biosynthesis**
FHGOJNDK_01572 FHGOJNDK_03168 FHGOJNDK_0327F HGOJNDK_06051	-	Glutamine transport ATP-binding protein GlnQ	**Glutamine transport**

1.4.2. Temperature stress resistance and spore formation and germination

The genome of *S. algeriensis* strain DSM 44581 contains several cold shock (csp) and heat shock (hsp) proteins, as well as other chaperone proteins involved in response to temperature stress (**Table V**). In addition, the genome contains several chaperones involved in heat stress response, including ClpC, ClpX, ClpB, DnaK, HemW, CopZ and ClpX, as well as other several small hsp genes. The genes of spore formation and germination involved in the process of spore formation and germination are presented in **Table VI**. Spore formation and germination genes are modest in the *S. algeriensis* strain DSM 44581 genome.

Table V. Genes involved in temperature stress resistance.

Locus tag	Gene	Product	Pathway
FHGOJNDK_02691 FHGOJNDK_02691	ClpC	ATP-dependent Clp protease ATP-binding subunit	**Temperature stress resistance**
FHGOJNDK_01850	ClpX	ATP-dependent Clp protease ATP-binding subunit	
FHGOJNDK_00057	-	18 kDa heat shock protein	
FHGOJNDK_03215 FHGOJNDK_05769	ClpB	Chaperone protein C	
FHGOJNDK_01771 FHGOJNDK_03210 FHGOJNDK_03695	DnaJ	Chaperone protein	
FHGOJNDK_01607 FHGOJNDK_02496 FHGOJNDK_03208 FHGOJNDK_03601	DnaK	Chaperone protein	
FHGOJNDK_01774	HemW	Heme chaperone	
FHGOJNDK_00872 FHGOJNDK_03369	CopZ	Copper chaperone	
FHGOJNDK_03189 FHGOJNDK_03627	-	10 kDa chaperonin	
FHGOJNDK_02608 FHGOJNDK_02947 FHGOJNDK_03628	-	60 kDa chaperonin	
FHGOJNDK_01851 FHGOJNDK_01852 FHGOJNDK_05718 FHGOJNDK_05719	-	ATP-dependent Clp protease proteolytic subunit	
FHGOJNDK_01850	ClpX	ATP-dependent Clp protease ATP-binding subunit	

Table VI. Genes involved in spore formation and germination.

Locus tag	Gene	Product	Pathway
FHGOJNDK_00556	SsgB	Sporulation-specific cell division protein	**Spore formation and germination**
FHGOJNDK_00938	kDa	55.5 kDa / 49.5 kDa sporulation protein	
FHGOJNDK_01329	SdpB	Sporulation-delaying protein S	
FHGOJNDK_01411	Spo0M	Sporulation-control protein spo0M	
FHGOJNDK_03057	Soj1	Sporulation initiation inhibitor protein Soj	
FHGOJNDK_03647	SsgB	Sporulation-specific cell division protein S	
FHGOJNDK_04403	Soj2	Sporulation initiation inhibitor protein	

1.4.3. Heavy metals resistance

Heavy metal transport/resistance genes are abundant in the *S. algeriensis* strain DSM 44581 genome (**Table VII**). These include several transport genes (for zinc, cobalt, copper, cadmium, manganese, arsenate, chromate, and fluoride) as well as genes encoding arsenate reductases and chromate transport. The large number of heavy metal resistance genes in the *S algeriensis* strain DSM 44581 genome suggests that this strain is able to deal with high levels of heavy metals.

Table VII. Genes involved in heavy metal transport and resistance.

Locus tag	Gene	Product	Pathway
FHGOJNDK_00731	-	Putative zinc-binding alcohol dehydrogenase	**Zinc transport**
FHGOJNDK_00988	-	Putative zinc-binding alcohol dehydrogenase	
FHGOJNDK_01342 FHGOJNDK_02615 FHGOJNDK_02711 FHGOJNDK_05893	-	ATP-dependent zinc metalloprotease FtsH	
FHGOJNDK_01753	-	Zinc uptake regulation protein	
FHGOJNDK_02200 FHGOJNDK_02201 FHGOJNDK_04368 FHGOJNDK_05354 FHGOJNDK_05356	-	Putative zinc protease	
FHGOJNDK_02673	ZnuC	Zinc import ATP-binding protein ZnuC	
FHGOJNDK_03837	Rip2	Putative zinc metalloprotease Rip2	
FHGOJNDK_03847	ZitB	Zinc transporter ZitB	
FHGOJNDK_04292	Rip1	Zinc metalloprotease Rip1	

FHGOJNDK_04472	-	Zinc carboxypeptidase	
FHGOJNDK_00104 FHGOJNDK_02235	CorA	Cobalt/magnesium transport protein CorA	**Cobalt-zinc-cadmium-Magnesium transport**
FHGOJNDK_01009	CobN	Aerobic cobaltochelatase subunit CobN	
FHGOJNDK_01783	-	Sirohydrochlorin cobaltochelatase	
FHGOJNDK_04725	MctB	Copper transporter MctB	**Copper resistance transport**
FHGOJNDK_03005	YcnJ	Copper transport protein YcnJ	
FHGOJNDK_02603	CutC	Copper homeostasis protein CutC	
FHGOJNDK_00872 FHGOJNDK_03369	CopZ	Copper chaperone CopZ	
FHGOJNDK_05167	ArsC	Arsenate reductase	**Arsenate reduction and transport**
GOJNDK_05168	-	Putative chromate transport protein	**Chromate reduction and transport**
FHGOJNDK_04860	LarMN	Putative fused nickel transport protein LarMN	**Nickel transport**
FHGOJNDK_02674	-	Manganese ABC transporter substrate-binding lipoprotein	**Manganese transport**
FHGOJNDK_02931	-	Putative manganese exporter	
FHGOJNDK_03201	MneP	Manganese efflux system protein	
FHGOJNDK_05055	-	Putative manganese catalase	
FHGOJNDK_05788	-	Manganese-dependent 2,3-dihydroxybiphenyl 1,2-dioxygenase	
FHGOJNDK_01702 FHGOJNDK_01703	CrcB	Putative fluoride ion transporter	**Fluoride transport**

1.4.4. Oxidative stress resistance

The *S. algeriensis* strain DSM 44581 genome contains multiple genes involved in the oxidative stress responses, including one Mn-Ni superoxide dismutase gene and a peroxide operon regulator (**Table VIII**). This collection of genes indicates a strong adaptation and resistance to reactive oxygen species (ROS), which are commonly produced and accumulated by bacteria under stressful conditions and are also known to be involved in plant defense responses.

Table VIII. Genes involved in the protection against oxidative stress.

Locus tag	Gene	Product	Pathway/function
FHGOJNDK_02549	[Mn]	Superoxide dismutase [Mn]	**Degradation of superoxide anion radicals**
FHGOJNDK_03890	[Ni]	Superoxide dismutase [Ni]	
FHGOJNDK_02806 FHGOJNDK_05754	KatE	Catalase	**Degradation of hydrogen peroxide**
FHGOJNDK_05853 FHGOJNDK_05853	HPII	Catalase HPII	
FHGOJNDK_05055	-	Putative manganese catalase	
FHGOJNDK_04797	DyP2	Multifunctional dye peroxidase DyP2	**Degradation of hydrogen peroxide and organic hydroperoxides**
FHGOJNDK_02019	-	Putative peroxiredoxin	
FHGOJNDK_02984	OsmC	Peroxiredoxin OsmC	
FHGOJNDK_02803	OhrA	Organic hydroperoxide resistance protein OhrA	
FHGOJNDK_01678	-	Alkyl hydroperoxide reductase E	
FHGOJNDK_03182	-	Organic hydroperoxide resistance transcriptional regulator	
FHGOJNDK_06073	-	Hydrogen peroxide-inducible genes activator	
FHGOJNDK_03195	-	Organic hydroperoxide resistance transcriptional regulator	
FHGOJNDK_05837	AhpD	Alkyl hydroperoxide reductase AhpD	
FHGOJNDK_04601	BpoC	Putative non-heme bromoperoxidase BpoC	
FHGOJNDK_03923	BtuE	Thioredoxin/glutathione peroxidase	
FHGOJNDK_02336 FHGOJNDK_05098	IspA	(2Z,6E)-farnesyl diphosphate synthase (2E,6E)-farnesyl diphosphate synthase	**Carotenoid biosynthesis 4,4′-diapophytoene and 4,4′-diaponeurosporene**
FHGOJNDK_00454	-	Phosphatidylinositol mannoside acyltransferase	
FHGOJNDK_00535	-	Putative diacyglycerol O-acyltransferase	
FHGOJNDK_00645	MnaT	L-amino acid N-acyltransferase MnaT	
FHGOJNDK_01651	-	1-acyl-sn-glycerol-3-phosphate acyltransferase	
FHGOJNDK_02907	-	Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase	
FHGOJNDK_00826	Tgs1	Putative diacyglycerol O-acyltransferase tgs1	

FHGOJNDK_01396	-	1-acyl-sn-glycerol-3-phosphate acyltransferase	
FHGOJNDK_04541	-	Putative acyltransferase	
FHGOJNDK_05089	-	1-acyl-sn-glycerol-3-phosphate acyltransferase	
FHGOJNDK_05682	-	Putative diacyglycerol O-acyltransferase	
FHGOJNDK_06003	-	Putative diacyglycerol O-acyltransferase	
FHGOJNDK_06003	-	Putative diacyglycerol O-acyltransferase	
FHGOJNDK_01058	-	UDP-3-O-(3-hydroxymyrisoyl) glucosamine N-acyltransferase	
FHGOJNDK_04066	-	4,4'-diapolycopene aldehyde oxidase	

1.4.5. Nitrogen, sulfur and phosphorus acquisition

S. algeriensis strain DSM 44581 possesses nitrate and nitrite reductase genes, as well as several genes involved in nitrate/nitrite transport (nirC). Phosphate transport and assimilation system genes pstC/pstA/phoD, as well as the sulfate reduction genes are present in the genome of strain DSM 44581 sat/cysH, sulfonates transport (ssuD) and degradation and account for its main organic dimethyl sulfone acquisition abilities (sfnG) (**Table IX**).

Table IX. Genes involved in nitrogen, sulfur and phosphorus metabolism.

Locus tag	Gene	Product	Pathway
FHGOJNDK_05644	NirC	Putative nitrate/nitrite transporter NarK2	**Nitrate/nitrite transporter**
FHGOJNDK_01778	Sat	Sulfate adenylyltransferase subunit 2	**Sulfate reduction**
FHGOJNDK_01777	CysH	Putative phosphoadenosine phosphosulfate reductase	
FHGOJNDK_05084	SsuD	Alkanesulfonate monooxygenase	**Sulfonates transport and degradation**
FHGOJNDK_01181	SfnG	FMNH(2)-dependent dimethylsulfone monooxygenase	**Dimethyl sulfone degradation**
FHGOJNDK_03339	PstC	Phosphate transport system permease protein PstC 1	**Phosphate transport**
FHGOJNDK_03340	PstA	Phosphate transport system permease protein PstA 1	
FHGOJNDK_00781 FHGOJNDK_00920 FHGOJNDK_03626	PhoD	Alkaline phosphatase D	**Phosphate assimilation**

1.4.6. Iron acquisition

The *S. algeriensis* strain DSM 44581 genome contains genes involved in ferric enterobactin (FepC, FepG, Fepd and BfrB), including iron-sulfat-fumarat complex (IdeR, FadF, PaoA and IscU), and also iron (3+) hydroxamate (YxeB), presented in (**Table X**).

Table X. Genes involved in iron acquisition.

Locus tag	Gene	Product	Pathway
FHGOJNDK_05981	FepC	Ferric enterobactin transport ATP-binding protein	**Ferric enterobatin transport-Ferritin upake-Ferritin**
FHGOJNDK_05980	FepG	Ferric enterobactin transport system permease protein	
FHGOJNDK_05979	FepD	Ferric enterobactin transport system permease protein	
FHGOJNDK_03324	-	Ferric uptake regulation protein	
FHGOJNDK_02950	BfrB	Ferritin	
FHGOJNDK_02055	FepC	Ferric enterobactin transport ATP-binding protein	
FHGOJNDK_00107	-	Cytochrome b6-f complex iron-sulfur subunit	**Iron-sulfat-Fumarat complex**
FHGOJNDK_01545	IdeR	Iron-dependent repressor	
FHGOJNDK_01593	-	Fumarate reductase iron-sulfur subunit	
FHGOJNDK_01645	-	Cytochrome bc1 complex Rieske iron-sulfur subunit	
FHGOJNDK_04150	-	Fumarate reductase iron-sulfur subunit	
FHGOJNDK_02254	-	Iron-sulfur cluster carrier protein	
FHGOJNDK_02781	-	Iron-sulfur cluster carrier protein	
FHGOJNDK_03203	FadF	Putative iron-sulfur-binding oxidoreductase	
FHGOJNDK_03719	-	Succinate dehydrogenase iron-sulfur subunit	
FHGOJNDK_05969	PaoA	Aldehyde oxidoreductase iron-sulfur-binding subunit	
FHGOJNDK_04937	-	Cytochrome b6-f complex iron-sulfur subunit	
FHGOJNDK_04938	-	Cytochrome b6-f complex iron-sulfur subunit	
FHGOJNDK_04986	IscU	Iron-sulfur cluster assembly scaffold protein	
FHGOJNDK_05143	-	Formate dehydrogenase, nitrate-inducible, iron-sulfur subunit	

FHGOJNDK_05978	YxeB	Iron(3+)-hydroxamate-binding protein	**Iron (3+) hydroxamate**

1.4.7. Genes involved in VOCs production/degradation

The *S. algeriensis* strain DSM 44581 genome contains genes involved in the acetoin biosynthesis, like acetolactate (IIBv1), and acetoin catabolism synthase (**Table XI**).

Table XI. Genes involved in VOCs production/degradation.

Locus tag	Gene	Product	Pathway
FHGOJNDK_04172	IlvB1	Acetolactate synthase large subunit	**Acetoin biosynthesis**
FHGOJNDK_04173	-	Acetolactate synthase small subunit	
FHGOJNDK_01542	AcuC	Acetoin utilization protein	
FHGOJNDK_01668 FHGOJNDK_01698 FHGOJNDK_04033	-	Dihydrolipoyl dehydrogenase	**Acetoin catabolism**

1.4.8. Genes involved in Quorum sensing

Quorum sensing. genes are abundant in the *S. algeriensis* strain DSM 44581 genome (**Table XII**). These include a number of genes like SsgB, kDa and Soj.

Table XII. Genes involved in Quorum sensing.

Locus tag	Gene	Product	Pathway
FHGOJNDK_04261 FHGOJNDK_04447	-	Signal peptidase I	**Quorum sensing**
FHGOJNDK_03424	SenX3	Signal-transduction histidine kinase	
FHGOJNDK_04244	FtsY	Signal recognition particle receptor	
FHGOJNDK_04252	-	Signal recognition particle protein	
FHGOJNDK_02443 FHGOJNDK_04244	MprB	Signal transduction histidine-protein kinase/phosphatase	

1.5. Genetic elements involved in *Saccharothrix algeriensis* strain DSM 44581 plant growth promotion activities

Several *Saccharothrix* strains, including *S. algeriensis* strain DSM 44581, are known to produce auxins, such as IAA, that play an important role in plant growth promotion and plant-microbe interactions (**Table XIII**).

Polyamines such as putrescine and spermidine have an important role in plant growth-promoting abilities. Furthermore, spermidine production by bacteria has been shown to decrease the activity of the tobacco ACC oxidase gene (ACO1) responsible for ethylene production, thereby lowering plant ethylene levels that affect a range of plant-microbe interactions. Several

polyamine metabolism and transport genes are found in the genome of *S. algeriensis* strain DSM 44581, including a potE gene (involved in putrescine production), also genes encoding spermidine synthase and several polyamine transport genes (potABCD and pupP), and tryptophan oxidase gene.

Involved in indole transformation and conversion (pyruvate carboxylase, AldA) consistent with an important role for polyamine metabolism in *S. algeriensis* strain DSM 44581 plant growth promotion abilities such as the gabT gene acting in the degradation of the beneficial molecule gamma-aminobutyric acid (GABA)which help respond to biotic and abiotic stresses. The DSM 44581 genome contained 81 genes predicted to be implicated in phytohormone production/modulation.

Table XIII. Genes involved in phytohormone production/modulation.

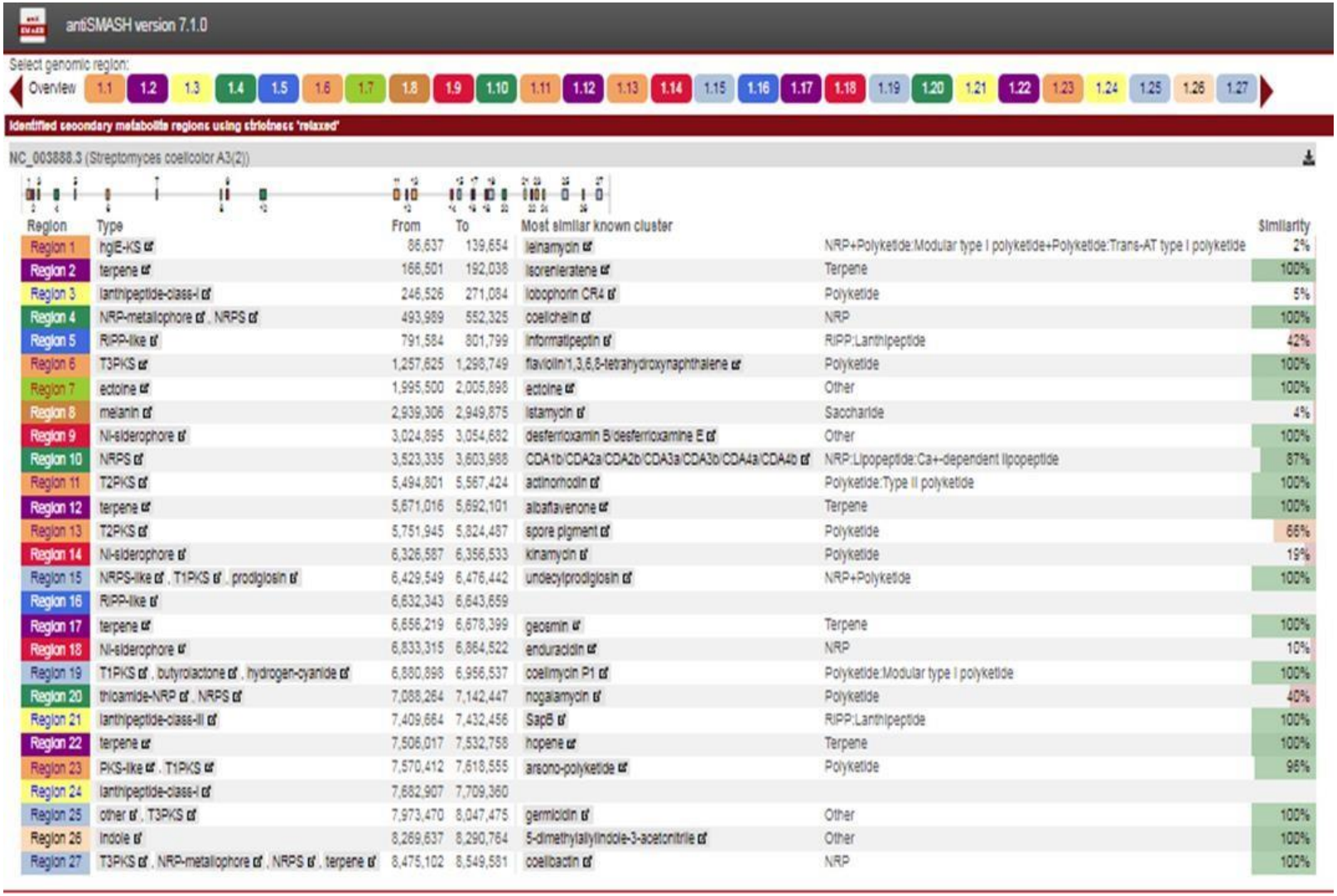
Locus tag	Gene	Product	Pathway
FHGOJNDK_00672	RebO	Flavin-dependent L-tryptophan oxidase	**L-tryptophan oxidase**
FHGOJNDK_05209	RebO	Flavin-dependent L-tryptophan oxidase	
FHGOJNDK_05794	-	4-aminobutyrate aminotransferase	**Tryptophan conversion to IAA intermediates**
FHGOJNDK_00137	-	Histidinol-phosphate aminotransferase	
FHGOJNDK_00189	-	Bifunctional aspartate aminotransferase and L-aspartate beta-decarboxylase	
FHGOJNDK_00444	-	Acetylornithine aminotransferase	
FHGOJNDK_00644 FHGOJNDK_02637 FHGOJNDK_03613 FHGOJNDK_05130	-	Glutamine-fructose-6-phosphate aminotransferase [isomerizing]	
FHGOJNDK_00664	-	Histidinol-phosphate aminotransferase	
FHGOJNDK_01300	DavT	5-aminovalerate aminotransferase	
FHGOJNDK_01421		Ornithine aminotransferase	
FHGOJNDK_01664	-	Branched-chain-amino-acid aminotransferase	
FHGOJNDK_01931	-	Serine-pyruvate aminotransferase	
FHGOJNDK_02187	-	Putative aminotransferase	
FHGOJNDK_02281	AlaC	Glutamate-pyruvate aminotransferase	
FHGOJNDK_02540		Phosphoserine aminotransferase	
FHGOJNDK_03262	-	Aspartate aminotransferase	
FHGOJNDK_02875		Putative N-succinyldiaminopimelate	
FHGOJNDK_03325	-	D-alanine aminotransferase	
FHGOJNDK_03392	DapC	Putative N-succinyldiaminopimelate aminotransferase	
FHGOJNDK_03535	-	Aspartate aminotransferase	
FHGOJNDK_04014	-	Diaminobutyrate-2-oxoglutarate	

		aminotransferase	
FHGOJNDK_04457	-	Adenosylmethionine-8-amino-7-oxononanoate	
FHGOJNDK_00137 FHGOJNDK_00664 FHGOJNDK_04474	-	Histidinol-phosphate aminotransferase	
FHGOJNDK_00444 FHGOJNDK_04703	-	Acetylornithine aminotransferase	
FHGOJNDK_02637 FHGOJNDK_03613 FHGOJNDK_05130	-	Glutamine-fructose-6-phosphate aminotransferase [isomerizing]	
FHGOJNDK_05794	-	4-aminobutyrate aminotransferase	
FHGOJNDK_01931	-	Serine-pyruvate aminotransferase	
FHGOJNDK_02187		Putative aminotransferase	
FHGOJNDK_02281	AlaC	Glutamate-pyruvate aminotransferase	
FHGOJNDK_02540	-	Phosphoserine aminotransferase	
FHGOJNDK_02875	-	Putative phenylalanine aminotransferase	
FHGOJNDK_03262	-	Aspartate aminotransferase Aminotransferase	
FHGOJNDK_03325	-	D-alanine aminotransferase	
FHGOJNDK_03392	DapC	Putative N-succinyldiaminopimelate aminotransferase	
FHGOJNDK_00189	-	Bifunctional aspartate aminotransferase and L-aspartate beta-decarboxylase	
FHGOJNDK_01421	-	Ornithine aminotransferase	
FHGOJNDK_01300	DavT	5-aminovalerate aminotransferase	
FHGOJNDK_01931	-	Serine-pyruvate aminotransferase	
FHGOJNDK_02187	-	Putative aminotransferase	
FHGOJNDK_02281	AlaC	Glutamate-pyruvate aminotransferase	
FHGOJNDK_04218	-	pyruvate carboxylase	**Indole-3-pyruvate transformation to indole-3-acetaldehyde**
FHGOJNDK_04496	-	Indole-3-glycerol phosphate synthase	**Indole-3-acetaldehyde conversion to IAA**
FHGOJNDK_00256	-	Phenylacetaldehyde dehydrogenase	
FHGOJNDK_00262	-	2-hydroxymuconic semialdehyde dehydrogenase	
FHGOJNDK_00461	-	Phenylacetaldehyde dehydrogenase	
FHGOJNDK_00530	-	N-succinylglutamate 5-semialdehyde dehydrogenase	
FHGOJNDK_00985	-	NAD/NADP-dependent betaine aldehyde	
FHGOJNDK_01944	-	Dehydrogenase fluoroacetaldehyde	

FHGOJNDK_03743 FHGOJNDK_03744 FHGOJNDK_04605 FHGOJNDK_05034 FHGOJNDK_05784 FHGOJNDK_05034 FHGOJNDK_05784	- - AldA - - - -	Aspartate-semialdehyde dehydrogenase NAD/NADP-dependent betaine Aldehyde dehydrogenase Putative aldehyde dehydrogenase Aldehyde dehydrogenase, thermostable Sulfoacetaldehyde dehydrogenase 2-aminomuconic 6-semialdehyde dehydrogenase	
FHGOJNDK_01580	MiaA	tRNA dimethylallyltransferase	**CK biosynthesis**
FHGOJNDK_05246	-	Biodegradative arginine decarboxylase	**Putrescine biosynthesis**
FHGOJNDK_00464	PotE	Putrescine transporter	
FHGOJNDK_04284 FHGOJNDK_04286	PotA PotB	Spermidine/putrescine import ATP-binding protein Spermidine/putrescine transport system permease protein	**Spermidine, putrescine biosynthesis**
FHGOJNDK_04553 FHGOJNDK_01079	-	S-adenosylmethionine decarboxylase proenzyme Spermine/spermidine acetyltransferase	**Spermidine biosynthesis**
FHGOJNDK_03895	-	N(4)-bis(aminopropyl)spermidine synthase	
FHGOJNDK_01089 FHGOJNDK_01937 FHGOJNDK_04552 FHGOJNDK_05120 FHGOJNDK_05122	-	Polyamine aminopropyltransferase	**Polyamine transport**
FHGOJNDK_04286	PotB	Spermidine/putrescine transport system permease protein	
FHGOJNDK_00214	PuuP	Putrescine importer	
FHGOJNDK_05794	GabT	4-aminobutyrate aminotransferase	**GABA degradation**

1.6. Biosynthetic gene clusters and secondary metabolites

The genome analysis of strain DSM 44581 using the antiSMASH software identified a gene cluster that encodes a specific product. The overview page of the analysis (interactive overview) provides a concise description of all regions present in the input submitted to antiSMASH (**Figure 23**). These regions are presented in 27 different region numbers, each associated with a specific product. Among these regions, there are 10 regions **(**Reg16: RiPP-like, Reg24: Ianthipepetide-classe-i, Reg1: hgIE-KS, Reg8: melanine, Reg3: Ianthipepetide-classe-I Polykitetide, Reg18: N-sidirophor, Reg20:Triomatide-NRP, Reg5: RiPP-like, Reg13: T2PS) that show a decrease in the percentage of cluster compounds. Promoting a novel molecule called "molicoler" and exposing additional regions that have a high degree of genetic similarity.



Region	Type	From	To	Most similar known cluster		Similarity
Region 1	hglE-KS	86,637	139,654	lelnamycin	NRP+Polyketide:Modular type I polyketide+Polyketide:Trans-AT type I polyketide	2%
Region 2	terpene	166,501	192,038	isorenieratene	Terpene	100%
Region 3	lanthipeptide-class-i	246,526	271,084	lobophorin CR4	Polyketide	5%
Region 4	NRP-metallophore , NRPS	493,989	552,325	coelichelin	NRP	100%
Region 5	RiPP-like	791,584	801,799	informatipeptin	RiPP:Lanthipeptide	42%
Region 6	T3PKS	1,257,625	1,298,749	flaviolin/1,3,6,8-tetrahydroxynaphthalene	Polyketide	100%
Region 7	ectoine	1,995,500	2,005,898	ectoine	Other	100%
Region 8	melanin	2,939,306	2,949,875	istamycin	Saccharide	4%
Region 9	NI-siderophore	3,024,895	3,054,682	desferrioxamin B/desferrioxamine E	Other	100%
Region 10	NRPS	3,523,335	3,603,988	CDA1b/CDA2a/CDA2b/CDA3a/CDA3b/CDA4a/CDA4b	NRP:Lipopeptide:Ca++-dependent lipopeptide	87%
Region 11	T2PKS	5,494,801	5,567,424	actinorhodin	Polyketide:Type II polyketide	100%
Region 12	terpene	5,671,016	5,692,101	albaflavenone	Terpene	100%
Region 13	T2PKS	5,751,945	5,824,487	spore pigment	Polyketide	66%
Region 14	NI-siderophore	6,326,587	6,356,533	kinamycin	Polyketide	19%
Region 15	NRPS-like , T1PKS , prodiglosin	6,429,549	6,476,442	undecylprodiglosin	NRP+Polyketide	100%
Region 16	RiPP-like	6,632,343	6,643,659			
Region 17	terpene	6,656,219	6,678,399	geosmin	Terpene	100%
Region 18	NI-siderophore	6,833,315	6,864,522	enduracidin	NRP	10%
Region 19	T1PKS , butyrolactone , hydrogen-cyanide	6,880,898	6,956,537	coelimycin P1	Polyketide:Modular type I polyketide	100%
Region 20	thioamide-NRP , NRPS	7,088,264	7,142,447	nogalamycin	Polyketide	40%
Region 21	lanthipeptide-class-iii	7,409,664	7,432,456	SapB	RiPP:Lanthipeptide	100%
Region 22	terpene	7,506,017	7,532,758	hopene	Terpene	100%
Region 23	PKS-like , T1PKS	7,570,412	7,618,555	arsono-polyketide	Polyketide	96%
Region 24	lanthipeptide-class-i	7,682,907	7,709,360			
Region 25	other , T3PKS	7,973,470	8,047,475	germicidin	Other	100%
Region 26	indole	8,269,637	8,290,764	5-dimethylallylindole-3-acetonitrile	Other	100%
Region 27	T3PKS , NRP-metallophore , NRPS , terpene	8,475,102	8,549,581	coelibactin	NRP	100%

Figure 23. Genome regions of *Saccharothrix algeriensis* strain DSM 44581 using antiSMASH.

Ranthipeptides are a recently discovered group of natural products that fall within the ribosomally synthesised and post-translationally modified peptide (RiPP) superfamily. They are characterised by their radical non-a thioether-containing structure. Ribosomally synthesised and post-translationally modified peptides (RiPPs) are a group of natural compounds that exhibit a wide range of structural characteristics and biological functions. These include antibacterial, anticancer, immunosuppressive, and antinociceptive effects.

1.7. Roary results

The outputs of summary result statistics of the analysis delineate a comparison between the pan-genome and core genome of the three strains isolated from Saharan soil and analysed in this study (*S. algeriensis*, *S. Saccharothrix saharensis* and *S. tamanrassetensis*). The pan-genome encompasses a total of 20387 genes, with the core genome accounting for merely 1% (293), while the shell genome constitutes 99% (20094), and the cloud genome presents a striking 0% (0), marking a substantial distinction among these species (**Table XIV and Figure 24**).

Table XIV. Summary results of comparative analyses.

Column 1	Column 2	Column 3
Core genes	(99% <= strains <= 100%)	293
Soft core genes	(95% <= strains < 99%)	0
Shell genes	(15% <= strains < 95%)	20094
Cloud genes	(0% <= strains < 15%)	0
Total genes	(0% <= strains <= 100%)	20387

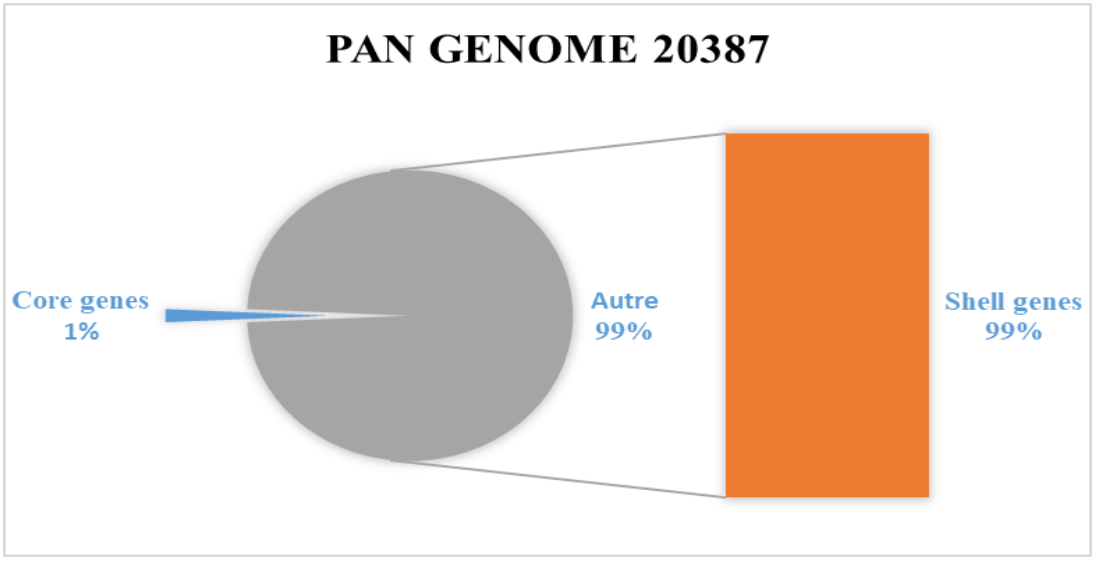


Figure 24. Schematic representation of results of comparative analyses.

2. Discussion

In our work, the confirmation of the taxonomic affiliation of a strain isolated from a Saharan environment in Algeria, which is *Saccharothrix algeriensis* strain DSM 44581, by genomic approaches was achieved. In addition to this, the exploration and prediction of the presence of genes of COGs, which are genetic elements involved in stress resistance, soil and plant colonization abilities, PGPB abilities, biosynthetic gene clusters and secondary metabolites production.

The use of TYGS server for the taxonomic study of the chosen strain assigned it with of *Saccharothrix algeriensis* in the phylogenetic tree based on the 16S rDNA and the phylogenomic tree based on the genome sequences (Meier-Kolthof and Göker, 2019). From the Genome BLAST Distance Phylogeny approach (GBDP), we obtained a distance formula d5, a pseudo-bootstrap support values >60% from 100 replications, with an average branch support of 93.7%. The phylogenomic tree was rooted at the midpoint (Meier-Kolthof *et al*., 2013). It is important to note that there is no difference between the phylogenetic and phylogenemic studies because the species *S. tamanrassetensis* CECT 8640 is linked the clade of *S. algeriensis* DSM 44581 and *S. austarliensis* DSH 43800 with the same Bootstrap, and *S. syringae* NRRL B-16468 leads to reclassify in the clade of *S. xinjiangensis* JCM 1232 with the same Bootstrap.

From DFAST server results, the general genome features of strain DSM 44581 consisted of 6,878,582 bps with a GC content of 74.1%. The genome carries CDSs 6,229 coding genes (including repetition of some genes in different places of the studied genome), as was reported by Tanizawa *et al.* (2019). The genome annotation shown in the circular diagram of BV-BRC presents the outcomes of position and order of the 94 assembled contigs, anti-microbial resistance genes, genes for transporters, drug-targets and GC content (Davis *et al*., 2019). Additionally, the COGs in the genome annotation were obtained using the RAST server with a total of 303 subsystems that were classified with 82% of subsystems coverage. Subsystem features belonged mostly to stress response genes (37 genes) that include oxidative stress (14 genes), stress response-no subcategory (18 genes), osmotic stress (2 genes), detoxification (2 genes) and periplasmic stress (1 gene), cofactors, vitamins, prosthetic groups, and pigments virulence, disease and defense (45 genes).

The properties of this actinobacteria show promising results, as this strain possesses plant growth-enhancing properties that have beneficial effects on plant growth and fight plant pathogens, as was reported by Al et Egamberdieva (2011). For example, many mechanisms have been revealed for explaining how PGPRs stimulate plant growth, including: production of antioxidants and modification of biotic and abiotic stresses. The ability to produce plant hormones such as auxine (IAA). Atmospheric nitrogen fixation. Dissolution of inorganic phosphate and other nutrients. Antagonism against microorganisms responsible for plant diseases through the production of antibiotics, enzymes that hydrolysate fungal components such as chitinase and siderophores. All these mechanisms have been reported by many researchers including; Goudjal *et al*. (2014), Boukaya *et al*. (2018), Wahyudi *et al*. (2019), Djemouai *et al*. (2022) and Dif *et al*. (2022).

Plants have a defense system called an antioxidant system, specific to the production of free radicals under stress conditions. A wide range of enzymes and antioxidant compounds are produced, such as superoxide dismutase (SOD), peroxydases (POX), catalase (CAT), polyphenol oxidase (PPO), which are involved in cellular protection as reported Thakker *et al*. (2013).

Many actinobacterial species have the ability to produce IAA, which is a crucial plant hormone that plays a significant role in promoting somatic growth by facilitating cell and tissue division (Duca *et al*., 2014). Many *Saccharothrix* strains, notably *S. algeriensis* strain DSM 44581, as reported by Benadjila *et al*. (2022), have the capability to synthesize auxins, specifically IAA. Polyamines, including putrescine and spermidine, have a significant impact on enhancing plant development. In addition, research has demonstrated that bacteria can produce spermidine, which in turn decreases the activity of the tobacco ACC oxidase gene (ACO1). This gene is responsible for the generation of ethylene in plants, so when its activity is reduced, the levels of ethylene in the plant decrease. Jafari *et al*. (2012) highlighted that this reduction in ethylene levels has an impact on several interactions between plants and microbes. The genome of *S. algeriensis* strain DSM 44581 contained various genes related to polyamine metabolism and transport. They play a role in the transformation and conversion of indole, aiding in the response to both biotic and abiotic stressors (El-Tarabily *et al*., 2020).

The antifungal activity of *Saccharothrix* species against various fungi, such as *Aspergillus*, *Fusarium*, *Penicillium*, and *Ascochyta*, has been previously documented (Boubetra *et al*., 2013; Zitouni *et al*., 2005). The research conducted by Compant *et al*. (2013) and Muzammil *et al.*

(2012) have also reported the efficacy of *S. algeriensis* NRRLB-24137 to suppress *Botrytis cinerea* both in laboratory conditions and in grapevine plants. The strain NRRL B-24137 exhibits notable antifungal capabilities due to its capacity to establish a presence in the roots of plants (Compant *et al*., 2013). Moreover, *S. algeriensis* NRRL B-24137 is recognized for its ability to generate dithiolopyrrolones, which are biologically active compounds believed to possess antifungal properties (Merrouche *et al*., 2010; Merrouche *et al*., 2011). Deketelaere *et al*. (2017) suggest that the most effective technique for using biocontrol agents is to place them in close proximity to the roots. In fact, Merrouche *et al*. (2017) conducted a study on the soil population density of strain NRRL B-24137 and *Fusarium oxxysporum* over a duration of 9 weeks. The results showed that the strain NRRL B-24137 significantly decreased the density of *Fusarium* in the soil.

The antiSMASH analysis attains a remarkably high level of accuracy in its annotations for each cluster identified in strain DSM 44581 genome. Furthermore, the analysis showed different regions for the production of some metabolic products, such as RiPP-like, Ianthipepetide-classe-I, hgIE-KS, melanine, polykitetide, N-sidirophor and triomatide-NRP. A novel molecule called panthipeptides is a recently discovered group of natural products that fall within the ribosomally synthesised and post-translationally modified peptide (RiPP) superfamily. They are characterized by non-a thioether-containing structures (Wu *et al*., 2021; Teixeira *et al*., 2021). The comparative analysis and Roary results show the outputs of the analysis that delineated a comparison between the pan-genome and the core genome of the compared three strains isolated from Saharan soil (*S. algeriensis*, *S. saharensis* and *S. tamanrassetensis*). The pan-genome encompassed a total of 20387 genes, with the core genome accounting for merely 1% (293). The cloud genome presents a striking 0%, marking a substantial distinction among these species.

Conclusion

This genomic study explains the significance of the actinobacterial strain *Saccharothrix algeriensis* DSM 44581 as possessing advantageous properties for plants and industry. The capacity to use this strain as an antagonist against pathogens comes from its substantial gene resources of biological gene clusters. This strain also promotes growth, supplies nutrients, and helps in the resistance to biotic and abiotic stresses. This strain may be also given exciting prospects in agricultural applications. It can be employed to increase and enhance the yield of crops and lower the use of chemical pesticide treatments. This can lead to guaranteeing the best possible way for food security in the years ahead, as using this strain is a sustainable approach to address resource shortages in the context of climate change.

In this study, the actinobacterial strain *Saccharothrix algeriensis* DSM 44581 was studied by genomic analysis which showed many genes that are involved in enhancement of the plant growth by different mechanisms. *S. algeriansis* strain DSM 44581 is guarded as a highly effective source of plant growth-promoting bacteria (PGPB) that may be employed as crop inoculum. This is due to the remarkable abilities they possess in terms of osmoprotection and enhancing plant development. Furthermore, the examination of the strain's genome identified a multitude of genes related to processes such as plant association and intense competition for nutrients and habitats. It indicated the creation of significant plant growth regulators, some of which also serve as stress protectors. The strain is believed to have the capacity to create osmolytes and may include several enzymatic antioxidant systems in response to harmful stress-induced overproduction. These activities are regarded as highly effective in reducing the negative effects of stress. Thus, within the framework of promoting technological advancement for sustainable agriculture, the biological methods would enable us to explore the economically viable growth of plants in Saharan settings. The strain DSM 44581 should be further evaluated *in vivo* in various Saharan soil conditions to confirm its potential as a PGPB and stress-protecting agent. The comparative analyses showed that strain DSM 44581 was different from the other two strains that were also isolated from Saharan soil, presenting a diminutive percentage 1% of the core genome.

In perspectives, a confirmation of some of the mechanisms that were highlighted in the genomic analysis can be done in laboratory conditions as well as with *in vivo* studies.

The possession of a national center for bioinformatics as that of many countries including Egypt (e.g Egyptian Network of Bioinformatics and Gnomic 2012), South Africa, and Tunisian.

References

References

1. Amador, L. A. A. (2017). Étude bioinformatique des génomes de porphyromonas (Doctoral dissertation), 222.
2. Aouiche, A., Sabaou, N., Meklat, A., Zitouni, A., Bijani, C., Mathieu, F., & Lebrihi, A. (2011). *Saccharothrix* sp. PAL54, a new chloramphenicol-producing strain isolated from a Saharan soil. *World Journal of Microbiology and Biotechnology*, *28*(3), 943-951. https://doi.org/10.1007/s11274-011-0892-2
3. Asif, R., Siddique, M. H., Hayat, S., Rasul, I., Nadeem, H., Faisal, M., Waseem, M., Zakki, S. A., Zitouni, A., & Muzammil, S. (2023). Efficacy of *Saccharothrix algeriensis* NRRL B-24137 to suppress *Fusarium oxysporum* f.sp. *vasinfectum* induced wilt disease in cotton. *PeerJ*, *11*, e14754. https://doi.org/10.7717/peerj.14754
4. Athalye, M., Goodfellow, M., Lacey, J., & White, R. P. (1985). Numerical classification of *Actinomadura* and *Nocardiopsis*. *International Journal of Systematic Bacteriology*, *35*(1), 86-98. https://doi.org/10.1099/00207713-35-1-86
5. Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H., Clément, C., Ouhdouch, Y., & Van Wezel, G. P. (2016). Taxonomy, physiology, and natural products of Actinobacteria. *Microbiology and Molecular Biology Reviews*, *80*(1), 1-43. https://doi.org/10.1128/mmbr.00019-15
6. Benadjila, A., Zamoum, M., Aouar, L., Zitouni, A., & Goudjal, Y. (2022). Optimization of cultural conditions using response surface methodology and modeling of indole-3-acetic acid production by *Saccharothrix texasensis* MB15. *Biocatalysis and Agricultural Biotechnology*, 39, 102271. https://doi.org/10.1016/j.bcab.2021.102271
7. Betsy Foxman. (2012). A primer of molecular biology. in molecular tools and infectious disease epidemiology (pp. P:53-78). Betsy Foxman, Academic Press. https://doi.org/10.1016/B978-0-12-374133-2.00005-8.
8. Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B. R., Metcalf, W. W., Helfrich, E. J., Van Wezel, G. P., Medema, M. H., & Weber, T. (2023). AntiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, *51*(W1), W46-W50. https://doi.org/10.1093/nar/gkad344
9. Boubetra, D., Sabaou, N., Zitouni, A., Bijani, C., Lebrihi, A., & Mathieu, F. (2013). Taxonomy and chemical characterization of new antibiotics produced by *Saccharothrix* SA198 isolated from a Saharan soil. *Microbiological Research*, 168(4), 223-230. https://doi.org/10.1016/j.micres.2012.11.005
10. Bouras, N., Merrouche, R., Lamari, L., Mathieu, F., Sabaou, N., & Lebrihi, A. (2008). Precursor-directed biosynthesis of new dithiolopyrrolone analogs by *Saccharothrix algeriensis* NRRL B-24137. *Process Biochemistry*, *43*(11), 1244-1252. https://doi.org/10.1016/j.procbio.2008.07.008
11. Chen, J., Xu, L., Zhou, Y., & Han, B. (2021). Natural products from Actinomycetes associated with marine organisms. *Marine Drugs*, *19*(11), 629. https://doi.org/10.3390/md19110629
12. Compant, S., & Mathieu, F. (2013). Use of DOPE- FISH tool to better visualize colonization of plants by beneficial bacteria? An example with *Saccharothrix algeriensis* NRRL B- 24137 colonizing grapevine plants. Molecular Microbial Ecology of the Rhizosphere, 929-931. https://doi.org/10.1002/9781118297674.ch87
13. Davis, J. J., Boisvert, S., Brettin, T., Kenyon, R. W., Mao, C., Olson, R., Overbeek, R., Santerre, J., Shukla, M., Wattam, A. R., Will, R., Xia, F., & Stevens, R. (2016). Antimicrobial resistance prediction in PATRIC and RAST. Scientific Reports, 6(1). https://doi.org/10.1038/srep27930
14. De Sousa, C. S., Hassan, S. S., Pinto, A. C., Silva, W. M., De Almeida, S. S., Soares, S. D. C., ... & Azevedo, V. (2018). Microbial omics: applications in biotechnology. In Omics Technologies and BioEngineering Academic Press. 3-20 pp.
15. Deketelaere, S., Tyvaert, L., França, S. C., & Höfte, M. (2017). Desirable traits of a good biocontrol agent against verticillium wilt. Frontiers in Microbiology, 8. https://doi.org/10.3389/fmicb.2017.01186
16. Dif, G., & Zitouni, A. (2023). Revealing of potential plant growth-enhancing traits through in Silico Genomic analysis of *Bacillus rhizoplanae* CIP111899. Basic and Applied Sciences - Scientific Journal of King Faisal University, 1-7. https://doi.org/10.37575/b/sci/230003
17. Dif, G., Belaouni, H. A., Yekkour, A., Goudjal, Y., Djemouai, N., Peňázová, E., Čechová, J., Berraf-Tebbal, A., Eichmeier, A., & Zitouni, A. (2021). Performance of halotolerant bacteria associated with Sahara-inhabiting halophytes *Atriplex halimus* L. and *Lygeum spartum* L. ameliorate tomato plant growth and tolerance to saline stress: From selective isolation to genomic analysis of potential determinants. *World Journal of Microbiology and Biotechnology*, 38(1). https://doi.org/10.1007/s11274-021-03203-2
18. Duca D, Lorv J, Patten CL, Rose D, Glick BR (2014) Indole-3-acetic acid in plant-microbe interactions. Anton Leeuw 106:85–125. https://doi.org/10.1007/s10482-013-0095-y
19. Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797. https://doi.org/10.1093/nar/gkh340
20. Farris, J. S., & Goloboff, P. A. (2008). Is REP a measure of "objective support"? *Cladistics*, *24*(6), 1065-1069. https://doi.org/10.1111/j.1096-0031.2008.00238.x
21. Flärdh, K., & Buttner, M. J. (2009). *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology*, *7*(1), 36-49. https://doi.org/10.1038/nrmicro1968

22. Garrity, G., & Cole, J. (2007). Taxonomic outline of the bacteria and Archaea. https://doi.org/10.1601/toba7.7
23. Harrison, A., Pearl, F., Mott, R., Thornton, J., and Orengo, C. 2002. Quantifying the similarities within fold space. *J. Mol. Biol.* 323 909–926.
24. Henz, S. R., Huson, D. H., Auch, A. F., Nieselt-Struwe, K., & Schuster, S. C. (2004). Whole-genome prokaryotic phylogeny. *Bioinformatics*, *21*(10), 2329-2335. https://doi.org/10.1093/bioinformatics/bth324
25. Hodkinson, B. P., & Grice, E. A. (2015). Next-generation sequencing: A review of technologies and tools for wound microbiome research. *Advances in Wound Care*, *4*(1), 50-58. https://doi.org/10.1089/wound.2014.0542
26. Innocenti, N., Golumbeanu, M., Fouquier d'Hérouël, A., Lacoux, C., Bonnin, R. A., Kennedy, S. P., Wessner, F., Serror, P., Bouloc, P., Repoila, F., & Aurell, E. (2015). Whole-genome mapping of 5′ RNA ends in bacteria by tagged sequencing: A comprehensive view in *Enterococcus faecalis*. *RNA*, *21*(5), 1018-1030. https://doi.org/10.1261/rna.048470.114
27. Kausar, A. & Gull, M. (2014). Effect of potassium sulphate on the growth and uptake of nutrients in wheat (*Triticum aestivum* L.) under salt stressed conditions. Journal of Agricultural Science, 6(8): 101-112.
28. Khamna, S., Yokota, A. and Lumyong, S. (2009). Actinomycetes isolated from medicinal plant rhizosphere soils: diversity and screening of antifungal compounds, indole-3-acetic acid and siderophore production. *World Journal of Microbiology and Biotechnology*, 25(4), 649–55. https://doi.org/10.1007/s11274-008-9933-x
29. Kreft, Ł., Botzki, A., Coppens, F., Vandepoele, K., & Van Bel, M. (2017). PhyD3: A phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics*, *33*(18), 2946-2947. https://doi.org/10.1093/bioinformatics/btx324
30. Kumar, K. R., Cowley, M. J., & Davis, R. L. (2019). Next-generation sequencing and emerging technologies. *Seminars in Thrombosis and Hemostasis*, *45*(07), 661-673. https://doi.org/10.1055/s-0039-1688446
31. Labeda, D. P., Hatano, K., Kroppenstedt, R. M., & Tamura, T. (2001). Revival of the genus *Lentzea* and proposal for Lechevalieria Gen. Nov. *International Journal of Systematic and Evolutionary Microbiology*, *51*(3), 1045-1050. https://doi.org/10.1099/00207713-51-3-1045
32. Labeda, D. P., Testa, R. T., Lechevalier, M. P., & Lechevalier, H. A. (1984). *Saccharothrix*: A new genus of the Actinomycetales related to *Nocardiopsis*. *International Journal of Systematic Bacteriology*, *34*(4), 426-431. https://doi.org/10.1099/00207713-34-4-426
33. Lacoste, C., Fabre, A., Pécheux, C., Lévy, N., Krahn, M., Malzac, P., Bonello-Palot, N., Badens, C., & Bourgeois, P. (2017). Le séquençage d'ADN a haut debit en pratique clinique. *Archives de Pédiatrie*, *24*(4), 373-383. https://doi.org/10.1016/j.arcped.2017.01.008
34. Lamari, L., Zitouni, A., Dob, T., Sabaou, N., Lebrihi, A., Germain, P., Seguin, E., & Tillequin, F. (2002). New Dithiolopyrrolone antibiotics from *Saccharothrix* sp. SA 233. II. Physicochemical properties and structure elucidation. *The Journal of Antibiotics*, *55*(8), 702-706. https://doi.org/10.7164/antibiotics.55.702
35. Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program: Table 1. *Molecular Biology and Evolution*, *32*(10), 2798-2800. https://doi.org/10.1093/molbev/msv150
36. Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., Fox, B. G., & Currie, C. R. (2016). Evolution and ecology of*Actinobacteria*and their Bioenergy applications. *Annual Review of Microbiology*, *70*(1), 235-254. https://doi.org/10.1146/annurev-micro-102215-095748
37. Liu, B. (2023). The annotation pipeline for the genome of a snake v1. https://doi.org/10.17504/protocols.io.4r3l27ez4g1y/v1,
38. Ludwig, W. Euzéby, J., Schumann, P., Busse, H., Trujillo, M. E., Kämpfer, P., & Whitman, W. B. (2012). Road map of the phylum Actinobacteria. *Bergey's Manual® of Systematic Bacteriology*, 1-28. https://doi.org/10.1007/978-0-387-68233-4_1
39. Marnix H. Medema, Kai Blin, Peter Cimermancic, Victor de Jager, Piotr Zakrzewski, Michael A. Fischbach, Tilmann Weber, Eriko Takano, Rainer Breitling, antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences, Nucleic Acids Research, Volume 39, Issue suppl_2, 1 July 2011, Pages W339–W346. https://doi.org/10.1093/nar/gkr466
40. Medema, M. H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., & Breitling, R. (2011). AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, *39*(suppl_2), W339-W346. https://doi.org/10.1093/nar/gkr466
41. Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10210-3
42. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, *14*(1). https://doi.org/10.1186/1471-2105-14-60

43. Meier-Kolthoff, J. P., Hahnke, R. L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B., Goodwin, L. A., Chertkov, O., Reddy, T., Pati, A., Ivanova, N. N., Markowitz, V., Kyrpides, N. C., Woyke, T., Göker, M., & Klenk, H. (2014). Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Standards in Genomic Sciences*, *9*(1). https://doi.org/10.1186/1944-3277-9-2
44. Meinel, D., Seth-Smith, H. M., & Egli, A. (2017). «Whole genome sequencing». *Forum Médical Suisse – Swiss Medical Forum*, *17*(1516). https://doi.org/10.4414/fms.2017.02917
45. Mercier, J. (2017). Logique paracohérente pour l'annotation fonctionnelle des génomes au travers de réseaux biologiques (Doctoral dissertation, Paris Saclay). P :172.
46. Merrouche, R., Bouras, N., Coppel, Y., Mathieu, F., Sabaou, N., & Lebrihi, A. (2011). New dithiolopyrrolone antibiotics induced by adding sorbic acid to the culture medium of *Saccharothrix algeriensis* NRRL B-24137. *FEMS Microbiology Letters*, *318*(1), 41-46. https://doi.org/10.1111/j.1574-6968.2011.02246.x
47. Merrouche, R., Yekkour, A., Lamari, L., Zitouni, A., Mathieu, F., & Sabaou, N. (2017). Efficiency of *Saccharothrix algeriensis* NRRL B-24137 and its produced Antifungal Dithiolopyrrolones compounds to suppress *Fusarium oxysporum*-induced wilt disease occurring in some cultivated crops. *Arabian Journal for Science and Engineering*, 42(6), 2321-2327. https://doi.org/10.1007/s13369-017-2504-4
48. Muzammil, S., Graillon, C., Saria, R., Mathieu, F., Lebrihi, A., & Compant, S. (2013). The Saharan isolate *Saccharothrix algeriensis* NRRL B-24137 induces systemic resistance in *Arabidopsis thaliana* seedlings against *Botrytis cinerea*. *Plant and Soil*, *374*(1-2), 423-434. https://doi.org/10.1007/s11104-013-1864-0
49. Nakkeeran, S., Fernando, W. D., & Siddiqui, Z. A. (2005). Plant growth promoting rhizobacteria formulations and its scope in commercialization for the management of pests and diseases. In PGPR: Biocontrol and biofertilization Springer, Dordrecht. P:257-296.
50. Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0997-x
51. Ortet, P. (2018). De l'annotation automatique des génomes à l'annotation experte pour la génomique environnementale, Aix-Marseille, P:139,Thèse Doctorat.
52. Ouargli Moufida. (2018). Les actinomycètes producteurs de molécules bioactives. *UNIVERSITE Badji Mokhtar-Annaba*, P:21, Thèse Doctorat.
53. Palaniyandi, S. A., Yang, S. H., Zhang, L., & Suh, J. (2013). Effects of actinobacteria on plant disease suppression and growth promotion. *Applied Microbiology and Biotechnology*, *97*(22), 9621-9636. https://doi.org/10.1007/s00253-013-5206-1
54. Pallavi Mishra, Ranjeet Maurya, Himanshu Avashthi, Shikha Mittal, Muktesh Chandra, & Pramod Wasudeo Ramteke. (2022). Genome assembly and annotation. In Bioinformatics: Methods and Applications (pp. P:49-66). Dev Bukhsh Singh, Rajesh Kumar Pathak, Bioinformatics, Academic Press. https://doi.org/10.1016/B978-0-323-89775-4.00013-4
55. Pathak, R. K., Singh, D. B., & Singh, R. (2022). Introduction to basics of bioinformatics. *Bioinformatics*, 1-15. https://doi.org/10.1016/b978-0-323-89775-4.00006-7
56. Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., & Stamatakis, A. (2010). How many bootstrap replicates are necessary? *Journal of Computational Biology*, *17*(3), 337-354. https://doi.org/10.1089/cmb.2009.0179
57. Pena, R. T., Blasco, L., Ambroa, A., González-Pedrajo, B., Fernández-García, L., López, M., Bleriot, I., Bou, G., García-Contreras, R., Wood, T. K., & Tomás, M. (2019). Relationship between quorum sensing and secretion systems. *Frontiers in Microbiology*, 10. https://doi.org/10.3389/fmicb.2019.01100
58. Saker, S., Almaksour, Z. A., Chorin, A., Lebrihi, A., & Mathieu, F. (2014). Enzymatic synthesis of Dithiolopyrrolone antibiotics using cell-free extract of *Saccharothrix algeriensis* NRRL B-24137 and biochemical characterization of two Pyrrothine N-acyltransferases in this extract. *Journal of Microbiology and Biotechnology*, *24*(1), 26-35. https://doi.org/10.4014/jmb.1309.09035
59. Salam, N., Jiao, J. Y., Zhang, X. T., & Li, W. J. (2020). Update on the classification of higher ranks in the phylum Actinobacteria. *International Journal of Systematic and Evolutionary Microbiology*, *70*(2), 1331-1355. https://doi.org/10.1099/ijsem.0.003920
60. Schippers, A., Bosecker, K., Willscher, S., Spröer, C., Schumann, P., & Kroppenstedt, R. M. (2002). *Nocardiopsis metallicus* sp. Nov., a metal-leaching actinomycete isolated from an alkaline slag dump. *International Journal of Systematic and Evolutionary Microbiology*, *52*(6), 2291-2295. https://doi.org/10.1099/00207713-52-6-2291
61. Selim, M. S., Abdelhamid, S. A., & Mohamed, S. S. (2021). Secondary metabolites and biodiversity of actinomycetes. *Journal of Genetic Engineering and Biotechnology*, *19*(1), 72. https://doi.org/10.1186/s43141-021-00156-9

62. Serraj, R., Krishnan, L., & Pingali, P. (2018). Agriculture and food systems to 2050: A synthesis. World Scientific Series in Grand Public Policy Challenges of the 21st Century, 3-45. https://doi.org/10.1142/9789813278356_0001
63. Setubal, J. C., Almeida, N. F., & Wattam, A. R. (2018). Comparative Genomics for Prokaryotes. In J. C. Setubal, J. Stoye, & P. F. Stadler (Éds.), Comparative Genomics (Vol. 1704, p. 55‑78). Springer New York. https://doi.org/10.1007/978-1-4939-7463-4_3.
64. Sharma, D., Singh, S., & Mittal, M. (2021). Bioinformatics and RNA. https://doi.org/10.1201/9781003107736
65. Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*. 11 739–747.
66. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313. https://doi.org/10.1093/bioinformatics/btu033
67. Subramani, R., & Aalbersberg, W. (2012). Marine actinomycetes: An ongoing source of novel bioactive metabolites. *Microbiological Research*, *167*(10), 571-580. https://doi.org/10.1016/j.micres.2012.06.005
68. Tanizawa, Y., Fujisawa, T., & Nakamura, Y. (2017). DFAST: A flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics*, *34*(6), 1037-1039. https://doi.org/10.1093/bioinformatics/btx713
69. Tanizawa, Y., Fujisawa, T., Arita, M., & Nakamura, Y. (2019). Generating publication-ready prokaryotic genome annotations with DFAST. *Methods in Molecular Biology*, 215-226. https://doi.org/10.1007/978-1-4939-9173-0_13
70. Teixeira, G. M., Mosela, M., Nicoletto, M. L. A., Ribeiro, R. A., Hungria, M.,Youssef, K., Higashi, A.Y., Mian, S., Ferreira, A.S., Goncalves, L.S.A., Pereira, U.P. and de Oliveira, A.G. (2021). Genomic insights into the antifungal activity and plant growth-promoting ability in *Bacillus velezensis* CMRP 4490. Frontiers in Microbiology, 11(n/a), 618415. https://doi.org/10.3389/fmicb.2020.618415
71. Thakker, J.N., Patel, S. and Dhandhukia, P.C. (2013). Induction of defense related enzymes in banana plants: Effect of live and dead pathogenic strain of *Fusarium oxysporum* f. sp. *cubense*. International Scholarly Research Notices, 2013(n/a), 1–6. https://doi.org/10.5402/2013/601303
72. Ulitsky, I. (2016). Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs. Nature Reviews. Genetics, 17(10), 601‑614. https://doi.org/10.1038/nrg.2016.85
73. Vannie,T. (2017). Dynamique de la structure des génomes et de leur biogéographie dans l'océan: analyses comparatives des données métagénomiques du projet Tara Oceans pour l'étude de la microalgue *Bathycoccus* et des communautés planctoniques globales. *Doctoral dissertation, Paris Saclay*, P:236.
74. Varshney, R. K., Singh, V. K., Kumar, A., Powell, W., & Sorrells, M. E. (2018). Can genomics deliver climate-change ready crops? *Current Opinion in Plant Biology*, *45*, 205-211. https://doi.org/10.1016/j.pbi.2018.03.007
75. Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & Van Sinderen, D. (2007). Genomics of*Actinobacteria*: Tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews*, *71*(3), 495-548. https://doi.org/10.1128/mmbr.00005-07
76. Vuolo, Francesco & Novello, Giorgia & Bona, Elisa & Gorrasi, Susanna & Gamalero, Elisa. (2022). Impact of Plant-Beneficial Bacterial Inocula on the Resident Bacteriome: Current Knowledge and Future Perspectives. Microorganisms. 10. 2462. https://doi.org/10.3390/microorganisms10122462
77. Wu, X., Wu, H., Wang, R., Wang, Z., Zhang, Y., Gu, Q., Farzand, A., Yang, X., Semenov, M., Borriss, R., Xie, Y. and Gao, X. (2021). Genomic features and molecular function of a novel stress-tolerant *Bacillus halotolerans* strain isolated from an extreme environment. Biology, 10(10), 1030. https://doi.org/10.3390/biology10101030
78. Yang, A.S. and Honig, B. 2000. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J. Mol. Biol*. 301 665–678.
79. Yassin, A. F., Rainey, F. A., Brzezinka, H., Jahnke, K., Weissbrodt, H., Budzikiewicz, H., Stackebrandt, E., & Schaal, K. P. (1995). Lentzea Gen. Nov., a new genus of the order Actinomycetales. *International Journal of Systematic Bacteriology*, *45*(2), 357-363. https://doi.org/10.1099/00207713-45-2-357
80. Zehnpfennig, J (2023). Mitochondrial genome assembly and annotation v1. https://doi.org/10.17504/protocols.io.kqdg3xeqqg25/v1.
81. Zhang, S., White, T. L., Martinez, M. C., McInroy, J. A., Kloepper, J. W., & Klassen, W. (2010). Evaluation of plant growth-promoting rhizobacteria for control of Phytophthora blight on squash under greenhouse conditions. *Biological Control*, 53(1), 129–135. https://doi.org/10.1016/j.biocontrol.2009.10.015
82. Zhi, X., Li, W., & Stackebrandt, E. (2009). An update of the structure and 16s rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *International Journal of Systematic And Evolutionary Microbiology*, *59*(3), 589-608. https://doi.org/10.1099/ijs.0.65780-0
83. Zitouni, A., Boudjella, H., Lamari, L., Badji, B., Mathieu, F., Lebrihi, A., & Sabaou, N. (2005). *Nocardiopsis* and *Saccharothrix* genera in Saharan soils in Algeria: Isolation, biological activities and partial characterization of antibiotics. *Research in Microbiology*, *156*(10), 984-993. https://doi.org/10.1016/j.resmic.2005.05.006

84. Zitouni, A., Boudjella, H., Mathieu, F., Sabaou, N., & Lebrihi, A. (2004). Mutactimycin PR, a new Anthracycline antibiotic from *Saccharothrix* sp. SA 103: I. Taxonomy, fermentation, isolation and biological activities. *The Journal of Antibiotics*, *57*(6), 367-372. https://doi.org/10.7164/antibiotics.57.367
85. Zitouni, A., Mathieu, F., Coppel, Y., Pont, F., Sabaou, N., & Lebrihi, A. (2004). Mutactimycin PR, a new Anthracycline antibiotic from *Saccharothrix* sp. SA 103: II. physico-chemical properties and structure elucidation. *The Journal of Antibiotics*, *57*(6), 373-378. https://doi.org/10.7164/antibiotics.57.373
86. Zitouni, A., Mathieu, F., Coppel, Y., Pont, F., Sabaou, N., & Lebrihi, A. (2004). Mutactimycin PR (I), a new Anthracycline antibiotic from *Saccharothrix* sp. SA 103. Part 2. physico- chemical properties and structure elucidation. *ChemInform*, *35*(49). https://doi.org/10.1002/chin.200449199

Sitography

- https://www.ncbi.nlm.nih.gov
- http://tygs.dsmz.de
- https://dfast.ddbj.nig.ac.jp/
- https://www.bv-brc.org
- https://rast.nmpdr.org
- https://usegalaxy.eu

Appendix

TYPE (STRAIN) GENOME SERVER

JOB SUMMARY

PRINT DATE: 2024-03-15 17:29:11 +0100

JOB ID: ed863e49-73ca-468e-87dd-5f39de75bd70

RESULT PAGE: https://tygs.dsmz.de/user_results/show?guid=ed863e49-73ca-468e-87dd-5f39de75bd70

Table 1: Phylogenies

Publication-ready versions of both the genome-scale GBDP tree and the 16S rRNA gene sequence tree can be customized and exported either in SVG (vector graphic) or PNG format from within the phylogeny viewers in your TYGS result page. For publications the SVG format is recommended because it is lossless, always keeps its high resolution and can also be easily converted to other popular formats such as PDF or EPS. Please follow the link provided above!

Table 2: Identification

The below list contains the result of the TYGS species identification routine. Explanation of

remarks that might occur in the below table:

remark [R1]: The TYGS type strain database is automatically updated on an almost daily basis. However, if a particular type strain genome is not available in the TYGS database, this can have several reasons which are detailed in the FAQ. You can request an extended 16S rRNA gene analysis via the 16S tree viewer found in your result page to detect not yet genome-sequenced type strains relevant for your study.

remark [R2]: > 70% dDDH value (formula d_4) and (almost) minimal dDDH values for gene-content formulae d_0 and d_6 indicate a potentially unreliable identification result and should thus be checked via the 16S rRNA gene sequence similarity. Such strong deviations can, in principle, be caused by sequence contamination.

remark [R3]: G+C content difference of > 1 % indicates a potentially unreliable identification result because within species G+C content varies no more than 1 %, if computed from genome sequences (PMID: 24505073).

Strain	Conclusion	Identification result	Remark
'Saccharothrix algeriensis DSM 44581 '	belongs to known species	*Saccharothrix algeriensis*	

Table 3: Pairwise comparisons of user genomes vs. type-strain genomes

The following table contains the pairwise dDDH values between your user genomes and the selected type- strain genomes. The dDDH values are provided along with their confidence intervals (C.I.) for the three different GBDP formulas:

- **formula d_0 (a.k.a. GGDC formula 1): length of all HSPs divided by total genome length**
- **formula d_4 (a.k.a. GGDC formula 2): sum of all identities found in HSPs divided by overall HSP length**
- **formula d_6 (a.k.a. GGDC formula 3): sum of all identities found in HSPs divided by total genome length**

Note: Formula d_4 is independent of genome length and is thus robust against the use of incomplete draft genomes. For other reasons for preferring formula d_4, see the FAQ.

Query	Subject	d_0	C.I. d_0	d_4	C.I. d_4	d_6	C.I. d_6	Diff. G+C Percent
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix algeriensis* DSM 44581	100.0	[100.0 - 100.0]	100.0	[100.0 - 100.0]	100.0	[100.0 - 100.0]	0.0
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix australiensis* DSM 43800	38.3	[34.9 - 41.8]	33.1	[30.6 - 35.6]	36.2	[33.3 - 39.3]	0.67
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix tamanrassetensis* CECT 8640	39.6	[36.2 - 43.0]	32.2	[29.8 - 34.7]	37.0	[34.1 - 40.1]	2.71
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix espanaensis* DSM 44229	32.0	[28.7 - 35.6]	29.5	[27.1 - 32.0]	30.3	[27.4 - 33.4]	1.94
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix obliqua* SC076	32.1	[28.7 - 35.7]	28.3	[25.9 - 30.8]	30.2	[27.2 - 33.3]	1.59
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix variisporea* DSM 43911	27.2	[23.9 - 30.9]	27.4	[25.0 - 29.9]	26.1	[23.2 - 29.2]	2.46
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix coeruleofusca* DSM 43679	26.7	[23.4 - 30.4]	26.4	[24.0 - 28.9]	25.5	[22.6 - 28.6]	1.56
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix syringae* NRRL B-16468	24.5	[21.2 - 28.2]	26.3	[23.9 - 28.7]	23.6	[20.8 - 26.7]	0.61
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix texasensis* DSM 44231	26.8	[23.5 - 30.5]	26.2	[23.8 - 28.7]	25.5	[22.7 - 28.6]	1.86
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix saharensis* DSM 45456	27.4	[24.0 - 31.0]	26.2	[23.9 - 28.7]	26.0	[23.1 - 29.1]	1.85
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix carnea* CGMCC 4.7097	26.6	[23.2 - 30.2]	26.1	[23.8 - 28.6]	25.3	[22.4 - 28.4]	2.64
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix xinjiangensis* JCM 12329	25.2	[21.9 - 28.9]	26.1	[23.8 - 28.6]	24.2	[21.4 - 27.3]	0.98
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix longispora* DSM 43749	27.7	[24.3 - 31.3]	26.1	[23.8 - 28.6]	26.2	[23.3 - 29.3]	0.99
'Saccharothrix algeriensis DSM 44581 .fna'	*Saccharothrix coeruleofusca* JCM 3313	26.8	[23.4 - 30.4]	26.1	[23.8 - 28.6]	25.5	[22.6 - 28.6]	1.46

Table 4: Strains in your dataset

Joint dataset of automatically determined closest type strains (if this mode was chosen), manually selected type strains (if selected accordingly) and the provided user strains, if provided (marked in yellow).

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Saccharothrix carnea CGMCC 4.7097	Liu et al. 2014	DSM 45878; NEAU yn17	*Saccharothrix carnea*	8918169	71.5	7960	Gp0197028	PRJNA370114	SAMN06295939	GCA_003014735	
Saccharothrix coeruleofusca DSM 43679	(Preobrazhenskaya and Sveshnikova 1974) Grund and Kroppenstedt 1990	NRRL B-16115; ATCC 35108; DSM 43679; JCM 3313; IFO 14520; NBRC 14520; VKM Ac-855; INA 1335	*Actinomadura coeruleofusca*; *Nocardiopsis coeruleofusca*; *Saccharothrix coeruleofusca*	7778610	72.6	7319	Gp0502536				2918321734
Saccharothrix longispora DSM 43749	(Preobrazhenskaya and Sveshnikova 1974) Grund and Kroppenstedt 1990	NRRL B-16116; ATCC 35109; JCM 3314; IFO 14522; NBRC 14522; VKM Ac-907; INA 10222	*Actinomadura longispora*; *Nocardiopsis longispora*; *Saccharothrix longispora*	8375440	73.1	7403	Gp0539492				2919027109
Saccharothrix saharensis DSM 45456	Boubetra et al. 2013	CCUG 60213; SA152	*Saccharothrix saharensis*	8924967	72.3	8018	Gp0305019	PRJNA547289	SAMN12025562	GCA_006716745	2818991416
Saccharothrix xinjiangensis JCM 12329	Hu et al. 2004	AS 4.1731; CGMCC 4.1731; DSM 44896; NBRC 101911; PYX-6	*Saccharothrix xinjiangensis*	9963838	73.2	8682		PRJDB10510	SAMD00645134		
Saccharothrix obliqua SC076	Azad et al. 2022	TBRC 14540; NBRC 115117	*Saccharothrix obliqua*	8034790	72.5	7023		PRJNA748151	SAMN20308612	GCA_019375475	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Saccharothrix espanaensis DSM 44229	Labeda and Lechevalier 1989	NRRL 15764; ATCC 51144; JCM 9112; VKM Ac-1969; LL-C19004-NS29	*Saccharothrix espanaensis*	9360652	72.2	8427	Gp0034502	PRJEB28	SAMEA2272236	GCA_000328705	2540341184
Saccharothrix australiensis DSM 43800	Labeda et al. 1984	NRRL 11239; ATCC 31497; JCM 3370; IFO 14444; NBRC 14444; VKM Ac-894; LL-BM782Ce82	*Saccharothrix australiensis*	7861373	73.5	6694	Gp0220509	PRJNA442832	SAMN08778305	GCA_003634935	2757320518
Saccharothrix tamanrassetensi s CECT 8640	Boubetra et al. 2015	DSM 45947; SA198	*Saccharothrix tamanrassetensi s*	8049770	71.4	7332	Gp0397079	PRJNA546730	SAMN12025052	GCA_014203665	
Saccharothrix coeruleofusca JCM 3313	(Preobrazhenska ya and Sveshnikova 1974) Grund and Kroppenstedt 1990	NRRL B-16115; ATCC 35108; DSM 43679; JCM 3313; IFO 14520; NBRC 14520; VKM Ac-855; INA 1335	*Actinomadura coeruleofusca*; *Nocardiopsis coeruleofusca*; *Saccharothrix coeruleofusca*	7668018	72.7	7214		PRJDB10510	SAMD00245354	GCA_014648515	
Saccharothrix syringae NRRL B-16468	(Gauze and Sveshnikova 1985) Grund and Kroppenstedt 1990 emend. Nouioui et al. 2018	ATCC 51364; DSM 43886; JCM 6844; IFO 14523; NBRC 14523; VKM Ac-1858; INA 2240	*Nocardiopsis syringae*; *Saccharothrix syringae*	10882771	73.5	9372	Gp0188076	PRJNA224116	SAMN02645315	GCF_000716755	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Saccharothrix variisporea DSM 43911	(ex Tomita et al. 1977) Kim et al. 2011	NRRL B-16296; ATCC 31203; JCM 3273; IFO 14104; NBRC 14104	*Dactylosporangium variesporum*; *Saccharothrix variisporea*	9408895	71.7	8413	Gp0220372	PRJNA460962	SAMN09070812	GCA_003634995	2778260942
Saccharothrix texasensis DSM 44231	Labeda and Lyons 1989	NRRL B-16134; ATCC 51593; JCM 9113; IFO 14971; NBRC 14971; VKM Ac-1968; LL-37U-77	*Saccharothrix texasensis*	9178199	72.3	8163	Gp0220505	PRJNA500558	SAMN10363104	GCA_003752005	2784132178
Saccharothrix algeriensis DSM 44581	Zitouni et al. 2004	NRRL B-24137; JCM 13242; NBRC 101915; SA 233	*Saccharothrix algeriensis*	6878582	74.1	6172	Gp0502543				2893654320
Saccharothrix algeriensis DSM 44581 .fna				6878582	74.1	6112					

Methods, Results and References

The genome sequence data were uploaded to the Type (Strain) Genome Server (TYGS), a free bioinformatics platform available under https://tygs.dsmz.de, for a whole genome-based taxonomic analysis [1]. The analysis also made use of recently introduced methodological updates and features [2]. Information on nomenclature, synonymy and associated taxonomic literature was provided by TYGS's sister database, the List of Prokaryotic names with Standing in Nomenclature (LPSN, available at https://lpsn.dsmz.de) [2]. The results were provided by the TYGS on 2024-03-14. The TYGS analysis was subdivided into the following steps:

Determination of closely related type strains

Determination of closest type strain genomes was done in two complementary ways: First, all user genomes were compared against all type strain genomes available in the TYGS database via the MASH algorithm, a fast approximation of intergenomic relatedness [3], and, the ten type strains with the smallest MASH distances chosen per user genome. Second, an additional set of ten closely related type strains was determined via the 16S rDNA gene sequences. These were extracted from the user genomes using RNAmmer [4] and each sequence was subsequently BLASTed [5] against the 16S rDNA gene sequence of each of the currently 20757 type strains available in the TYGS database. This was used as a proxy to find the best 50 matching type strains (according to the bitscore) for each user genome and to subsequently calculate precise distances using the Genome BLAST Distance Phylogeny approach (GBDP) under the algorithm 'coverage' and distance formula d_5 [6]. These distances were finally used to determine the 10 closest type strain genomes for each of the user genomes.

Pairwise comparison of genome sequences

For the phylogenomic inference, all pairwise comparisons among the set of genomes were conducted using GBDP and accurate intergenomic distances inferred under the algorithm 'trimming' and distance formula d_5 [6]. 100 distance replicates were calculated each. Digital DDH values and confidence intervals were calculated using the recommended settings of the GGDC 4.0 [2,6].

Phylogenetic inference

The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.6.1 including SPR postprocessing [7]. Branch support was inferred from 100 pseudo- bootstrap replicates each. The trees were rooted at the midpoint [8] and visualized with PhyD3 [9].

Type-based species and subspecies clustering

The type-based species clustering using a 70% dDDH radius around each of the 14 type strains was done as previously described [1]. The resulting groups are shown in Table 1 and 4. Subspecies clustering was done using a 79% dDDH threshold as previously introduced [10].

Results

Type-based species and subspecies clustering

The resulting species and subspecies clusters are listed in Table 4, whereas the taxonomic identification of the query strains is found in Table 1. Briefly, the clustering yielded 13 species clusters and the provided query strains were assigned to 1 of these. Moreover, user strains were located in 1 of 13 subspecies clusters.

Figure caption SSU tree

Figure 1. Tree inferred with FastME 2.1.6.1 [7] from GBDP distances calculated from 16S rDNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula d_5. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 77.7 %. The tree was rooted at the midpoint [8].

Figure caption genome tree

Figure 2. Tree inferred with FastME 2.1.6.1 [7] from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5. The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 91.5 %. The tree was rooted at the midpoint [8].

References

[1] Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. 2019;10: 2182. DOI: 10.1038/s41467-019-10210-3

[2] Meier-Kolthoff JP, Sardà Carbasse J, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acid Res. 2022;50: D801–D807. DOI: 10.1093/nar/gkab902

[3] Ondov BD, Treangen TJ, Melsted P, et al. Mash: Fast genome and metagenome distance estimation using MinHash. Genome Biol 2016;17: 1–14. DOI: 10.1186/s13059-016-0997-x

[4] Lagesen K, Hallin P. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. Oxford Univ Press; 2007;35: 3100–3108. DOI: 10.1093/nar/gkm160

[5] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. DOI: 10.1186/1471-2105-10-421

[6] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14: 60. DOI: 10.1186/1471- 2105-14-60

[7] Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol. 2015;32: 2798–2800. DOI: 10.1093/molbev/msv150

[8] Farris JS. Estimating phylogenetic trees from distance matrices. Am Nat. 1972;106: 645–667.

[9] Kreft L, Botzki A, Coppens F, Vandepoele K, Van Bel M. PhyD3: A phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. Bioinformatics. 2017;33: 2946–2947. DOI: 10.1093/bioinformatics/btx324

[10] Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. Complete genome sequence of DSM 30083^{T}, the type strain (U5/41^{T}) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci. 2014;9: 2. DOI: 10.1186/1944-3277-9-2