Diversity of Root Associated Bacteria of Orchids and Phylotaxogenomics of a Few Bacteria

Thesis submitted for the degree of
Doctor of Philosophy

By

Anusha Rai
(Reg. No. 16LPPH13)

Research Supervisor
Prof. Ch. Venkata Ramana

Department of Plant Sciences
School of Life Sciences
University of Hyderabad
Hyderabad – 500 046
INDIA

April 2022

UNIVERSITY OF HYDERABAD

(A Central University established in 1974 by an Act of the Parliament) HYDERABAD 500046

प्रतिष्ठित संस्थान
INSTITUTION OF EMINENCE

CERTIFICATE

This is to certify that **Ms. Anusha Rai** has carried out the research work embodied in the present thesis under the supervision and guidance of **Prof. Ch. Venkata Ramana** for the full period prescribed under the Ph.D. ordinances of this University. We recommend her thesis entitled **"Diversity of Root Associated Bacteria of Orchids and Phylotaxogenomics of a Few Bacteria"** for submission for the degree of Doctor of Philosophy to the University.

HEAD 13/04/2022
Dept. Plant Sciences

अध्यक्ष / HEAD
वनस्पति विज्ञान विभाग / Dept. of Plant Sciences
जैविक विज्ञान संकाय / School of Life Sciences
हैदराबाद विश्वविद्यालय / University of Hyderabad
हैदराबाद / Hyderabad-500 046, भारत / INDIA

DEAN
School of Life Sciences

DEAN
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

12/4/2022

SUPERVISOR
Prof. Ch.V. Ramana

Prof. Ch. Venkata Ramana Ph.D
Department of Plant Sciences School of Life Sciences
University of Hyderabad, P.O. Central University,
Hyderabad - 500 046, India

UNIVERSITY OF HYDERABAD
(A Central University established in 1974 by an Act of the Parliament) HYDERABAD 500046

INSTITUTION OF EMINENCE

DECLARATION

I, **Anusha Rai**, hereby declare that this thesis entitled **"Diversity of Root Associated Bacteria of Orchids and Phylotaxogenomics of a Few Bacteria"** submitted by me under the guidance and supervision of **Prof. Ch. Venkata Ramana** is an original and independent research work. I, hereby declare that this work is original and has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Anusha Rai
Anusha Rai
(Ph.D. student)

Ch Ramana 12/4/2022
Supervisor
Prof. Ch. V. Ramana
Prof. Ch. Venkata Ramana Ph.D
Department of Plant Sciences School of Life Sciences
University of Hyderabad, P.O. Central University,
Hyderabad - 500 046, India

UNIVERSITY OF HYDERABAD
(A Central University established in 1974 by an Act of the Parliament) HYDERABAD 500046

प्रतिष्ठित संस्थान
INSTITUTION OF EMINENCE

CERTIFICATE

This is to certify that the thesis entitled "**Diversity of Root Associated Bacteria of Orchids and Phylotaxogenomics of a Few Bacteria**" submitted by **Ms. Anusha Rai**, bearing registration number **16LPPH13**, in partial fulfilment of the requirements for the award of Doctoral of Philosophy in Department of Plant Sciences, School of Life Sciences, University of Hyderabad, is a bonafide work carried out by her under my guidance and supervision. This thesis is free from plagiarism and has not been submitted in any part or in full to this or any other University or Institute for the award of any degree or diploma.

Parts of the thesis have been:

A. Authored in the following publications:

1. Rai A, Indu, Smita N, Deepshikha G, Gaurav K, Dhanesh K et al (2019) Emerging Concepts in Bacterial Taxonomy. In: Satyanarayana T, Johri B, Das S (eds.) Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications. Springer, Singapore.
2. Rai A, Smita N, Suresh G, Shabbir A, Deepshikha G, Sasikala Ch, Ramana ChV (2019) *Paracoccus aeridis* sp. nov., an indole producing bacterium isolated from the rhizosphere of an orchid, Vanda sp. Int J Syst Evol Microbiol 70:1720-1728.
3. Rai A, Smita N, Shabbir A, Jagadeeshwari U, Keertana T, Sasikala Ch, Ramana ChV (2021) *Mesobacillus aurantius* sp. nov., isolated from an orange-colored pond near a solar saltern. Arch Microbiol 203:1499-1507.

4. Rai A, Jagadeshwari U, Gupta D, Smita N, Sasikala Ch, Ramana ChV (2021) Phylotaxogenomics for the reappraisal of the genus *Roseomonas* with the creation of six new genera. Front Microbiol 12:1787.

B. Presented at the following conferences:

1. Presented a poster at the "8th International conference on Photosynthesis and Hydrogen Energy Research for Sustainability" (2017), UoH, Hyderabad
2. Presented poster at the "59th Annual Conference of Association of Microbiologists of India & International Symposium on host-pathogen interaction" (2018), UoH, Hyderabad
3. Presented an oral presentation at "National Conference on Frontiers in Plant Biology" (2019), UoH, Hyderabad

Further, the student has passed the following courses towards the fulfilment of the coursework requirement for Ph.D. degree.

S. No.	Course	Name	Credits	Pass/Fail
1	PL-801	Analytical techniques	4	Pass
2	PL-802	Research ethics, data analysis and biostatistics	3	Pass
3	PL-803	Lab work & seminar	5	Pass

HEAD 13/04/2022
Dept. Plant Sciences

अध्यक्ष / HEAD
वनस्पति विज्ञान विभाग / Dept. of Plant Sciences
जैविक विज्ञान संकाय / School of Life Sciences
हैदराबाद विश्वविद्यालय University of Hyderabad
हैदराबाद / Hyderabad-500 046, भारत / INDIA

DEAN
School of Life Sciences

DEAN
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

12/4/2022
SUPERVISOR
Prof. Ch. V. Ramana

Prof. Ch. Venkata Ramana Ph.D
Department of Plant Sciences School of Life Sciences
University of Hyderabad, P.O. Central University,
Hyderabad - 500 046, India

ACKNOWLEDGEMENTS

It is said that there are the five types of people you should always surround yourself with; they are the inspired, the open-minded, the motivated, the grateful and the passionate. It can invariably change the course of your life.

***Prof. Ch. Venkata Ramana**, is someone who continuously inspires and motivates his students. His passion for research is immeasurable. We, as students, are always in awe by seeing the effort and positive energy that he puts into his work. Therefore, I would like to extend my heartfelt gratitude to my supervisor for his persistent encouragement and support, without which, this journey would not have been possible. Working under his guidance has given me a sense of scientific freedom for understanding and helped broaden my scientific purview. I am grateful for his insightful guidance in scientific aspects like writing, sampling, communication and presentation skills. Above all, I feel grateful for his unwavering support at difficult times during my doctoral research. Importantly, his positive perspective towards life, failure and success has made a major impact on my daily life and his philosophy on life will remain etched in my mind forever. I am fortunate to have been supervised by him.*

*I sincerely thank my doctoral committee members, **Prof. S. Rajagopal** and **Prof. K. Gopinath,** for their judicious comments on my work. I thank them for permitting me to use their lab facilities. I also thank **Prof. Apparao**, **Prof. Reddy**, **Prof. Padmaja** and **Prof. Raghavendra** for extending their lab facilities for their use.*

*I am grateful to the former Deans, **Prof. P. Reddanna** and **Prof. Ramaiah**, the current Dean, **Prof. S. Dayananda**, School of Life Sciences; and the former Head, **Prof. Ch. VenkataRamana** and present Head, **Prof. Padmaja**, Dept. of Plant Sciences, for rendering their support in all possible ways.*

*I extend my heartfelt gratitude to **Prof. Ch. Sasikala**, CEN, IST, JNTU for letting me use the lab facilities in her lab. I equally thank her for sharing her research experiences and expertise in every scientific aspect possible. Her insightful comments on my work have aided me in the successful completion of my PhD work.*

I am extremely grateful to my teachers from Saptashri Gyanpeeth, Loreto Convent, SSIHL and UoH. They nurtured my scientific temperament and undoubtedly helped me in inculcating moral values and ethics in life.

*I thank **Dr. Jeewan Singh** (Botanical Survey of India) and **Prof. Bernhard Schink** (University of Konstanz, Germany) for identifying orchids and correcting the nomenclature protologues, respectively.*

I thank the Department of Science (DST) for providing "Innovation in Science Pursuit for Inspired Research" (INSPIRE) for the research fellowship. The infrastructure support provided by UGC-SAP, DBT-CREBB,DBT-Builder and DST-FIST to the Dept. of Plant Sciences, is highly acknowledged. Financial support from CSIR, DBT-TATA and IOE projects are acknowledged.

I thank my former and senior lab members (Dr. Ganesh M, Dr. Laksmi Prasuna, Dr. Mujahid M, Dr. Subhash Y, Dr. Ramprasad Dr. Tushar L, Dr. Suresh G, Dr. Ashif Ali, Dr. Dhanesh K, Dr. Shabbir A, Dr. Indu, Dr. Deepshikha G, Dr. Gaurav K) for their genuine guidance and scientific discussions.

It was equally a great pleasure working with my current lab-mates. They are inspiring, open-minded, innovative and motivated; my learning experience was enjoyable and memorable with them. I thank them (Smita, Sreya, Ipsita, Kim, Krishaniah, Vignesh, Ria, Amulya, Mahima, Sushmita, Ramadu) for maintaining a healthy and learning environment. Sravya and Raima are also acknowledged.

I am grateful to lab mates and seniors from JNTU-H (Jagadeeshwari U, Sangeetha Mam, Dr. Shivani) for assisting in several projects and related works.

It is rightly said that "A sweet friendship refreshes the soul" as my life stands testimony to the said proverb. I have had a great opportunity of having friends disguised in the form of seniors/juniors. They were Indu, Anusree, Moubani, Rutuparna, Smita. Whatever problems were there, I always looked upon them for refuge and still I do. My other important companions were Divya, Dr. Norbu, Pragyansini, Ranjeeta, Mounica with whom I could discuss my problems and share moments of unfeigned happiness. My UoH friends (*Prodosh, Deepak, Meetali, Sourav*) *and from schools (SGP, LC) are also acknowledged. I am truly grateful for your friendship as it gave me strength to face the world at some point of my life. I equally thank the other research scholars of the School of Life Sciences for their cooperation and for helping me throughout the research periods.*

The help and cooperation of the non-teaching staff in SLS and administrative building (Madhavi Mam) is highly acknowledged.

My journey would not have been meaningful without my parents and grandparents blessings. They have given me the wings to fly by always supporting my decisions and believing me. Without their love and sacrifice I would have been nowhere. I am grateful to my aunties, uncles, sisters and brothers for always cheering and standing up for me.

I thank God, the Almighty, for helping me cross all the hurdles of life. I am always grateful for His/Her divine presence throughout.

Lastly, I consider myself fortunate to have been surrounded by all five types of people…..

Anusha Rai

“I was taught that the way of the progress was neither swift nor easy”

- Madame Marie Curie

Dedicated to my beloved parents and my supervisor

CONTENTS

LIST OF ABBREVIATIONS

Abbreviations	**Expansion**
AA	Auxiliary activity
AAI	Average amino acid identity
ACN	Acetonitrile
AFLP	Amplified fragment length polymorphism
AIC	Akaike information criterion
AL	Aminolipid
alkJ	Alcohol dehydrogenase gene
AMD	Acid mine drainage
ANI	Average nucleotide identity
antiSMASH	Antibiotics & Secondary Metabolite Analysis Shell
ATCC	American Type Culture Collection
atpD	ATP synthase subunit beta gene
BP	Beibl and Pfennig media
BCC	BIOTEC Culture Collection
BGC	Biosynthetic gene clusters
BGI	Beijing Genomics Institute
BLASTn	Nucleotide Basic Local Alignment Search Tool
BLASTp	Protein Basic Local Alignment Search Tool
BPGA	Bacterial Pan Genome Analysis tool
BRC1	Bacterial Rice Cluster 1
cas	CRISPER-associated protein
CBM	Carbohydrate-binding module
CCUG	Culture Collection University of Göteborg
CDS	Coding DNA sequences
CE	Carbohydrate esterase
CECT	Colección Española de Cultivos Tipo
CGE	Center for Genomic Epidemiology
CIP	Institute Pasteur Collection
C/N	Carbon/Nitrogen
COG	Clusters of Orthologous Groups of Proteins
CPR	Candidate Phyla Radiation
CRISPR	Clustered regularly interspaced short palindromic repeats
CSI	Conserved signature indel
CSP	Conserved signature protein
cyp	Cytochrome P450 ω-hydroxyl gene
*d*DDH	Digital DNA-DNA Hybridisation
DDJB	DNA databank of Japan
dnaK	Chaperon protein DnaK
DPG	Diphosphatidylglycerol
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH
EDTA	Ethylenediaminetetraacetic acid
ENA	European Nucleotide Archive
EST	Expressed sequence tags
FAME	Fatty acid methyl ester
FESEM	Field scanning electron microscope
FTP	File transfer protocol
gL^{-1}	Gram per litre
G+C	Guanine+cytosine
GDMCC	Guangdong Microbial Culture Collection Center
GGDH	Genome to genome distance hybridization
GH	Glycosyl hydrolase
GL	Glycolipid
GO	Gene ontology
GPS	Global positioning system
gryB	DNA gyrase subunit B
GT	Glycosyl transferase
h	Hour

HMM	Hidden Markov model
HPLC	High pressure liquid chromatography
HSP	High scoring segment pair
IAA	Indole-3-acetic acid
IBBNT	International Bulletin of Bacteriological Nomenclature and Taxonomy
ICSB	International Committee on Systematics of Prokaryotes
IJSEM	International Journal of Systematic and Evolutionary Microbiology
IPP	Isopentenyl pyrophosphate
IUCN	International Union for Conservation of Nature
JCM	Japan Collection of Microorganisms-RIKEN BioResource Center
KACC	Korean Agricultural Culture Collection
KEGG	Kyoto Encyclopedia of Gene and Genomes
KEMP	Korea Environmental Microorganisms Bank
L	Litre
LCB	Local collinear blocks
LC-MS	Liquid chromatography liquid chromatography
LMG	Laboratorium voor Microbiologie
MAFFT	Multiple Alignment using Fast Fourier Transform
Mbp	Mega base pair
MCCC	Marine Culture Collection of China
MEGA	Molecular evolutionary genetics analysis
MEP	2-C-methyl-d-erythritol 4-phosphate/1-deoxy-d-xylulose 5-phosphate
µg	Microgram
μg^{-l}	Microgram per litre
mg	Milligram
MHB	Mycorrhizal helper bacteria
ML	Maximum likelihood
MLSA	Multilocus sequence analysis
µl	Microlitre
ml	Milliliter
µm	Micrometer
mM	Millimolar
MUM	Maximally unique matches
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NBRC	NITE Biological Resource Center
NCBI	National center for biotechnology information
NDMS	Non-metric multi-dimensional scaling
nm	Nanometer
NNI	Nearest-neighbour-interchange
NRPS	Non-ribosomal peptide synthetases
OGRI	Overall genomic relatedness indices
OMF	Orchid Mycorrhizal Fungi
OMM	Oligotrophic mineral media
OPA	o-Phthaldialdehyde
OrthoANI	Orthologous average nucleotide identity
OTU	Operational taxonomic unit
PATRIC	PathoSystems Resource Integration Center
PCR	Polymerase chain reaction
PDA	Photodiode array detector
PFGE	Pulse field gel electrophoresis
PGAP-X	Pan-genome analysis pipeline
PGAweb	Pan genomes analyses webserver
PGPR	Plant growth promoting rhizobacteria
PhoH	Phosphate starvation-induced protein
PL	Phospholipid
POCP	Percentage of conserved protein
psi	Pound per square inch
Q	Ubiquinone
RAB	Root associated bacteria
RAPD	Randomly amplified polymorphic DNA

RAST	Rapid Annotations using Sub-Systems Technology
RDP	Ribosomal Database Project
recA	DNA recombination and repair protein
RFLP	Restriction fragment length polymorphism
RiPPs	Ribosomally generated and post-translationally modified peptides
rpoB	DNA-directed RNA polymerase subunit beta
rpoD	RNA polymerase sigma factor
R_f	Retention Factor
RT-PCR	Real time-polymerase chain reaction
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscopy
sqhC	Squalene cyclase gene
SR1	Sulfur river 1
SRA	Sequence read archive
SSC	Sodium saline citrate
TAE	Tris-base, acetic acid and EDTA
TE	Tris-EDTA
TLC	Thin layer chromatography
TPKS	Type 1 polyketide synthases
TrkA	Potassium uptake protein
tryB	Tryptophan synthase subunit beta
TYGS	Type Strain Genome server
UBCG	Up-to-date bacterial core gene
UPLC	Ultra-performance liquid chromatography
v/v	Volume per volume
w/v	Weight per volume
YpjP	Uncharacterized protein
ZM	Zobell Marine
2D	Two dimensional

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1. INTRODUCTION

Microorganisms are an indispensable part of the environment and are essential for the general functioning of our planet. They survive in remarkably diverse environmental conditions and often exist in an interdependent network of communities. Similarly, organisms are adapted to various environmental conditions because of their specialized niches (Zeyad et al. 2021). Understanding the structural and functional dynamics by which they exist could lead us to a generation of valuable resources for potential metabolites with various biotechnological applications (Mishra et al. 2021). This vast microbial diversity is the primary source of undiscovered biological diversity. To better understanding of this vast, yet primarily unknown diversity, microbial taxonomy plays a vital role in unearthing its unrealized enzymatic, genetic and industrial potential.

1.1. Isolation sources

1.1.1. Orchids

Orchids are one of the most prominet families of angiosperms consisting of 25,000 - 30,000 species in the world (Christenhusz and Byng 2016). They come under the phylum of Magnoliophyta, class Liliopsida, order Aspargales and family Orchidaceae (Sut et al. 2017). They are found diversely in South America, Mexico and Central America, West and Central Africa and other Asian countries. These plants have several economical, medical and horticultural properties (Gross et al. 2016). They are varied, exotic and prolific growing plant species (Zhang et al. 2018). They grow chiefly in nutrient-deficient habitats. Most varieties of orchid species are epiphytic. According to the IUCN (International Union for Conservation of Nature), orchids are categorized under the vulnerable section as they are very responsive to rapid climatic and environmental changes (Favre-Godal et al. 2020). Orchids inhabit a wide range of habitats, including epiphytes on trees, terrestrial growth on the ground or rocks like lithophytes (Phillips et al. 2020).

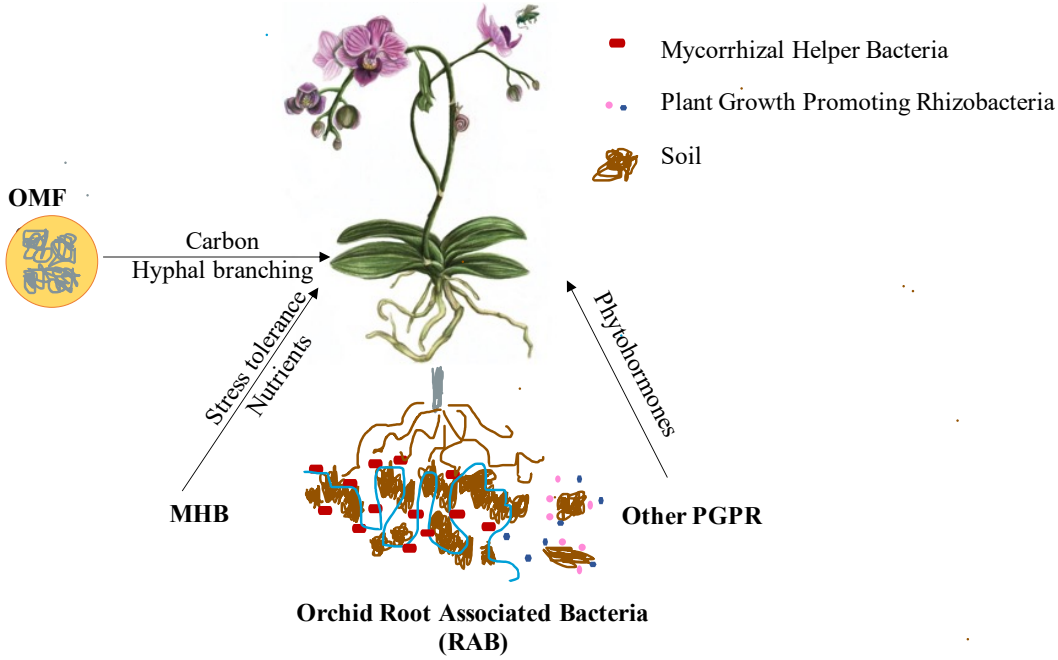


Fig. 1. The dynamics of the tripartite relationship between orchids, orchid mycorrhizal fungi (OMF) and mycorrhizal helper bacteria (MHB) (Source information adapted from Kaur and Sharma 2021)

Most orchid seeds in nature can germinate only if infected by a fungus (Favre-Godal et al. 2020). Their life cycle involves an inevitable tripartite interaction between plant systems, mycorrhizal fungi and plant growth-promoting rhizobacteria (PGPR) (Li et al. 2021; Fig. 1). In general, plants are always facing biotic and abiotic stress in the environment, which significantly influences the growth and development of plant species (Martin et al. 2017). In this context, plants scrupulously mobilize root-associated bacteria (RABs) in response to the production of root secretion, stress and host genotype. RABs are those bacteria that populate the rhizosphere/endorhizal area in a host plant (Kaur and Sharma 2021). In orchid plants, RABs are often observed as having roles of PGPR and mycorrhizal helper bacteria (MHB), components of hologenomes and holobionts (Kaur and Sharma 2021). Sections of RABs promote development (as PGPRs), prevention of pathogens and stress alleviation (Kaur and Sharma 2021; Mohanty et al. 2021). RABs are noted for enhancing plant growth by amassing nutrients (Herrera et al. 2021), nitrogen

fixation (Afzal et al. 2019), hormonal allocation (Tsavkelova et al. 22016), protection from pathogens (Herrera et al. 2021) and mitigation of abiotic stresses (Shah et al. 2021). Most studies retrieved in the literature focus on the fungal aspect, whereas the studies concerning the bacterial diversity associated in terms of composition and functions are less explored (Kaur and Sharma 2021). Metagenomic studies have shown that genera like *Pseudomonas, Burkholderia, Sphingomonas, Bacillus, Comamonas* were common in different orchid varieties (Alibrandi et al. 2020; Li et al. 2017).

1.1.2. Salt pan

Salt pans are part of hypersaline environments, kernels for remarkable ecological habitats with varying growth conditions (salinity, pH and temperature; Oren 2002). Salt pans are broad, expansive areas covered with salt and minerals; they are formed with exceeding rates of evaporation of saline water bodies. These hypersaline areas support extremophiles from the domains of Archaea, Bacteria and Eukarya with distinctive biotic functions owing to the ecological and physiological adaptations (Das et al. 2019). Microorganisms from these habitats can be used as exceptional biological models for application-oriented biotechnological and industrial research (Karray et al. 2018). The isolation and scrutiny of the bacterial members from these extreme hypersaline environments are of prime importance because of their metabolic and physiological re-modelling mechanisms; which makes them an excellent prospect for many industrial applications (Mokashe et al. 2018; Zhu et al. 2020). Over the last few years, researchers have successfully isolated different halophilic species for biotechnological applications. *Salinibacter ruber*, *Halomonas elongata*, *Paenibacillus tarimensis* and *Nesterenkonia* sp. are known for their carotenoid production, ectoine, cellulases and amylases, respectively (Gunde-Cimerman et al. 2018; Zhu et al. 2020).

1.1.3. Beach sand

Marine habitats are composed of a broad spectrum of environments, each supporting a complex assemblage of potentially unique macroorganisms and microorganisms. Specific habitats like open ocean beaches and intertidal areas are essential as they form the interface between terrestrial/freshwater and marine systems (Patel 2021). Around 40 % of the global coastline consists of gravel and sand beaches (Bird 2000). As the beaches are subjected to various tidal and fluvial processes, it continuously shapes their geochemistry, mineralogy and biological processes. Molecular studies of bacterial diversity have shown that they occupy an important ecological niche in the marine environment (Mineta and Gojobori 2016). These organisms are pervasive, and their ecological niches can be studied as an archetype for detecting biosignatures and reconstructing evolutionary occurrences (Walker et al. 2005). These bacteria have remarkable ways of survival in severe conditions like solar radiance, nutrient deficiency and hyper-arid conditions (Mezzasoma et al. 2021). They are critical in biogeochemical cycle regulation, bioremediation processes and stromatolite formations (Reid et al. 2000; Wierzchos et al. 2015). Therefore, these marine habitats are currently being explored mainly for bacteria for various biotechnological applications (Guan et al. 2020). Novel species of bacteria have been described: *Ciceribacter thioxidans* (nitrate reduction) (Deng et al. 2017); *Rhizobium yantingense* (mineral weathering) (Chen et al. 2015), and *Barrientosiimonas endolithica* (Parag et al. 2015).

1.1.4. Acid mine drainage

Acid mine drainage (AMD) is water that has been contaminated by mining activity and is usually related to coal mining. It's a principal source of water pollution in locations where mining has already occurred (Dhir 2018). Mining waste is highly acidic and contains a lot of sulphate and heavy metals (Guo et al. 2022). Despite its extreme nature, a various

of eukaryotic and prokaryotic life forms live there and are classified as extremophiles. To survive and grow in such harsh environments, these extremophilic bacteria can execute a range of metabolic processes, including pH adjustment, Fe/S oxidation, C/N fixation, and metal bio-remediation (Chen et al. 2016). The microbial method employs these extremophilic bacteria to treat the AMD wastewater (Guo et al. 2022). Therefore, novel microorganisms from such unique habitats would play a pivotal role in bio-remediation and industrial applications.

1.1.5. Hot springs

Hot springs are geothermally heated groundwater and are hosts to diverse and prolific microbial communities. There are 400 geothermal springs identified in India, spread across 7 geothermal sites, out of which only 28 have been scientifically explored (Poddar and Das 2018). They are also powerful chemical reactors and mineral deposition sites (Sharma et al. 2009). Based on these habitats, identification of novel organisms for their bio-prospective studies (antimicrobials, enzymes, bio-remediation) has been carried out (Mahajan and Balachandran 2017; Panda et al. 2013). Even though hot springs are found widely in India, have inherent scientific interest, microbial investigation of these springs has received little attention.

1.2. Metagenome studies

The advancement in next-generation sequencing (NGS) technologies has transformed the entire scenario in the scientific world. NGS has provided novel insights into the functional and structural nexus of the microbiome (Ning and Tong 2019). With constant effort and upgradation of sequencing platforms, the role of these microbiomes has become lucid in plants, animals, humans and environments (Liu et al. 2021). This kind of state-of-art has led to the establishment of “metagenomics” and was first coined by Handelsman in 1998 (to denote a detection method for biosynthetic gene clusters). In

metagenomics, the genetic material present in an environmental sample is directly analyzed even without the prior cultivation of the cultures (Oulas et al. 2015). The metagenomic studies have allowed the bioprospection and detection of potential enzymes and secondary metabolites for biotechnological and industrial purposes (Eze et al. 2021; Kang et al. 2021). NGS methods for microbiome analysis are amplicon/marker gene-based and shotgun metagenomics.

1.2.1. Metagenomic approaches

1.2.1.1. Amplicon based

The introduction of marker-based molecular techniques has resulted in the augmentation of studies related on microbial distribution in the environment. For amplicon-based, DNA is extracted, and a specific region of marker gene like the rRNA gene is amplified and sequenced (Oulas et al. 2015). 16S rRNA (bacteria), 18S rRNA (eukaryotes) and internal transcribed regions (fungi) are often used as it gives a high resolution for species differentiation (Woese and Fox 1977). The identification of these genes is based on the gene identity to the references databases like SILVA (Carlton et al. 2002), Greengenes (DeSantis et al. 2006) and RDP (Cole et al. 2005). It is cost-effective and can screen a large consortium of bacteria and eukaryotes in any sample. 16S rRNA amplicon-based analysis includes the following processes quality checks, de-noising, detection of chimeric sequences, clustering into 'Operational Taxonomic Units (OTUs) based on sequence identity and final taxonomic assignment to these OTUs (Callahan et al. 2016; Caporaso et al. 2010; Schloss et al. 2009). This approach is largely applied for the profiling studies of microbial communities and the detection of rare taxa (Stefanini and Cavalieri 2018). Mothur is one of the most notably used software for microbial community analysis (Schloss et al. 2009).

1.2.1.2. Shotgun based

Shotgun metagenomics allows the untargeted sequencing of all the genomes of microorganisms present in a sample (uncultured as well) (Quince et al. 2017). It has multiple functions of profiling taxonomic populations, functional annotation of microbial communities, and generating whole genome sequences (Abraham et al. 2020). The shotgun studies involve the following processes (a) quality control and elimination of host contaminated genes (Bolger et al. 2014); (b) assembling the sequences for profiling functional, taxonomic and genomic attributes (Huson et al. 2016; Ye et al. 2019); (c) statistical analysis for data interpretation; (d) validation. For the assembly process, contigs are arranged either by reference-based (using the pre-existing genome as a map) or *de novo*-based assembly (no references/solely by computational method) (Oulas et al. 2015). Binning is an integral part of the assembly wherein contigs are assigned to specific groups either by comparing the k-mer sequences (genomic signatures) or by homology-based (alignment algorithms) (Sharpton 2014). Shotgun metagenomics has been successfully applied to study resistome hotspots (Chakraborty et al. 2020), rhizosphere microbial communities (Enebe and Babalola 2021), the gut microbiome (Chen et al. 2018), infectious studies (Gu et al. 2019), bioremediation and drug discovery (Datta et al. 2020).

1.2.2. Sequencing platforms

Sanger sequencing was introduced in 1977 by Frederick Sanger and has been used for more than three decades (Meera et al. 2019). Due to increased innovation in science, sequencing technologies have drastically improved over the years. Various sequencing methods have emerged which have allowed nimble, comprehensive and cost-effective genomic information, of a given sample. Different sequencing technologies and other information are given in Table 1. Comparative studies using different platforms have revealed that results were comparable amongst all the platforms with slight differences in

coverage depth and diversity/abundances (Allali et al. 2017). Therefore, selecting the NGS platform and analysis tools/software becomes critical for impactful research.

Platform	Sequencing Principle	Model	Read length (bp)	Output data	Company
		FIRST GENERATION (1977)			
Sanger sequencer	Fluorescent dideoxy terminator	3730 xl	≈ 900	<3 Mb	Applied Biosystems
		SECOND GENERATION (2004-2009)			
454	Pyrosequencing light emission	GS FLX	≈ 650	700Mb	Roche
Ion Torrent	Hydrogen ion semiconductor	Ion 316 chip	100	100Mb/chip	Life Technologies
SOLiD™	Sequence by ligation	5500 xl	75	30 Gb	Life Technologies
Illumina	Fluorescent single-molecule	HiSeq 2500/MiSeq	50-150	55 Gb	Illumina, Inc.
		THIRD GENERATION (2010-till date)			
PacBio	SMRT™ technology	PacBio RS	>1000	500Mb	Pacific Bioscience
Oxford Nanopore	Strand sequencing	GridION™/MinION™	500-2300	5-10Gb	Oxford Nanopore Technologies

Table 1. Different characteristics of first, second and third sequencing technologies

1.3. Bacterial taxonomy

Microbiology is an essential branch of biology divided into various subjects and segregated into sections like taxonomy, agriculture, medicine, environment, virology, bacteriology, based on its utility. Like any field of science, the comprehensive study calls for a basic understanding of the matter for analysis. The term “basic” implied here is alluded to implementing “taxonomy” in bacteriology. At the outset, bacterial taxonomy aides as a platform for identifying and affiliating a species to a particular taxon along with correlation to its phylogenetic properties. It concerns the identification and classification of the novel taxa (if new) and their nomenclature, which is done in accordance with the ordinances given in the Prokaryotic Code 1990 (revised in 2008) by the International Code of Nomenclature of Prokaryotes (Parker et al. 2019). It is an essential field as it imparts a scientific foundation for the salient discernment of the bacterial species. Taxonomy and systematics have often been applied reciprocally; however, there lies a variance between the two. Taxonomy is one aspect of the broader field of systematics. Taxonomy is built on

theory but governed by practical classification, whereas systematics is an evolutionary study of a diverse set of related organisms (Hamilton et al. 2021). In the current scenario, the novelty of the bacterial taxa is resolved by the phylogenetic analysis of the 16S rRNA gene inconsistency with the phenotypic, chemotaxonomic and genomic studies of the species. Thus, leading to the discovery and rearrangement of taxa with higher biotechnological implications, which has given microbiology a new dimension.

1.3.1. Advancement in bacterial taxonomy

The early 1670s could be claimed as the time of origination of the science of microbiology. A segment of microbiology, i.e., bacterial taxonomy, became highly necessary to accommodate and arrange the new bacterial information, resulting in an accumulation of knowledge. Further, brisk and precise techniques were developed for the taxonomic studies. Around 90 % of the species published in 'Bergey's Manual of Determinative Bacteriology' were eventually reduced, whereas those appended on the approved lists of bacterial names were considered valid (Garrity 2016; Skerman et al. 1980). To date, different approaches have been included for the advancement of the bacterial taxonomy. Therefore, these have given rise to the emergence of bacterial taxonomy. The evolution of bacterial taxonomy can be graded into the following phases:

1.3.1.1. Phase I (Year 1600–1900)

Taxonomic descriptions during this time were based on morphological and biological observation. During this period (early 1670), Antonie *van* Leeuwenhoek and Robert Hooke were the first to observe 'animalcules' like microscopic forms by the single-lens microscope (Gest 2004). Koch invented the agar plate method to cultivate bacterial cultures, which made the study of bacterial species easy (Blevins and Bronze 2010). Ferdinand Cohn, in 1872, proposed the hierarchal designation of bacterial entities into genus and species (Drews 2000). Different methods like Gram staining and acid-fast

staining were developed by Christian Gram and Paul Ehrlich for understanding the cell morphology of bacteria. In addition, R. J. Petri developed petri plates for practical usage in laboratories (Blevins and Bronze 2010). These landmarks paved other pathways for the advancement of molecular-based taxonomy.

1.3.1.2. Phase II (Year 1900–1970)

Between 1900-1970, microbiologists started applying physiological and biochemical attributes to taxa studies. This was done following the report given by the Society of American Microbiologists (later revised into the American Society for Microbiology) in 1923 (Bennett 2011). Bergey's Manual of Determinative Bacteriology (first edition) by David Hendricks Bergey was based on this report (Sentausa and Fournier 2013). The code of nomenclature for bacteria was proposed in 1930 at the First International Congress of Microbiology and later changed to the International Committee on Systematics of Prokaryotes (ICSB) (Parker et al. 2019). The International Bulletin of Bacteriological Nomenclature and Taxonomy (IBBNT) was constituted in 1950. It was later changed to the International Journal of Systematic and Evolutionary Microbiology (IJSEM) for the valid description of taxa (Oren 2015). In bacterial taxonomy, numerical taxonomy was widely applied because of the generation of massive datasets (Schleifer et al. 2009). In subsequent years, the application of semantides was made widely. By the 1960s, many groups had started using DNA-DNA hybridization as the golden standard for describing species (Schildkraut et al. 1961). Colwell, in 1970, first used the term "polyphasic" for denoting genotypic and phenotypic characters for describing genus *Vibro* members. Finally, in 1977, Carl Woese showed the implication of marker genes like 16S rRNA gene sequences to identify archaea and bacteria (Woese, 2004).

1.3.1.3. Phase III (Year 1970- till date)

This phase is still in progress due to the rapid changes in genomic and proteomic research areas. The sequencing of DNA first began in 1977 by Frederick Sanger by the chain termination method (Heather and Chain 2016). In the following years, different DNA typing methods like amplified fragment length polymorphism (AFLP) (Gzyl et al. 2005), restriction fragment length polymorphism (RFLP) (Iaganowska and Kaznowski 2004), randomly amplified polymorphic DNA (RAPD) (Gzyl et al. 2005), pulse-field gel electrophoresis (PFGE) (Xu et al. 2009) and multilocus sequence typing (MLSA) (Maiden et al. 1998) were applied for demarcating bacterial taxa (Stackebrandt et al. 2002). The breakthrough in genome sequencing of *Haemophilus influenzae* in 1995, marked the genomic era (Fleischmann et al. 1995). It had a major effect on bacterial taxonomy. NGS techniques were majorly introduced in the late years of 2000 and they provided a platform for rapid sequencing of prokaryotic genomes (Besser et al. 2018). As per the EzBioCloud database (Yoon et al. 2017), 63587 16S rRNA gene sequences have been archived, with 191909 genomes being sequenced. Of these, 26.2 % (50319) are valid, while 2 % (4030) are invalid. "*Candidatus"* taxa are accounted for 1 % (1151), while the remaining 71 % (136409) are the total taxa of the bacterial phylotype (www.ezbiocloud.net/). A total of new species names validly published from 2020 to 2021 was 2508 (https://www.ezbiocloud.net/dashboard).

1.3.2. Significance of bacterial taxonomy

Bacterial taxonomy sets the foundation upon which all relevant biological research is conducted as it aids scientific communication, therefore allowing seamless identification, classification of a bacteria and its nomenclature. Therefore, accurate identification is a basic necessity in any field of science (food industry, biotechnology, clinical studies). The following points summarize its importance in brief:

- To understand evolution, ecology, genomics and more, taxonomy helps us understand the basic unit of biology constituting the tree of life and assists in picturizing the evolutionary history and its congruent relationship with the closest organisms
- It has created standard tools to identify and classify taxa, providing a universal platform for studying and disseminating authentic and reproducible information about thousands of species across nations; in the long run, avoiding redundancies

A species is ascribed to a genus following the binary system (italicized), e.g., *Escherichia coli* (genus species). Species are named after a person, either a famous person in the field or a discoverer; the place they are isolated from or after the specific property or the functional role of a bacterium. Bacterial taxonomy helps differentiate even at the strain level by considering factors like heterogenous metabolism. For example, *Escherichia coli,* a well-studied organism, is poorly classified as some strains of this taxon are identical only by 20 % of their genomes. Similarly, "*Shigella*" strains are an effective subspecies of *E. coli*, although, for some medical reason, kept under a separate genus (Lan and Reeves 2002). Being highly diverse, these taxa should have been given a different taxonomic rank; however, to avoid disconcertment in the medical milieu, it remains unchanged (Abram et al. 2021). Therefore, the identity of any organism in-depth (up to the strain level) is fundamental before it is taken up for industry/research purposes. There are similar other instances wherein species have been identified and have caused errors in classification (Flores-Félix et al. 2019; Lalucat et al. 2021). Therefore, circumspective identification of an organism is pertinent before it is for any kind of study to avoid any kind of taxonomic error.

1.3.3. Components of bacterial taxonomy

Bacterial taxonomy generally started to resolve the physical affiliation of a bacteria to its phylogeny, thus corroborating its genomic imprints. The polyphasic based approaches are still relevant and effectively-being applied to the taxonomic studies (Tindal et al. 2010). As per the report by the ad hoc committee of the International Committee for Systematic Bacteriology issued in 1897 (Wayne et al. 1987), the following criterion could be applied for valid taxa description: (a) genotypic (b) phenotypic and (c) chemotaxonomic properties.

1.3.3.1. Genotypic analysis

With the advancement of molecular techniques and reduction in the price of sequencing technologies, the genetic typing of microorganisms has become a very useful tool in modern taxonomic studies. The paradigm growth in the application of gene markers to genome sequencing in taxogenomic studies has led to the establishment of standard protocols for genesis of novel taxa.

1.3.3.1.1. Marker gene-based

1.3.3.1.1.1. 16S rRNA

The 16S rRNA gene is an element of the prokaryotic ribosome (30S), highly conserved in the bacterial system. The length of the 16S rRNA gene is around 1500 bp, divided into nine variable regions dispersed between highly conserved regions (Jay et al. 2015). Therefore, strains are delineated by the 16S rRNA gene sequence; the strains exhibiting more than 98.2 to 99 % sequence identity were considered different species. However, with the excellent quality sequences, this value has been re-evaluated up to 98.7–99 % (Sackebrandt and Ebers 2006; Janda and Abbott 2007). DNA-DNA hybridization (DDH) is equally necessary when strains share more than 98.7 % 16S rRNA gene sequence identity. The threshold value of 16S rRNA gene identity at the genus, family and phylum level is 94.5 %, 86.5 % and 75 %, respectively (Yarza et al. 2014). The locus of ribosomes,

such as the internal transcribed spacer region positioned between its 16S and 23S rRNA genes, has also been scrutinized for phylogenetic applications. Although this technique can circumscribe at the species/strain level, it remains inconsistent at the lower level (Garcia-Martinez et al.1999).

1.3.3.1.1.2. Multilocus sequence analysis (MLSA)

Multilocus sequence analysis is often applied in epidemiology and pathological purposes; however, still found useful in bacterial taxonomy. It has an advantage over the conventional process for resolving genomic relatedness at intra/inter-species levels by profiling important housekeeping gene sequences (Maiden et al. 1998). The superiority of this method lies not only for the cultivable species but also for those which are cumbersome to isolate (Martens et al. 2008). In this study, 6–11 housekeeping genes of the species are studied (~470 bp long and stable), concatenated and applied to the phylogenetic trees. Various members of genera *Afifella* (Buddhi et al. 2019), *Thioclava* (Liu et al. 2017), *Salinivibrio* (López-Hermoso et al. 2017), *Bacillus* (Liu et al. 2015), *Brucella* (Ashford et al. 2020), *Phaeochromatium* (Shivali et al. 2012) and several taxa, researchers have applied these MLSA markers for species delineation. The cut-off values (% identity) for MLSA among the given species members were between 95.6-100 %. Shivali et al. (2012) have applied an innovative method of MLSA barcoding for the genus *Marichromatium* and reclassified *M. fluminis* into *Phaeochromatium fluminis*.

1.3.3.1.2. Genome-based

The polyphasic taxonomy augmented with molecular fingerprinting techniques like RFLP, AFLP and others have assisted in delineating of the taxa for several years. As a result of the arbitrary the topology of single gene-based analysis, significant gaps were generated in the definition of taxa. Therefore, genome-based taxonomy has been that missing link. Therefore, to overcome this shortcoming of single marker-based study,

researchers have established the field of "Phylogenomics". The word phylogenomics was first coined by Jonathan Eisen to predict gene functions (Eisen 1998). As per the current standard, phylogenomics can be described as the genome-based inference of a taxonomic standing of an organism (Szarvas et al. 2020). Its application in the bacterial systematics world started in the late 2000s. As thousands of genome sequences of bacteria are accessible only a few hundred belong to that of the type strains. Therefore, significantly limiting the usage of genomic data for the comparative study in bacterial taxonomy (Chun and Rainey 2014). Many studies have identified core genes ranging from 64 to 104 for constructing phylogenomic through various programs like UBCG (Na et al. 2018), CVTree3 (Zuo and Hao 2015), and bcgTree (Ankerbrand and Keller 2016). In recent practices, overall genomic relatedness indices (OGRIs) like ANI, AAI, *in silico* GGDH have been applied for taxa delineation.

1.3.3.1.2.1. Melting temperature difference (ΔT_m)

Melting temperature (T_m) is a fundamental thermodynamic property of DNA (Matsuda 2017). It is outlined as the temperature at which half of the DNA is in a state of a single-strand (ssDNA) depending on the DNA length and nucleotide sequence (Breslauer et al. 1986). For species definition, approximately 70 % or greater value of DDH corroborated with <5 °C (ΔT_m) was considered (Wayne et al. 1987).

1.3.3.1.2.2. *in-silico/Digital* DNA-DNA hybridization

DDH is essential when a novel taxon shows more than 98.7 % 16S rRNA gene sequence identity (Tindall et al. 2010). This technique accounts for the ability of DNA between two genomes (related members) to denature and renature. Therefore, Schildkraut et al. (1961) demonstrated the DNA-DNA relatedness between two organisms using this fundamental approach (Janda and Abbott 2007). DDH values of ≥70 % have been recommended for delineating species (Wayne et al. 1987). The significant disadvantages

of wet-lab DDH were cumbersome and non-replicable. With the increase in genome sequencing, researchers have replaced wet lab DDH with computational based *digital*/*in-silico* DDH. However, DDH has its disadvantages as it is not relevant to all the genera of prokaryotes, e.g. *Rickettsia* (Drancourt and Raoult 1994; Sentausa and Fournier 2013).

1.3.3.1.2.3. G+C content (mol %)

Genomic DNA base composition (G+C content) influences the genome stability, functioning and evolution of the bacterial process (Du et al. 2018; Šmarda et al. 2014). GC content is the calculated percentage of GC in the genome, varying from one organism to another. It is often used in the taxonomic delineation of genera and species (Meier-Kolthoff et al. 2014) and was first proposed by Sueoka (1962). Between the prokaryotic members, G+C content may vary in the range of 17 % to 80 % (Almpanis et al. 2018). Two approaches have been adapted to calculate G+C content in bacteria, traditional (thermal denaturation, HPLC, RT-PCR) and *in silico* (genome-based) analyses. Meier-Kolthoff et al. (2014) demonstrated that G+C content, if calculated from genome sequences, differed by 1 % within species and not by 3-5 %.

1.3.3.1.2.4. Average nucleotide identity (ANI)

Average nucleotide identity (ANI) is a computational program introduced by Goris et al. (2007) and is being used to define species of archaea and bacteria. It was devised as a substitute for replacing DDH. It checks the percentage identity between a pair of genomes at nucleotide level. The identity is calculated by splitting the query sequences (*in silico*) of 1020 bp against the homologous regions of the subject genome and aligning at least 70 % in length (Yoon et al. 2017). The average identity value between these regions is the final ANI value (Yoon et al. 2017). Various alignment programs for ANI identity like BLASTN (ANIb) (OrthoANIb), MUMMer (ANIm), and USEARCH (OrthoANIu) are used (Altschul et al. 1990; Lee et al. 2016). The studies revealed that 70 % DDH value corroborated 95 %

of ANI value (Goris et al. 2007; Yoon et al. 2017). ANIm and OrthoANIu process was faster than BLASTN (ANIb, OrthANIb) program-based calculator. However, OrthoANIu and ANIb are recommended for comparative studies (Yoon et al. 2007). Also, the ANI value between the range of 93 to 96 % should be cautiously studied, and phenotypic traits should be applied for differentiation of taxa (Rosselló-Móra and Amann 2015).

1.3.3.1.2.5. Average amino acid identity (AAI)

Average amino acid identity (AAI) was introduced by Konstantinidis and Tiedje (2005) as a chief benchmark for resolving taxa at a higher level. AAI can be defined as the average amino acid sequence percent identity between two genomes of species under study. It is calculated in the same way as mentioned for AAI, except that the amino acid sequence is applied. It is observed that it has higher resolution power for delineating the closely related species and good evolutionary relatedness (Rosselló-Móra 2005). Further, genome size and horizontal gene transfer are likely to have less influence on AAI than other gene-based indexes. The reason behind this is that here a more significant number of genes (at the minimum 50 genes or >500 genes) is considered for the calculation (Rosselló-Móra 2005). AAI values can be considered when the gene content is varying (<80 %) and divergent and used chiefly in supporting genus delineation.

1.3.3.1.2.6. Percentage of conserved proteins (POCP)

Percentage of conserved proteins (POCP) is a genome-related index applied for the demarcation of the genus boundary in the prokaryotic groups. According to Qin et al. (2014), for members of the same genus, the pairwise POCP values should be more than 50 %. This value could vary between 0 % to 100 %. However, many studies later showed that the given cut-off of 50 % POCP value is not suitable for demarcating many genera as it is highly conservative for genus level delineation (Liu et al. 2020; Orata et al. 2018; Wirth and Whitman 2018).

Therefore, based on the OGRI values, species and genus boundaries are given as follows. A bacterial species could be described as a group of strains exhibiting ≥70 % *in silico* DDH value, >95 % gene identity on multiple concatenated genes, >95 % of ANI and AAI values (Wayne et al. 1987; Konstantinidis and Tiedje 2005; Rosselló-Móra and Amann 2015) (Fig. 3). The rank of subspecies can be given to close and genetically related strains with divergent phenetic characters (Wayne et al. 1987). The threshold values of 60-80 % of AAI and 40-69 % of POCP are currently considered as the genus boundaries (Aliyu et al. 2016). The laying of standards and thresholds for demarcating taxonomic rank will always be discretional, even with genome-based novel methods' incorporation. Nevertheless, consistency in the whole genome derived studies will be achieved in the near future with continuous effort.

1.3.3.1.2.7. Pan genome analysis

Pan genome was first used as a designated term for an assembly of orthologous and distinctive genes in a set of organisms (Costa et al. 2020). The term was first coined by Tettelin et al. (2005) for the gene makeup of several *Streptococcus agalactiae* strains. Core genes are the indispensable part that codes for major metabolic pathways and genetic processes, in contrast, accessory genes are the ones that have been acquired in response to environmental conditions. Pan genome are applied in various studies like in vaccine development for *Leptospira interrogans* (Zeng et al. 2017), defense systems of antiphage (Doron et al. 2018) and evolutionary studies (Lebreton et al. 2017). Pan genome analysis also forms a major part of the phylogenomic analysis and speciation studies (Livingstone et al 2018; Gonzales-Siles et al. 2020). Pan genomes can be either closed (the size is fixed even after the addition of genomes) (Fig. 2A) or open (no. of the genome is directly proportional to the size of the pan genome) (Fig. 2B). Various online software like PGAP-X (http://pgaweb.vlcc.cn/pgapx/) (Zhao et al. 2018), PGAweb (http://pgaweb.vlcc.cn)

(Chen et al. 2018), and BPGA (https://sourceforge.net/projects/bpgatool/) (Chaudhari et al. 2016) are available for the pan genome analysis.

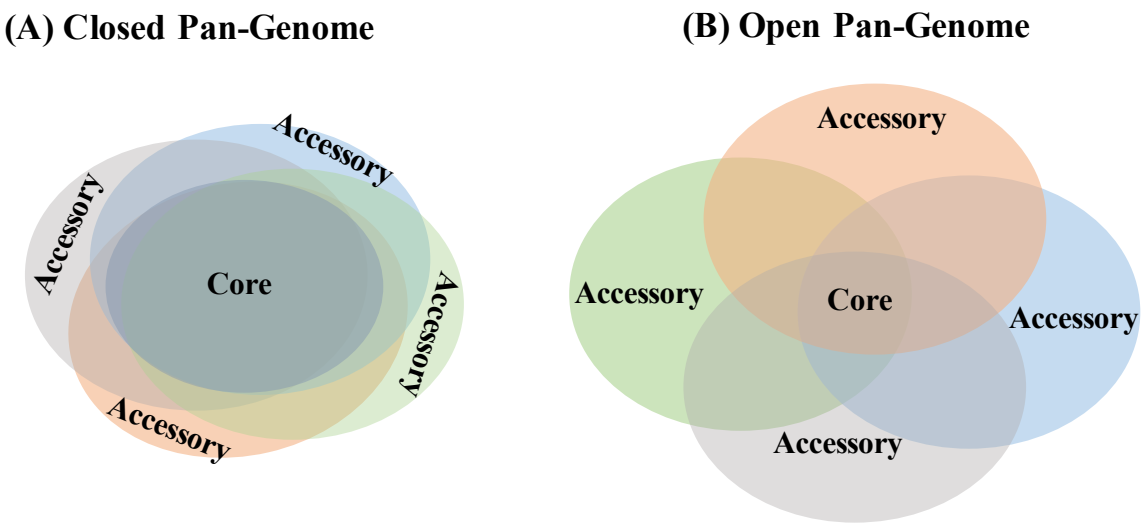


Fig. 2. Venn diagram representing closed pan-genome
(A) Open pan-genome
(B) Closed pan-genomes have more genes as compared to the open pan-genome whereas open pan-genomes have more accessory genes (Source information adapted from Costa et al. 2020)

1.3.3.1.2.8. Conserved signature indels

Protein signatures are unique and reliable molecular markers that are applied for aiding the prokaryotic classification and aid as a biochemical/molecular marker. The approaches that necessitate identification of these signatures are conserved signature indels (CSIs) applied in many re-classifications of prokaryotes (Naushad et al. 2014). The aligned regions of proteins where a specific insertion or deletion has occurred in members of taxa but not in the other taxa can be defined as conserved signature indels (CSIs) (Gupta 1998). Further, these regions should be flanked by conserved regions to rule out the possibility of an artefact. They are distinctive genetic occurrences at stages of evolution, therefore, can be implemented for the classification of taxa (Dobritsa et al. 2019; Hu et al. 2018).

1.3.3.1.2.9. Genome annotation

Genome annotation is described as determining functional segments within the nucleotide sequence of a genome (Emmanuel et al. 2021). In the initial stage often, it is focused on the annotation of genes, location of coding genes and their regulatory regions, function predictions. The two main components of genome annotation are structural annotation and functional annotation (Ejigu and Jung 2020). Structural annotation includes predicting features of DNA like introns, exons, promoters, non-coding RNA genes (tRNA, rRNA, miRNA), pseudogenes, CRISPR/cas, etc. (Zhang et al. 2016). The first measure in the structural annotation is repeated masking followed by gene (protein-coding/regulatory) identification. The gene identification is made by *ab initio*, homology-based or combined methods (Ejigu and Jung 2020; Stein 2001). *Ab initio* procedure predicts genes based on nucleotide sequences and statistical/mathematical models (hidden Markov model, HMM) to identify different features (Scalzitti et al. 2020).

On the contrary, homology-based prediction does so by aligning sequences to the expressed sequence tags (EST), existing proteins, or identities of genes through databases (Ejigu and Jung 2020). The combined methods use both *ab initio* and homology-based gene prediction (Mathé et al. 2002). The European Nucleotide Archive (ENA) (Brooksbank et al. 2014), GenBank (Sayers et al. 2019), and DNA Databank of Japan (DDBJ) (Kodama et al. 2018) are some of the nucleotide and protein sequence databases for structural annotation. Functional annotation is defined as the meaningful association of biological data to the gene or protein sequences that have been predicted by the structural annotation (Ejigu and Jung 2020). It is achieved by using the BLASTP program for the high-score local alignment (above some threshold value) against the protein databases, after which the function is assigned. Various databases like Gene Ontology (GO) (Ashburner et al. 2000), Kyoto Encyclopedia of Gene and Genomes (KEGG) (Kanehisa and Goto 2000) are

currently applied for functional annotation. Rapid Annotations using Subsystems Technology (RAST) is an online algorithm for the genome annotation of prokaryotes (Aziz et al. 2008). It recognizes genes using a combined homology method and subsystems for gene calling (Overbeek et al. 2005). Glimmer (although other options like Prodigal/GeneMark are also present) are mainly applied (Aziz et al. 2008). Furthermore, genome annotation services in PathoSystems Resource Integration Center (PATRIC) use the RAST toolkit to interpret genomic features (Davis et al. 2020). Finally, the general steps in phylogenomic studies are presented in Fig. 3.

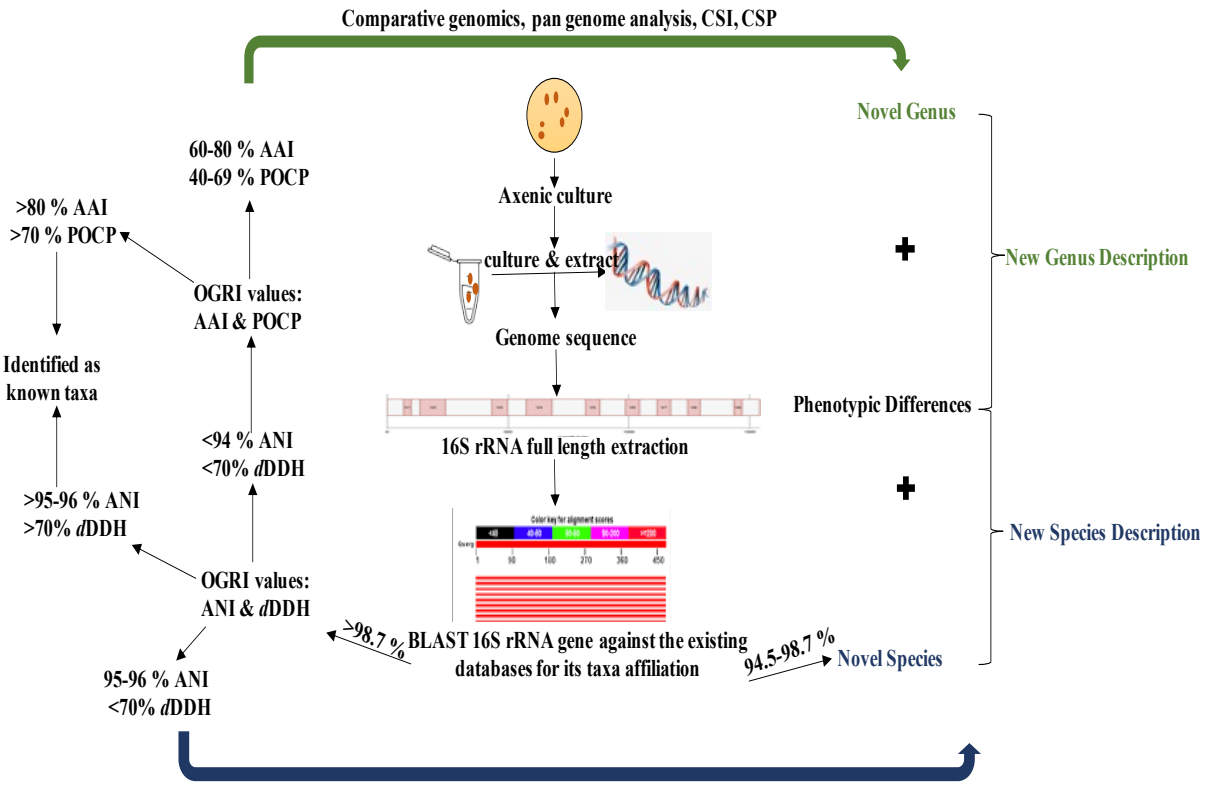


Fig. 3. Genome based delineation of taxa at genus and species level (Source information adapted from Chun et al. 2018)

1.3.3.2. Phenotypic analysis

Phenotypic characteristics are the discernible traits observed in an organism under study. Its distinguishing characteristics should vehemently substantiate the variations in the genome sequences. The phenotypic studies form the foundation for the formal circumscription of taxa, from subspecies and species up to the family level (Garrity 2016). The precedence of phenotypic characterization at the outset is that it is observed effortlessly, tracked, measured and economical. However, the same characteristics may vary according to different conditions under which the genes are expressed, limiting their application (Richter and Rosselló-Móra 2009). Therefore, phenotypic information should be studied under identical growth conditions along with the type strain of closely related organisms (Logan et al. 2009; Tindal et al. 2010). Such phenotypic characters are described below in detail.

1.3.3.2.1. Morphological characterization

The morphological characteristics are those traits that can be analyzed in the laboratory under specified or controlled environments. The morphological traits observed and applied for taxonomic studies are colony morphology (shape, size, colour), Gram-staining properties, flagellation, sporulation, etc.

1.3.3.2.2. Metabolic characterization

Metabolic characterization involves both physiological and biochemical studies. The conventional phenotypic tests applied in classical microbiological laboratories are the growth of organisms on different substrates (carbon, nitrogen, vitamin) with different growth ranges in various conditions such as pH, temperature, salinity, oxygen conditions (aerobic/anaerobic). Other metabolic characteristics like enzyme activity, fermentation studies and susceptibility to different antibiotic stress were tested as described (Tindall et al. 2010).

1.3.3.3. Chemotaxonomic analysis

Chemotaxonomy concerns the application of various analytical techniques based on various chemical constituents of the bacterial cells for classification (Tindal et al. 2010). Over the years, standard and reliable chemotaxonomic markers have been established to accurately identify and classify bacteria. The most widely used chemotaxonomic markers comprise cell wall components such as peptidoglycan, polar lipid composition, comparable ratios of fatty acids, sugars, etc. Other important chemotaxonomic markers are polyamines, isoprenoid quinones, carotenoids, chlorophyll composition, indole derivatives and fermentation products.

1.3.3.3.1. Polar lipid

Polar lipids are amphiphilic with a hydrophobic tail and a hydrophilic head associated with membrane integrity (Zheng et al. 2019). The primary function of polar lipids is to maintain cell homeostasis, signaling and other developmental processes (Bale et al. 2019). The membrane lipid composition varies widely between groups of bacteria (Nguyen and Kim 2017). Therefore, polar lipids are considered important chemotaxonomic markers for taxa distinguishing in prokaryotes (Tindall et al. 2010). Phosphatidylcholine, phosphatidylserine, phosphatidylinositol, phosphatidylglycerol and phosphatidylethanolamine are some glycerophospholipids (FA glycerol backbone) (Dugail et al. 2017). Glycolipids and amino lipids are non-phosphorus/neutral lipids (Dugail et al. 2017). The merging of the three species of *Krokinobacter* into the genus *Dokdonia* (Yoon et al. 2011) and reappraising some members of the genus *Rhodobium* into a new genus marker (Urdiain et al. 2008) serves as a classical instance of resolving taxonomic status by applying polar lipids as chemotaxonomic.

1.3.3.3.2. Cellular fatty acid

Cellular fatty acids are carboxylic acids with saturated/unsaturated aliphatic chains obtained after hydrolysis of fats. As mentioned above, the membranes of bacteria consist of lipid bilayers. Its principal function is to maintain membrane stability and functionality of the cell (Singer 1972) and was first shown by Michel Eugène Chevreul as *graisse acide*/ acidic fat (Chevreul 1823). FA are conserved because of their role in maintaining the fluidity of the cell. Under stressful and specific conditions, significant changes in bacterial carotenoid composition, hopanoids, proteins and lipid composition occur (de Carvalho and Caramujo 2018). This may occur as a prime generation of metabolic energy occurs in the membrane; therefore, membrane stability is important (Bajerski et al. 2017). For microbial identification, FAs of varying lengths of 9 to 20 carbon atoms, saturated/unsaturated, branched FA (iso, aneiso, methylated), hydroxy FA, cyclo-FA are surveyed for the classification by the process of fatty acid methyl ester (FAME) analysis (Kunitsky et al 2006). FAME analysis was first developed by Myron Sasser and was done using a technique called Sherlock® microbial identification system (MIDI Inc) (Kunitsky et al. 2016). It can be described as the phenotypic profiling of the total saponifiable lipids in bacteria and identifying by gas chromatography (Sharmili and Ramasamy 2016). This involves quantification and specific analysis of FA by hydrolysis of the lipids followed by methylation, the retention time of and the products (methylated esters) assists in identifying the FA (Kunitsky et al. 2006). Therefore, justifying the taxonomic application of cellular FA as a powerful tool for bacterial taxonomy.

1.3.3.3.3. Isoprenoid quinones

Isoprenoid quinones consist of a benzene core with 2 hydrogen atoms substituted by oxygen atoms, resulting in the formation of 2 carbonyl bonds (El-Najjar et al. 2011). The respiratory isoprenoid quinones are present in the cytoplasmic membrane of the

bacteria and are an essential part of the electron transport systems, along with having antioxidant properties (Nowicka and Kruk 2010). In nature, quinones are majorly divided into naphthoquinones and benzoquinones (evolutionary younger). Menaquinone and phylloquinone come under naphthoquinones whereas, ubiquinone and plastoquinone come under benzoquinones. They serve as a diagnostic tool in bacterial taxonomy because the quinone molecules differ in saturation, side-chain length and hydrogenation in different taxa (Nowicka and Kruk 2010). Initially, the studies of Bishop et al. (1962) and Crane (1965) also showed that inherent structural differences could be applied to distinguishing bacterial taxa. Subsequently, Hirashi et al. (1984) studied the distribution of quinones in the members of the family of *Rhodospirillaceae* and divided them into five categories. It is one of the major chemotaxonomic markers used (Collins and Jones 1981).

1.3.3.3.4. Cell wall amino acid

Peptidoglycan (PG) is an essential component of the bacterial cell wall, with its primary function being to maintain cell integrity (Vollmer et al. 2008). PGs are linear stands of glycan (alternating *N*-acetylglucosamine and *N*-acetylmuramic acid linked by β-1-4 bonds) cross-linked by short peptides (Rogers et al. 1980). Variations in cross-linkages and PG structures were found in various bacteria. Schleifer and Kandler (1972) included a tri-digital system for PGs classification and subsequently showed the correlation between peptidoglycan types and taxonomic groups. Therefore, serve as an excellent marker for chemotaxonomic analysis.

1.3.3.3.5. Polyamines

The distribution of polyamines in bacteria varies significantly in quantitative/qualitative ways, due to which it has been applied as a chemotaxonomic marker. Polyamines are small polycationic compounds with numerous cellular functions related to DNA, protein, RNA complexes and modulation of ionic channels (Carriche et al.

2021; Igarashi and Kashiwagi 2010). Polyamine patterns are visualized for differentiating and defining taxonomic groups (Busse and Auling 1988; Hamana and Matsuzaki 1992). 24 linear and 4 branched types are known, of which molecules like putrescine, spermidine, 2-OH putrescine, syn-homospermin and spermin are highly studied and applied in chemotaxonomy. Polyamine patterns wherein the absence/presence of norspermidine, homospermidine or spermidine; 2-OH putrescine or diaminopropane are investigated for distinguishing the members belonging to the classes (alpha, beta, gamma and delta) of *Proteobacteria* (Hamana and Matsuzaki 1992). Also, the presence of syn-norspermidine in the members of the genus *Vibrio* distinguished it from the other genera (Yamamoto et al. 1986).

1.3.3.3.6. Hopanoids

Hopanoids are pentacyclic triterpenoid sterol-like compounds found in the membrane lipids of diverse bacteria (Belin et al. 2018). Hopanoids are considered molecular fossils and first discovered by petroleum geologists. Therfore, they hold source-specific information, and can be associated with distinct taxonomic groups. Hopanoids are chemotaxonomic markers and used in some current chemotaxonomic analysis (Cvejic et al. 2000; Lodha et al. 2015).

1.3.3.3.7. Carotenoids

Carotenoids are tetra-terpene pigments/molecules (8 isoprene units) distributed and are abundantly present in bacteria, archaea, fungi and plants (*de novo* synthesis). Around 850 naturally formed carotenoids are reported (Maoka 2020). Carotenoids are categorized into carotenes (hydrogen, carbon) and xanthophylls (hydrogen, carbon, oxygen). Bacterial carotenoids are markedly diverse and used as a distinguishing characteristic in chemotaxonomy. Initially, Imhoff and Caumette (2004) recommended incorporating carotenoid analysis for the description of phototrophic species. In addition, many

taxonomic descriptions have included carotenoid pigments to differentiate at the species level (Anil et al. 2009; Kim et al. 2020).

1.3.3.3.8. Indole derivatives

Indole is a bicyclic heterocyclic molecule with the formula of C_8H_7N, consisting of a benzene ring fused with a pyrrole ring. The indole derivatives (ID) are derived from tryptophan or an indole molecule. Naturally, IDs are widely distributed and are produced abundantly by bacteria. Mostly functioning as signaling compounds, resistance to drugs, plant growth promotion, virulence, etc. (Kumari and Singh 2019; Lee and Lee 2010). Certain bacterial taxa produce economically important ID molecules like auxin (indole-3-acetic acid), melanin (neurohormone), bacillamide (algicidal), Xiamycin (antiviral) (Netz and Opatz 2015). Under varying conditions, species produce unique indole profiles, which can be used to differentiate taxa. Indole derived molecules have been used as a chemotaxonomic marker in various studies for delineating species (Huang et al. 2021; Zhang et al. 2021).

1.4. Definition of the problem

India is one of the 17 mega biodiverse countries in the world and houses approximately 47000 plants and 96000 animal species (Singh and Das 2018; Dar et al. 2020). This diversity may be attributed to the unique geographical location and majorly divided into hotspot regions of the Himalayas, Western Ghats, Indo-Burma and Sundaland (Tiwari et al. 2022). Despite the vast richness, the microbial world is understudied and less explored. Therefore, discovering microbial wealth from these hotspots areas may lead us to discover novel microorganisms with potential application in biotechnology and industry. Therefore, isolation sources for bacterial diversity studies are critical for the successful acquisition of application-oriented bacterial isolates. One such interesting area of study from these hotspots are orchid plants, solar saltpans, sand beaches, hot springs and acid

mine, having unique modes of interaction with the microbiome (Chap.1.1.1). Therefore, scrutinizing these bacteria will play a crucial role in understanding their physiological processes and interactions. Indexing the communities by deliberated taxonomic knowledge can be significant in species conservation and flower cultivation (in the case of orchid plants). Therefore, our main aim would be to catalogue their bacterial diversity and infer their potential pathogenic and economic capabilities. As for the statistics, between 2020-2021, fifty-six species were described from India, which forms 2.5 % of the global species description (2371) of valid names (https://lpsn.dsmz.de); which is negligible as compared to the substantial microbial diversity which our country holds. Therefore, notable efforts are to be made to study the bacterial diversity for unearthing specialized microbial communities. Based on this understanding, the present study was undertaken with the following objectives.

1.5. Objectives

1. To study the root-associated bacterial diversity of a few epiphytic orchids
2. To discover the cultivable bacterial diversity of epiphytic orchids together from other habitats and describe new species, if any

MATERIALS AND METHODS

2. MATERIALS AND METHODS

2.1. Labware and chemicals

2.1.1. Laboratory wares

Laboratory glassware and plastic wares manufactured by different companies (Borosil, SCHOTT Duran®, Tarson, Riviera, Labsystems, Eppendorf, Thermo Fisher, and Anumbra) like Erlenmeyer flasks, round bottom flasks, test tubes, petri-plates, cylinders for measuring, beakers, slides, screw cap test tubes, storage bottles, pipettes, spreaders, PCR tubes were used.

2.1.2. Chemicals, solvents, and laboratory kits

The analytical chemicals and solvents from different firms (Sigma-Aldrich, HiMedia, Qualigens, SR LIFESCIENCES, Amresco, Merck, Thermo Fisher) were purchased and used. Ready-to-use kits for DNA extraction (Nucleopore Fungus Bacteria Kit, Genetix brand) and gel purification (SureExtract PCR kit, Genetix brand) were used.

2.1.3. Distilled, double-distilled, and milli-Q water

For routine media preparation, buffer and DNA/RNA work distilled, double-distilled, and milli-Q water was used and procured from the millipore facility at the School of Life Sciences, University of Hyderabad.

2.2. Buffers, standard solution, and pH determination

Double distilled water and Milli-Q were used to prepare of buffer solutions and standard solutions. Phosphate buffer (KH_2PO4/K_2HPO4), bicarbonate buffer ($NaHCO_3/NaOH$), phosphate buffer saline (PBS) (Na_2HPO4/KH_2PO4) at different pH were prepared and sterilized by autoclaving. Tris-EDTA (TE) buffer [1 ml of 1M Tris-HCl (pH 8), 0.2 ml of 0.5M EDTA (pH 8) with volume made up to 100 ml] and 50X Tris-Acetate-EDTA (TAE) buffer [24.2 g Tris base, 5.71 ml glacial acetic acid, 10 ml 0.5 mM EDTA (pH 8), final volume 100 ml) was prepared. 20X Sodium saline citrate (SSC) buffer

was prepared as 1.75 g of NaCl, 0.77 g of sodium citrate, and a final volume of 100 ml (pH 7, adjusted with HCl). pH was determined using Digisun Electronics (DI-707) and adjusted accordingly.

2.3. Culture media

In the study, the following media was used and composition are given thereof:

i) Nutrient broth (NB) (gL^{-1}): beef extract (1), yeast extract (YE) (1.5), peptone (5), NaCl (5); pH adjusted to 7 and autoclaved

ii) Beibl and Pfenning media (BP) (gL^{-1}): Sodium pyruvate (3), YE (0.12), Ferric citrate [5 ml (0.1 % w/v stock)], Vitamin B_{12} [1 ml (0.2 mg/10 ml stock)], SL7 (1ml), KH_2PO_4 (0.5), $MgSO_4.7H_2O$ (0.2), NaCl (0.4), NH_4Cl (1.2), $CaCl_2.2H_2O$ (0.05); SL7 consists of the following (mg/L); $CoCl_2.6H_2O$ (200); $MnCl_2.4H_2O$ (100); $NaMoO_4$ (40); $ZnCl_2$ (70); H_3BO_3 (60); $CuCl_2.2H_2O$ (20); $NiCl_2.6H_2O$ (20); HCl 1ml (25 % v/v); Vitamin B_{12} to be sterilized by filtration and pH adjusted to 7

iii) Zobell Marine Media (ZM) (gL^{-1}): peptone (5), YE (1), Ferric citrate (0.1), NaCl (19.5), $MgCl_2.6H_2O$ (8.8), $Na_2SO_4.10H_2O$ (3.2), $CaCl_2.2H_2O$ (1.8), KCl (0.55), $NaHCO_3$ (0.16), KBr (0.08), $SrCl_2$ (0.034), H_3BO_3 (0.22), Na_2SiO_3 (0.004), NH_4NO_3 (0.0016), Na_2HPO_4 (0.008), NaF (0.0024); finally pH adjusted to 7

iv) R2A media (gL^{-1}): Casein hydrolysate (0.5), YE (0.5), peptone (0.5), dextrose (0.5), starch (0.5), sodium pyruvate (0.3), K_2HPO_4 (0.3), $MgSO_4$ (0.024); final pH adjusted to 7

v) ATCC Medium (2039) (gL^{-1}): Solution A [$MgSO_4.7H_2O$ (2), K_2HPO_4 (0.4), $(NH_4)_2SO_4$ (0.8)], *Wolfe's mineral solution (5 ml) and dist. water (800 ml)]. Solution A pH adjusted to 2.3 with H_2SO_4 and sterilized by filtration. Solution B [$FeSO_4.7H_2O$ (20), dist. water (200 ml)], stir for dissolving and filter sterilize.

Add solution A and B. *Wolfe's mineral solution [$MnSO_4.H_2O$ (0.5), $MgSO_4.7H_2O$ (3), $FeSO_4.7H_2O$ (0.1), NaCl (1), $CaCl_2.2H_2O$ (0.1), $CoCl_2.6H_2O$ (0.1), $ZnSO_4.7H_2O$ (0.1), $CuSO_4.5H_20$ (0.01), H_3BO_3 (0.01), $Na_2MoO_4.2H_20$ (0.01), dist. water (1000 ml)]

vi) Oligotrophic mineral media (OMM) (gL^{-1}): $MgCl_2.6H_2O$ (0.030), KCl (0.03), $NaHCO_3$ (0.05), KH_2PO_4 (0.18), NH_4Cl (0.115), Na_2SO_4 (0.15), YE (0.2), Trace elements (mg) [$MnCl_2.4H_2O$ (180), $Na_2B_4O_7.10H_2O$ (450), $ZnSO_4.7H_2O$ (22), $CuCl_2.2H_2O$ (5), $Na_2MoO_4.2H_2O$ (3), dist. water (1000 ml)]

2.4. Cleaning and sterilization

The essential glassware was first rinsed and cleaned with detergent (Teepol solution), after which the rinsed wares were soaked in the chromic acid solution overnight. Finally, the wares were rinsed thoroughly, dried, autoclaved, and used. The sterilization process was carried out by autoclaving (SA52, Ketan) of the heat-stable materials/solutions at 121° C, 15 psi for 15 min. The heat-labile solutions were filtered by 0.22 µm membrane filters manufactured by the Millipore company.

2.5. Sample collection

Roots of epiphytic and terrestrial orchids from various geographical regions of Sikkim, West Bengal, Orissa, and Karnataka (Fig. 4); sediment/sand from coastal regions of India were collected. The roots were collected and put in sterile centrifuge tubes (50 ml). Characteristically, the sampled orchids had a healthy appearance. For other samples like sediments and sands, the pH, salinity (AD310 conductivity meter, Adwa, Romania), and GPS locations of the site were recorded during the sampling. The samples were transported at the earliest to the lab for analysis, and for further use, refrigerated at 4 °C.

Bulbophyllum sp.
Chitrey, Sikkim
27. 10N 88.46E

Agrostophyllum sp.
Chitrey, Sikkim
27.11N 88.47E

Cymbidium sp.
Lachung, Sikkim
27. 95 N 73.99E

Dendrobium sp.
Darjeeling, West Bengal
27..07N 88.38E

Dendrobium sp.
Darjeeling, West Bengal
27.07N 88.39′E

Bulbophyllum sp.
Darjeeling, West Bengal
27.07N 88.39E

Thunia sp.
Darjeeling, West Bengal
27.07N 88.41E

Rhynchostylis sp.
Devarakund, Orissa
21.70N 86.44E

Aeridis sp.
Amboli, Karnataka
15.95N 73.99E

Vanda sp.
Amboli, Karnataka
15.85N 73.99E

Fig. 4. Description of orchids (name, place, location) collected for metagenome at the sampling site from various regions

2.6. Amplicon based metagenomic studies

2.6.1. DNA extraction from orchid samples

For the metagenome, 1 g of the roots of the ten orchids were pooled from each sample and crushed in 0.1 % (w/v) saline water. DNA was extracted from ten samples with the help of MO BIO's PowerSoil DNA Isolation kit (QIAGEN) as per the manufacturer's instruction. DNA concentration of each sample was checked by measuring with a NanoDrop spectrophotometer (A_{260}/A_{280} ratio) to see that similar concentrations was considered for amplification. Simultaneously, the DNA samples were visualised in 1 % (w/v) agarose gel under a UV transilluminator (Universal Hood II, BIO-RAD). DNA samples were preserved at -20 °C until further analysis.

2.6.2. Library preparation and Illumina MiSeq sequencing of the 16S rRNA gene

The extracted DNA was then outsourced to Eurofins Bangalore, India, for sequencing. The amplicon-based library (2x300 MiSeq) from the above samples were constructed using the Nextera XT Index kit (Illumina inc.) as per the 16S metagenomic sequencing library preparation code. The amplicons were amplified by i5 and i7 primers, adding multiplexed index sequences and standard adapters for cluster generation (P5 and P7) according to the standard Illumina protocol. The library was finally purified by 1X AMPureXP beads. V1-V3 specific primers were designed to amplify the regions of DNA, and the primers used were F-V1-V3 (F-AGAGTTTGATCMTGGCTCAG) and R-V1-V3 (R-ATTACCGCCGGCTGTGG). The amplicon library was added in MiSeq at a concentration (10-20 pM) for cluster generation and sequencing. Paired-end sequencing allows the template fragments to be sequenced on both forward/reverse sides. In the paired-end flow cell, the kit reagents help in binding the sample complementary adapter oligos. The adapter carefully cleaves the forward strands after the re-synthesis of the reverse strand while the opposite end of the copied reverse strand is used for sequencing.

2.6.3. Downstream analysis and taxonomic assessment

The adapter sequences, ambiguous reads (nucleotides "N" >5 %), sequences (reads with > 10% quality threshold, < 20 phred score), sliding window of 20 bp, and a minimum length of 100 bp were trimmed using Trimmomatic v 0.35 (Bolger et al. 2014). All the reads generated were submitted to the SRA database through the FTP server at NCBI in the database (www.ncbi.nlm.nih.gov). The bio-project was created at the NCBI portal to submit the raw reads. After the bio-project's approval, a bio-sample was created by filling in relative information of the samples. All the sequence data were given NCBI accession numbers. The fastq reads obtained from sequencing were further processed using an open-source bioinformatics pipeline, mothur (Schloss et al. 2009). The sequences were first stitched into single-end reads/contigs (command: make.contigs) and screened. Sequence length below 350 bp with homopolymers over 8 bp containing ambiguous base calls and incorrect primers were removed (command: screen.seqs). Common sequences were binned to create a set of unique sequences to allow for the counting in each sample (command: unique.seqs, count.seqs). Next, unique sequences were aligned (V1-V3 regions) (command; align.seqs) against the SILVA database (Quast et al. 2013).

After filtering of alignment (command: filter.seqs), reads were de-noised using a pre-clustered routine to remove sequences that were likely due to sequencing errors (command: pre.cluster). Putative chimeric sequences were detected and removed via the chimera VSEARCH algorithm (command: chimera.vsearch, remove.seqs) (Rognes et al. 2016). After the successful quality checking and filtering, sequences from each sample were clustered using RDP Bayesian classify with a minimum confidence of 0.8 (command: classify.seqs), with each cluster representing a species (Wang et al. 2007). The resulting sequences were clustered into OTUs using dist.seqs and cluster. Finally, a taxonomy for each OTU (classify.otu) was generated. Rarefaction curves were represented using MG-

RAST (Meyer et al. 2008). The bacterial diversity analysis and comparison of taxa were carried out using Microsoft Excel and software (PAST V3.26, Morpheus, Graphpad Prism). The unclassified sequences from each dataset were extracted and aligned by CLUSTAL W (with representative/type species of 53 phyla). The trees were visualized in the iTOL software (www.itol.embl.de), and novel phylotypes were deduced accordingly.

2.7. Characterization of bacterial strains

2.7.1. Enrichment, isolation, and purification

The samples (0.1g) from industrial acid drainage and hot-springs were enriched in 15 ml of ATCC medium and OMM broth in test tubes (15×150 mm), respectively. For hot-springs and acid drain samples it was incubated at 45±2 °C for 15 days and 32±2 °C for 3 to 4 days, respectively. On turbidity in the test tubes for industrial acid drainage and hot-springs, serially diluted aliquots (10^{-4}) were plated in ATCC medium and OMM agar, respectively. For the orchid studies, 1g of the root macerate was taken under sterile condition, after which it was suspended in saline [0.8 % (w/v) NaCl]. It was serially diluted, followed by plating on the Biebl and Pfenning (BP) media, which contained pyruvate (3 gL^{-1}) as the source of carbon and root extract (1 % v/v). The water sample collected from the saltpan was pasteurized at 80 °C for 15 min to eliminate of vegetative cells and diluted (10^{-4}) (Garabito et al. 1998). 100 µl of the dilution was spread on ZM agar added with 2 % (w/v) NaCl, followed by incubation at 30 °C for 3 days. For all the samples, the distinct colonies appearing were picked and subsequently purified by repeated streaking on the media mentioned above.

As for the sand from the beach sample, with NaOCl (4 % v/v), H_2O_2 (3 % v/v) and $HgCl_2$ (0.5 % w/v) (1:1:1), 1g of layer 2 sand was surface sterilized. The sample was incubated overnight to remove the epilithic microorganisms. The sand was washed thoroughly with sterile distilled water. The sample extraction buffer was added (Horath and

Bachofen 2009) and crushed thoroughly using sterile mortar-pestle and sonicated under sterile conditions for 30 min to liberate all the microorganisms. 100 µl of this extraction buffer (sonicated) was used for enrichment in an (1/10 strength) oligotrophic Biebl and Pfennig (BP) media with pyruvate [0.001 % w/v] and NH_4Cl [0.001 % w/v] as carbon and nitrogen sources, respectively (Biebl and Pfennig 1981). Following 10 days of incubation, 100 µl of the turbid enrichment was spread and plated on ZM agar at 30 °C. The pink colonies were purified by subsequent streaking on the NA plate.

2.7.2. Culture preservation

The purified cultures were maintained either in glycerol stocks or as agar stabs. For glycerol stocks, 1ml of axenic culture (at exponential growth phase) grown in nutrient broth (NB) were added to 1ml of 100 % autoclaved glycerol. The stocks were finally stored at 4 °C for 12 hours and then transferred to -20 °C for longer preservation. Agar stabs were prepared in plastic tubes (5ml), wherein 3 ml of NB with 2 % agar was added and finally allowed to solidify in a slant. The cultures were streaked on the surface and incubated at room temperature for 3-4 days. It was then kept at 4 °C for preservation for a longer period. The cultures were periodically checked (once every 3 months) for purity and viability.

2.7.3. Phenotypic studies

2.7.3.1. Morphological analysis

2.7.3.1.1. Colony morphology

The morphological properties of colonies like colour, shape, size, elevation, surface, margin, and opacity of all bacterial strains were observed in NA media grown at optimum conditions. For comparative studies of strains with their closely related type strains, those strains were ordered/exchanged from other culture collections as mentioned.

2.7.3.1.2. Cell morphology

2.7.3.1.2.1. Microscopic observation

The morphology of the cell (size, shape, reproduction/division) was studied by microscopic techniques like a light microscope (Alphaphot-2 YS2, Nikon), phage contrast microscope (Olympus-B201), confocal microscope (Zeiss LSM 880), and field scanning electron microscope (FESEM) (Philips XL3O). For the bacterial cells to be visualised under FESEM, 1 ml of the exponential phase culture was centrifuged at 8000 g for 10 min at room temperature. The cell pellet was further washed with phosphate buffer (0.05 mM, pH 7.2) and centrifuged (identical condition as mentioned). The washed pellet was finally added to a 0.25 % (v/v) glutaraldehyde mixture and incubated overnight at 4 °C. The cells were pelleted again and dehydrated by re-washing sequentially with increased ethanol concentration (10-100 % at an intervals of 10 %). Samples (dissolved in 100 % ethanol) were applied on mini round-shaped coverslips and further kept on SEM stabs. SEM stabs were given for gold sputtering and finally visualised under FESEM.

2.7.3.1.2.2. Staining methods

Staining methods like Gram Stain and Schaeffer-Fulton staining method were used to study the cell wall properties and check endospore. HiMedia Gram stain kit (K001-1KT) was used for staining. For the Schaeffer-Fulton staining process, malachite green (primary) was added to the smear of bacterial cells and heated, leaving for 15 min. The smear was decolourised with distilled water and counterstained by safranin (0.5 %) for 1 min. It was finally rinsed with distilled water and observed under the light microscope (100x).

2.7.3.2. Physiological characterization

2.7.3.2.1. Determination of growth mode

Photoorganoheterotrophic mode of growth was determined by inoculating culture into completely filled screw cap tubes (10×100 mm) with BP media. Pyruvate (0.3 % w/v)

was used as a carbon substrate and was anaerobically incubated under light (2400 lux) at 30±2 °C. Chemoorganoheterotrophic mode of growth was checked in an Erlenmeyer flask (250 ml) with 100 ml BP media supplemented with pyruvate as carbon source (0.35 % w/v) and incubated aerobically at 30±2 °C (dark). For checking the fermentative mode of growth, BP media supplemented with pyruvate (0.3 % (w/v) as a carbon source in screw cap tubes (fully filled, 10×100 mm) was used, followed by incubation anaerobically in the dark at 30±2 °C. At 540 nm, growth was checked turbidometrically using a colorimeter (Systronics) against the blank (media without culture). The optical density, if increased, was considered as an increase in growth.

2.7.3.2.2. Growth in various conditions

i) Temperature: Culture was grown under the temperature range of 10, 15, 20, 25, 30, 35, and 40 °C in NB and incubated for 3-4 days.

ii) pH: Similarly, growth at pH range 5-11 was checked in R2A/NB media using a buffer system: pH 5-8, KH_2PO_4/K_2HPO_4 (phosphate buffer); pH 9-11, $NaHCO_3$/NaOH (bicarbonate buffer). It was then incubated at 30±2 °C.

2.7.3.2.3. Growth at different NaCl %

Bacterial cultures to be tested were inoculated in NB added with 0-15 % NaCl (w/v) (at an interval of 1 %) and incubated at 30±2 °C for 3-4 days. Growth turbidity was measured as mentioned above (Section 2.4.3.2.1).

2.7.3.2.4. Vitamin requirement

For the bacterial culture under study, vitamins [thiamine-HCl (300 μgl^{-1}), biotin (50 μgl^{-1}), calcium pantothenate (10 μgL^{-1}), pyridoxine-HCl (15 μgL^{-1}), nicotinic acid (300 μgL^{-1}), *p*-aminobenzoic acid (200 μgL^{-1}) and vitamin B_{12} (15 μgL^{-1})] were tested in BP media by replacing yeast extract with filter sterilized solution of vitamins. The cultures

were sub-cultured without the vitamins repeatedly for three times to check its absolute requirement.

2.7.3.2.5. Carbon utilization

As mentioned above, the various carbon sources (D-glucose, pyruvate, D-mannitol, rhamnose, D-galactose, ribose, D-mannose, maltose, sucrose, sorbitol, inositol, L-alanine, histidine, gluconate, L-proline, D-xylose, trehalose, citrate, methanol, lactose, acetate, propionate, fructose, glycerol, ethanol, malate, arabinose) were added to BP media by replacing pyruvate in BP media at 0.3 % (w/v).

2.7.3.2.6. Nitrogen utilization

The nitrogen sources requirement for growth was tested by replacing NH_4Cl with various nitrogen sources (peptone, sodium nitrate, ammonium sulphate, sodium glutamate, casamino acid) at 7 mM. Growth was checked turbidometrically, as mentioned above.

2.7.3.3. Biochemical characterization

2.7.3.3.1. Oxidase and catalase test

Oxidase discs from HiMedia (DD018) were used in accordance with the manufacturer's instructions. The catalase test, was assessed by observing bubble generation upon the addition of H_2O_2 (3 % v/v).

2.7.3.3.2. Nitrate reduction and H_2S production

Nitrate reduction was tested based on HiMedia (DD041) discs as per the manufacturer's instruction. The production of H_2S was determined on sulphide indole motility agar medium. The blackening of the media along the stab line of inoculation suggests a positive result.

2.7.3.3.3. Hydrolysis of compounds

Hydrolysis of gelatin (3 % w/v), starch (0.2 % w/v), urea (2 % w/v) were carried out in gelatin agar (HiMedia M920), starch agar (HiMedia M1075) and urea broth

(HiMedia M112), respectively. For other compounds like Tween 20, Tween 40, Tween 80, casein, chitin, and cellulose, the hydrolytic capability was checked at 0.1 % (w/v) substituted as substrates in BP media.

2.7.4. Chemotaxonomic studies

Cells were cultured in nutrient broth aerobically (3-4 days at 30 °C). Cells were pelleted (10000 g at 4 °C for 10 min) when the culture growth was attained at around 70 % of the maximum optical density at the logarithmic phase. The pellet was utilized for the following analysis.

2.7.4.1. Polar lipid profiling

For all the strains, 1g of the lyophilised cells were used to extract the polar lipids as described by Kates (1986). To this 1g, 2 ml of NaCl (0.3 % w/v) was added and kept for 15 min at room temperature. 10 ml of methanol was further added to the mixture and vortexed well. The mixture was then steam heated till it was evaporated entirely and was allowed to cool. For this dried cell part, 10 ml of chloroform and 6 ml of 0.3 % NaCl was added and vortexed well. After the centrifugation, a well-separated aqueous phase was observed. The lower layer (chloroform layer) was taken to the round bottom flask and concentrated using a rotary flash evaporator (LABOROTA 4001, Heildolph). 0.3 ml of chloroform and methanol (2:1) was finally used for dissolving the dried part consisting of polar lipid. Finally, the polar lipids were distinguished by using silica gel TLC plates (Kieselgel 60 F254, Merck) by a 2-dimensional TLC system. In the 1st and 2nd dimensions, a solvent mixture of chloroform/ methanol/water (65:25:4) and chloroform/methanol/acetic acid/water (80:12:15:4) was used, respectively. The lipids were visualized as spots by applying ethanolic molybdophosphoric acid (5 % w/v) on the TLC plate and heat drying. The polar lipids were further distinguished into specific groups as follows:

i) Dragendorff's reagent: It was prepared as follows; solution A, 1.7 g of bismuth nitrate (basic), was added to 100 ml of acetic acid, and H_2O (1:4); solution B, 40 g of potassium iodide to be dissolved in H_2O. Finally, the mixture of solution A (5 ml), solution B (5 ml), acetic acid (20 ml) and H_2O (70 ml) was used for spraying on plates. It was used to differentiate the choline-containing lipid molecules (developed as orange spots).

ii) Ninhydrin reagent: 0.1% (w/v) of ethanolic ninhydrin was used as a reagent for spraying, followed by heating. The development of pink spots on the plates suggested the presence of amino-containing lipids.

iii) α-Naphthol reagent: 3.2 % (w/v) of the α-naphthol compound in a solvent mixture of methanol, H_2O, and H_2SO_4 (25:1.5: 3 v/v/v) was used for spraying. If pink to purple spots were formed on heat drying, it showed glycolipids' presence.

iv) Molybdenum blue reagent: It was prepared as follows: solution A, 40.1 g of molybdenum trioxide was added to 1 L of 25 N H_2SO_4; solution B, 1.78 g of powdered molybdenum was added to 0.5 L of 25 N H_2SO and slowly boiled for 20 min. The solutions were stored in reagent bottles after cooling. Finally, the reagent mixture of solution A (20 ml), solution B (20 ml), and distilled water (40 ml) was prepared and used for spraying. Blue spots, if developed, suggested the presence of phosphate-containing lipids.

2.7.4.2. Cellular fatty acid profiling

Cellular fatty was methylated, segregated, and identified by gas chromatography as per the specifications of the Microbial Identification System (Microbial ID; midi version 6.0 version; method, rtsba6; www.midi-inc.com). It was outsourced to Royal Life sciences Pvt. Ltd, Secunderabad, India.

2.7.4.3. Quinone analysis

Cultures grown in NB for 72 hours aerobically were centrifuged. To 1g of fresh pellet, 10 ml of acetone, and 2 ml of distilled autoclaved H_2O. The mixture was sonicated (50% power) for one hour, followed by the addition of 180 ml acetone. The mixture was further incubated overnight at 4 °C (maximum extraction). The solution was centrifuged at 1000 g for 20 minutes maintained at 4 °C. Using the Whatman filter paper, the mixture was filtered and evaporated with a rotary flash evaporator. To the dried filtrate, 180 ml of n-hexane and 90 ml of autoclaved distilled water were added and vortexed well (20 min at 1000 g at room temperature). After which, the solution was allowed to phase out into 2 layers for 20-30 min. n-Hexane layer was gently poured, and Na_2SO_4 was added to eliminate the remanent water. Finally, the n-hexane layer was dried, and 0.5 ml of ethanol was added to dissolve the quinone and analysed. It was then analysed as per the procedure given by Xie and Yokota (2003).

2.7.4.4. Cell wall amino acid analysis

As described, cell wall amino acids were analysed (McKerrow et al. 2000; Schleifer 1985). The fresh pellet was washed with 50 mM phosphate (pH 7.2), suspended in phosphate buffer, and sonicated at 50 % power for 20 minutes at 4 °C. For 5 times, the severed cell pellets were washed with phosphate buffer and each time centrifuged (8000 g at 4 °C for 10 min) at room temperature. It was then followed by washing pellets with autoclaved milli-Q water 3 times. The total lysate was centrifuged for 60 min at 10000 g for collecting the disrupted cell wall. Further, the cell supernatant was gathered and centrifuged (60 min at 100 g). The cell wall remanent was solvated in 2-3 ml of 12 N HCl, hydrolysed (12 hours autoclaved), and transferred into a beaker to evaporate HCl (hot-plate). Finally, the remainder was dissolved in 1 ml of sterile milli-Q water and neutralized (NaOH). The cell wall hydrolysed amino acid, and standard amino acids were derivatized

concurrently by the o-phthaldialdehyde (OPA) reagent. OPA reagent was prepared as follows:

i) 9 mg sodium tetraborate (HiMedia) dissolved in 4.5 ml H_2O with gentle heating

ii) 66 mg of N-acetyl-L-cysteine and 27 mg of o-phthaldehyde (27mg) (Sigma Aldrich), both dissolved in 0.5 ml methanol, then added in (i)

Finally, amino acids were derivatized with an OPA reagent (1:1 ratio), followed by filtration (0.2 µl filter, Icon pall supor@200). In HPLC, the derivatized samples were analysed with C_{18} column, phosphate buffer (50mM pH 7.2): acetonitrile (ratio 85:15) as the mobile phase, with a flow rate of 0.4 ml/min (UV array of detector 334nm).

2.7.4.5. Polyamine analysis

According to Dion and Herbst (1970), polyamines from cultures were extracted. Cultures at the exponential phase were harvested by centrifugation (10000 g at 4 °C for 10 min), and the pellet was washed (3x) with 50 mM phosphate buffer (pH 7.5). The fresh pellet was added in 400-600 µl of 0.2 M perchloric acid and sonicated (50% power for 15 min). The disrupted cells were neutralised with 1 M of $NaHCO_3$ and centrifuged at 4 °C for 10 min. Polyamines in the supernatant were derivatized with dansyl chloride (3 $mgml^{-1}$), whereas free dansyl chloride was removed by adding proline (50 mM). Then, it was extracted with toluene (3 times) and dried (rotary flash evaporator). It was finally dissolved in methanol (HPLC grade). Polyamines were determined by HPLC equipped with a photodiode array detector (PDA), C_{18} Luna column (5 µm, 250×4.6 mm). A linear gradient system was applied to separate polyamines. Solvent A (acetic acid in water 1 % v/v) and solvent B [acetonitrile (ACN): methanol at ratio 4:1] was used as a mobile phase. The flow rate of 0.8 ml/min was used. The presence of polyamines was demonstrated by the peak absorption of standards and retention time, simultaneous with the spiking using internal standards.

2.7.4.6. Hopanoid analysis

For hopanoid extraction, cultures were grown aerobically at 30±2 °C for 4-5 days (1L) and harvested by centrifugation (10000 g for 15 min at 4 °C). The cells were washed with phosphate buffer (0.05 mM, pH 7.2), kept overnight at -80 °C, and lyophilised. The cell pellets were suspended in 10 ml of methanol/dichloromethane (DCM)/water (10:5:4) and sonicated (50 % power) for 15 min at room temperature; repeated twice. For this pooled supernatant, 10 ml of DCM and water were added, mixed, and centrifuged (identical condition as mentioned); the DCM phase was taken at this step (twice). The total hopanoid extract was dried under the rotary evaporator (Heidolph) and dissolved in 0.5 ml of DCM (HPLC grade). Finally, hopanoids were determined by 2-dimension TLC wherein DCM (30 ml) was used as a mobile phase in both directions. R_f values of known hopanoid molecules were considered for identification.

2.7.4.7. Identification of indoles

The culture was aerobically cultured in the presence/absence of tryptophan at 1mM concentration. Cultures were harvested after reaching the stationary phase by centrifugation (10000 g at 4 °C for 10 min), followed by acidification (pH 2) with 5 N HCl. Supernatants collected were extracted (3 times) with ethyl acetate and pooled. The pooled extracts were evaporated using a rotary flash evaporator (LABOROTA 4001, Heidolph) under a vacuum. Finally, the dehydrated extract was dissolved in 0.5 ml methanol (HPLC-grade), filtered with a 0.22 µm membrane, and tested for indoles using a para-diaminobenzaldehyde reagent. For HPLC analysis Gradient technique was applied to separate metabolites for 30 min with mobile phase as acetic acid (1 % v/v) and acetonitrile (ACN). 0-55 % of ACN for 25 min with step increase to 100 % ACN kept for 3 min and finally returned to 1 % wash by acetic acid. HPLC (Shimadzu, LC-20AT), C_{18} column (5 µm, 250×4.6 mm), with a flow rate of 1.5 ml/min, 20 µl as injection volume and detected at 230, 280, and 350 nm.

2.7.4.8. Pigment profiling

2.7.4.8.1. Whole cell absorption spectrum

The whole cell absorption of the bacterial cells was carried out as given by Truper and Pfennig in 1981. 5 ml of culture grown aerobically (30±2 °C for 3-4 days) was centrifuged at 10000 g at 4 °C. The pellet was dissolved in 3.5 ml of 5 % (w/v) sucrose solution (HiMedia) and mixed well. Finally, the absorption spectrum was measured from 300-1100 nm on a spectrophotometer (Spectronic Genesys 2), sucrose solution without inoculation was used as a blank.

2.7.4.8.2. Carotenoid identification

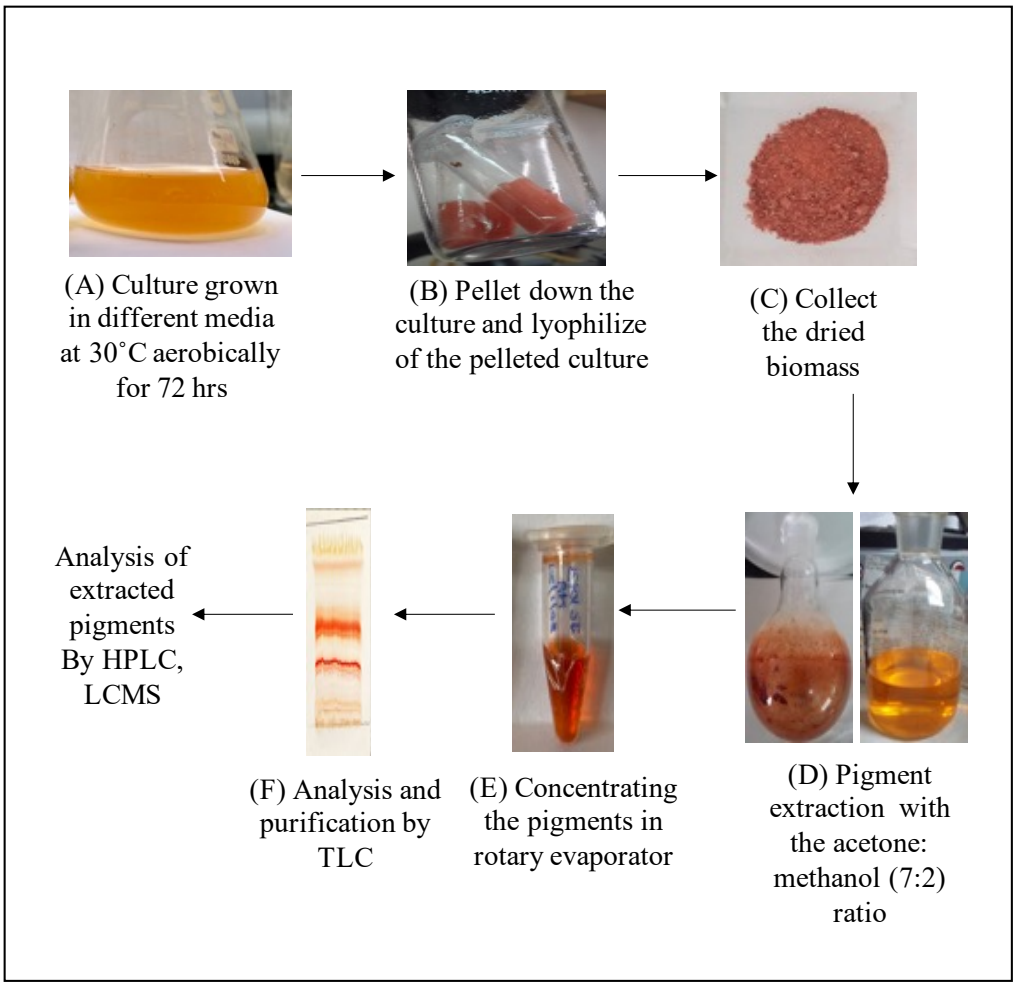


Fig. 5. Extraction procedure of the carotenoid pigments from a bacterial strain

i) Culturing and harvesting: For primary inoculum, a loop of bacterial cells inoculated into 100 ml of the BP media at pH 7.5, followed by incubation with shaking at 30 °C and 140 rpm for 72 hrs. 10 % inoculum from primary culture were added to 1L of NB, ZM, and BP media to check the increased biomass production. Each media was supplemented with 1% (w/v) NaCl. The pH of the media was maintained in the range of 7-8, and the temperature was maintained at 30±2 °C. The culture was incubated aerobically for 72 hrs at 30 °C. After the incubation period, the cells were pelleted by centrifugation (10000 g for 10 min) and washed with saline (PBS buffer, 50 mM) (Fig. 5A, B).

i) Extraction: The lyophilized bacterial biomass (10 g) (Fig. 5C) was used for extraction of the pigments with a solvent mixture of acetone and methanol (7:2, v/v), incubated with 0.1 M NaOH overnight and followed by sonication and centrifugation (10000 g for 10 min) for extraction (thrice) of secondary metabolites and pigments (Fig. 5D). The supernatant was used as a source of pigments and was dried. It was finally dissolved in 500 ml of methanol (HPLC grade) (Fig. 5E).

ii) Analysis: This previous step yielded crude secondary metabolites (carotenoid compounds). Carotenoids were determined by HPLC (LC-20AT Shimadzu, Japan) with a PDA detector using the C_{18} column (Luna, 5 µm, 250×4.6 mm). Pigments were separated using an isocratic solvent system of ACN, methanol, and tetrahydrofuran (58:35:7 v/v/v) as a mobile phase at a flow rate of 1 ml min^{-1}. The compounds were determined at 450 nm and in between the range of wavelengths of 300-800 nm. Further, the crude fraction was analysed by LC-MS (outsourced to CSIR-IICT, Hyderabad). LC-MS was used for analysis and was performed on EXACTIVE Orbitrap high-resolution mass spectrometry system (Thermo) equipped with Accela 600 UPLC system. The carotenoid pigments obtained were

further purified by the thin layer chromatography using a solvent system of chloroform and methanol (8:2) (Fig. 5F). The methanolic extracts of each fraction were centrifuged, filtered, and used for mass spectrometry analysis (Dr. Reddy's Institute of Life Sciences).

2.7.5. Genotypic studies

2.7.5.1. Marker based phylogenetic studies

2.7.5.1.1. 16S rRNA gene amplification and identification

i) Primers used: The following primers were utilized for 16S rRNA gene amplification.

Name	Primer Orientation	Sequence 5' to 3'
F8	Forward	AGAGTTTGATCCTGGCTCAG
F27	Forward	GTTTGATCCTGGCTCAG
R1492	Reverse	GGTTACCTTGTTACGACT
R1525	Reverse	AGAAAGGAGGTGATCCAGCC

ii) 16S rRNA gene amplification and identification protocol: A single isolated colony was added to 20 µl of milli-Q water (200 µl PCR tube) and denatured at 95 °C for 20 min in a PCR machine (BIO-RAD MJ Mini). A combination of one of the following primer sets of F8/R1492, F8/R1525, F27/R1492, F27/1525 was used to amplify the 16S rRNA gene (Lane et al. 1985; Liesack et al. 1991). A master mix from TAKARA (EmeraldAmp® GT PCR) was used for the amplification. The reaction mixture was formulated as follows:

Component	Volume (µl)
Master Mix	25
Forward Primer	2
Reverse Primer	2
Milli-Q	16
Template DNA*	5

*from denatured colony/ genomic DNA isolated

The following conditions were pursued for 16S rRNA gene amplification:

Step	Time (min)	Temp (°C)
Initial denaturation	10	94
Denaturation	1	94
Annealing	1	50-58
Elongation	1.5	72
	No. of cycles: 35	
Final elongation	10	72

The 16S rRNA gene amplified product was checked on 1 % agarose gel (made in TAE buffer) and DNA marker (1Kb, Thermo Scientific) at 50 volts for 30 min. The bands were visualised by ethidium bromide staining and observed under a UV transilluminator (Universal Hood II, BIO-RAD). The selected amplified band was further extracted and purified by the SureExtract PCR kit (Genetix). The purified PCR product was then sourced to Agrigenome Labs Pvt. Ltd. (Kerela) for sequencing. The 16S rRNA gene was sequenced using ABI 3730 XL DNA Analyser automated sequencer. The raw sequences (.ab1 format) derived were assembled by implementing DNA Lasergee SeqMan Pro software. The contig assembled (1300-1500 bp) sequence identity was searched by BLAST algorithm in the EzBioCloud database (Yoon et al. 2017). The multiple sequences were aligned using MUSCLE in MEGA X (Edgar 2004), whereas the pairwise nucleotide sequence alignment tool for taxonomy (www.ezbiocloud.net/tools/pairAlign) was used for the determination of pairwise gene identity.

2.7.5.1.2. Phylogenetic studies

The phylogeny of the taxa under study was determined with MEGA7/MEGA X software (Kumar et al. 2016, 2018). The 16S rRNA gene sequences of all strains under consideration were derived from NCBI and EzBioCloud databases. The sequences were submitted for multiple sequence alignments using MUSCLE algorithms integrated into MEGA7 or MEGA X. Appropriate nucleotide/amino acid substitution models were determined using MEGA7 software under the lowest Akaike information criterion (AIC).

Kimura two-parameter was used for computing the distances in a pairwise deletion manner (Kimura, 1980). For the construction of the phylogenetic tree, neighbor-joining (NJ) was applied along with maximum parsimony (MP) and maximum likelihood (ML) in the MEGA 7/X software. As given by Felsenstein (1985), the percentage support values were derived by implementing the bootstrap procedure.

2.7.5.2. Genome based phylogenetic studies

2.7.5.2.1. Genomic DNA isolation

Both conventional and kit methods were used for isolating DNA from axenic cultures. For the conventional method, 200 ml cultures were centrifuged (10000 g at 4 °C for 10 min), and the pellet was washed thrice with milli-Q water. 500 µl of lysis buffer was added, and incubated at 60 °C for 1 hour. Lysis buffer consists as follows: 60 µl SDS (10 %), 10 µl protease K (20 mg ml^{-1}), 10 µl lysozyme (0.5 mg ml^{-1}) in 420 µl of TE buffer (pH 8). An equal volume of tris buffer saturated with phenol, chloroform, and isoamyl alcohol (25:24:1) was added and incubated at 37 °C for 30 min. The mixture was centrifuged at 8000 g for 15 min at room temperature. An aqueous layer was transferred to the fresh tube, 0.1 ml sodium acetate (pH 4.8), and 1 ml of chilled ethanol (100 %) was added, followed by incubation for 30 min at room temperature. After incubation, it was centrifuged for 10 min at 8000 g at 4 °C, and the pellet was washed with ethanol (70 % v/v). The pellet was air-dried and finally dissolved in the TE buffer and stored at -20 °C. As by the kit method, genomic DNA was extracted using Nucleopore gDNA Fungal/Bacterial mini kit (Genetix) as per the manufacturer's protocol.

2.7.5.2.2. Genome sequencing and assembly

The genomic DNA isolated from different strains was sent to M/s. Agrigenome Labs Pvt. Ltd (Kochi) and BGI (GCM 10 K type strain sequencing project) in China (Wu and Ma 2019) for sequencing. DNA samples in M/s. Agrigenome labs were sequenced

employing Illumina HiSeq 2500 platform, resulting in paired-end sequencing of the fragment library (100×2 bp), resulting in about 8.4×106 quality filter reads. The genomes were assembled using the velvet software quast 4.0. Whereas in BGI, genomes were sequenced using the Illumina XTen platform. The assembly of the reads was done using the SOAPdenovo v1.05 software (Luo et al. 2012).

2.7.5.2.3. Multilocus sequence analysis

For the MLSA analysis, the complete length sequences of nucleotide and amino acids of 8 protein-coding housekeeping genes (*atpD, dnaK, gryB, recA, rpoA, rpoB, rpoD, tryB*) were retrieved from the genomes present (NCBI database) of the members of the taxa under study, concatenated and aligned by MUSCLE program. For the 16S rRNA gene, phylogenetic trees (ML) were constructed for the MLSA. The concatenated gene sequence identity values between strains were also calculated.

2.7.5.2.4. ΔT_m determination

DNA-DNA relatedness (ΔT_m) between the bacterial species was studied by determining the thermal denaturation temperatures of hybrid and homologous genomic DNA as given by Gonzalez and Saiz-Jimenez (2005). DNA before experimentation was adjusted to the same concentration (0.5 µg each), and thermal denaturation was carried out in the 0.1 sodium saline citrate (SSC) buffer. The quantification of thermal denaturation was done (Eppendorf Thermal cycler). The thermal condition for denaturation was at 95 °C for 60 min, with a progressive drop by 10 °C until it reached 15 °C. The hybrid and homologous DNA were stored at 4 °C. Prior to thermal denaturation, SYBR Green solution (1:100000 final conc.) was added to the PCR plate as duplicates. The program was as follows, 25 °C for 10 min, followed by an increase from 25-99 °C at 0.2 °C/s. During this ramp, fluorescence was measured at each step, and melting curves were calculated as temperature corresponding to a 50 % decrease in fluorescence.

2.7.5.2.5. *in-silico*/*digital* DNA-DNA hybridization

Traditional DNA-DNA hybridisation has been replaced by *in-silico*/digital DNA-DNA hybridisation based on genome sequences of the species. The following procedures are pursued firstly using the tool of BLAST. The pair of genomes of X and Y are aligned, resulting in the yield of inter-genomic high-scoring segment pairs (HSPs). After which, the HSPs data are converted to a single genome to genome distance, followed by the conversion of obtained distance into percentage-wise identity s(*d*). It is carried out by utilizing the following distance formula: $s(d) := m*d + c$; where d=obtained distance values; value for slope (m) and intercept (c) obtained by robust line fitting procedure (Auch et al. 2010). In our study, DDH values were calculated from Genome-to-Genome Distance Calculator (GGDC) 3.0 (http://ggdc.dsmz.de/ggdc.php) (Meier-Kolthoff et al. 2022).

2.7.5.2.6. Average nucleotide identity

The ANI calculator (http://enve-omics.ce.gatech.edu/ani/) from Kostas Lab was used to calculate ANI values between two organisms (Rodriguez and Konstantinidis 2014). Similarly, ANI values for larger datasets were calculated by the OrthoANI algorithm, which employs USEARCH (instead of BLAST) for improved iteration (Yoon et al. 2017).

2.7.5.2.7. Average amino acid identity

The AAI calculator from Kostas lab was used to estimate average amino acid identity values between the given two species (http://enve-omics.ce.gatech.edu/aai/) (Rodriguez and Konstantinidis 2014).

2.7.5.2.8. Percentage of conserved proteins

The first step of analysis in POCP would be to calculate sequence identity and identification of conserved regions (by BLASTP) (Qin et al. 2014). The query genome proteins were viewed as conserved if the BLAST result showed an E value <1e-5 (sequence match of >40 %) and more than 50 % alignment part. Finally, the POCP value was

calculated by the formula of [(C1+C2)/(T1+T2)] *100%, C1 and C2 stand for the conserved number of proteins, and T1 and T2 represent the total of proteins in a pair of genomes under study.

2.7.5.2.9. UBCG based phylogenomic tree

Up-to-date bacterial core gene (UBCG) tool is programmed by Chun's lab for the phylogenomic tree using genome sequences of organisms (Na et al. 2018). It involves three primary steps, i) identification of the core genes, ii) multiple alignments of the sequences, and iii) visualization of the phylogenetic tree. The prodigal program predicts coding DNA sequences (CDS) from the whole genome sequence. Based on the CDS identified, hidden Markov models (HMM) searches identify the marker genes/conserved genes in the dataset. As in the second step, each gene sequence is aligned by the MAFFT program, after which the alignments are concatenated, and gaps are filtered containing positions. The final step involves the inference of a phylogenetic tree for each gene and concatenated (92 core genes) sequences and is called as UBCG tree (Fast-Tree/RAxML). MEGA X was used for tree construction with the FASTA-generated alignment files.

2.7.5.2.10. Pan-genome analysis

Pan-genome studies were carried out by implementing the Bacterial Pan Genome Analysis tool (BPGA), a mega-fast pan-genome analysis tool (Chaudhari et al. 2016). Here, first step involves pre-processing raw files into a single input file for clustering (second step). Clustering of genes is based on sequence identity into orthologous clusters by USEARCH, followed by pan-matrix generation. The binary matrix is generated based on the absence/presence of clusters. The fourth step consists of pan-genome profiling, which computes genes shared after the stepwise addition of each genome. It is plotted as core/pan-genome profile curves. It presents users with statistical data on pan-genome (genome wisecore, unique, accessory, and exclusively absent gene numbers and phylogeny

construction (based on concatenated core genes, pan-matrix data), pathway, and function analysis.

2.7.5.2.11. Conserved signature indels

The conserved signature indels (CSIs) in protein sequences discreet to each of the proposed clades were carried out as described (Gupta 1998; Naushad et al. 2014). The proteins were BLASTP searched against the NCBI database, and the proteins with a high scores (E values, <1e -20) were retrieved in all the members from the given clade and other bacterial taxa. The alignment of these protein sequences was employed by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and was examined for the presence of signature indels (insertion/deletion).

2.4.5.2.12. Genome annotation and comparison

Genomes were annotated using various platforms like RAST Server, NCBI, KBase, and PATRIC bioinformatics tools. The annotation data for each species under the study were downloaded from the respective servers as excel files and studied for the following:

i) Orthologous-based gene content comparison: This is the simplest form of comparison for a set of genomes as to which genomes have what genes in families (Setubal et al. 2018). It has been represented as absence/presence gene matrix or venn diagram with the help of OrthoVenn2 online software (Xu et al. 2019).

ii) Whole genome alignments: A method to compare genomes is by alignment, wherein a query genome is used as a reference and is aligned against the given set of genomes. Mauve (genome alignment) built into the PATRIC database was used to align the genomes to identify multiple maximal matches and local collinear blocks (LCBs).

iii) Function-based gene content comparison: Further, gene comparison between genomes based on functions can be achieved if protein description systems like

subsystems (Overbeek et al. 2015) and COG (Galperin et al. 2015) are used during gene annotation. The structural families of related carbohydrate-binding and catalytic modules were annotated by the dbCAN meta-server (http://bcb.unl.edu/dbCAN2).

iv) Secondary metabolite biosynthetic gene analysis: For the elucidation of secondary metabolites biosynthetic gene clusters (BCGs), antiSMASH 5.0 bacterial version was used (https://antismash.secondarymetabolites.org) (Blin et al. 2019). The BCG families were annotated by the HMM, and the strictness level was set to "relaxed."

v) Virulence factors and pathogen-associated genes: To study the pathogenicity, genomes were submitted in the .gbk format to the IslandViewer 4 website (Bertelli et al. 2017), with *Salmonella enterica* Serovar *Typhimurium* LT2 as the reference genome. Genome sequences (.fas format) were further tested for virulence factors against four *Listeria* taxa using the VirulenceFinder 2.07 by the Center for Genomic Epidemiology (CGE) against 4 genera of *Escherichia*, *Listeria*, *Staphylococcus,* and *Enterococcus*. The ID % between the input and the database matching gene was set at 90 %, with a minimum length of 60 %.

2.7.5.3. Genome sequences used

The strains of different taxa (genus *Paracoccus*, *Mesobacillus*, *Microbacterium*, *Roseomonas*, *Alcanivorax,* metagenome assisted genomes) were downloaded from various databases like NCBI, and IMG to be used for the study. The names of the microorganisms, accession numbers, and sources of isolation have been presented in Table 2.

Table 2. Microorganisms' name, strain number, source of isolation, and accession number

Species	Isolation Source	Genome accession no.
Strain JC501^T	Orchid roots	SELD00000000
Paracoccus marinus NBRC 100637^T	Coastal seawater	VJYZ00000000
P. contaminans LMG 29738^T	Airborne contaminant	CP020612
P. chinensis KS-11^T	Sediment	FNGE00000000
P. solventivorans DSM 6637^T	Soil	FRCK00000000
P. alkenifer DSM 11593^T	Biofilter waste treatment	FNXG00000000
P.sediminis DSM 26170^T	Marine sediment	FZNM00000000
Rhodobacter capsulatus DSM 1710^T	Sewage plant	QKZO00000000
Strain JC1013^T	Salt pan	JAAVUM000000000
Mesobacillus selenatarsenatis JCM 14380^T	Contaminated sediment	BASE00000000
M. subterraneus ATCC 136^T	Thermal aquifer	JXIQ00000000
M. jeotgali KCCM 41040^T	Fermented seafood	CP025025
M. boroniphilus DSM 17376^T	Soil	BAUW00000000
Mesobacillus persicus B48^T	Hypersaline lake	FOBW01000039
Mesobacillus foraminis CV53^T	Non-saline alkaline groundwater	CP033044
Mesobacillus campisalis SA2-6^T	Solar saltern	LAYY01000001
Mesobacillus mediterraneensis Marseille-P2366	Human gut microbiota	FOJL01000001
Mesobacillus zeae JJ-247^T	Maize rhizosphere	QWVT01000008
Bacillus mycoides ATCC 6462^T	Type material	CP009692
Niallia circulans NBRC 13626^T	Type material	BCVE00000000
Bacillus tropicus LM1212-W3^T	Marine sediments	CP041071
Robertmurraya siralis 171 544^T	Silage	SWLZ01000001
Bacillus halotolerans ATCC 25096^T	Type material	LPVF01000001
Bacillus badius NBRC 15713^T	Type material	BCVF01000001
Bacillus cereus ATCC 14579^T	Type material	CP004722
Neobacillus drentensis NBRC 102427^T	Grassland	BCUX01000001)
Neobacillus cucumis V32-6	Environment	PGVE01000001
Cytobacillus oceanisediminis 2691	Sediment of ocean	CP015506
Neobacillus drentensis NBRC 102427^T	Grassland	BCUX01000001
Neobacillus vireti DSM 15602^T	Grassland	(LDNB01000001)
Geobacillus galactosidasius DSM 18751^T	Compost	(NDYL01000001)
Roseomonas sp. KCTC 32190^T	Beach sand	JABBKX00000000
Roseomonas oryzicola KCTC 22478 ^{T}	Rhizosphere of rice	JAAVUP00000000
Roseomonas lacus CGMCC 1.3617 ^{T}	Freshwater lake sediment	BMKW00000000
Roseomonas alkaliterrae DSM 25895 ^{T}	Geothermal soil	JACIJE00000000
Roseomonas stagni DSM 19981^T	Sediment of pond water	FOSQ00000000
Roseomonas frigidaquae JCM 15073^T	Water cooling system	JAAVTX00000000
Roseomonas frigidaquae JCM	Water cooling system	JAATJR000000000
Roseomonas selenitidurans BU-1 ^{T}	Urban soil	JAAVNE00000000
Roseomonas algicola PeD5 ^{T}	Green alga	JAAIKB000000000
Roseomonas bella CQN31 ^{T}	Lake sediment	QGNA000000000
Roseomonas gilardii subsp. *rosea* ATCC BAA-691^T	Blood	JADY00000000
Roseomonas gilardii subsp. *rosea* NCTC13290 ^{T}	Blood	UGVO01000005
Roseomonas gilardii U14-5	Water	CP015583
Roseomonas aerilata DSM 19363 ^{T}	Air sample	JONP00000000
Roseomonas deserti M3 ^{T}	Oil contaminated desert soil	MLCO00000000
Roseomonas rosea DSM 14916 ^{T}	Indoor building material	FQZF00000000
Roseomonas mucosa NCTC 13291 ^{T}	Blood	UGVN00000000
Roseomonas vastitatis CPCC 101021 ^{T}	Badain desert	QXGS00000000
Roseomonas oryzae KCTC 42542 ^{T}	Rhizosphere soil	VUKA00000000

Species	Isolation Source	Genome accession no.
Roseomonas cervicalis ATCC 49957 ^{T}	Blood	ADVL00000000
Roseomonas rhizosphaerae YW11 ^{T}	Soil	PDNU00000000
Roseomonas aestuarii JR1/69-1-13	Marine water sample	PDOA00000000
Roseomonas aerophila NBRC 108923 ^{T}	Air	JACTVA00000000
Roseomonas ludipueritiae DSM 14915 ^{T}	Indoor building material	JACTUZ00000000
Roseomonas coralli M0104 ^{T}	Gorgonian Coral	SNVJ00000000
Roseomonas pecuniae N75 ^{T}	Surface of coin	JACIJD00000000
Roseomonas wenyumeiae Z23^{T}	Feces of Tibetan antelope	RFLX00000000
Alcanivorax pacificus W11-5^{T}	CP004387	Deep Sea
Alcanivorax indicus SW127^{T}	QGMP00000000	Sea water
Alcanivorax hongdengensis A-11-3^{T}	AMRJ000000000	Surface seawater
Alcanivorax profundi MTEO17^{T}	QYYA00000000	Deep seawater
Alcanivorax nanhaiticus 19-m-6^{T}	ARXV00000000	Deep sea sediment
Alcanivorax sediminis PA15-N-34^{T}	WIRE00000000	Deep sea sediment
Alcanivorax jadensis T9^{T}	ARXU00000000	Intertidal sediment
Alcanivorax borkumensis SK2^{T}	AM286690	Sea water/sediment
Alcanivorax xenomutans JC109^{T}	OBMO00000000	Shrimp pond
Alcanivorax dieselolei B5^{T}	CP003466	Seawater/sediment
A. dieselolei CGMCC1.3690^{T}	BMKY00000000	Seawater/sediment
Alcanivorax gelatiniphagus MEBiC08158^{T}	VCQT00000000	Tidal flat sediments
Alcanivorax profundamaris ST75FaO-1^{T}	JABJWH00000000	Deep sea water
Alcanivorax venustensis ISO4^{T}	ARXR00000000	Deep sea sediment
Ketobacter alkanivorans GI5^{T}	CP022684	Sea water
Acetobacter aceti NBRC 14818^{T}	Vinegar	BAMU01000001
Acetobacter ascendens LMG 1590 ^{T}	Wine	CP015164
Acetobacter cerevisiae LMG 1625 ^{T}	Beer	LHZA01000001
Acetobacter cibinongensis 4H-1 ^{T}	Fruit and curd	BAMV01000001
Acetobacter ghanensis LMG 23848T ^{T}	Fermented cocoa	LN609302
Acetobacter indonesiensis 5H-1 ^{T}	Fermented products	BAMW01000001
Acetobacter malorum LMG 1746 ^{T}	apple	LHZC01000001
Acetobacter nitrogenifigens DSM 23921^{T}	Tea	AUBI01000001
Acetobacter okinawensis JCM 25146 ^{T}	Sugarcane	BAJU01000127
Acetobacter orientalis 21F-2^{T}	Fermented food	BAMX01000001
Acetobacter orleanensis JCM 7639^{T}	Vinegar	BAMY01000001
Acetobacter oryzifermentans dm	Fermented rice	CP022374
Acetobacter pasteurianus Ab3	Vinegar	CP012111
Acetobacter peroxydans NBRC 13755	Ditch water	BJMV01000001
Acetobacter persici JCM 25330^{T}	Peach	BAJW01000236
Acetobacter senegalensis LMG 23690 ^{T}	Mango	LHZU01000001
Acetobacter syzygii 9H-2^{T}	fruit	BAMZ01000000
Acetobacter tropicalis LMG 19825^{T}	coconut	LHZQ01000001
Acetobacter fabarum KR	Cocoa beans	NCXK01000001
Acetobacter sp. 46_36	Human gut	MNTT00000000
Acetobacter sp. DmW_043	Kitchen fruit fly	JOMN01000006
Acetobacter sp. DsW_063	Garden fruit fly	JOPL01000001
Acetobacter sp. UBA5411	mud	DHNR01000022
Asaia astilbis JCM 15831	flowers	BAJT01000049
Asaia bogorensis NBRC 16594^{T}	Flower of orchid	AP014690
Asaia platycodi JCM 25414^{T}	Balloon flower	BAKW01000035
Asaia prunellae JCM 25354^{T}	*Prunella* sp. flowers	BAJV00000000
Ameyamaea chiangmaiensis LMG 27010	Red ginger	JABXXR00000000
Strain AT-5844	Skin	AGEZ00000000
Strain DB1506	Mixed culture	QLIX00000000

Species	Isolation Source	Genome accession no.
Strain UBA6159	Waste water	DIUL00000000
Strain KEBCLARHB70R	*Cladonia arbuscula*	VCDI00000000
Strain PAMC 26569	*Cladonia borealis*	CP053708
Strain SCN 69-10	Thiocyanate bioreactor	MEFR00000000
Strain bog_908	Permafrost soil	PKZM00000000
Acidicaldus organivorans DX-1	Geothermal site	JPYW00000000
Acidiphilium angustum ATCC 35903^{T}	Acid mine	JNJH01000001
Acidiphilium cryptum JF-5 ^{T}	Coal mine sediment	009484
Acidiphilium multivorum AIU301 ^{T}	Mine drainage	015186
Acidiphilium rubrum ATCC 35905^{T}	Mine drainage	FTNE01000074
Acidisphaera rubrifaciens HS-AP3^{T}	Acidic environment	BANB01001239
Acidocella aminolytica DSM 11237^{T}	mines	FQVJ01000109
Acidocella facilis ATCC 35904 ^{T}	mines	JHYG01000001
Acidocella aminolytica DSM 11237 ^{T}	Acidic coal mine drainage	FQVJ00000000
Acidisphaera rubrifaciens HS-AP3 ^{T}	Acidic hot springs and mine	BANB00000000
Acidomonas methanolica NBRC 104435	sludge	BAND01000546
Ameyamaea chiangmaiensis LMG 27010	Flowers of red ginger	JABXXR010000001
Belnapia moabensis DSM 16746	Soil crust	JQKB00000000
Belnapia rosea CPCC 100156	Soil	FMZX00000000
Bombella apis SME1	Bee hive	WHNS01000001
Bombella intestini R-52487^{T}	Bumblebee crop	JATM01000005
Commensalibacter intestini A911^{T}	Drosophila gut	AGFR01000026
Caenispirillum bisanense USBA 140	unspecified	OCNJ01000040
Crenalkalicoccus roseus YIM 78023 ^{T}	Alkaline hot spring	SJDM00000000
Dankookia rubra JCM30602	Sediment of stream	SMSJ00000000
Endobacter medicaginis LMG 26838^{T}	Nodules in an acidic soil	JABXXQ00000000
Eliorae thermophila YIM 72297^{T}	Hot springs	QMDI01000001
Gluconacetobacter aggeris LMG 27801	soil	JABEQD010000001
Granulibacter bethesdensis CGDNIH4^{T}	Lymph node cultures	CP003182
Gluconacetobacter azotocaptans LMG 21311	plant	JABEQF010000001
Gluconacetobacter diazotrophicus PA1 5^{T}	Sugarcane roots	011365
Gluconacetobacter entanii AV429	vinegar	JABJWD010000072
Gluconacetobacter johannae LMG 21312	plant	JABEQH010000001
Gluconacetobacter liquefaciens DSM 5603^{T}	Dried fruit	QQAW01000001
Gluconacetobacter sacchari LMG 19747	plant	JABEQJ010000001
Gluconacetobacter asukensis LMG 27724	stone	JABEQE010000001
Gluconobacter albidus TMW2	water	CP014689
Gluconobacter cerinus NBRC 3267^{T}	cherry	BEWM01000001
Gluconobacter frateurii NBRC 3264^{T}	strawberry	BEWN01000001
Gluconobacter japonicus NBRC 3271^{T}	Fruit of *Myrica* sp.	BEWO01000001
Gluconobacter kanchanaburiensis NBRC 103587^{T}	Spoiled fruit	BJVA01000001
Acetobacter fabarum KR	Cocoa beans	NCXK01000001
Acetobacter sp. 46_36	Human gut	MNTT00000000
Gluconobacter morbifer G707^{T}	Gut of Drosophila	AGQV01000019
Gluconobacter oxydans H24	Industry	019396
Gluconobacter roseus NBRC 3990^{T}	Persimmon fruit	BJLY01000001
Gluconobacter sphaericus NBRC 12467^{T}	Fresh grape	BJMK01000001
Gluconobacter thailandicus HD924	Wheat	CP043043
Granulibacter bethesdensis CGDNIH4^{T}	Lymph node	CP003182
Humitalea rosea DSM 24525^{T}	Soil	QKYU00000000
Komagataeibacter cocois WE7^{T}	Contaminated coconut milk	QEXL01000001
Komagataeibacter diospyri MSKU15	Sapodilla fruit	BDLW01000001
Komagataeibacter europaeus SRCM101446	food	CP021467

Species	Isolation Source	Genome accession no.
Komagataeibacter hansenii ATCC 23769	Vinegar	CM000920
Komagataeibacter intermedius AF2	Kombucha tea	JUFX02000232
Komagataeibacter kakiaceti JCM 25156^T	Fruit vinegar	BAIO01000947
Komagataeibacter maltaceti LMG 1529^T	Vinegar	POTC01000001
Komagataeibacter melomenusus AV436^T	Apple cider vinegar	JABJWC010000001
Komagataeibacter nataicola RZS01	unspecified	CP019875
Komagataeibacter oboediens BPZTR01	Homemade vinegar	CP043481
Komagataeibacter pomaceti AV446	Vinegar fermentation sludge	PRCW01000102
Komagataeibacter rhaeticus ENS	Kombucha tea	CP050139
Komagataeibacter saccharivorans JH1	fruitfly	CP036404
Komagataeibacter sucrofermentans LMG 18788^T	vinegar	NKUA01000082
Komagataeibacter swingsii LMG 22125^T	Italian apple fruit	NKUB01000099
Komagataeibacter xylinus E25	unspecified	CP004360
Kozakia baliensis DSM 14400^T	Palm brown sugar and ragi	CP014674
Neoasaia chiangmaiensis NBRC 101099	Flower of red ginger	CP014691
Neokomagataea tanensis AH13^T	Candle bush	CP032485
Neokomagataea thailandica NBRC 106555^T	Lantana plant	BCZB01000001
Nguyenibacter vanlangensis LMG 31431^T	Rice	JABXXP010000001
Paracraurococcus ruber JCM9931	Soil	SMOA00000000
Parasaccharibacter apium G7_7_3c	Hindgut of bee	CP020554
Strain S9	Oil contaminated soil	RCZP00000000
Strain sp. Z24	Feces of Tibetan antelope	RAQU00000000
Strain AR75	Sediment	STGB00000000
Strain HF4	Sediment	STGD00000000
Strain 18066	Roots of beech	CACSJM010000001
Strain B5	Soil	ALOX01000001
Strain FDAARGOS_362	Blood	CP024588
Strain TAS13	Activated sludge	BDLP01000001
Rhodovarius lipocyclicus CCUG 44693 ^{T}	Industrial hygiene control	JAAABL00000000
Rubritepida flocculans DSM 14296 ^{T}	Hot spring	AUDH00000000
Rhodospirillales bacterium 20-60-12	Mine water	NCAP00000000
Rhodospirillales bacterium 69-11	Ammonium sulfate bioreactor	MKTV00000000
Rhodospirillales bacterium Ga0074136	Water treatment plant	LNEW01000000
Rhodopila globiformis DSM 161	Sulphur spring	NHRY00000000
Rhodovastum atsumiense DSM 21279	Paddy field	VWPK00000000
Rhodospirillales bacterium 70-18	Ammonium sulfate bioreactor	MKWE00000000
Stella vacuolata ATCC 43931^T	Horse manure	AP019702
Stella humosa ATCC 43930 ^{T}	Soil	AP019700
Siccirubricoccus deserti SYSU D8009	Desert sample	JACOMF00000000
Swaminathania salitolerans NBRC 104436	Plant	BJVC01000001
Saccharibacter floricola DSM 15669^T	pollen	KB899333
Swingsia samuiensis AH83^T	Plant	CP038141
Tanticharoenia sakaeratensis NBRC 103193^T	Soil	BALE01000094
Zavarzinia compransoris DSM 1231^T	environment	QGLF00000000
Strain S71-1-4	Superficial marine water	AQOR0000000000
Strain CSSed165cm_249	Sea water	SLJS00000000
Uncultured DS1.3.55	Hyper saline soda lake	PNLU00000000
Strain JB21	Drinking water system*	JAIOUB00000000
UnculturedSRR6193124.bin.10 MetaBAT	Soda alkali saline lake	CAJXLA00000000
Strain DP30	Sea sand	WTUY00000000
Strain P2S70	Beach sand*	AXBZ000000000
Strain MD8A	Marine sediment	AQOQ000000000

Species	Isolation Source	Genome accession no.
Uncultured oa1	Protist cell culture*	CP051240
Uncultured oa2	Protist cell culture*	CP051253
Uncultured HI0013	Sea water*	LWEM00000000
Uncultured AG-333-G18	Marine layer*	CABYMG00000000
Uncultured Bin_5	Sediment core*	JABURH00000000
Uncultured ESRF-bin82	Seawater*	NTLH00000000
Uncultured C13	Deep sea water*	QAZC00000000
Uncultured C17	Deep sea water*	QAZB00000000
Uncultured CPC95	Marine pelagic biome*	NYSM00000000
Strain DSM 26293	Deep water plume	FNUM00000000
Uncultured HT1	Marine sediment*	CP080267
Uncultured HI0003	Sea water*	LWEI00000000
Uncultured HI0007	Sea water*	LWEJ00000000
Uncultured HI0011	Sea water*	LWEL0000000
Uncultured HI0033	Sea water*	LWET00000000
Uncultured HI0035	Sea water*	LWEU00000000
Uncultured ih2	Protist cell culture*	CP051252
Uncultured NORP347	Marine benthic biome*	JAESUS00000000
Uncultured NP94	Marine water*	PBRW00000000
Strain NBRC 102024	seawater	BDAS00000000
Uncultured NP13	Marine water*	PCAQ00000000
Uncultured SRR3933375_bin.7_MetaBAT_v2.12.1_MAG	Marine water*	CAJXUR00000000
Uncultured SP60	Marine water*	PAZZ00000000
Uncultured SP122	Marine pelagic biome*	PAQZ00000000
Uncultured UBA1900	Saline marine water*	DDGO00000000
Uncultured UBA2692	Saline marine water*	DELQ00000000
Uncultured UBA3183	Saline marine water*	DEXN00000000
Unculture UBA9185	Saline marine water*	DNMT00000000
Strain ABS183	Hypersaline sediment	LKAP00000000
Uncultured C15_100_9	Marine seawater*	JADFAH00000000
Uncultured C15_98_8	Marine seawater*	JADFAG00000000
Strain DG881	Dinoflagellate	ABRW00000000
Strain E4	Settlings	CP032351
Uncultured ESRF-bin1	Seawater*	NTLV00000000
Uncultured ESRF-bin2	Seawater*	NTLS00000000
Uncultured ESRF-bin8	Seawater*	NTLD00000000
Uncultured ESRF-bin36	Seawater*	NTLN00000000
Uncultured HI0083	Sea water*	LWFX00000000
Uncultured HI0044	Sea water*	LWEZ00000000
Uncultured NORP25	marine aquifer*	NVWY0000000
Strain NBRC 101098	Sea water	AP014613
Strain NBRC 102028	Sea water	BDAT00000000
Uncultured NP90	Marine seawater*	PBSH00000000
Uncultured SRRR3933283_bin.5_MetaB	Marine water*	CAJXVB00000000
Uncultured SZUA-1036	Hydrothermal plume*	DQLP00000000
Strain VBW004	hydrothermal vent	WNHC00000000
Strain 97CO-6	Yellow sea sediment	PIXQ00000000
Strain 97CO-5	Yellow sea sediment	AZYR00000000
Strain 24	Plastic Marine debris	SNUA00000000
Strain 6-D-6	Superficial sea water	AQPF00000000
Strain KX64203	Deep sea sediment	LVIC00000000
Strain KS-293	Sea water	CZHF00000000
Uncultured MCMED-G36	Marine seawater*	JACETG00000000
Uncultured NORP70	Marine biome*	NVVF00000000

Species	Isolation Source	Genome accession no.
Strain P40	Deep seawater	CP012331
Strain PN-3	Beach sand*	AXBX00000000
Strain 43B_GOM-46	Seawater	JADM00000000
Strain Alg238-V104	Coastal water body	UNRD00000000
Strain N3-2A	Coastal sediment	CP022307
Strain NMRL4	Oil polluted seawater	JADDOL00000000
Uncultured UBA1913	Saline marine water*	DDGB00000000
Uncultured UBA3989	Saline marine water*	DGCB00000000
Uncultured EAC41	Marine pelagic biome	NZRG00000000
Uncultured ERR599032_bin.125_MetaBAT_v2.12.1_MAG	Marine water*	CAJWDJ00000000
Strain FXH-223	Coastal seawater	JAJGNA00000000
Strain IO_7	Seawater	CP053420
Uncultured NP12	Marine pelagic biome*	PCBJ00000000
Strain SY10-13	Marine sediment	JAEKJB00000000
Uncultured SAT63	Marine pelagic biome*	PAKR00000000
Uncultured Sp32	Marine pelagic biome*	PBEO00000000
Uncultured UBA1092	Marine pelagic biome*	DCCB00000000
Uncultured UBA4150	Marine pelagic biome*	DFVW00000000
Uncultured UBA8877	Marine pelagic biome*	DPCU00000000
Strain ZXX171	Seawater	WXVX00000000
Strain 521-1	Marine sediment	ARXX00000000
Uncultured Bin_18	Sediment core*	JABURE00000000
Uncultured CPC11	Marine pelagic biome*	NYXH00000000
Uncultured DT_1	Marine metagenome*	JABSOF00000000
Strain DSM 26295	Deep water plume	FXTF00000000
Uncultured ERR599044_bin.85_meta	Marine water*	CAJWYX00000000
Uncultured ih1	Protist cell culture*	CP051239
Uncultured IN928	Marine pelagic biome*	NZKK00000000
Uncultured MES4	Brine pool sample*	VENH00000000
Uncultured M30B6	Saline lake water*	CP073758
Uncultured MCMED-G37	Marine seawater*	JACETH00000000
Uncultured NORP288	Subseafloor aquifer*	JASESSO00000000
Uncultured NORP281	Subseafloor aquifer*	JAESSI00000000
Uncultured NAT75	Marine water sample*	NZUW00000000
Uncultured NP156	Marine water sample*	PBZB00000000
Uncultured Nap_24	Marine water sample*	LSMO00000000
Uncultured SP64	Marine pelagic biome*	PAZW00000000
Uncultured SZUA-1011	Hydrothermal plume*	DQKV00000000
Uncultured SAT1356	Marine water sample*	PAWL00000000
Uncultured SRR3933184_bin.8_MetaBAT_v2.12.1_MAG	Marine water*	CAJXVB000000000
Uncultured UBA679	Saline marine water*	DBNF00000000
Uncultured UBA1475	Saline marine water*	DCSD00000000
Uncultured UBA1504	Saline marine water*	DCRA00000000
Uncultured UBA1608	Saline marine water*	DCNA000000000
Uncultured UBA1614	Saline marine water*	DCMU00000000
Uncultured UBA1911	Saline marine water*	DDGD00000000
Uncultured UBA1972	Saline marine water*	DDDU00000000
Uncultured UBA2114	Saline marine water*	DCYI00000000
Uncultured UBA2509	Saline marine water*	DDNX00000000
Uncultured UBA2512	Saline marine water*	DDNU00000000
Uncultured UBA2685	Saline marine water*	DELX00000000
Uncultured UBA2694	Saline marine water*	DELO00000000
Uncultured UBA3194	Saline marine water*	DEXC00000000
Uncultured UBA3353	Saline marine water*	DEQZ00000000

Species	Isolation Source	Genome accession no.
Uncultured UBA3585	Saline marine water*	DFMU00000000
Uncultured UBA 4152	Saline marine water*	DFVU00000000
Uncultured UBA4485	Saline marine water*	DGNT00000000
Uncultured UBA4488	Saline marine water*	DGNQ00000000
Uncultured UBA4521	Saline marine water*	DGMJ00000000
Uncultured UBA4527	Saline marine water*	DGMD00000000
Uncultured UBA4543	Saline marine water*	DGLN00000000
Uncultured UBA9369	Saline marine water*	DMUV00000000
Uncultured UBA9453	Saline marine water*	DMEC000000000
Uncultured UBA9462	Saline marine water*	DLVX000000000
Uncultured UBA9466	Saline marine water*	DLVT00000000
Uncultured UBA9685	Saline marine water*	DOIN00000000
Uncultured UBA9692	Saline marine water*	DOAE00000000
Uncultured UBA9874	Saline marine water*	DQAU00000000
Uncultured UBA10233	Saline marine water*	DPIQ00000000
Uncultured UBA10370	Saline marine water*	DOGO00000000
Uncultured UBA10373	Saline marine water*	DOGP00000000
Uncultured UBA10377	Saline marine water*	DOGM00000000
Uncultured UBA10381	Saline marine water*	DNZX00000000
Uncultured UBA10698	Saline marine water*	DQDM00000000
Uncultured UBA10724	Saline marine water*	DMGJ00000000
Uncultured UBA10813	Saline marine water*	DORU00000000
Uncultured UBA10824	Saline marine water*	DPSU00000000
Uncultured UBA11597	Saline marine water*	DPZZ00000000
Uncultured UBA11623	Saline marine water*	DONL00000000
Uncultured UBA12379	Saline marine water*	DPAJ00000000
Uncultured Bin12	Marine water*	PDUM00000000
Ketobacter alkanivorans GI5^{T}	Sea water	CP022684

*Metagenome assembled genomes

RESULTS

3. RESULTS

3.1. Analysis of root associated bacteria of orchids

3.1.1. Sampling of orchids

Roots of orchids (10 species spanning 8 genera) were sampled from four different regions of India [Sikkim, Darjeeling (Himalayan), Orissa (Eastern Ghats), Karnataka (Western Ghats] (Fig. 5). With limited availability, 1-4 roots per specimen were collected. The GPS locations of the sites are given in Fig. 5. The orchids were identified to the genus level with the help of Dr. Jeewan Singh Jalal (Botanical Survey of India) and orchid databases (https://orchidbliss.com/how-to-identify-orchids/). The roots collected were separately processed for metagenome and culture studies.

3.1.2. Metagenome sequencing output and quality control

DNA from 10 different root samples of orchids were successfully isolated with concentrations ranging from 50.3 to 176.9 ng ml^{-1}. The V1-V3 region of 16S rRNA genes was targeted and amplified from these samples. The mean of the library fragment size distribution ranged from 599 bp to 661 bp. The metagenomic sequences obtained were then submitted to NCBI-SRA (Sequence Read Archive), and the accession numbers for each of the samples are given in Table 3. Illumina Miseq500 sequencing generated 2,548,301 raw reads whereas, after quality and length filtering, 200,583 reads remained.

Table 3. Bio-samples, accession numbers and reads generated from each sample used in orchid metagenomic studies

Sample no.	Name of the orchid	Bio-sample no.	Accession no.	Reads/sample
191	*Bulbophyllum* sp. 1	SAMN09829936	SRX4557107	235102
209	*Agrostophyllum* sp.	SAMN09748320	SRX4497023	218422
210	*Cymbidium* sp.	SAMN09829870	SRX4556892	342323
211	*Dendrobium* sp. 1	SAMN0982971	SRX4556891	151468
213	*Dendrobium* sp. 2	SAMN09829935	SRX4557108	401915
215	*Bulbophyllum* sp. 2	SAMN09748319	SRX4497018	255353
218	*Thunia* sp.	SAMN10180266	SRX4802390	200689
265	*Rhynchostylis* sp.	SAMN09812579	SRX4556878	262877
W2	*Aeridis* sp.	SAMN10181218	SRX4802389	207171
W3	*Vanda* sp.	SAMN10250717	SRX4897570	272981

3.1.3. Bacterial composition and diversity analysis

For identifying root associated bacteria of orchids, 97 % sequence identity value was used for obtaining OTUs in which 200,583 reads were grouped into 6389 OTUs spanning 17 phyla. The rare-fraction curves of 10 samples were generated, and the plots levelled off into a plateau (Fig. 6). This suggested that good sampling depth and saturation indicated a good representation of the bacterial community in the study. Their distribution studies were considered at the taxonomic levels of phylum and genus.

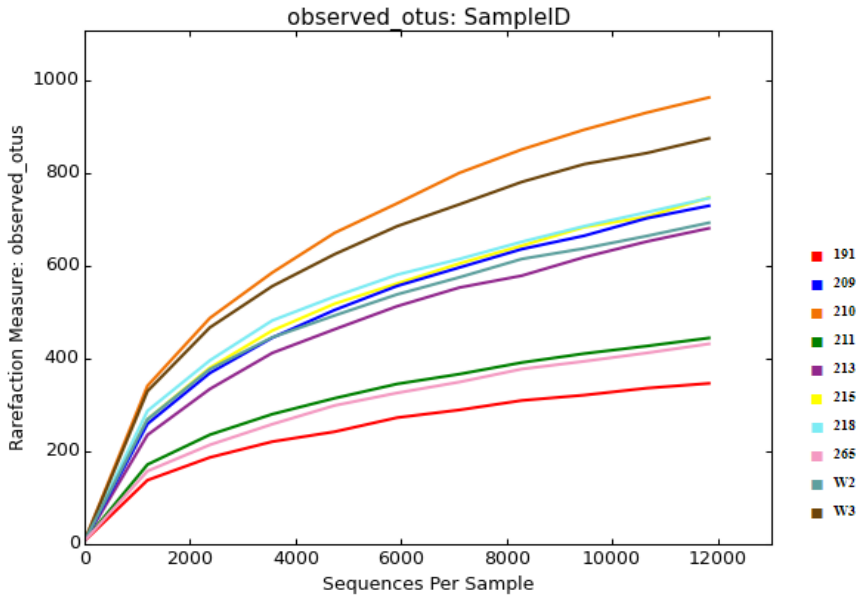


Fig. 6. Rare-fraction curve of the ten orchid samples for V1-V3 region of 16S rRNA gene, clustered at 97 % sequence identity across various samples

The three most abundant phyla were *Proteobacteria* (29 to 42 %), Actinobacteria (8 to 33 %), and Bacteroidetes (2 to 20 %) (Fig. 7). The abundance in this study denotes the percentage of reads in the total OTUs present. Phylum *Acidobacteria, Firmicutes,* and *Planctomycetes* comprised of 7 % of the total OTUs observed. Compared to the other phyla, *Armatimonadetes* (1 %) and *Chloroflexi* (0.7 %) were represented less but distributed in all the samples. Phylum *Proteobacteria* were most abundant in *Dendrobium* sp.2 (42 %), *Bulbophyllum* sp. (41 %), and *Agrostophyllum* sp. (40 %). Similarly, *Actinobacteria* were

dominant in *Dendrobium* sp. and *Rhynchostylis* sp. constituting 33 % each whereas, *Bacteroidetes* were abundant in *Aeridis* sp. (20 %) and *Rhynchostylis* sp. (19 %). Four different "*Candidatus*" phyla, BRC1, OD1, SR1, WS3, and TM7 were also detected in the OTUs. In contrast, phylum *Verrucomicrobia* was less common except in *Dendrobium* sp.2 (0.5 %), *Aeridis* sp. (1.4 %), and *Vanda* sp. (2 %). As for the sequences represented as OTUs below 97 % were not classified by the RDP classifier, hence were labelled as unclassified. In our study, unclassified OTUs at the phylum level consisted of 10 to 20 % of the total OTUs observed. These sequences were of greater interest and therefore extracted for in-depth analysis.

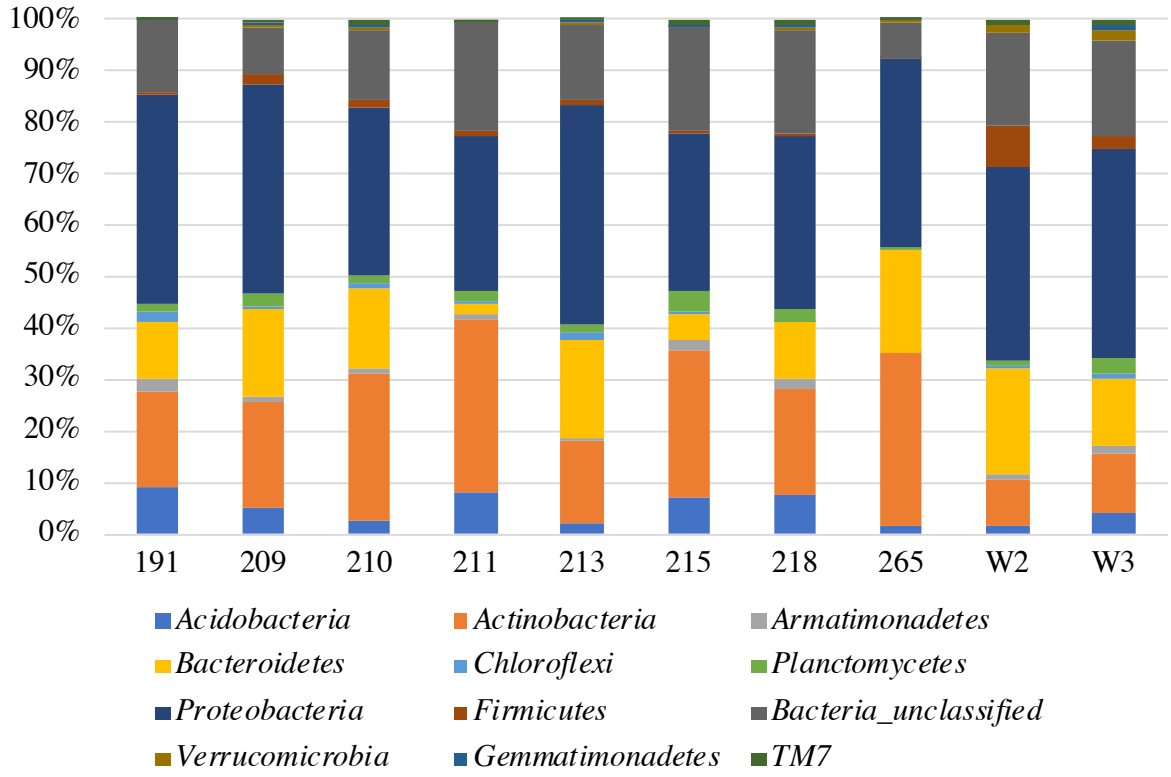


Fig. 7. Relative abundance of bacterial taxa distribution in the orchid root microbiome with pre-dominant bacterial phyla (>1 %) as obtained by 16S rRNA gene amplicon based metagenome analysis

At the taxonomic level of the genus, the distribution in each of the samples is illustrated in Fig. 8. The abundance and distribution of genera based OTUs (>0.1 %) were used for creating heatmaps by a single hierarchical clustering (Fig. 8A). Genera like *Mucilaginibacter* and *Burkholderia* were distributed in all the samples, with abundant in *Bulbophyllum* sp. at 6.6 % and 6.7 %, respectively. Additionally, 13 classified taxa of the genus *Streptomyces* (2.4 %), *Rhizobium* (1.8 %), *Methylobacterium* (0.74 %), *Sphingomonas* (0.5 %), *Solirubrobacter* (0.46 %), *Gemmata* (0.43 %), *Conexibacter* (0.2 %), *Spirosoma* (0.41 %), *Mycobacterium* (0.42 %), *Pseudonocardia* (0.4 %), *Brevundimonas* (0.4 %) *Caulobacter* (0.31 %) and *Actinoplanes* (0.2 %) were present commonly amongst all the samples (Fig. 8A). Unclassified OTUs at the genus level consisted of 50 to 89 % of the total OTUs observed (Fig. 8B). Further, the distribution of common and unique taxa at the genus level for orchid species are presented in the Venn diagram (Fig. 9) and Table 4.

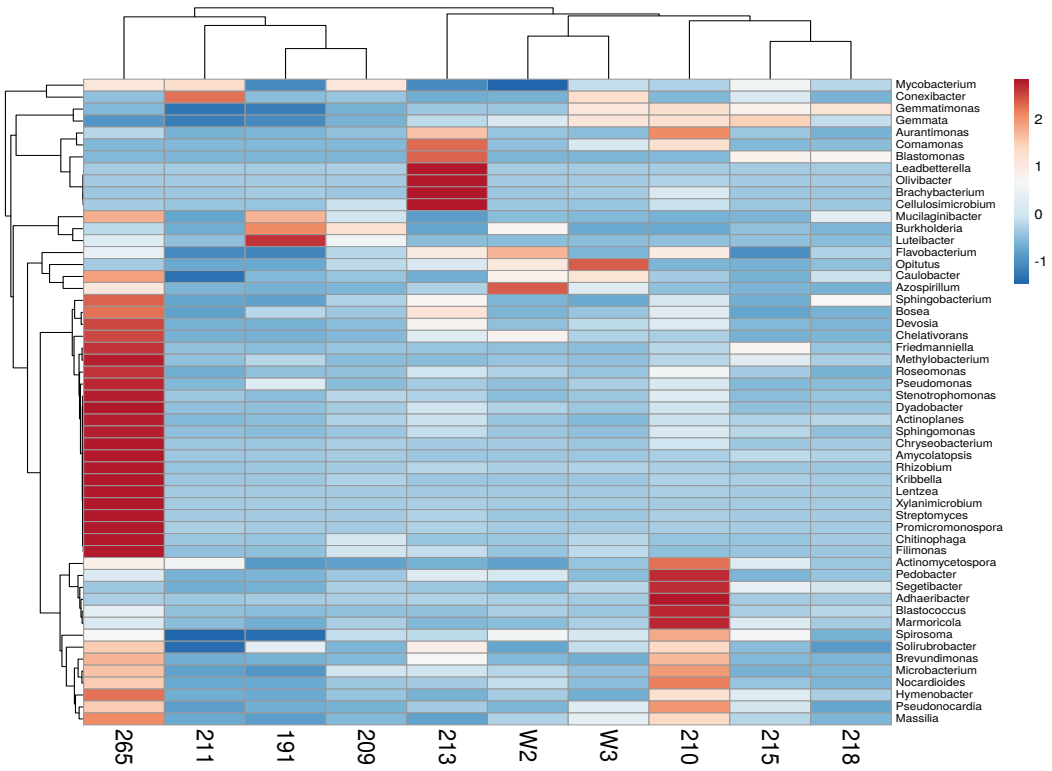


Fig. 8A. Relative abundance of bacterial taxa of orchid root samples showing the pre-dominant bacterial genera (>0.1 %) as obtained by 16S rRNA gene amplicon-based metagenome analysis of OTUs at the genus level

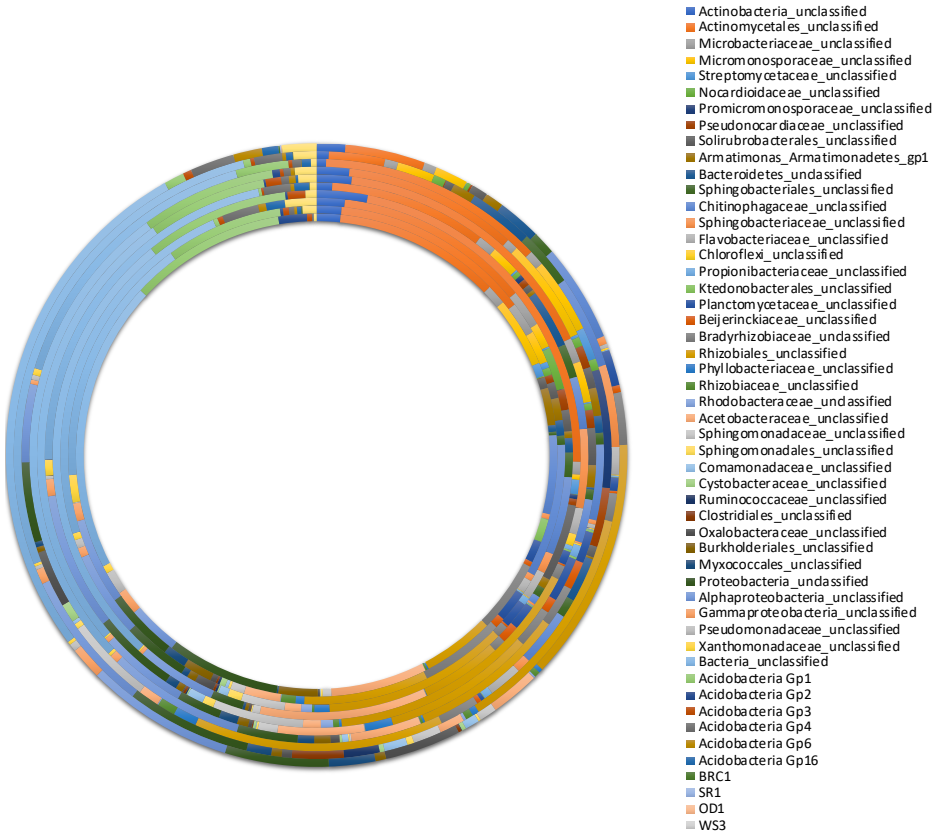


Fig. 8B. Relative abundance of bacterial unclassified taxa at the level of genus of orchid root microbiome (>0.1 %) as obtained by 16S rRNA gene amplicon based metagenome analysis (inner to outer, 191–W3)

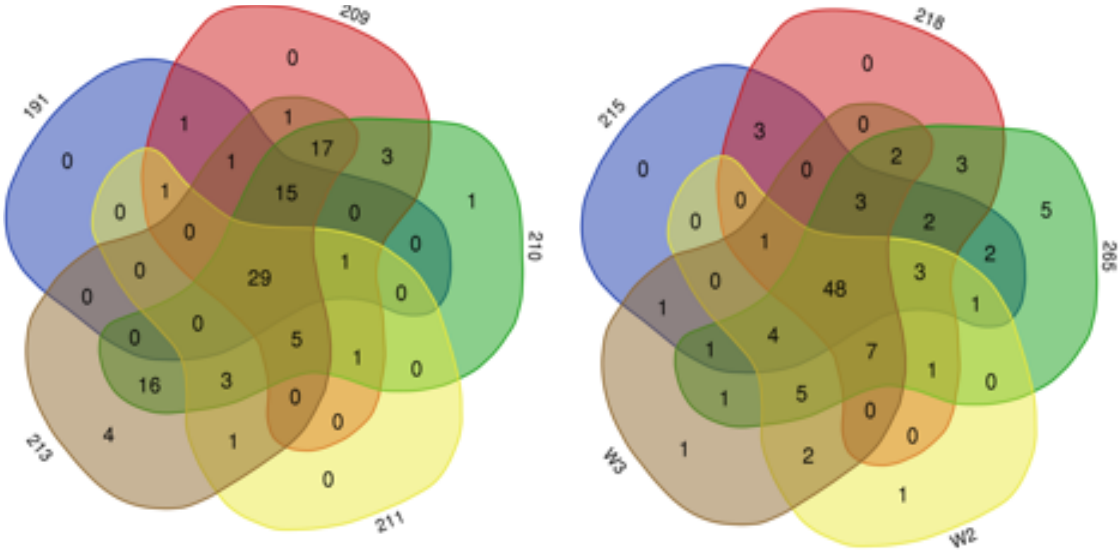


Fig. 9. Distribution of bacterial genera amongst samples (designation of the sample no. are given in Table 3)
(A) between 191, 209, 210, 211, 213
(B) between 215, 218, 265, W2, W3

Table 4. The distribution of common and unique bacterial genera in different orchids

Sample No.	Total genera	Taxa
191 209 210 211 213 215 218 265 W2 W3	23	*Pseudonocardia*, *Methylobacterium*, *Sphingomonas*, *Burkholderia*, *Rhizobiales*_unclassified, *Mycobacterium*, *Acidobacteria*_Gp1, *Acidobacteria*_Gp3, *Mucilaginibacter*, *Microbacteriaceae*_unclassified, *Solirubrobacterales*_unclassified, *Beijerinckiaceae*_unclassified, *Micromonosporaceae*_unclassified, *Actinomycetales*_unclassified, *Proteobacteria*_unclassified, Armatimonas_*Armatimonadetes*_gp1, *Alphaproteobacteria*_unclassified, *Gammaproteobacteria*_unclassified, *Acetobacteraceae*_unclassified, *Bradyrhizobiaceae*_unclassified, *Chitinophagaceae*_unclassified
191 209 210 213	15	*Sphingomonadales*_unclassified, *Rhizobiaceae*_unclassified, *Pseudomonadaceae*_unclassified, *Oxalobacteraceae*_unclassified, *Xanthomonadaceae*_unclassified, *Sphingobacteriaceae*_unclassified, *Bacteroidetes*_unclassified, *Roseomonas*, *Gemmata*, *Pseudomonas*, *Solirubrobacter*, *Rhizobium*, *Caulobacter*, *Spirosoma*, *Bosea*
209 210 211 213	5	*Chitinophaga*, *Marmoricola*, TM7_genus_incertae_sedis, *Stenotrophomonas*, *Actinomycetospora*
191 209 213	1	*Ktedonobacterales*_unclassified
209 210 211	1	*Massilia*
209 210 213	17	*Flavobacterium*, *Chryseobacterium, Brevundimonas*, *Segetibacter*, Gp4, *Aurantimonas*, *Friedmanniella*, *Actinoplanes*, *Gemmatimonas*, *Sphingobacterium*, *Opitutus*, *Pedobacter*, *Microbacterium*, *Nocardiodes*, *Cellulosimicrobium*, *Devosia*, *Dyadobacter*
210 211 213	3	*Pseudonocardiaceae*_unclassified, *Myxococcales*_unclassified, *Promicromonospora*
191 209	1	*Luteibacter*
209 210	3	*Hymenobacter, Kribbella*, *Amycolatopsis*
210 213	16	*Rhodobacteraceae*_unclassified, *Cystobacteraceae*_unclassified, *Phyllobacteriaceae*_unclassified, *Ruminococcaceae*_unclassified, *Flavobacteriaceae*_unclassified, *Propionibacteriaceae*_unclassified, *Clostridiales*_unclassidied, *Comamonadaceae*_unclassified, *Nocardioidaceae*_unclassified, *Olivibacter*, *Chelativorans*, *Comamonas*, *Adhaeribacter*, *Brachybacterium*, Gp16
215 218 265 W2	3	*Actinoplanes*, *Hymenobacter*, *Brevundimonas*
215 218 265 W3	3	*Streptomyces*, *Segetibacter*, *Pseudonocardiaceae*_unclassified
215 218 W2 W3	1	Gp6
215 265 W2 W3	4	*Roseomonas*, Gp16, *Sphingomonadaceae*_unclassified, *Sphingomonadales*_unclassified
218 265 W2 W3	7	*Opitutus*, *Dyadobacter*, *Flavobacterium*, *Chitinophaga*, *Rhizobiaceae*_unclassified, *Comamonadaceae*_unclassified, *Flavobacteraceae*_unclassified
215 218 265	2	*Friedmanniella*, *Amycolatopsis*
215 265 W2	1	*Aurantimonas*
215 265 W3	1	*Conexibacter*
218 265 W2	1	*Chryseobacterium*
218 265 W3	2	*Filimonas*, *Stenotrophomonas*
265 W2 W3	5	*Pseudomonas*, *Chelativorans*, *Devosia*, *Azospirillum*, *Pseudomonadaceae*_unclassified
215 218	3	*Blastomonas*, Gp2, *Ktedonobacterales*_unclassified
215 265	2	*Kribbella*, *Propionibacteriaceae*_unclassified
215 W3	1	*Chloroflexi*_unclassified
218 265	3	*Sphingobacterium*, *Blastococcus*, *Streptomycetaceae*_unclassified
265 W3	1	*Bosea*
W2 W3	2	*Rhodobacteraceae*_unclassified, *Clostridiales*_unclassified
265	5	*Promicromonospora*, *Xylanimicrobium*, *Luteibacter*, *Lentzea*, *Promicromonosporaceae*_unclassified,
W2	1	*Ruminococcaceae*_unclassified
W3	1	*Comamonas*
210	1	*Blastococcus*

The alpha diversity of the RABs in each orchid sample was calculated by considering Shannon (H), Equitability (E), Simpson (D) and Inversion Simpson (I) diversity indices (Table 5). *Cymbidium* sp. microbiota showed the highest Shannon, Equitability, and Inverse Simpson index (H = 7.9 ± 0.02, E = 7.2 ± 0.08, I = 25.0 ± 0.0). The results showed that *Cymbidium* sp. has the most bacterial diversity showed by Shannon diversity and Equitability (considering the even distribution of OTUs). Further, to understand the community level relationship between all the samples, beta diversity was studied. Here, non-metric multidimensional scaling (NMDS) analysis using Bray-Curtis distance method was carried out (Fig. 10). The NMDS plot showed the formation of 4 distinct groups based on the overlap of microbiota (Fig. 10). The first distinct group consisted of 5 orchid species of *Bulbophylum* sp. 1, *Bulbophylum* sp. 2, *Dendrobium* sp. 1, *Agrostophyllum* sp., and *Thunia* sp.; the second group consisted of *Aeridis* sp. (W2) and *Vanda* sp. (W3); the third group consisted of *Cymbidium* sp. and *Dendrobium* sp. 2 (213); the fourth group consisting of a single species is *Rhynchostylis* sp.

Table 5. Comparison of alpha-diversity values of orchid root samples [i. Shannon index (H), equitability index (E), >0, higher the value more the diversity; ii. Inverse Simpson index (I) is reciprocal of Simpson index (D), higher the value higher the diversity]

Orchid	**Shannon (H)**	**Equitability (E)**	**Inv. Simpson (I)**
Bulbophyllum sp. 1	5.2±0.03	6.6±0.03	2.6±0.03
Agrostophyllum sp.	6.9±0.04	6.9±0.05	11.1±0.01
Cymbidium sp.	7.9±0.02	7.2±0.08	25.0±0.0
Dendrobium sp. 1	5.5±0.06	6.7±0.1	10.0±0.0
Dendrobium sp. 2	6.9±0.07	7.0±0.09	20.0±0.0
Bulbophyllum sp. 2	6.9±0.06	6.8±0.04	20.0±0.0
Thunia sp.	7.1±0.01	6.7±0.07	14.2±0.1
Rhynchostylis sp.	5.1±0.06	6.0±0.04	12.5±0.0
Aeridis sp.	7.0±0.01	6.6±0.07	14.2±0.09
Vanda sp.	7.5±0.07	7.0±0.09	20.1±0.0

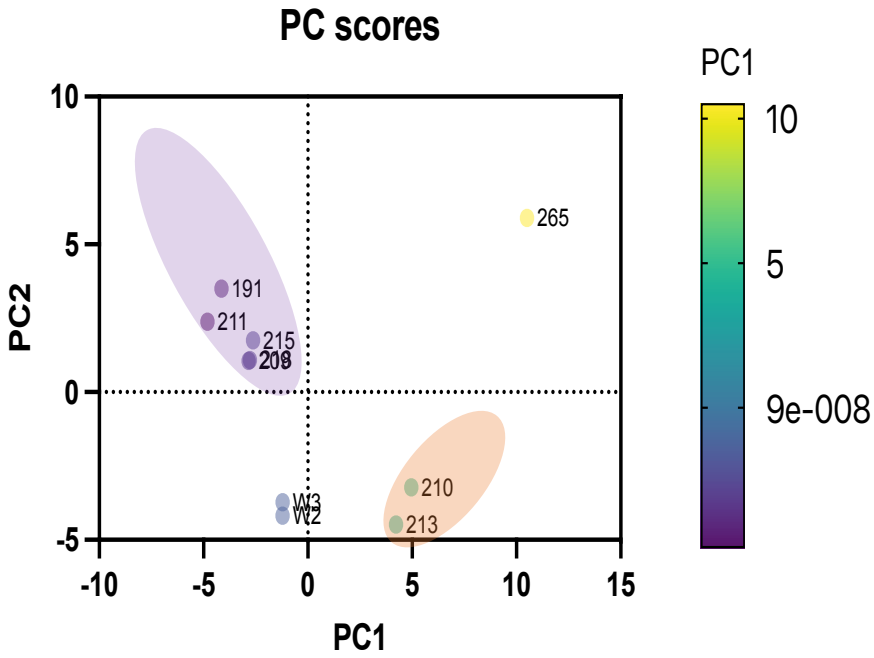


Fig. 10. Beta diversity analysis with non-metric multidimensional scaling (NMDS) plot measured by Bray-Curtis distance between the orchid samples; the distance between the samples here demonstrates the differences in bacterial diversity

3.1.4. Unclassified and unique bacterial phylotypes

At the phylum level, unclassified OTUs in the orchids consisted in between the range of 10 to 20 % of the total OTUs observed. These unclassified sequences indicate that a certain percentage of unknown bacterial diversity of RABs prevail, which may have a significant role in the holistic development of the orchid plants. Therefore, the phylogenetic analysis of these unclassified sequences was carried out against the reference sequences of 16S rRNA genes of validly published/*Candidatus* (*Ca.*) members (IRPCM 2004). *Bulbophyllum* sp. had the lowest number of unique phylotypes ("*Ca.* UH24") whereas *Vanda* sp. (W3) had the highest number of 10 ("*Ca.* UH26", "*Ca.* UH31", "*Ca.* UH40", "*Ca.* UH50", "*Ca.* UH52", "*Ca.* UH55-UH59"). The shared distribution of these unique

phylotypes is given in Table 6. The information on the total number of phylotypes and unique phylotypes detected for each sample is given in Fig. 11A-J. The unclassified sequences were grouped into 108 phylotypes with, 37 (34 %) being unique. An overview illustrating the root bacterial tree of life of all orchid plants is presented in Fig. 12. Further, the comparison of unique phylotypes between the orchids showed that no unique phylotypes represented the core root microbiome or were shared commonly between the orchids.

Table 6. Common and unique phylotypes shared between orchids root microbiome

Sample identity	Total	Unique phylotypes
209 210 265 W2 W3	1	UH26
W2 W3 210	1	UH31
191 210 211	1	UH24
215 W2	1	UH44
215 W3	1	UH40
218 265	1	UH45
265 W3	1	UH52
209	4	UH25, UH28, UH29, UH27
210	1	UH30
211	1	UH32
213	7	UH39, UH33, UH38, UH34, UH37, UH36, UH35
215	3	UH43, UH41, UH42
218	3	UH48, UH46, UH47
265	3	UH51, UH49, UH53
W2	1	UH54

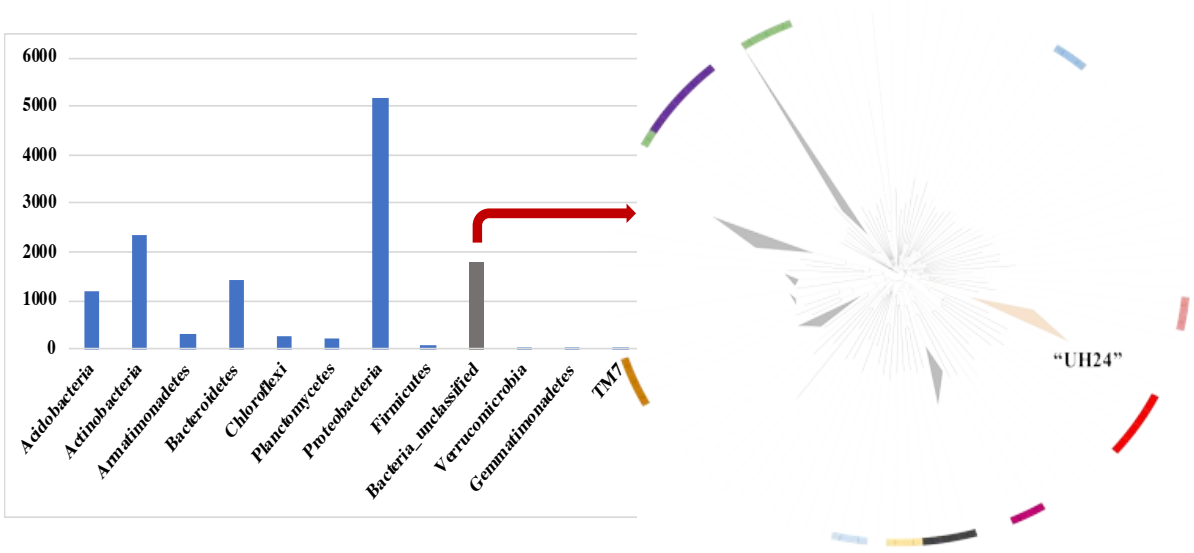


Fig. 11A. For *Bulbophyllum* sp. 1, distribution of 7 total and 1 unique phylotypes ("*Ca.* UH24") at the phylum level

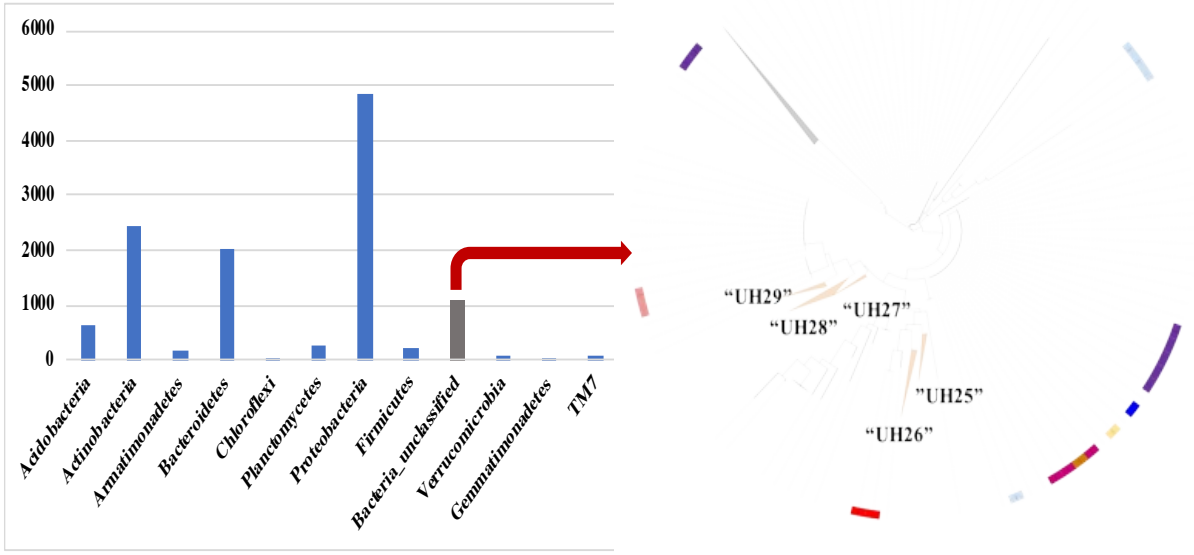
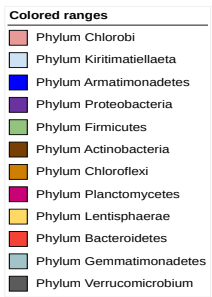


Fig. 11B. For *Agrostophyllum* sp., distribution of 6 total and 5 unique phylotypes ("*Ca.* UH25-UH29") at the phylum level



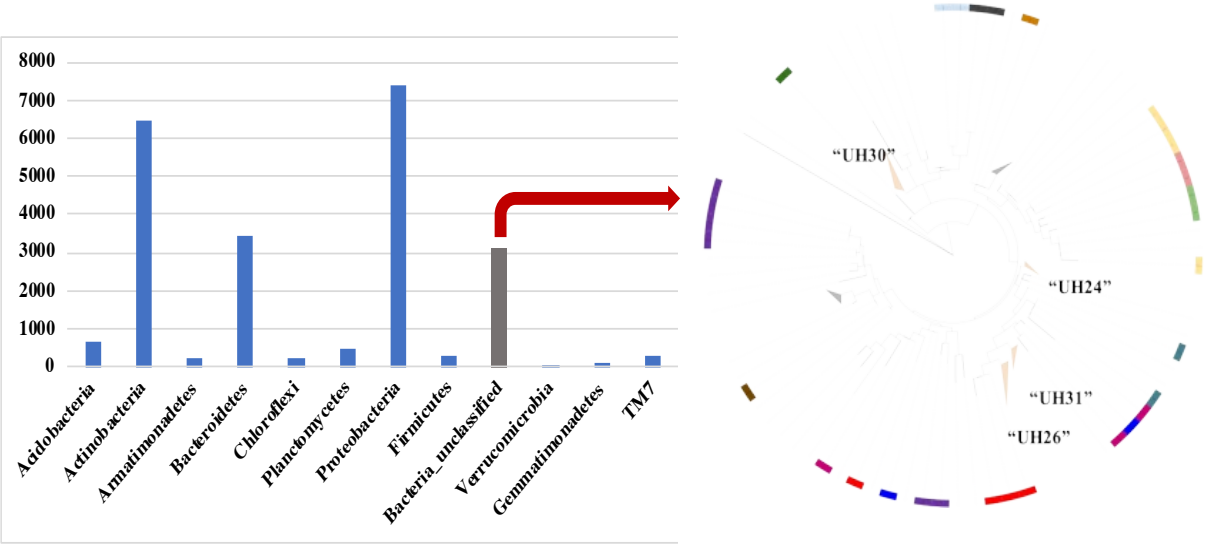


Fig. 11C. For *Cymbidium* sp., root associated bacterial tree of life with 6 total and 4 unique phylotypes ("*Ca.* UH24", "*Ca.* UH26", "*Ca.* UH30-UH31") at the phylum level

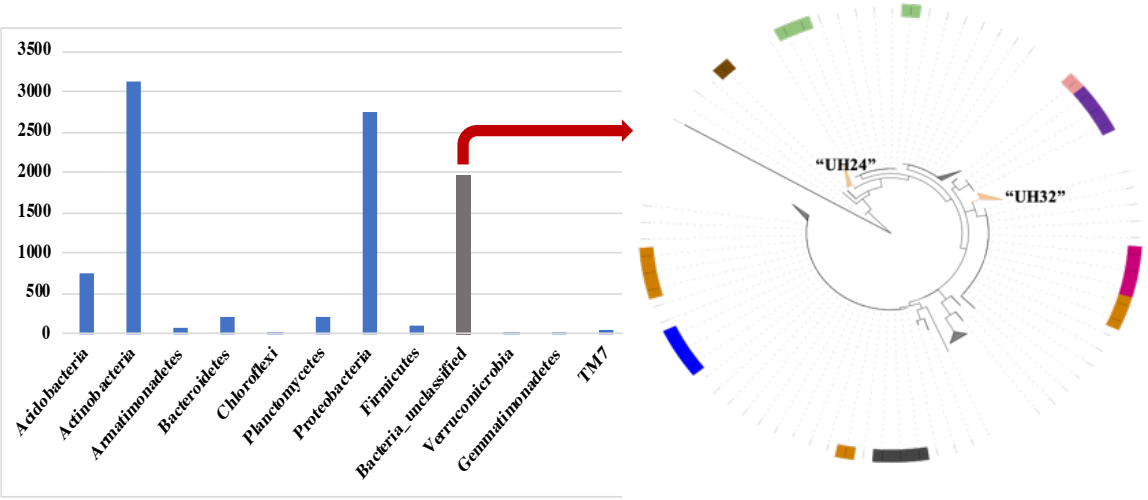
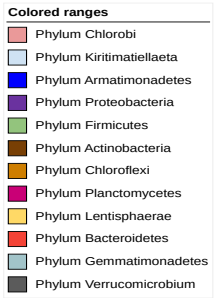


Fig. 11D. For *Dendrobium* sp. 1, root associated bacterial tree of life with 5 total and 2 unique phylotypes ("*Ca.* UH24", "*Ca.* UH32") at the phylum level



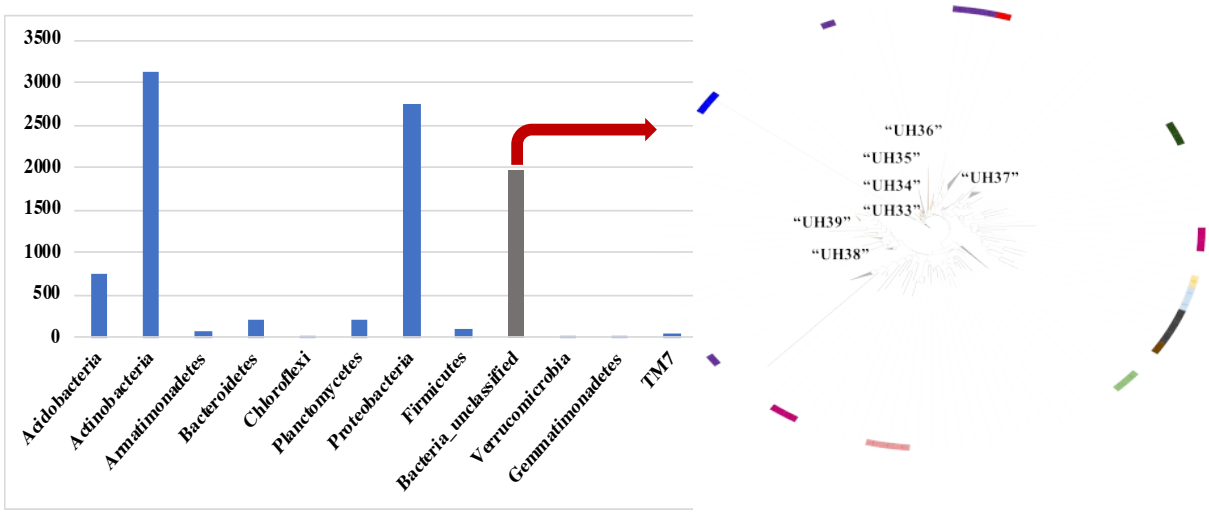


Fig. 11E. For *Dendrobium* sp. 2, root associated bacterial tree of life with 13 total and 7 unique phylotypes ("*Ca.* UH33-39") at the phylum level

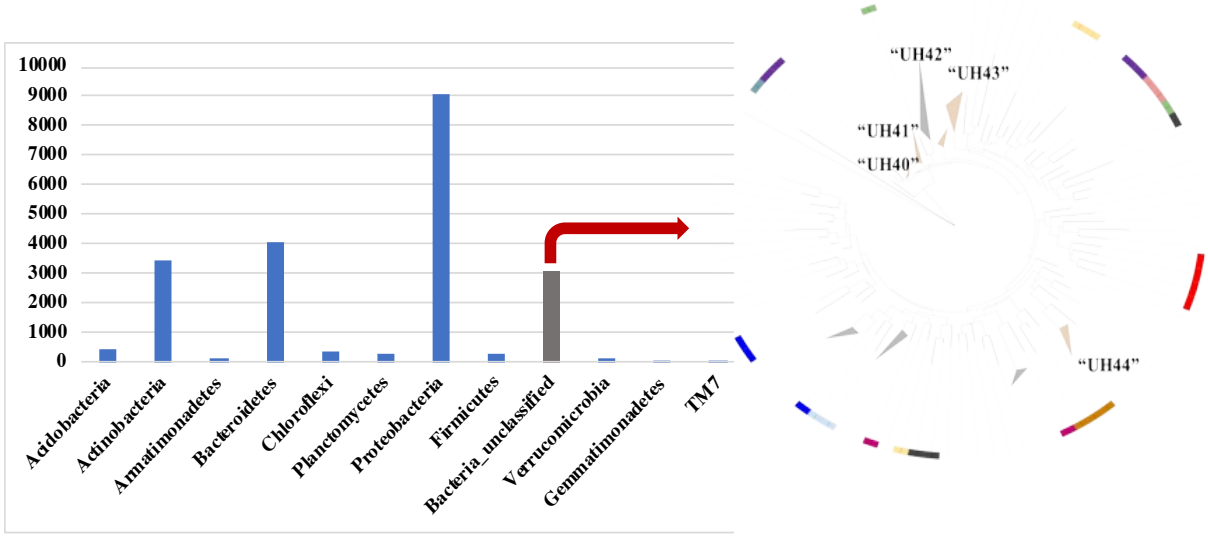
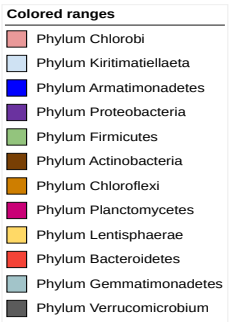


Fig. 11F. For *Bulbophyllum* sp., root associated bacterial tree of life with 9 total and 5 unique phylotypes ("*Ca.* UH40-UH44") at the phylum level



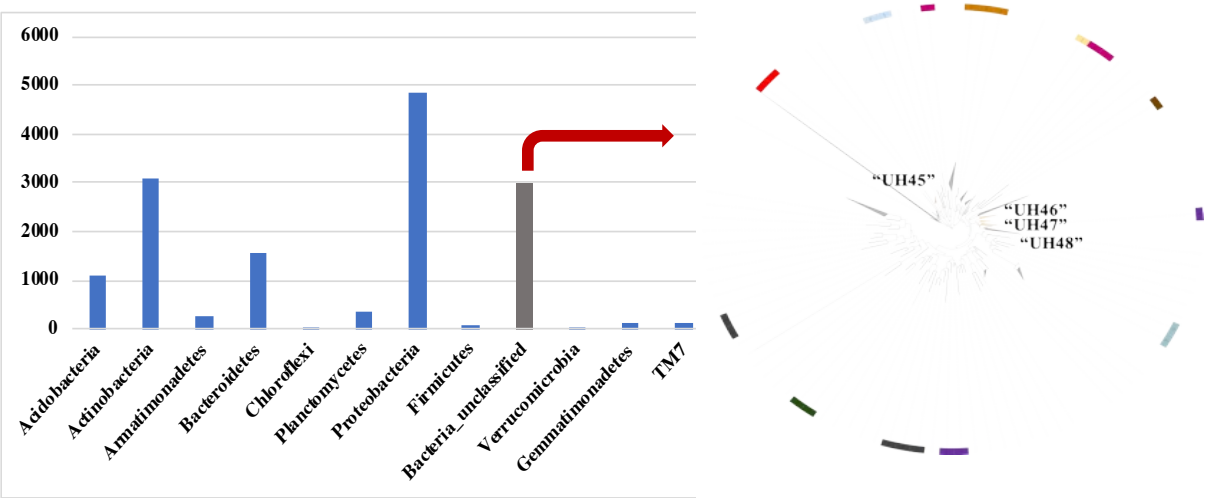


Fig. 11G. For *Thunia* sp., root associated bacterial tree of life with 12 total and 4 unique phylotypes ("*Ca.* UH45-UH48") at the phylum level

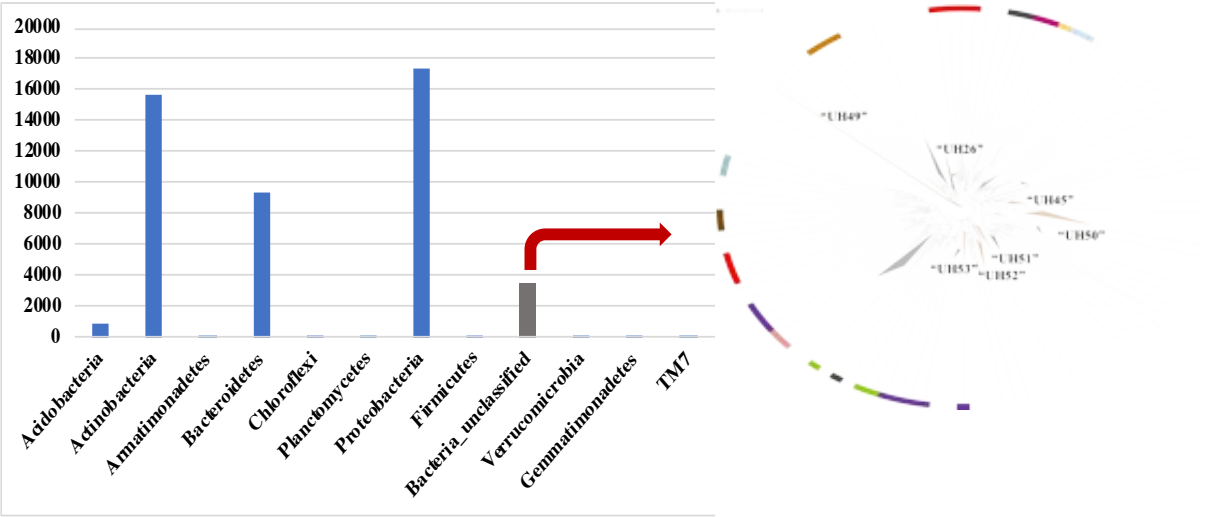
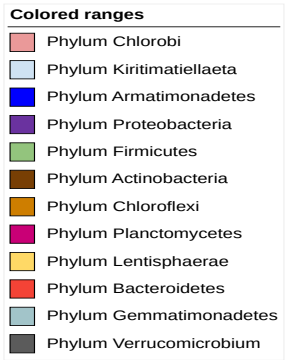


Fig. 11H. For *Rhychostylis* sp., root associated bacterial tree of life with 15 total and 7 (*"Ca.* UH26*" "Ca.* UH45*" "Ca.* UH49-UH53*"*) unique phylotypes at the phylum level



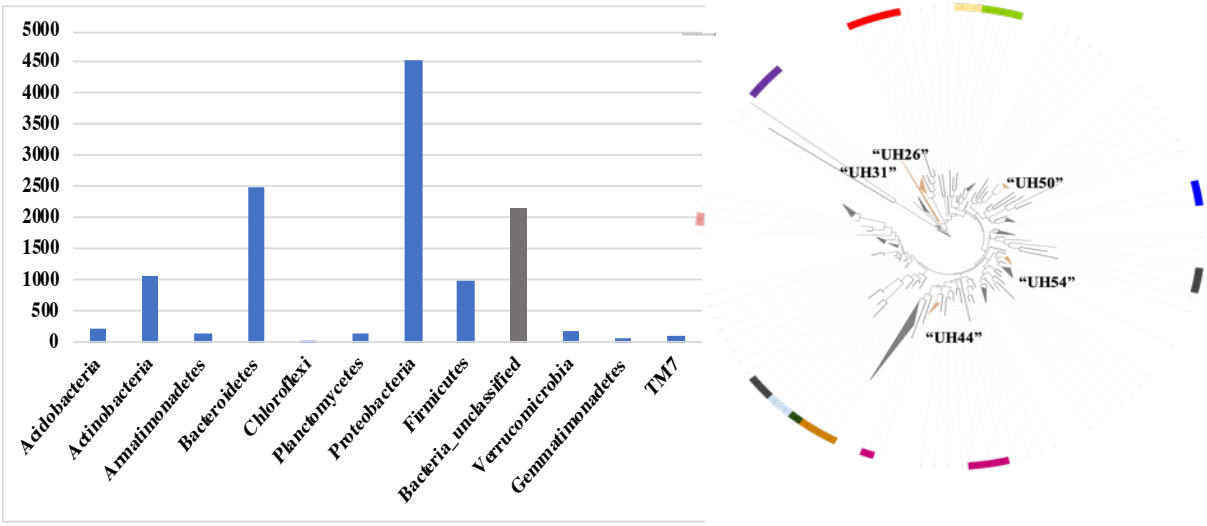


Fig. 11I. For *Aerides* sp., root associated bacterial tree of life with 13 total and 5 unique phylotypes ("*Ca.* UH26", "*Ca.* UH31", "*Ca.* UH44", "*Ca.* UH50", "*Ca.* UH54") at the phylum level

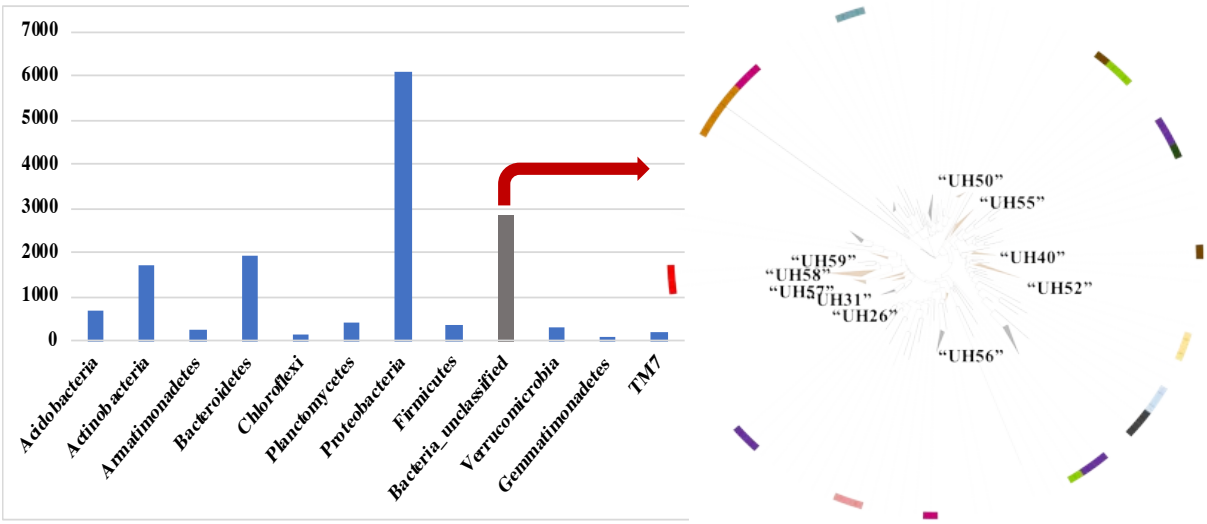
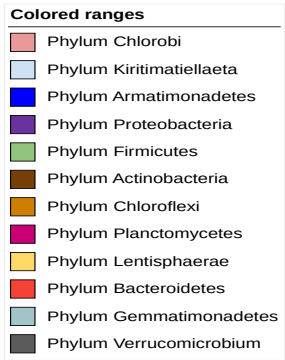


Fig. 11J. For *Vanda* sp., root associated bacterial tree of life with 17 total and 10 unique phylotypes ("*Ca.* UH26", "*Ca.* UH40", "*Ca.* UH50", "*Ca.* UH52", "*Ca.* UH55-UH59") at the phylum level



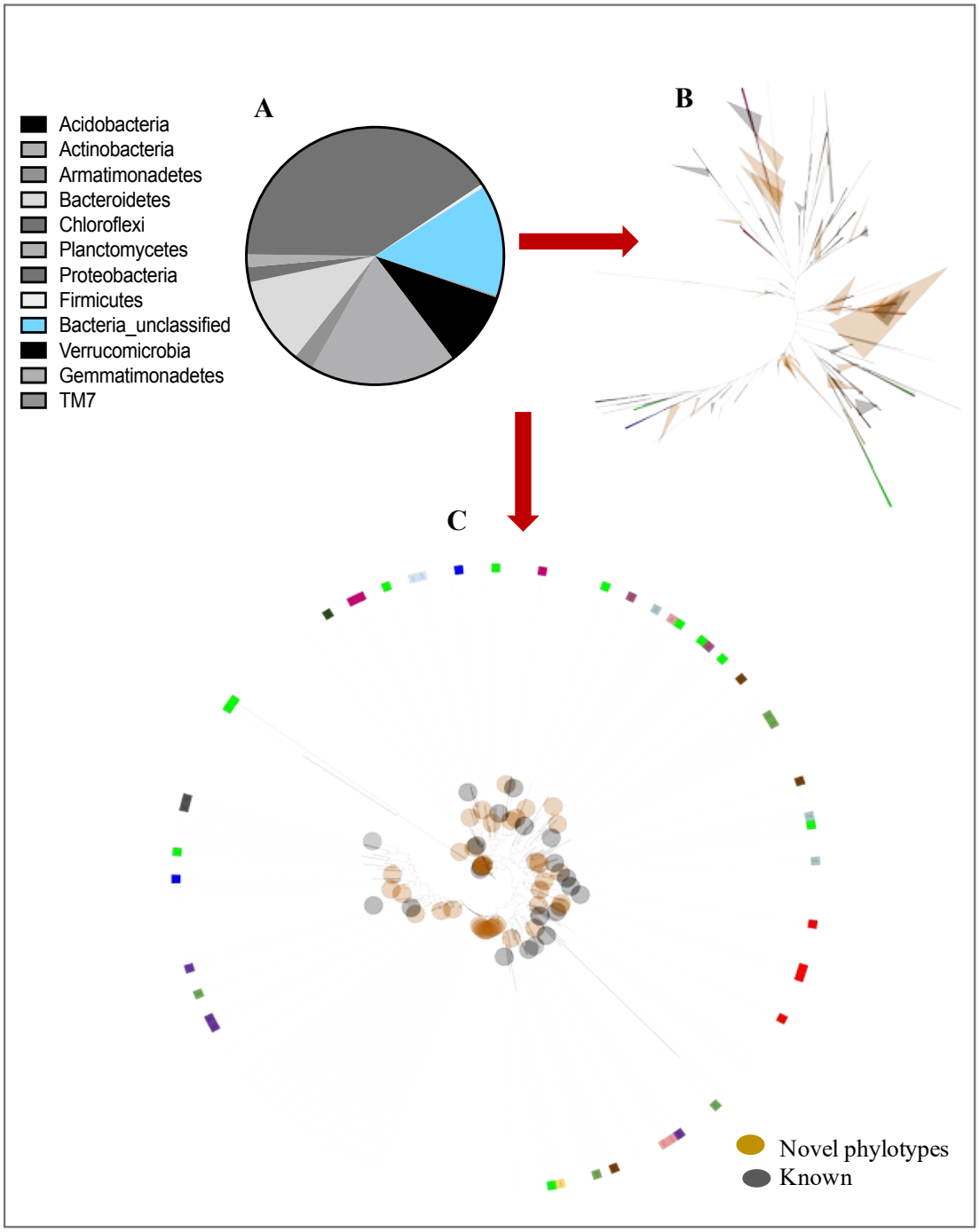


Fig. 12. An overview representing the root associated bacterial tree of life of orchid microbiome
(A) Percentage of unclassified sequences at the phylum level
(B) Unrooted tree representing the novel phylotypes (yellow triangles)
(C) At the phylum level, around 108 phyloclades were identified with 37 unique phylotypes identified at the level of phylum

3.2. Phylogenomics and polyphasic studies of bacteria from various sources

The samples of orchid roots, water/sediments/sand (hot springs, saltpans), and industrial effluents were collected, enriched, cultivated, and subjected to isolation. Turbidity was observed from the enrichment of water/sediment (saltpan/hot-springs) and effluents (acid drainage) samples. 98 colonies were selected from the above enriched and orchid root samples, which were further purified after repeated streaking. The initial screening was based on colony/cell morphology and pigmentation for identification. A total of 22 isolates was sent for 16S rRNA gene sequencing, followed by the sequence identity analysis by BLAST in EzBioCloud. Out of the total strains isolated, eighteen strains were isolated from different orchid roots and majorly belonged to the phylum of *Actinobacteria, Proteobacteria,* and *Firmicutes*. Four strains (JC497, JC498, JC500, JC505) had the highest sequence identity (>99 %) with members of genus *Streptomyces*, similarly, four strains (JC502, JC536, JC537, JC539) with members of the genus *Methylobacterium*. Strain JC162 (beach sand), JC560 (hot spring) were members of the genus *Roseomonas,* whereas strain JC496 (industrial acid mine) and JC1013 (salt pan) belonged to the genus of *Acidithiobacillus* and *Mesobacillus,* respectively. All the strains were preserved as 50 % glycerol stock at -20 °C. The phylogenetic affiliation along with other information are given in Table 7.

Table 7. The phylogenetic affiliation of the isolated strains along with its source, location and accession numbers

Strain No.	Identity (%)	Closely related species	Accession no.	Isolation Source	Place	GPS
JC497	99.9	*Streptomyces olivaceus* NRRL B-3009^T	LT799397	*Cymbidium* sp.*	Darjeeling, West Bengak	27.03N 88.40N
JC498	99.9	*Streptomyces badius* NRRL B-2567^T	LT799398	*Cymbidium* sp.*	Peshok, West Bengal	27.07N 88.26E
JC501	98.5	*Paracoccus marinus* KKL-5^T	LT799401	*Aeridis* sp.*	Amboli, Karnataka	15.85N 73.99E
JC504	99.8	*Streptomyces badius* NRRL B-2567^T	LT799404	*Cymbidium* sp.*	Pakyong, Sikkim	27.22N 88.58E
JC505	100	*Streptomyces cirratus* NRRL B-3250^T	LT799405	*Cymbidium* sp.*	Pakyong, Sikkim	27.22N 88.58E
JC500	99.2	*Streptomyces swartbergensis* HMC13^T	LT799400	*Cymbidium* sp.*	Darjeeling, West Bengal	27.03N 88.40N
JC502	99.5	*Methylobacterium radiotolerans* JCM 2831^T	LT799402	*Vanda* sp.*	Amboli, Karnataka	15.85N 73.99E
JC536	99.8	*Methylobacterium oryzae* CBMB20^T	LT907971	*Vanda* sp.*	Amboli, Karnataka	15.85N 73.99E
JC537	99.6	*Methylobacterium phyllostachycos* BL47^T	LT907972	*Dendrobium* sp.*	Peshok, West Bengal	27.07N 88.26E
JC539	99.7	*Methylobacterium oryzae* CBMB20^T	LT907974	*Vanda* sp.*	Amboli, Karnataka	15.85N 73.99E
JC500	99.2	*Microlunatus orkiwanensis* DSM 21744^T	LT799399	*Dendrobium* sp.*	Peshok, West Bengal	27..07N 88.38E
JC538	99.7	*Agrobacterium purense* LMG 25623^T	LT907970	*Cymbidium* sp.*	Lachung, Bengal	27. 95 N 73.99E
JC540	99.6	*Bacillus muralis* DSM 16288^T	LT907971	*Rhyncostylis* sp.*	Devkund, Orissa	21.70N 86.44E
JC559	100	*Priestia aryabhattai* B8W22^T	LR535809	*Rhyncostylis* sp.*	Devkund, Orissa	21.70N 86.44E
JC633	97.8	*Corynebacterium suicordis* CECT 5724^T	LR535807	*Dendrobium* sp.*	Peshok, West Bengal	27.07N 88.26E
JC701	98.8	*Microbacterium testaceum* NBRC 12675^T	LR656263	Orchid root*	Peshok, West Bengal	27.17N 88.26E
JC703	97.4	*Microbacterium proteolyticum* RZ36^T	LR746173	Orchid root*	Lachung, Sikkim	27. 95 N 73.99E
JC708	98.3	*Janibacter melonis* CM2104^T	LR746172	Orchid root*	Peshok, West Bengal	27.17N 88.26E
JC560	99.3	*Roseomonas cervicalis* ATCC 49957^T	LT996832	Hot spring	Atri, Orissa	20.21N 85.50E
JC162	99.7	*Roseomonas oryzicola* YC6724^T	HE984358	Beach Sand	Rameshwaram, Tamil Nadu	9.16N 79.13E
JC1013	98.8	*Mesobacillus selenartesenatis* SF-1^T	LS998022	Salt pan	Vedaranyam, Tamil Nadu	10.21N 79.50E
JC496	99.4	*Acidithiobacillus albertensis* DSM 14366^T	LT799396	Industrial Acidmine	Jantia Hills, Meghalaya	25.39N 93.01E

*root part of the orchid

3.2.1. Characterization of strain JC501^T

3.2.1.1. Home habitat, colony, and cell morphology

Strain JC501^T was isolated from the root sample of a flowering *Aeridis maculosa,* an epiphytic orchid. It was collected on June 2017 from Amboli (Western Ghat) Karnataka (GPS: 15° 95′ 69″ N, 73° 99′ 47″ E), India (Fig. 13A). Strain JC501^T colonies were 1–2mm (diameter), circular, smooth, translucent, raised with irregular margins, pale orange, and spread across the petri-plate when grown on nutrient agar (half-strength) (HiMedia M001) (Fig. 13B). Cells of strain JC501^T were coccoid in shape (0.9–1 µm in width) and divided by binary fission (Fig. 13C). Cells were non-motile and Gram-stain-negative.

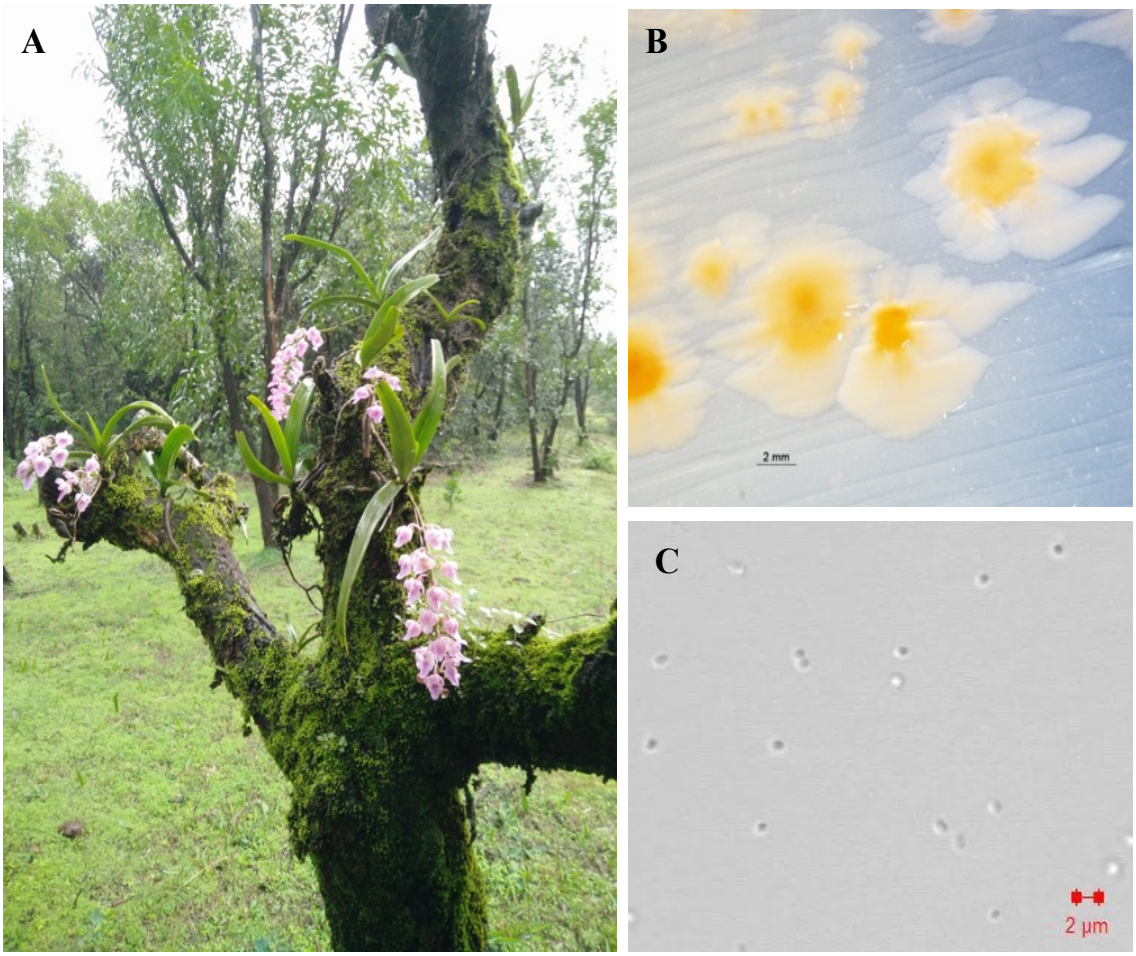


Fig. 13. (A) An epiphytic orchid (*Aerides maculosa*) which was flowering during the month of June 2017 was collected from Amboli, Western Ghats, Karnataka (15°95′69″N and 73°99′47″E), India. Colony morphology of
(B) Colony morphology of strain JC501^T;
(C) Confocal microscopic image of the strain JC501^T

3.2.1.2. Phylogeny based on 16S rRNA, MLSA, and whole genome

The analysis of the 1456 bp length of the 16S rRNA gene by BLAST (EzBioCloud) showed that strain JC501^T is phylogenetically affiliated to the members of the genus *Paracoccus* with its nearest type strain *P. marinus* KKL-A5^T (98.9 %) and other members of the genus *Paracoccus* (<97.3 %). The maximum likelihood (ML) (model tested under the lowest Akaike information criterion) tree was generated using 32 type strains and 9 non-type strains, showing the highest 16S rRNA gene sequence identity. Kimura's two-parameter + gamma distribution + invariable site (K2+G+I) model with uniform rates, and the heuristic search algorithm nearest-neighbour-interchange (NNI) with complete deletion. The ML tree confirmed the relatedness of strain JC501^T with members of the genus *Paracoccus* and distinctly segregated from the closest phylogenetic neighbour *P. marinus* KKL-A5^T along with related cultivated strains and clones (Fig. 14). In the ML tree, strain formed a distinct clade; henceforth, this clade shall be designated as *P. marinus* clade.

Additionally, the full nucleotide and amino acyl sequences of 8 protein-coding housekeeping genes (*recA, gryB, rpoA, rpoB, rpoD, tryB, dnaK, atpD*) were taken from the genomes of the members of the *Paracoccus* clade, concatenated and aligned. For the 16S rRNA gene, the ML tree was constructed for the MLSA (Fig. 15A, B). The gene sequence identity (concatenated) between strain JC501^T and *P. marinus* NBRC 100637^T was 84.6 % (with other members of the genus <72.9 %). The RAXML-based phylogenomic tree (UBCG tool) also proved that the strain was distinctly affiliated with the genus *Paracoccus* (Fig. 16). The Type (Strain) Genome server (http://tygs.dsmz.de) also confirmed strain JC501^T as a 'potential new species.' Therefore, all dendrograms (16S rRNA gene, MLSA, whole genome based) showed that strain JC501^T clustered distinctly with *P. marinus* clade.

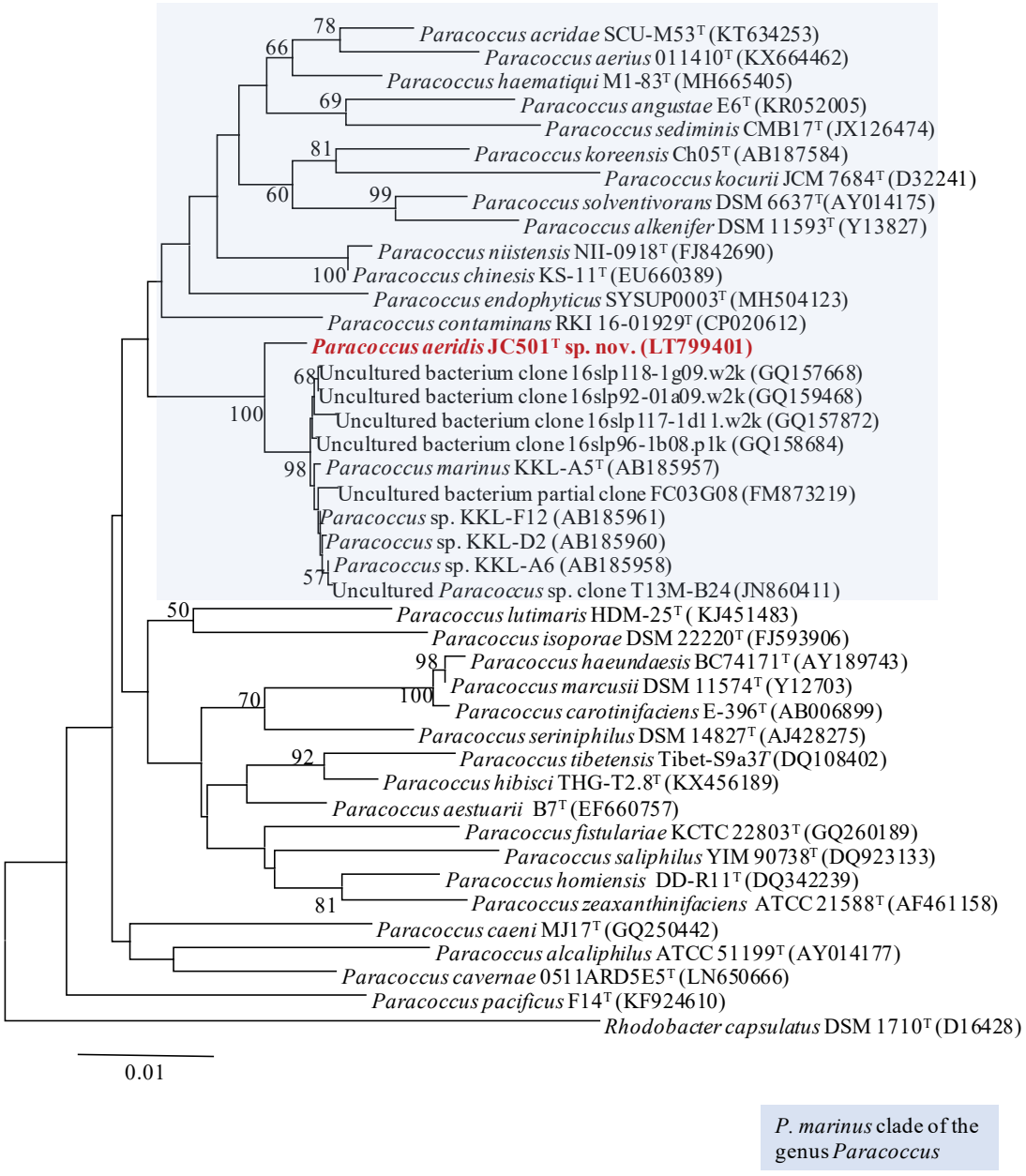


Fig. 14. Phylogenetic tree based on the16S rRNA gene sequence comparisons, using the maximum-likelihood (ML) (model tested under AIC under Kimura 2 parameter). MEGA7 software was used for computing. *Rhodobacter capsulatus* DSM 1710^T (D16428) was used as an outgroup. Numbers at the nodes represent the bootstrap values. Bar, 0.01 accumulated changes per nucleotide.

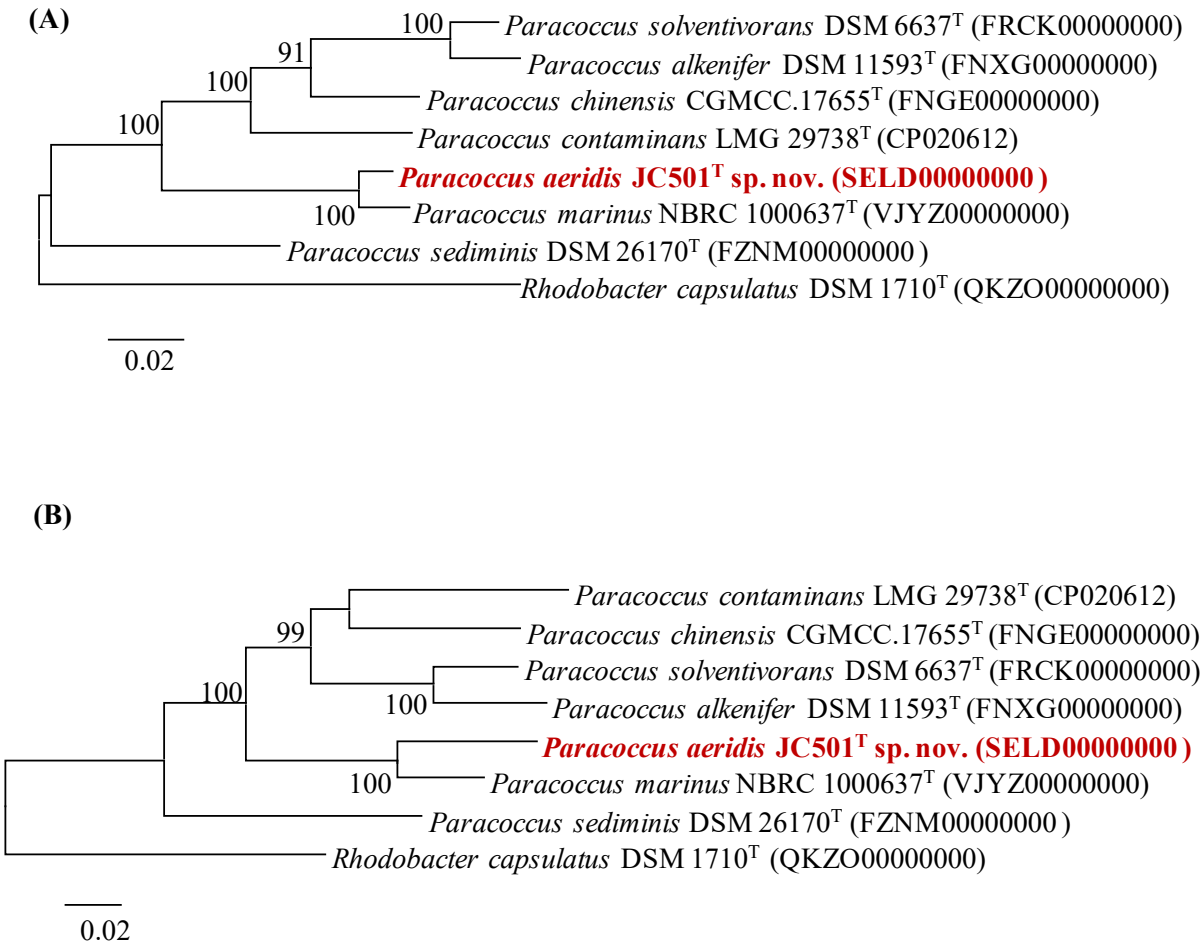
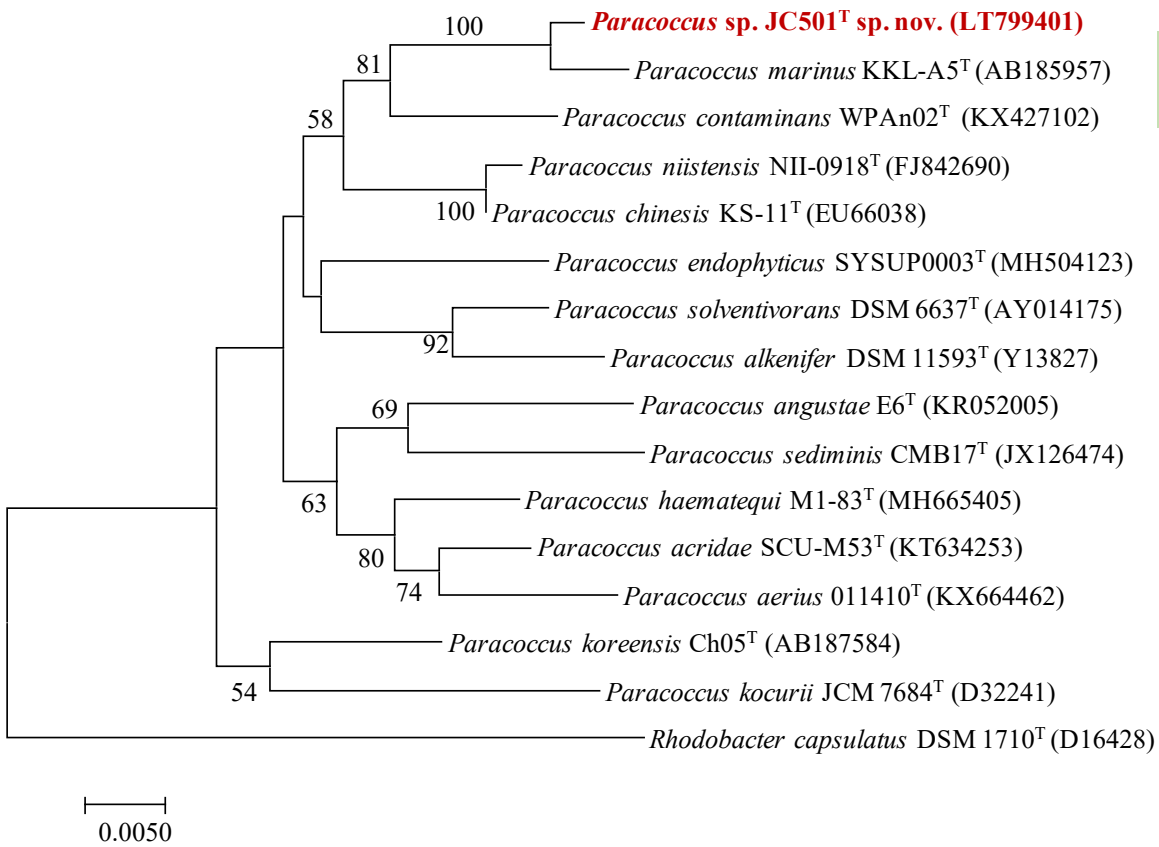


Fig. 15. ML based phylogenetic tree (*recA, gyrB, rpoA, rpoB, , rpoD, trpB, dnaK, atpD)* (Model tested under AIC under LG+G+I+F and GTR+G+I parameter respectively) MEGA7 software was used for computing. *Rhodobacter capsulatus* DSM 1710^T (QKZO00000000) was used as outgroup. Numbers at the nodes represent the bootstrap values. On the nodes bar, 0.02 accumulated changes per nucleotide and amino acid respectively.
(A) MLSA concatenated amino acid gene sequences
(B) MLSA concatenated nucleotide gene sequences



OGRI

16S rRNA similarity(%)	ANI (%)	*dDDH (%)
100	100	100
98.9	85.9	29.8
97.3	77.6	21.9
97.0	-	-
96.9	78.8	22.5
96.9	78.5	22.0
96.2	78.1	21.9
95.7	77.2	21.2
95.2	-	-
95.0	75.6	20.7
97.2	80.0	21.0
96.8	-	-
95.8	-	-
95.5	-	-
94.7	-	-

Fig. 16. UBCG software based phylogenomic tree based on 92 concatenated genes with *Rhodobacter capsulatus* DSM 1710^T as the outgroup; along with the overall genomic related index (OGRI). Bar, 0.01 accumulated changes per nucleotide

3.2.1.3. Genomic characterization

The genome size of strain JC501^T was 3.3 Mbp, whereas DNA G+C content was 69.4 mol %. Similarly, for *P. marinus* NBRC 100637^T, it was 3.2 Mbp and 69.3 mol %, respectively. ANI and *d*DDH values between strain JC501^T and *P. marinus* NBRC 100637^T were 85.9 % and 29.8 %, respectively. The *d*DDH and ANI values with the other species of the *P. marinus* clade are provided in Fig. 16. Additionally, genome sequences of *P. contaminans* LMG 29738^T, *P. chinensis* CGMCC.17655^T, *P. solventivorans* DSM 6637^T, *P. alkenifer* DSM 11593^T, and *P. sediminis* DSM 26170^T were downloaded from NCBI and used for the analysis.The genomic characteristics are given in Table 8 and clearly distinguish strain JC501^T from other neighbouring members of the genus *Paracoccus*. Against genomes of other six members, strain JC501^T was used as a reference genome. There was a difference in the alignment of the LCBs of *Paracoccus* species (Fig. 17A). The comparison of proteomes showed that protein sequence identity of 10-30 % was shared with those of other *P. marinus* NBRC 100637^T clade members (Fig. 17B). The anti-SMASH results showed that strain JC501^T had five biosynthetic gene clusters for secondary metabolites [hser-lactone (homoserine lactone), T1PKS/NRPKS like (type I polyketide synthases/non-ribosomal peptide synthases), beta-lactone (formicamycin), terpene (carotenoid) and T3PKS (type III polyketide synthases)].

Table 8. Genome characteristics of the strain JC501^T, *P. marinus* NBRC 100637^T and the members of the *P. marinus* clade of the genus *Paracoccus*

Species	Strain	Genome Size (Mbp)	Scaffold	G+C content (mol %)	DDBJ/EMBL/ GenBank
Strain JC501^T	NBRC 113644^T	3.3	40	69.4	SELD00000000
P. marinus	NBRC 100637^T	3.1	38	69.4	VJYZ00000000
P. contaminans	LMG 29738^T	2.9	2	68.7	CP020612
P. chinensis	KS-11^T	3.6	84	68.1	FNGE00000000*
P. solventivorans	DSM 6637^T	3.4	27	68.7	FRCK00000000*
P. alkenifer	DSM 11593^T	3.2	19	67.3	FNXG00000000*
P. sediminis	DSM 26170^T	3.7	40	66.0	FZNM00000000*

All organisms are free living; *, DOE-Joint Genome Institute, USA

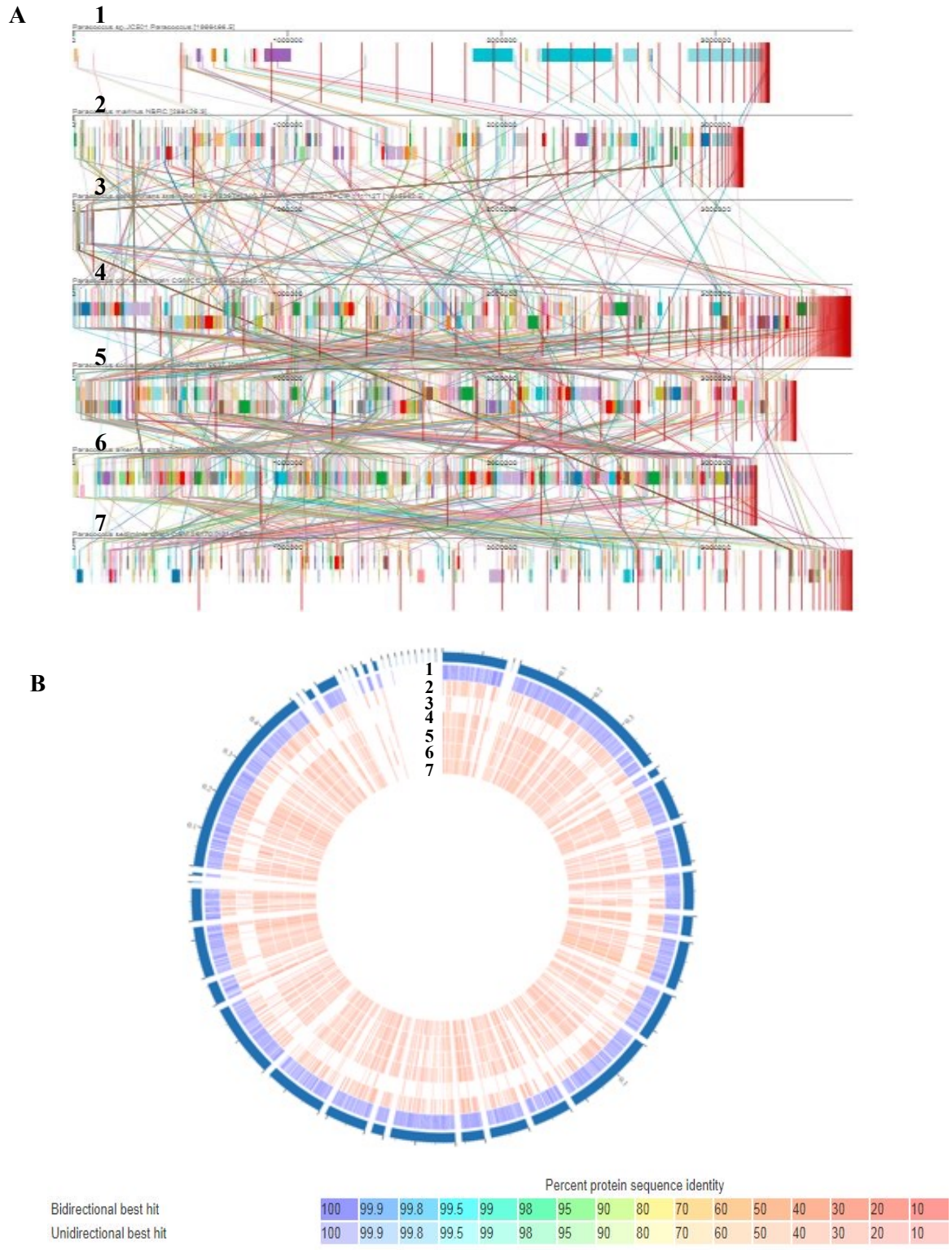


Fig. 17. (A) Genome organization of *Paracoccus* spp., The alignment of the local collinear blocks (LCBs) of the *Paracoccus* species differ from each other variably
(B) Protein map based on protein sequence identity between strain JC501^T and other closely related species of *Paracoccus.*
Genome of strain JC501^T was used as a reference. (1) strain JC501^T (2) *P. marinus* NBRC 100637^T (3) *P. contaminans* LMG 29738^T (4) *P. chinesis* CGMCC 1.7655^T (5) *P. solventivorans* DSM 6637^T (6) *P. alkenifer* DSM 11593^T (7) *P. sediminis* DSM 2160^T

3.2.1.4. Metabolic characterization

The gene constitution of strain JC501^T varied as compared to the other closely related members of *P. marinus* clade (Table 9). Strain JC501^T (*P. marinus* clade) contains genes putatively responsible for the metabolism of tryptophan into indole-3-acetic acid (IAA) via the indole-3-acetamide pathway/route (IAM), while *P. marinus* NBRC 100637^T, does not encode those genes. *P. contaminans* LMG 29738^T can produce indole-derived molecules, but its pathway remains obscure as the genes are absent. *P. marinus* clade members have a 2-C-methyl-d-erythritol 4-phosphate/1-deoxy-d-xylulose 5-phosphate (MEP) pathway. Further, the differences in the gene composition for tryptophan metabolism, carotenoid biosynthesis, urea decomposition, iron metabolism, siderophore, and keto-adipate pathway are presented in Table 9.

Strain JC501^T differed from other members of the *P. marinus* clade in having purine and benzoate-degrading genes. Strain JC501^T differed from *P. marinus* NBRC 100637^T in containing genes belonging to galactose, glutathione, glycerophospholipid, amino acid (thiamine, serine) metabolism, and styrene degradation pathways (Fig. 18A). Likewise, strain JC501^T varied from *P. contaminans* LMG 29738^T in constituting genes that were part of amino acid (alanine, glutamate, aspartate, cysteine), glycerolipid, ascorbate, and carotenoid biosynthetic metabolism (Fig. 18B). Metabolic pathway genes of amino acid (phenylalanine, alanine), and pyrimidine were observed in strain JC501^T but absent in *P. chinesis* CGMCC 1.7655^T (Fig. 18C). Carotenoids, folate biosynthesis pathways, and amino acids (tyrosine, methionine), glycerolipid, arachidonic acid metabolism were present in JC501^T but absent in *P. solventivorans* DSM 6637^T (Fig. 18D). Similarly, strain JC501^T varied from *P. alkenifer* DSM 11593^T and *P. sediminis* DSM 2160^T in genes belonging to carotenoid biosynthesis, amino acid (alanine, phenylalanine) metabolism (Fig. 18E) and glutathione, carotenoid and folate biosynthesis pathways, respectively (Fig. 18F).

Table 9. Comparison of the selected genes of the strain JC501^T and its other closely related members of the subclade of the genus *Paracoccus* 1. Strain JC501^T, 2. *P. marinus* NBRC 1000637^T, 3. *P. contaminans* LMG 29738^T, 4. *P. chinensis* KS-11^T 5. *P. solventivorans* DSM 6637^T , 6. *P. alkenifer* DSM 11593^T and 7. *P. sediminis* DSM 26170^T

Putative Genes	1	2	3	4	5	6	7
Typtophan metabolism							
Anthranilate synthase, aminase component (EC 4.1.3.27)	+	+	+	+	+	+	+
Anthranilate synthase, amidotransferase component	+	+	+	+	+	+	+
Anthranilate phosphoribosyltransferase	+	+	+	+	+	+	+
Phosphoribosylanthranilate isomerase	+	+	+	+	+	+	+
Indole-3-glycerol phosphate synthase	+	-	-	+	-	-	-
Tryptophan synthase alpha chain	+	+	+	+	+	+	+
Tryptophan synthase beta chain	+	+	+	+	+	+	+
Aldehyde dehydrogenase	+	+	+	+	+	+	+
Acetamidase (EC 3.5.1.4)	+	-	-	+	-	-	-
Tryptophan 2,3-dioxygenase (EC 1.13.11.11)	+	+	-	-	-	-	-
Kynureninase	-	+	+	+	+	+	+
Carotenoid Biosynthesis							
Phytoene synthase	+	+	-	+	-	-	+
Phytoene desaturase	-	+	-	-	-	-	+
Phytoene dehydrogenase	+	+	-	+	-	-	-
Lycopene cyclase	+	+	-	+	-	-	+
Beta-carotene hydroxylase	+	+	-	-	-	-	+
Beta-carotene ketolase	+	+	-	-	-	-	-
Zeaxanthin glucosyl transferase	-	-	-	-	-	-	+
Urea decomposition							
Urea ABC transporter, ATPase protein *UrtE*	-	-	-	+	-	-	+
Urea ABC transporter, ATPase protein *UrtD*	-	-	-	+	-	-	+
Urea ABC transporter, permease protein *UrtC*	-	-	-	+	-	-	+
Urea ABC transporter, permease protein *UrtB*	-	-	-	+	-	-	+
Urea ABC transporter, urea binding protein	-	-	-	+	-	-	+
Urease accessory protein *UreG*	-	-	-	+	-	-	+
Urease accessory protein *UreF*	-	-	-	+	-	-	+
Urease accessory protein *UreE*	-	-	-	+	-	-	+
Urease alpha subunit (EC 3.5.1.5)	-	-	-	+	-	-	+
Urease beta subunit (EC 3.5.1.5)	-	-	-	+	-	-	+
Urease gamma subunit (EC 3.5.1.5)	-	-	-	+	-	-	+
Urease accessory protein *UreD*	-	-	-	+	-	-	+
Urea channel *UreI*	-	-	-	-	-	+	-
Iron metabolism							
Ferric iron ABC transporter, iron-binding protein	+	-	-	+	+	+	+
Ferric iron ABC transporter, permease protein	+	-	-	+	+	+	+
Ferric iron ABC transporter, ATP-binding protein	-	-	-	+	+	+	+
Siderophore_Aerobactin							
Iron-chelator utilization protein	+	+	+	+	-	-	+
Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), permease component *FhuB*	-	-	+	+	-	-	+
Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), ATP-binding protein *FhuC*	+	+	+	+	-	-	+
Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), periplasmic substrate binding protein *FhuD*	+	+	+	+	-	-	+
Ferrichrome-iron receptor	+	+	+	-	-	-	+
Catechol_branch_of_beta-ketoadipate_pathway							
Catechol 1,2-dioxygenase 1 (EC 1.13.11.1	-	-	-	+	-	-	+
Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)	+	+	+	+	+	+	+
Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)	+	+	+	+	+	+	+
Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)	+	+	+	+	+	+	+

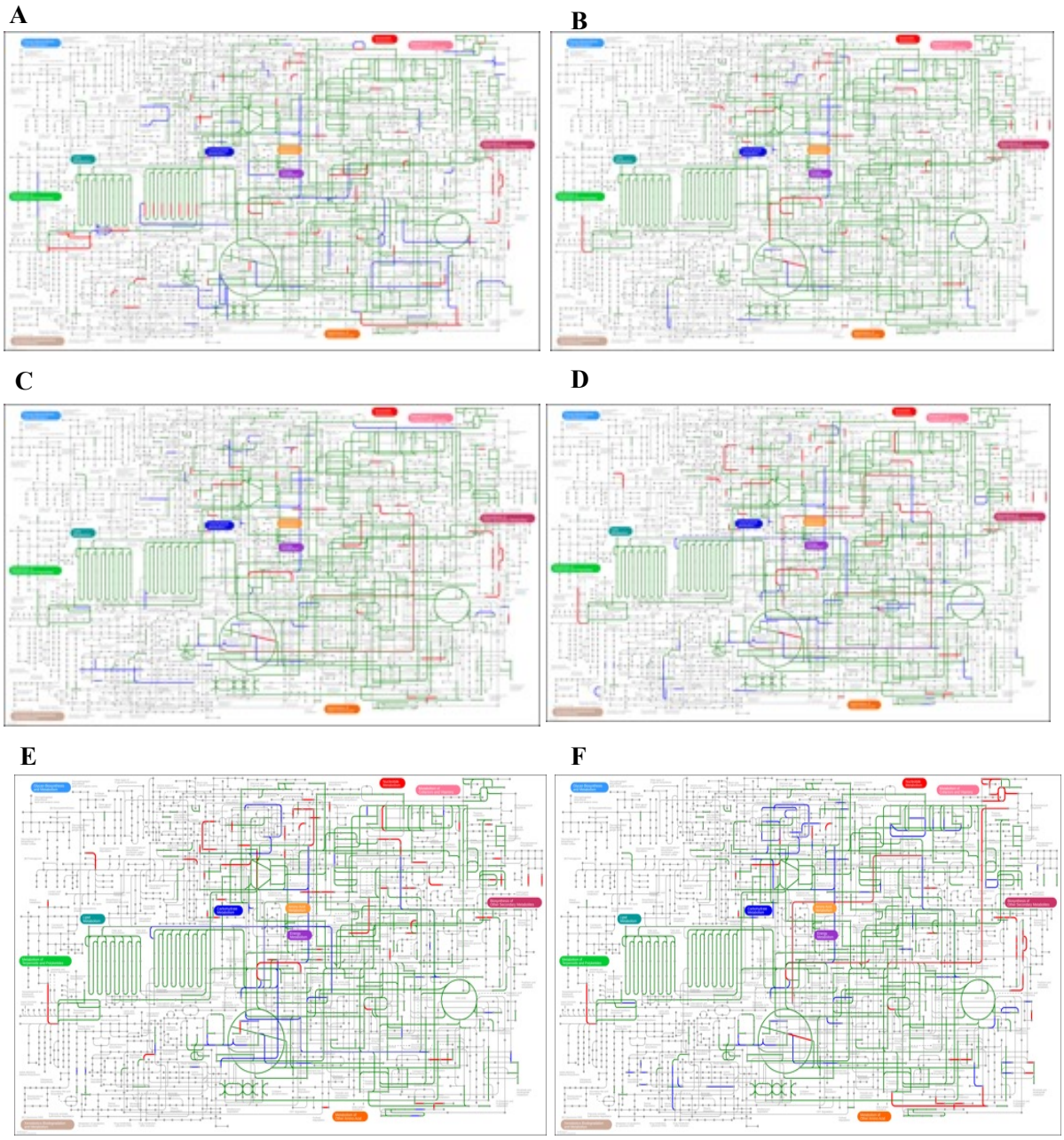


Fig. 18. Common and unique biochemical pathways between strain JC501^T and
(A) *P. marinus* NBRC 100637^T
(B) *P. contaminans* LMG 29738^T
(C) *P. chinesis* CGMCC 1.7655^T
(D) *P. solventivorans* DSM 6637^T
(E) *P. alkenifer* DSM 11593^T
(F) *P. sediminis* DSM 2160^T

3.2.1.5. Physiological and biochemical analysis

Both strains (strain JC501^T and *P. marinus* NBRC 100637^T) could grow well in half-strength nutrient broth, R2A, BP media, and tryptic soy agar (HiMedia M011) at 30 °C for 3 days. Therefore, nutrient broth (1/2 strength) was used for other unless otherwise mentioned. Strain JC501^T could grow well between 25–35 °C and pH 6.0–7.5 (optimum 30 °C and pH 7.0), while *P. marinus* NBRC 100637^T grew at 10–35 °C with pH 6-9 (optimum of 35 °C and pH 7.5). Strain JC501^T did not require NaCl for its growth while tolerated up to 1.0 % (w/v), whereas *P. marinus* NBRC 100637^T tolerated up to 5.0 %. Strain JC501^T could grow well in several of carbon substrates like D-glucose, D-mannose, rhamnose, pyruvate, D-mannitol, D-galactose, maltose, sorbitol, ribose, sucrose, inositol, histidine, and alanine. Weak growth was observed in proline, citrate, trehalose, methanol, acetate, lactose, gluconate, and D-xylose, while no growth was seen in fructose and propionate. Strain JC501^T can utilize only peptone and not ammonium sulphate, sodium nitrate, casamino acids, and sodium glutamate. The carbon substrate utilization of strain JC501^T differed from other members of *P. marinus* clade significantly and is given in Table 10. Both strains did not require additional supplements of vitamins for growth; however, growth yields were enhanced with the addition of yeast extract (0.01 %). The two strains did not hydrolyse gelatin, starch, chitin, urea, cellulose, casein, or Tweens (20/40/80), whereas they negative for H_2S, nitrate reduction, and acid production. Both strains were positive for the activities of catalase and oxidase.

Table 10. Differential characteristics showing the differences between 1.Strain JC501^T 2. *P. marinus* NBRC 100637^T 3. *P. contaminans* LMG 29738^T 4. *P. haematiqui* M1-83^T(Kämpfer et al. 2019) 5. *P. niistensis* NII-0918^T(Dastager et al. 2011)

Characteristics	**1***	**2***	**3***	**4**	**5**
Source	Roots of *Aerides sp.*	Coastal sea-water	Contaminated axenic water	Horse blood	Forest Soil
Colony colour	Shiny and bright orange	Dull orange	Beige	Orange coloured	Vivid-orange
Cell diameter (µm)	0.9-1	0.5-0.8	1-1	1-1	0.5-0.7
Temperature (˚C) Range (Optimum)	25-35(30-35)	10-35(25-35	10-45 (28-30)	10-45 (28-30)	10-40 (28-30)
pH range (Optimum)	6-7(7)	6-9 (7-8)	5-7.5 (6-7.5)	5-7.5(6-7.5)	6-12 (7-8)
Growth range for NaCl (%)	0-1	>5	0-1	0-1	0-7
Indole production	+	-	+	nd	-
Urea Hydrolysis	-	-	-	-	+
Carbon substrate					
Maltose	+	-	+	+	nd
Sucrose	+	-	-	-	-
Inositol	+	-	-	+	-
Mannitol	+	-	-	+	-
D-Sorbitol	+	-	+	-	-
Lactose	w+	-	+	nd	-
Acetate	w+	-	+	+	-
Mannose	+	-	+	+	-
Alanine	+	+	+	-	-
Cellular Fatty acid					
$C_{18:1}\omega7c$/ $C_{18:1}\omega6c$	3^+	3^+	3^+	3^+	3^+
$C_{10:0}$3-OH	(-)	1^+	(-)	(-)	(-)
$C_{13:0}$	(-)	(-)	(-)	(-)	1^+
$C_{16:0}$	2^+	2^+	(-)	1^+	(-)
$C_{17:0}$	2^+	tr	(-)	1^+	Tr
$C_{18:0}$	1^+	2^+	2^+	2^+	2^+
$C_{19:0}$	1^+	1^+	(-)	(-)	Tr
anteiso-$C_{13:0}$	2^+	2^+	(-)	(-)	1

*;data obtained are performed in the authors laboratory,

`; genomic G+C; Phoshpatidylcholine, diphosphatidylglycerol, phosphatidylglycerol and unidentified aminolipids are the major polar lipids in all the strains. Organic substrate utilization was tested for chemoorganoheterotrophic growth. Glucose, galactose and pyruvate were utilized by all the strains. Propionate was not utilized by any of the strains. Sodium nitrate, ammonium sulphate, sodium glutamate and casamino acids were not utilised as nitrogen sources by the strains except for the peptone. All strains were negative for the hydrolysis of starch, gelatin, Tween 20, 40, 80, chitin and casein; negative for H_2S production and NO_3 reduction; +, good growth; -, no growth; w+, weak growth; 1^+,1-5; 2^+,5-15;3^+, >15; (-), absent; tr; trace, nd; no data available

3.2.1.6. Cellular fatty acid, polar lipid, and quinones

The major fatty acids, which are >20 % of the total fatty acids observed in strain JC501^T were $C_{16:0}$, $C_{17:0}$, summed feature 8 ($C_{18:1}$ ω7c and/or $C_{18:1}$ ω6c), and anteiso-$C_{11:0}$. The fatty acid profile differed variably between strain JC501^T and other members is presented in Table 10. Major polar lipids were phosphatidylcholine, phosphatidylglycerol, glycolipid, diphosphosphotidylglycerol, four unidentified phospholipids (PL1, 2, 3, 4), two unidentified amino lipids (AL1, 2), and two unidentified lipids (L1, 2) (Fig. 19A). Strain JC501^T and *P. marinus* NBRC 100637^T differed with the presence of unidentified amino lipid (AL2) and two unidentified lipids (L1, 2) in strain JC501^T (Fig. 19A, B). Both strains had ubiquinone 10 as the major isoprenoid quinone.

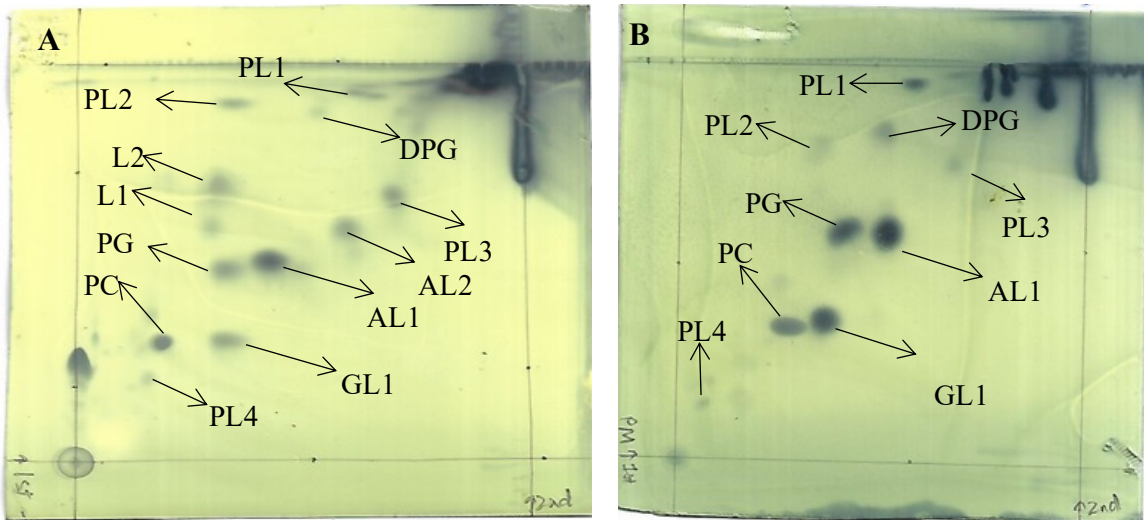


Fig. 19. Two dimensional chromatograms showing polar lipids of
(A) strain JC501^T
(B) *P. marinus* NBRC 100637^T
For polar lipids, chloroform-methanol-water (75:32:4 [v/v]) was used as the solvent in the first direction, and chloroform-acetic acid-methanol-water (86: 16: 15:4 [v/v]) was used in the second direction. PG, phosphatidylglycerol; DPG, diphosphotidylglycerol; PC, phosphatidylcholine; AL1,2, unidentified aminolipids; PL1-4, unidentified phospholipids; GL1, unidentified glycolipid, L1-2, unidentified lipid.

3.2.1.7. Carotenoid, hopanoids, and indole metabolism

For carotenoid analysis, seven HPLC chromatogram peaks were observed for strain JC501^T of which peak 1 was adonixanthin diglucoside and peak 5 to β-carotene (Fig. 20). Adonixanthin diglucoside was also observed in *P. marinus* NBRC 100637^T, whereas β-carotene and unidentified peaks 5–7 were absent. *P. marinus* NBRC 100637^T, tested negative whereas strain JC501^T tested positive for indole production from L-tryptophan (Fig. 21A). The variation in HPLC chromatogram of strain JC501^T between the control (L-tryptophan absence) and test (L-tryptophan present) is shown in Fig. 21B. In the test, 4 indole-derived molecules were identified based on UV absorption spectra; of these peak 4 corresponded as indole 3-acetic acid, whereas others remained unidentified (Fig. 21B). Further, the LC-MS analysis of the indolic extract could identify molecules such as 3-methyleneoxindole, 3-indolepropionic acid, indole 3-methylacetate, and 3-methyoxindole (based on retention time, m/z, mass spectral data) (Fig. 21C) (Table 11). These results were further supported by the genome studies given above.

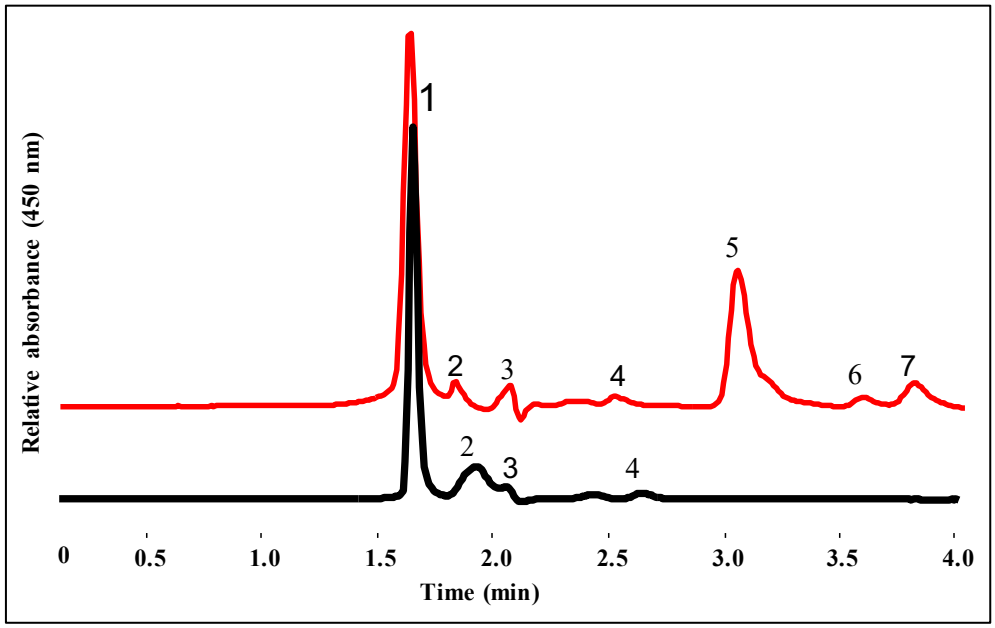


Fig. 20. HPLC chromatogram showing carotenoid peaks of strain JC501^T (red line) and *P. marinus* NBRC 100637^T (black line). 1, adonixanthin diglucoside, 2-4,6,7 as unidentified carotenoids; 5, β- carotene

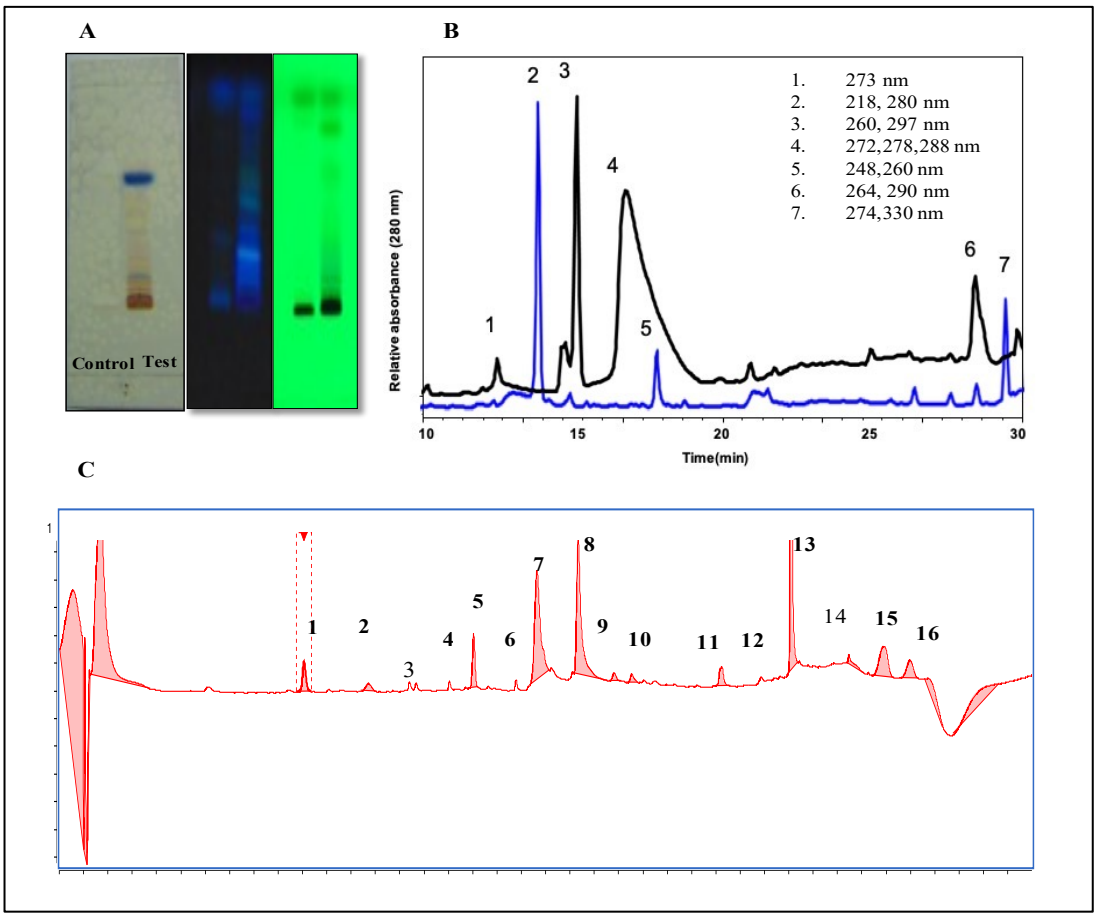


Fig. 21. Strain JC501^T tested positive for indole test whereas *P. marinus* NBRC 1000637^T tested negative

(A) TLC plate developed using Van Urk-Salkowski reagent after the sample (Strain JC501^T control vs treated)

(B) HPLC chromatogram showing L-tryptophan catabolites produced by strain JC501^T (black line); control (L-tryptophan un-fed; blue line). 1-7, absorption maxima (nm) taken from the absorption spectra of the peak. Peaks which have not shown absorption spectra are not labelled. 1-3, 5-7, unidentified metabolites; 4, indole-3 acetic acid

(C) LC-MS was done by the ESI-Q-ToF. 16 peaks were observed with

Table 11. LC-MS data of some indole derivatives found in strain JC501^T

Peak No.	Retention time	m/z	Compound formula	Compound name	Mass Spectral
1	9.923-10.230	145	C_9H_7NO	3-Methyleneoxindole	146.0600, 118.0650
2	12.458-12.958	189	$C_{11}H_{11}NO_2$	3-Indolepropionic acid	190.0857, 130.0645
6	16.914-17.124	189	$C_{11}H_{11}NO_2$	Indole 3-methylacetate	190.0860, 130.0650
9	22.645-23.033	147	C_9H_9NO	3-Methyloxindole	149.0261, 105.0372

3.2.2. Characterization of strain JC701^T

3.2.2.1. Home habitat, colony, and cell morphology

Strain JC701^T was isolated from a root sample of a flowering epiphytic orchid. It was collected in March 2020 from Darjeeling, West Bengal (GPS: 27° 07′ 69″ N, 88° 26′ 47″ E), India (Fig. 22A). When grown ion nutrient agar, strain JC701^T colonies were 0.5–1 mm in diameter, translucent, circular, raised, smooth and pale yellow (HiMedia M001). Cells of strain JC501^T were rod-shaped (0.8–1 µm) and divided by binary fission (Fig. 22B, C). Cells were Gram-stain-positive, non-motile, and non-endospore forming.

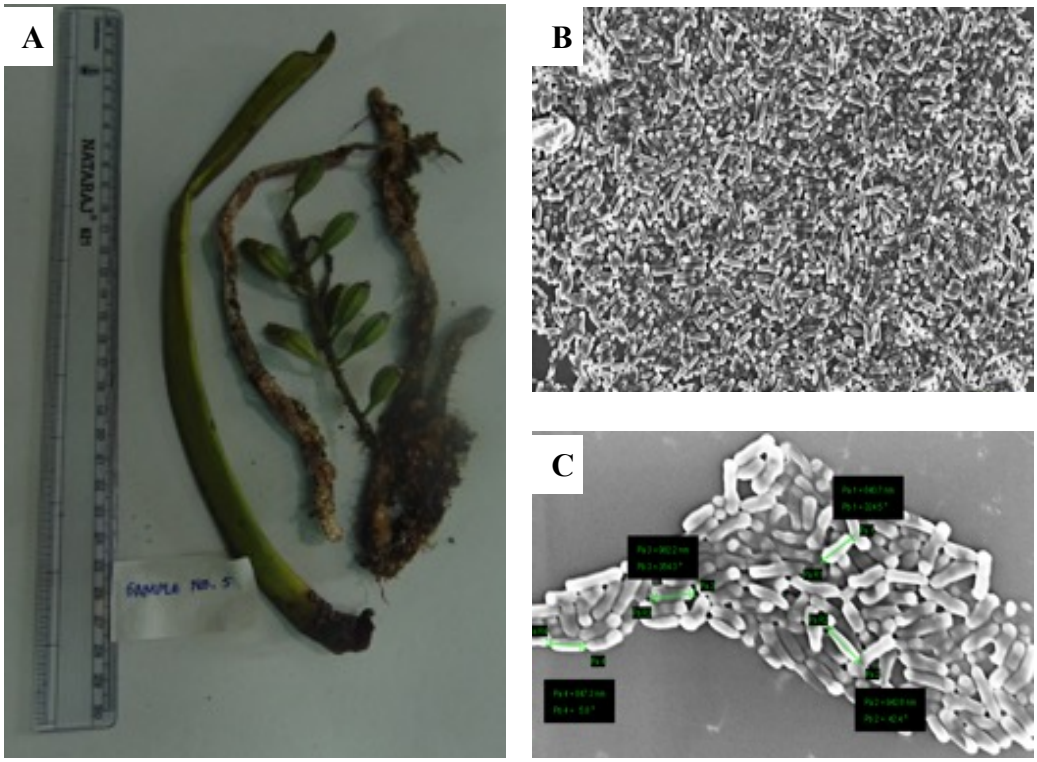


Fig. 22. (A) An epiphytic orchid (post-flowering) which was collected from Darjeeling, West Bengal
(B) SEM image of the cells of the culture grown in NA media for 4 days at 30˚C (C) Rod shaped cells of strain JC701^T (size, 0.8-1 µm)

3.2.2.2. Phylogeny based on 16S rRNA and whole genome

The analysis of the 1552 bp length of the 16S rRNA gene by BLAST analysis (EzBioCloud) showed that strain JC701^T is phylogenetically affiliated to the genus *Microbacterium,* with its nearest type strain being *M. testaceum* NBRC 12675^T (99.3 %) and other members of the genus *Microbacterium* (<98.7 %). The 16S rRNA gene-based phylogenetic tree suggested that strain JC701^T formed a distinct clade with members like *M. testaceum*, *M. hydrothermale*, *M. resistens* and *M. kyungheense* (Fig. 23). Likewise, a comparable clustering order was observed when the phylogenetic tree was constructed (92 concatenated sequences) by the UBCG tool (Fig. 24).

3.2.2.3. Genomic characterization

The genome size of strain JC701^T and *M. testaceum* NBRC 12675^T was 3.5 Mbp and 3.6 Mbp, respectively. The genomic characteristics are further shared in Table 12 and clearly distinguish strain JC701^T from *M. testaceum* NBRC 12675^T. Similarly, DNA G+C content was 70.5 mol % for strain JC701^T,whereas for *M. testaceum* NBRC 12675^T was 69.7 mol %. The *d*DDH, ANI, and AAI values between the two strains (strain JC701^T and *M. testaceum* NBRC 12675^T) were 29.4 %, 85.2 %, and 86.6 %, respectively. Further, the OrthoVenn2 results showed that 2573 clusters were shared between both strains, with 22 unique clusters belonging to strain JC701^T and 26 clusters belonging to *M. testaceum* NBRC 12675^T. 22 unique clusters of strain JC701^T were grouped into biological processes of cellular, nitrogen compound, aromatic compound, and RNA metabolism. For *M. testaceum* NBRC 12675^T, 26 unique clusters were grouped into carbohydrate, cellular, drug, developmental, macromolecule, and heterocyclic compound metabolism. The anti-SMASH software predicted four biosynthetic gene clusters for secondary metabolites, which were beta-lactone (microansamycin), terpene (carotenoid), T3PKS (herboxidiene), and NAPAA (non-alpha poly-amino acids) like qinichelins.

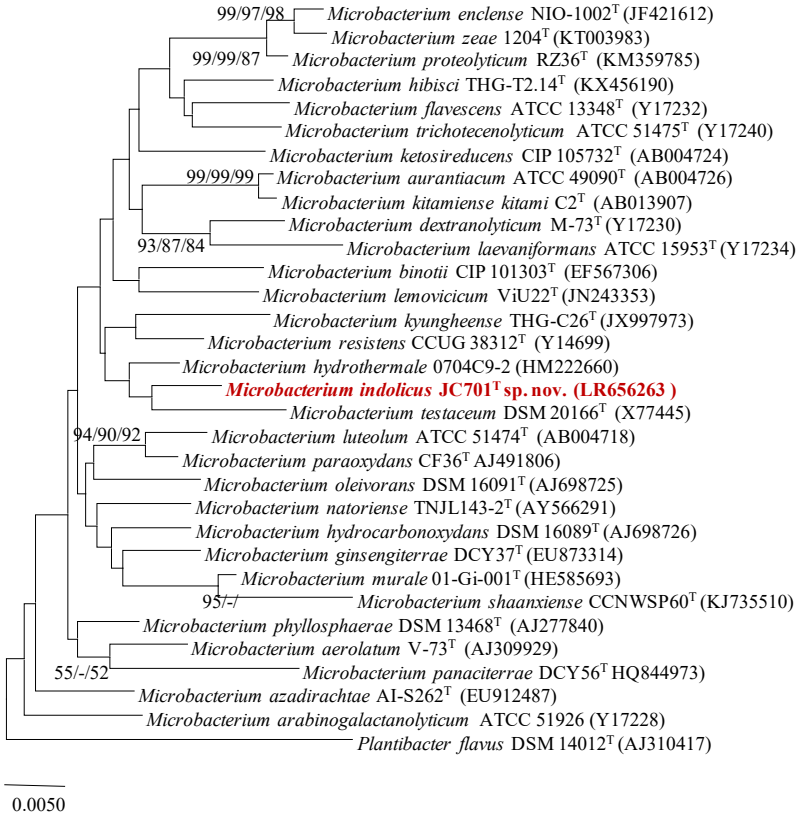


Fig. 23. Phylogenetic tree based on 16S rRNA gene sequences showing affiliation of *Microbacterium* sp. JC701^T with the other members of the genus *Microbacterium* using NJ algorithm and additional algorithms (ML and MP) *Plantibacter flavis* DSM 14012^T (AJ310417) is an outgroup.

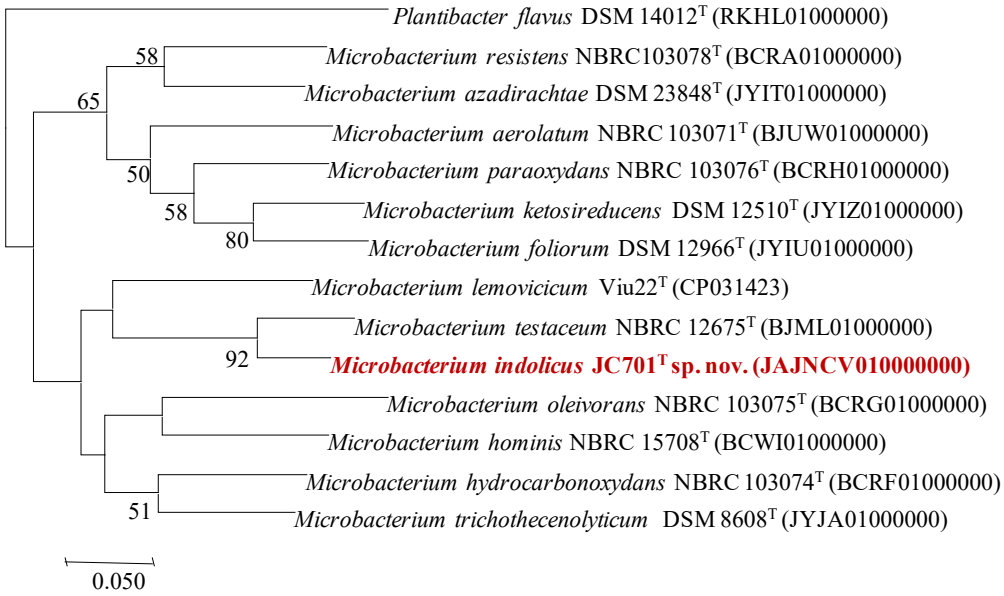


Fig. 24. Phylogenomic tree based on 92 concatenated genes (UBCG tool) showing affiliation of *Microbacterium* sp. JC701^T with the other members of the genus *Microbacterium*. *Plantibacter flavis* DSM 14012^T (RKHL00000000) is an outgroup.

3.2.2.4. Physiological and biochemical analysis

Both strain JC701^T and *M. testaceum* NBRC 12675^T could grow well in nutrient broth and BP media at 30 °C within 3 days. Strain JC701^T could grow well in the range of 14 to 37 °C (optimum 30 °C), whereas *M. testaceum* NBRC 12675^T could grow in a range of 20 to 37 °C (optimum 30 °C). As for pH, strain JC701^T, and *M. testaceum* NBRC 12675^T required an optimum pH 7 (pH range 5.0 to 8.0) and 7.5 (range of 8 to 10), respectively. Strain JC701^T did not require NaCl for its growth while tolerated up to 6.0 % (w/v), whereas *M. testaceum* NBRC 12675^T tolerated up to 7.0 %. Strain JC701^T showed positive growth in a number of carbon sources like D-glucose, ribose, sorbitol, sucrose, inositol, pyruvate, arabinose, lactose, fructose, galactose, maltose, and rhamnose (Table 12), whereas no growth was observed with carbon sources like mannose. Strain JC701^T could utilize both peptone and ammonium sulphate as nitrogen sources. The carbon substrate utilization by strain JC701^T differed greatly from *M. testaceum* NBRC 12675^T and are presented in Table 12. Both strains did not require additional supplements of vitamins for growth; however, growth yields were enhanced with an addition of yeast extract (0.01 %). The two strains under the study could not hydrolyse starch and gelatin; they were similarly negative for H_2S production. Both strains were positive for oxidase activity. The differential characteristics for both are presented in Table 12.

Table 12. Differential characteristics between 1. Strain JC701^T 2. *M. testaceum* NBRC 100637^T

Characteristics	1	2
Colony morphology	Yellow, shiny	Orange
Isolation Source	Orchid roots	Rice grain
Motility	Non-motile	Motile
Catalase	-	+
Cell shape	Rod	Coccus
Cell size (µm)	1	0.9-1
Optimum Temperature (°C) (range)	30 (14-37)	32 (20-37)
Optimum pH (range)	7 (5-8)	7.5 (8-10)
NaCl concentration	0-6 %	0-7 %
Carbon substrates		
D-Glucose	+	-
D-Ribose	+	-
D-Sorbital	+	-
Sucrose	+	-
Inositol	+	-
Mannose	-	+
Arabinose	+	-
D-Lactose	+	-
Rhamnose	+	-
NO_3 reduction/Indole	-/+	+/-
Cellular Fatty Acid		
anteiso-$C_{15:0}$	34.6	43.3
anteiso-$C_{17:0}$	29.1	27.2
iso-$C_{16:0}$	22.5	23.1
iso-$C_{14:0}$	0.9	0.8
iso-$C_{17:0}$	5.6	2.1
iso-$C_{15:0}$	4.5	2.5
$C_{16:0}$	2.1	0.5
anteiso-$C_{17:1}$ A	0.8	-
Genome characteristics		
Genome Size (Mbp)	3.5	3.6
Scaffolds	9	28
G+C content (mol %)	70.5	69.7
DDBJ/EMBL/GenBank	JAJNCV00000000	BJML00000000

Both strains are oxidase positive. Organic substrate utilization was tested for chemoorganoheterotrophic, for both strains positive growth in fructose, galactose, maltose (as carbon substrate; positive growth in ammonium sulphate and peptone (nitrogen source). Both the strains are H_2S production negative, gelatin and starch hydrolysis negative

3.2.2.5. Cellular fatty acid, polar lipid, quinone, and indole analysis

The major fatty acids, which is >20 % of the total fatty acids in strain JC701^T were anteiso-$C_{15:0}$, anteiso-$C_{17:0}$, and iso $C_{16:0}$. The differences in the fatty acid profile between strain JC701^T and *M. testaceum* NBRC 12675^T are in Table 12. The major polar lipids of strain JC701^T was phosphatidylglycerol, diphosphosphotidylglycerol, an unidentified phospholipid, and three unidentified glycolipids (GL1, 2, 3) (Fig. 25A). Strain JC701^T differed from *M. testaceum* NBRC 12675^T in having an unidentified phospholipid and unidentified glycolipid. The major isoprenoid quinone in both were menaquinone-10 and menaquinone-11. The major polyamine is spermidine. Strain JC701^T tested positive for indole production from L-tryptophan, whereas *M. testaceum* NBRC 12675^T tested negative (Fig. 25B). The differences in HPLC chromatogram of strain JC701^T between the control (L-tryptophan absence) and test (L-tryptophan present) are shown in Fig. 25B. In the test sample, 4 indole-derived molecules were identified on UV absorption spectra. Peak 3 was identified as indole 3-acetic acid, whereas others (peak 1, 2, 4) are unidentified (Fig. 25B).

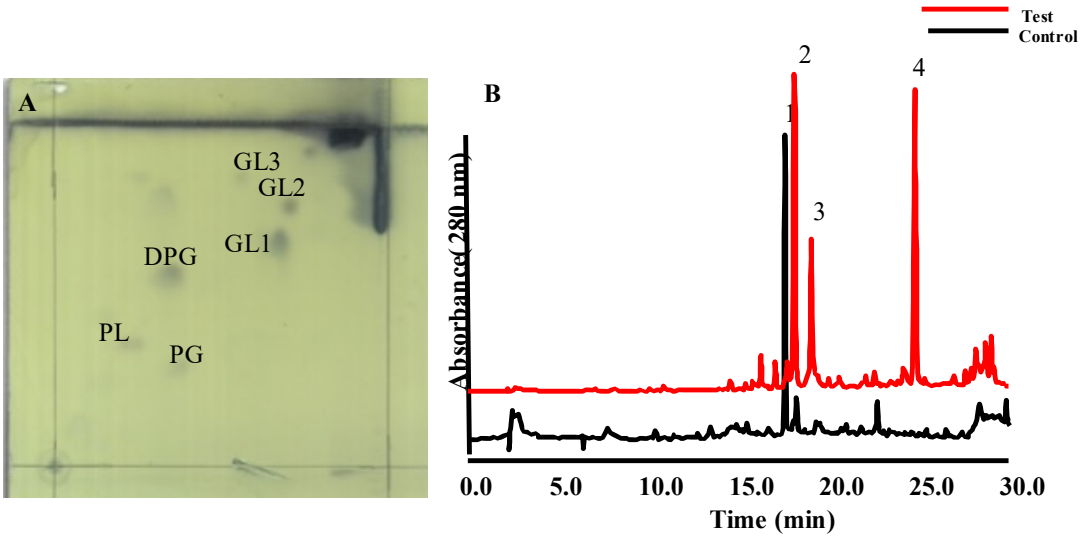


Fig. 25. (A) 2 dimensional chromatograms showing polar lipids of strain JC701^T: PG, phosphatidylglycerol;
(B) HPLC chromatogram showing L-tryptophan catabolites produced by strain JC701^T (red line), control (L-tryptophan un-fed; black line) at 280 nm. Peaks which have not shown absorption spectra are not labelled. 1-2, 4, unidentified metabolites; 3, indole-3 acetic acid

3.3.3. Characterization of strain JC1013^T

3.3.3.1. Home habitat, colony, and cell morphology

In May 2018, an orange-colored pond found in a saltern area on the coast of Tamil Nadu, India (Fig. 26A) was chosen as the sampling source (home habitat) for bacterial isolation (GPS positioning: 10°21′11″N and 79°50′41″E). At the time of sample collection, the pond's pH and salt concentrations were 7.5 and 80 g.l^{-1}, respectively. After a few days, orangish pink colonies developed, which were purified by streaking on nutrient agar (NA) containing 2 % (w/v) of NaCl. Glycerol stocks were made from pure culture, kept at -80 °C. Strain JC1013^T was cultivated in NB (2% NaCl) unless otherwise mentioned. JC1013^T colonies were 2 mm in diameter, orange in hue, dispersed irregularly, and flat (Fig. 26B). Gram-positive, endospore-forming cells are strain JC1013^T (Fig. 26C). The cells of strain JC1013^T were rod-shaped and non-motile, measuring 2–3 m in length and 0.1–0.2 m in breadth (Fig. 26D, E). In nutrient agar (HiMedia M001), tryptic soy agar (HiMedia M011), and BF media, strain JC1013^T grew well.

3.3.3.2. Phylogeny based on 16S rRNA gene and whole genome

EZBioCloud BLAST study of the 16S rRNA gene sequence (1445 bp) indicated that strain JC1013^T had the highest identity (%) with the members of the genus *Mesobacillus*; *M. selenatarsenatis* SF-1^T (98.7 %) and *M. subterraneus* DSM 13966^T (97.8 %) being the nearest members. According to the 16S rRNA gene-based phylogenetic neighbour joining tree, strain JC1013^T formed a separate cluster with its closely related species of the genus *Mesobacillus,* which was further combined with ME and ML (Fig. 27). JC1013^T. This clustering order was comparable to the RAxML-based UBCG phylogenomic tree (Fig. 28). Strain JC1013^T was also confirmed as a "possible new species" by the Type (Strain) Genome server (TYGS) (http://tygs.dsmz.de).

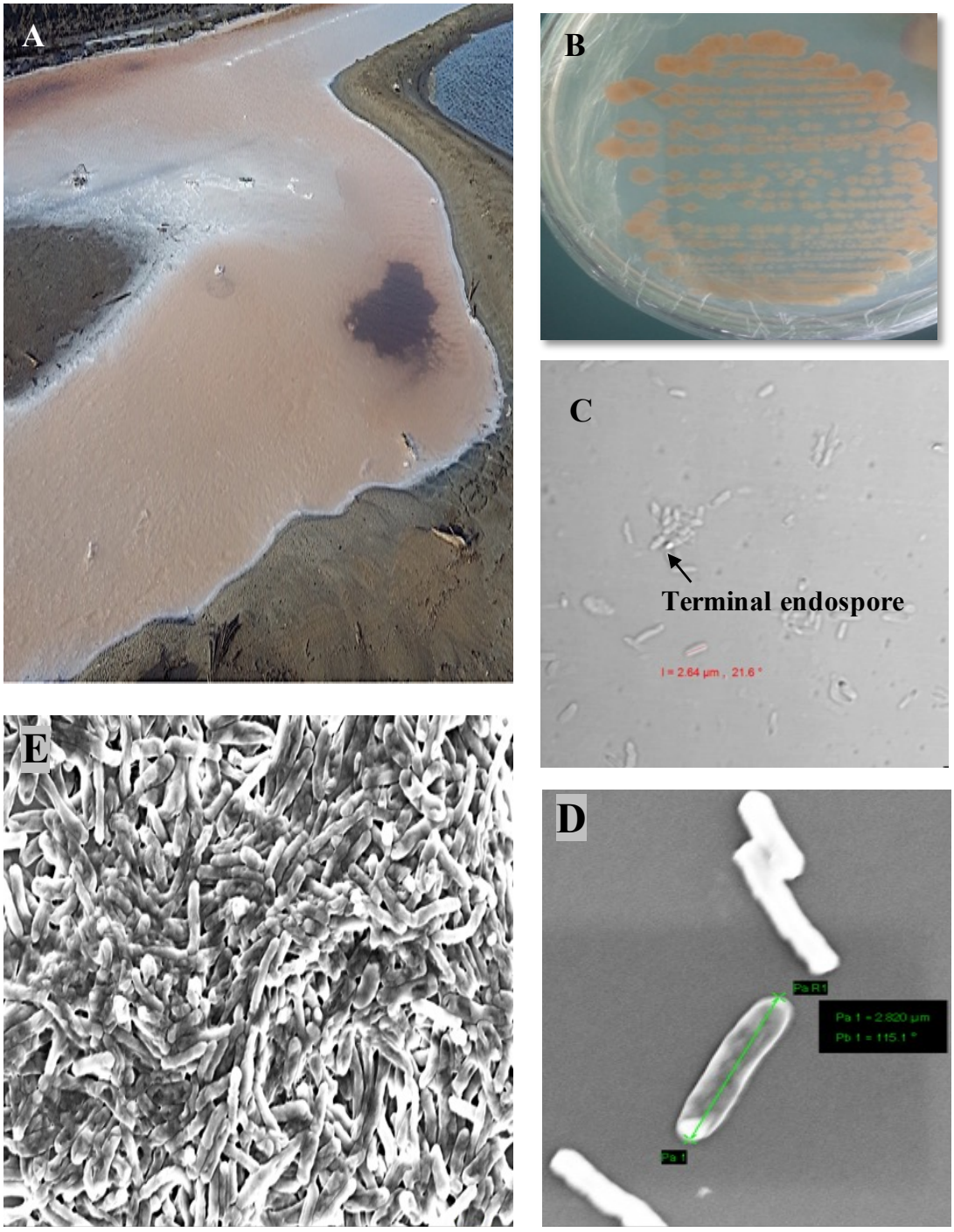


Fig. 26. (A) Sampling site
(B) Colony morphology of strain JC1013^T in NA
(C) Confocal microscopic image of strain JC1013^T
(E) Cell image of strain JC1013^T under FESEM
(F) Rod-shaoed morphology of strain JC1013^T

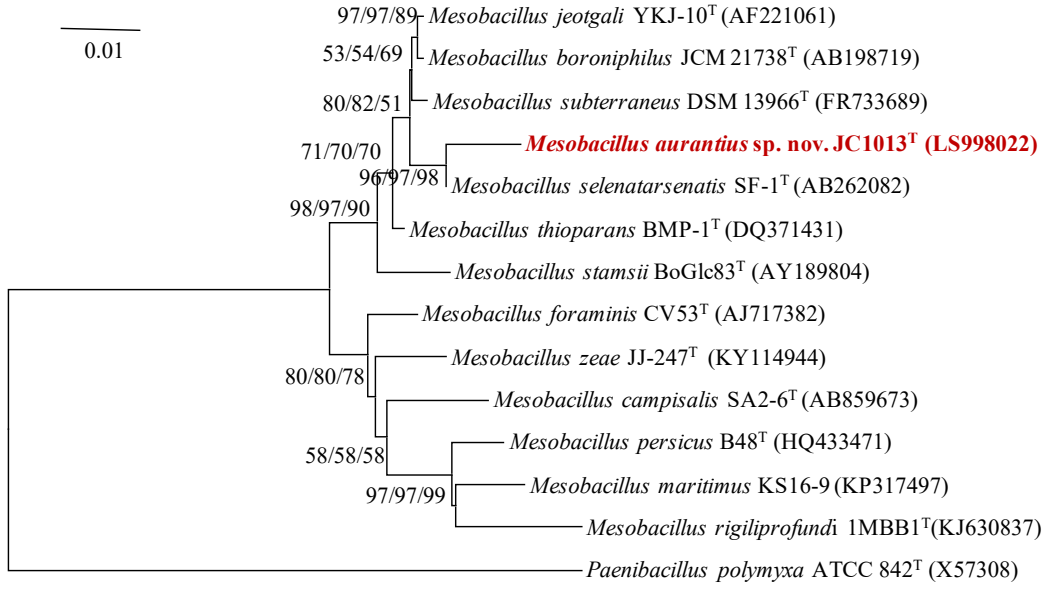


Fig. 27. Phylogenetic tree based on 16S rRNA gene sequence comparisons, using the Neighbor-joining algorithm (NJ) and the additional algorithms like ME and ML. MEGA7 software was used for computing. *Paenibacillus polymyxa* ATCC 842^T (AFOX01000032) was used as an outgroup. Numbers at the nodes represent the bootstrap values. On the nodes Bar, 0.01 accumulated changes per nucleotide.

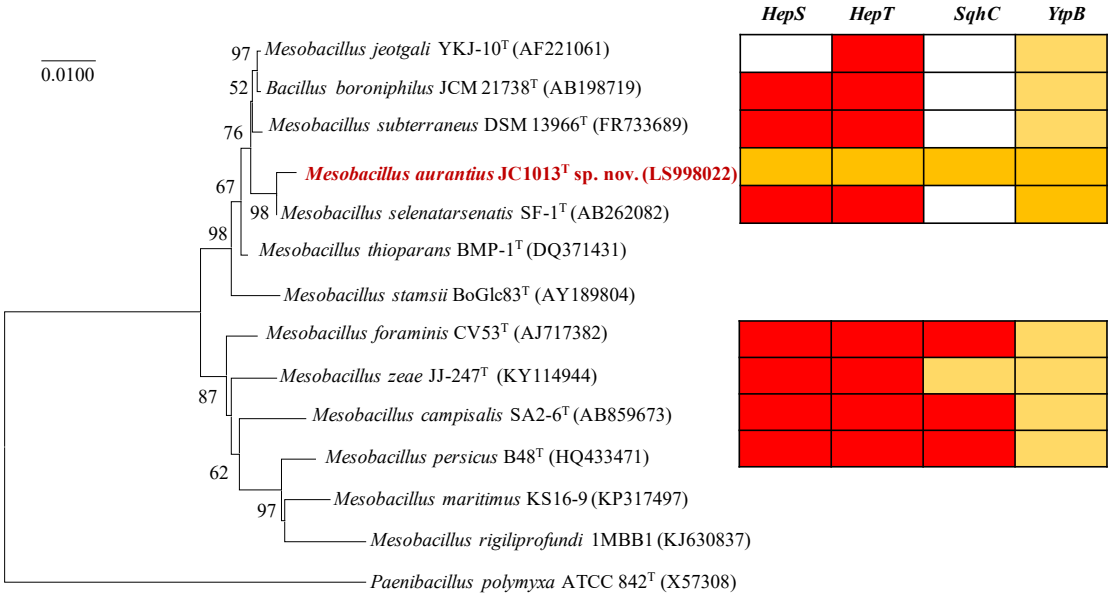


Fig. 28. The phylogenomic tree of genus *Mesobacillus* based on 92 concatenated genes representing the distribution of sporulene synthesis genes. *HepS*- Heptaprenyl diphosphate synthase component I; *HepT*- Heptaprenyl diphosphate synthase component II; *Sqhc*- Sporulenol synthase (Squalene hopene cyclase); *YtpB*- Tetraprenyl beta curcumene synthase; Squares with color refers to identity of genes involved in sporulene synthesis compared to *B. subtilis* genes. ≥50 % (red color); <50 % (yellow color); Not present (white).

3.3.3.3. Genomic characterization

Strain JC1013^T had a genome size of 4.6 Mbp, 43 scaffolds, and an N_{50} value of 2,17,424 bp. Strain JC1013^T and *M. selenatarsenatis* SF-1^T have 42.7 % and 42.1% G + C content in their genomes, respectively. Table 13 lists the genomic features of strain and other closely related relatives. Between strain JC1013^T and *M. selenatarsenatis* SF-1^T, the ANI and AAI scores were 84.6 %, and 88.5 %, respectively. For these two strains, the *d*DDH levels were 29 %. With other members, the ANI, AAI, and *d*DDH values were below 28.9 %, 88.3 % and 84.5 %, respectively. When compared to the other genomes, the genomic arrangement of strain JC1013^T looked to be reordered and inverted (perhaps because of rearrangements, recombination, and horizontal transfer) (Fig. 29A). The LCBs of the *Mesobacillus* genus varied greatly from one another (Fig. 29A). The PATRIC server's proteome comparison service was also used to compare the proteome map based on protein sequence identity. The results revealed that the bulk of the proteins in the strain JC1013^T are 10–90 % compared to those in other *Mesobacillus* species (Fig. 29B). This indicated that strain JC1013^T was distinct from other *Mesobacillus* strains.

Table 13. Genome characteristics of the strain JC10131^T, 2, *M. selenatarsenatis* SF-1^T ;3, *M. subterraneus* DSM13966^T; 4, *M. jeotgali* DSM18226^T; 5, *M. boroniphilus* JCM 21738^T

Species	**Strain**	**Genome Size (Mbp)**	**G+C content (mol %)**	**DDBJ/EMBL/ GenBank**
Strain JC1013^T	=NBRC 114146^T =KACC 21451^T	4.6	42.7	JAAVUM00000000
M. selenatarsenatis	=JCM 14380^T =DSM18680^T	4.8	42.8	BASE00000000*
M. subterraneus	=ATCC 136^T =DSM13966^T	4.2	43.9	JXIQ00000000*
M. jeotgali	=KCCM 41040^T =JCM 10885^T	4.7	41	CP025025*
M. boroniphilus	=DSM 17376^T =ATCC BAA 1204^T	4.4	42	BAUW0000000*

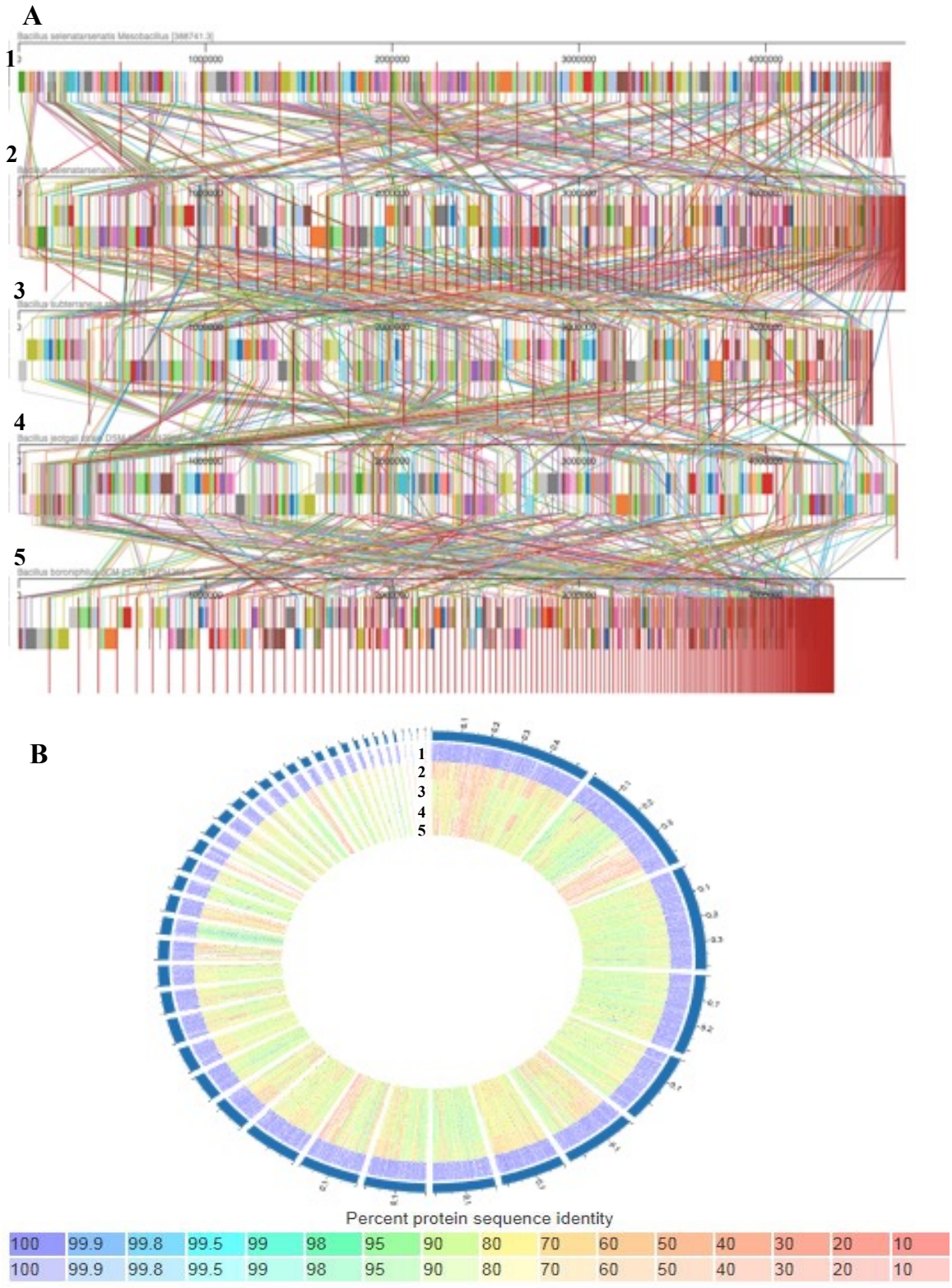


Fig. 29. (A) Genome organization of *Mesobacillus* spp. Genome of strain JC1013^T was used as a reference
(B) Genome map based on protein sequence identity between strain JC1013^T and other closely related species of *Mesobacillus*. 1. Strain JC1013^T; 2. *M. selenatarsenatis* SF-1^T; 3, *M. subterraneus* DSM 13966^T; 4, *M. jeotgali* DSM 18226^T; 5, *M. boroniphilus* JCM 21738^T

Further, strain JC1013^T, *M. selenatarsenatis* SF-1^T, *M. subterraneus* DSM 13966^T, *M. jeogtali* DSM 18226^T, and *M. boroniphilus* JCM 21738^T were examined for pan-genome analysis (Fig. 30). A total of 6023 genes were evaluated for this study, with 2154 designated as core genes. *M. selenatarsenatis* SF-1^T and strain JC1013^T possess 482 and 118 accessory genes. The number of accessory genes in *M. subterraneus* DSM 13966^T, *M. jeogtali* DSM 18226^T, and *M. boroniphilus* JCM 21738^T are given in Fig. 30A. The pan and core of strain JC1013^T under investigation differed significantly from its related members (Fig. 30B, C).

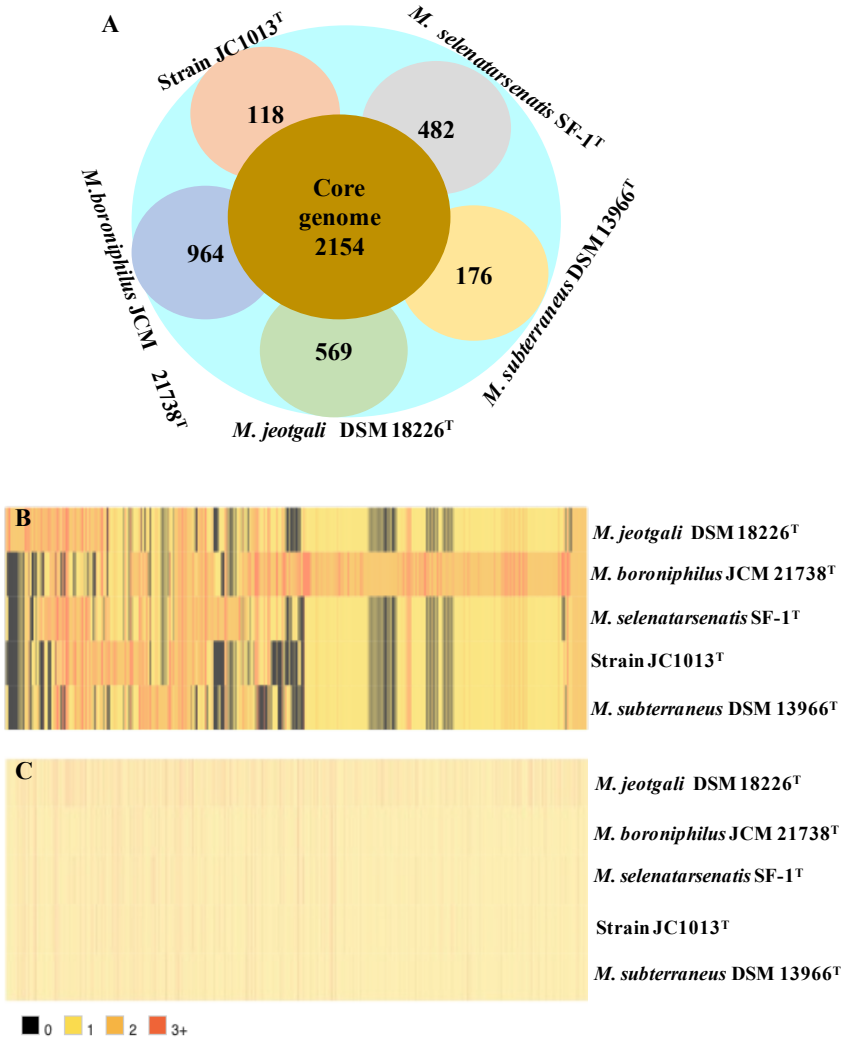


Fig. 30. Pan genome analysis of strain JC1013^T and its closely related members
(A) Diagrammatic representation of the core genome and accessory genome (outer circle),
(B) Bar plot showing the pan-genome of strain JC1013^T and closely related members
(C) Bar plot representing the core genomes of strain JC1013^T and closely related members

3.3.3.4. Metabolic characterization

Only strain JC1013^T has all of the genes encoding the sporulene production pathway, according to the in-depth genomic investigation of *Mesobacillus* members (Fig. 28). The important enzyme sporulenol synthase (EC 4.2.1.137) was missing among the other 4 members (*M. selenatarsenatis, M. jeogtali, M. subterraneus, and M. boroniphilus*). Sporulenes are a type of triterpenoid found in *Bacillus subtilis* spores. Their function is unknown; however, strain JC1013^T could be a possible generator of sporulenes. Furthermore, strain JC1013^T was distinguished from other *Mesobacillus* species by the presence of genes coding for styrene degradation, carotenoid biosynthesis, benzoate degradation, and xylene degradation. The gene encoding catechol 2, 3-dioxygenase (EC 1.13.11.2) in strain JC1013^T alone indicates that it can grow on aromatic hydrocarbons. The presence of carotenoid biosynthetic genes in strain JC1013^T can be ascribed to the existence of coloured pigments, whereas absent in *M. selenatarsenatis* SF-1^T.

3.3.3.5. Conserved signature indels and carbohydrate-active enzyme annotation

CSIs of two amino acid insertion in PhoH family protein (Fig. 31A), one amino acid deletion in YpjP family protein and one amino acid insertion in TrkA family potassium uptake protein (Fig. 31B) were also found in strain JC1013^T. (Fig. 31C). The annotated data of carbohydrate-active enzymes was also utilised to create Venn diagrams (Fig. 32). Six glycosyltransferases (GT) families, four carbohydrate esterase (CE) families, twenty-three glycosyl hydrolase (GH) families, five carbohydrate-binding modules (CBM) families, and two auxiliary activity (AA) families were discovered after annotating all of the genomes (Fig. 32A). Two families, AA2 of the auxiliary activity family (Fig. 32B) and CE1 of the carbohydrate esterase family, were unique to strain JC1013^T (Fig. 32C). Glycosyl-o-hydrolase (Fig. 32D), carbohydrate-binding modules (Fig. 32E), and glycosyltransferases are among the other type of enzyme families represented (Fig. 32F).

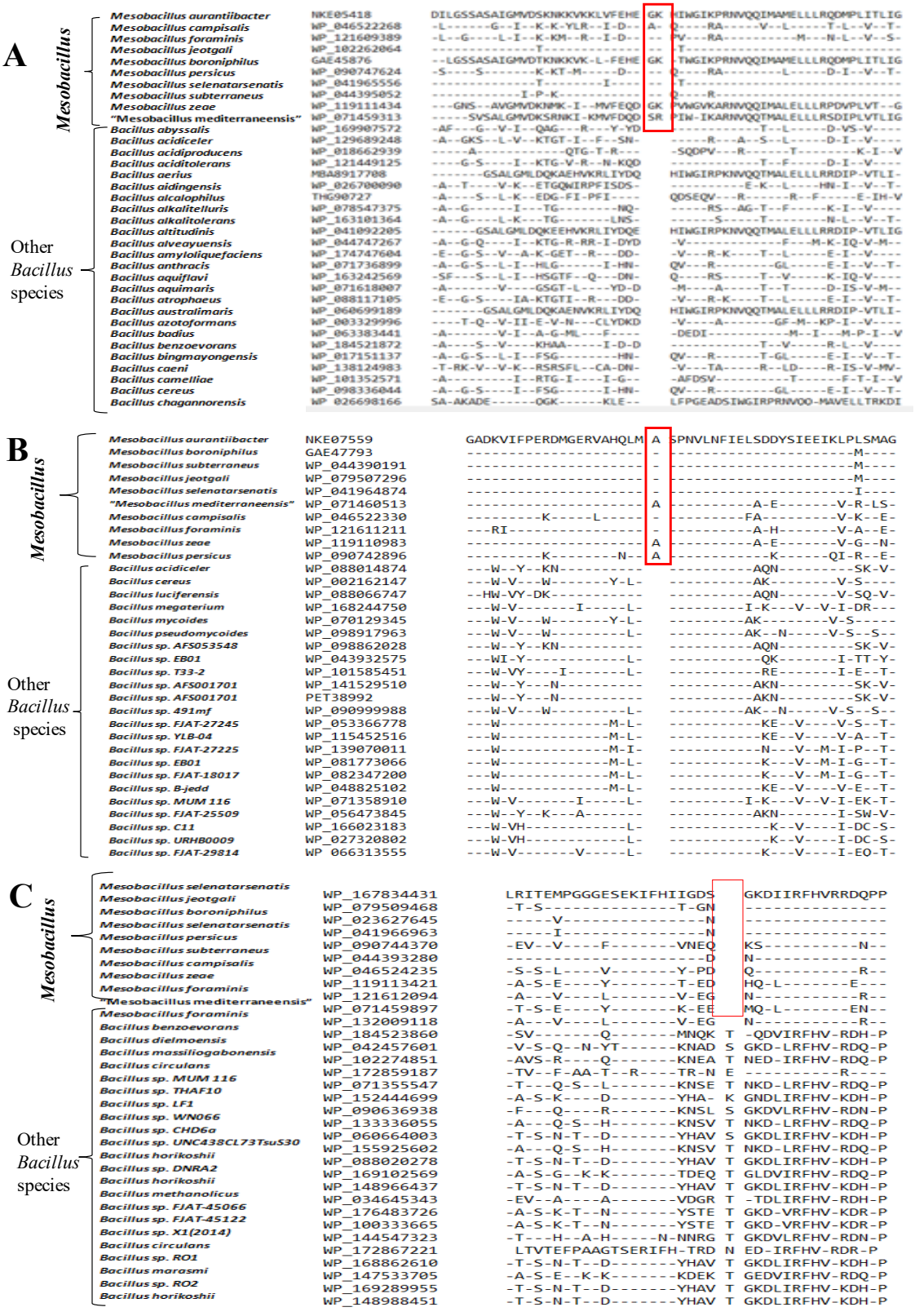


Fig. 31. (A) Partial sequence alignments of PhoH (phosphate starvation-induced) family protein showing a two amino acid insert
(B) TrkA family (potassium uptake protein) containing a one amino acid insertion
(C) YpjP (uncharacterized) family protein showing a 1 amino acid deletion

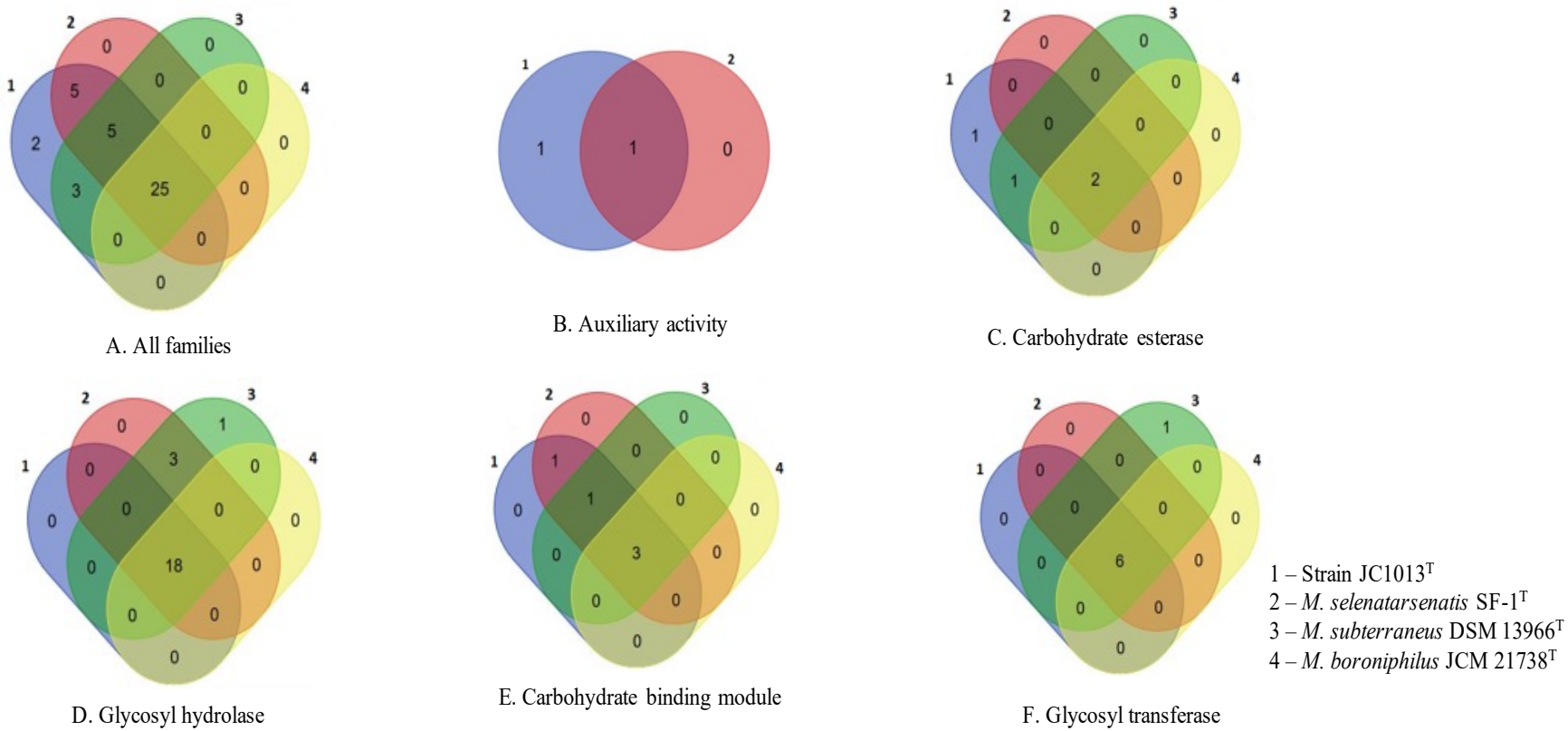


Fig. 32. Venn diagram showing the comparative carbohydrate-active enzyme annotation of strain JC1013^T with its closely related members

3.3.3.6. Physiological and biochemical analysis

Strain JC1013^T grew at temperatures ranging from 25 to 40 °C (optimum at 30 °C) and a pH range of 6.0–8.5. (optimally at 7.0). In the presence of 0–7 % w/v NaCl, strain JC1013^T grew. When lactose, inositol, pyruvate, sorbitol, and mannose were used as the only carbon sources, the strain JC1013^T thrived well; however, fructose had no effect. With carbon sources like galactose, glucose, and arabinose, strain JC1013^T grew slowly. *M. selenatarsenatis* MTCC 9470^T grew well anaerobically; however, strain JC1013^T did not. Both cultures could utilise ammonium chloride and tryptophan as nitrogen sources. Both strains (JC1013^T and *M. selenatarsenatis* MTCC 9470^T) did not require any additional vitamins to grow; however, yeast extract improved the growth. Table 14 shows how strain JC1013^T used substrate differently than *M. selenatarsenatis* MTCC 9470^T. The gelatin hydrolysis test was negative for strain JC1013^T, but the indole test was positive. Galactosidase, esterase, alkaline phosphatase, esterase lipase, and naphthol-AS-BI-phosphohydrolase were all found positive in strain JC1013^T. Leucine arylamidase, lipase, cysteine arylamidase, valine arylamidase, trypsin, acid phosphatase, α-chymotrypsin, α-galactosidase, α-glucosidase, β-glucuronidase, and β-glucosidase were all negative in both strains. *M. selenartsenatis* MTCC 9470^T was responsive to all antibiotics except penicillin, while strain JC1013^T was sensitive to everything except chloramphenicol and tetracycline.

Table 14. Characteristics that differentiate JC1013^T from its closely related species. Strains: 1, JC1013^T; 2, *M. selenatarsenatis* SF-1^T; 3, *M. subterraneus* DSM 13966^T (Kanso et al. 2002);4, *M. jeotgali* DSM 18226^T (Yoon et al. 2017)

Characteristics	**1***	**2***	**3**	**4**
Source	Pond, salt lantern	Sediment, effluent drain	Bore-well water	Fermented sea food
Colony colour	Orangish pink	white	white	Cream yellow
Oxygen requirement	aerobic	facultative anaerobic	facultative anaerobic	facultative aerobe
Cell diameter (µm)	0.9-1	0.5-0.8	0.5-0.8	0.8-1.1
Temperature (˚C) range for growth (Optimum)	28-37 (30)	25-40 (40)	Room temperature to 45	30-35
pH range for growth (Optimum)	6-8.5 (7)	7.5-9 (8)	7-9	7-8
Growth range for NaCl (%)	0-7	0-5	0-9	0-14
Carbon substrate:				
Galactose	w+	+	-	-
Arabinose	w+	-	-	-
Glucose	w+	-	+	+
Sucrose	-	+	+	+
Mannitol	-	+	-	-
Indole	+	-	-	-
Gelatin hydrolysis	-	+	+	+
Nitrate reduction	+	+	+	+
Cellular FA				
iso-$C_{15:0}$	38.1	46.5	46	49.3
iso-$C_{16:0}$	3.8	1.5	5.4	2.3
iso-$C_{14:0}$	4.0	2.5	4.3	1.9
anteiso-$C_{12:0}$	0.6	-	-	-
anteiso-$C_{15:0}$	7.3	3.1	6.4	8.8
$C_{16:0}$ 3-OH	7.8	-	-	-
$C_{16:0}$ ω5c alcohol	7.0	-	8.5	4.5
iso- $C_{17:0}$ ω10c	7.4	16.8	5.7	7.5
4 (*iso*-$C_{17:1}$ I/ *anteiso* B)	5.8	5.1	4.1	6.6
Genomic G+C mol %	42.7	42.8	43±1	41

*data presented were performed in the author's laboratory;

Both strains were positive for the hydrolysis of starch, CMC; nitrate reduction; H2S production whereas negative for hydrolysis of casein, urea, chitin and absence of lipase, leucine arylamidase, valine arylamidase, cysteine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, α-galactosidase, β-glucoronidase, α-glucosidase, β-glucosidase. Major polar lipids in both the strains were diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine. Pyruvate, lactose, inositol, mannose, and sorbitol were utilised as carbon substrates by both the strains whereas fructose was not utilized

3.3.3.7. Cellular fatty acid, polar lipid, and quinone analysis

Iso-$C_{15:0}$, anteiso-$C_{15:0}$, $C_{16:0}$ 3-OH, iso-$C_{17:0}\omega10c$ and summed feature 4 (iso-$C_{17:1}$ I/ anteisoB) were the primary fatty acids found in strain JC1013^T (> 20 % of the total fatty acids). Table 15 shows the differences in fatty acid profiles between strain JC1013^T and *M. selenatarsenatis* MTCC 9470^T. Diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, two unidentified aminolipids (AL1, 2), and four unidentified phospholipids (PL1, 2, 3, 4) are all found in strain JC1013^T. (Fig. 33A). Diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, and nine unidentified phospholipids (PL1-9) were found as polar lipids in *M. selenatarsenatis* MTCC 9470^T (Fig. 33B). The quinone system in both strains was menaquinone-7.

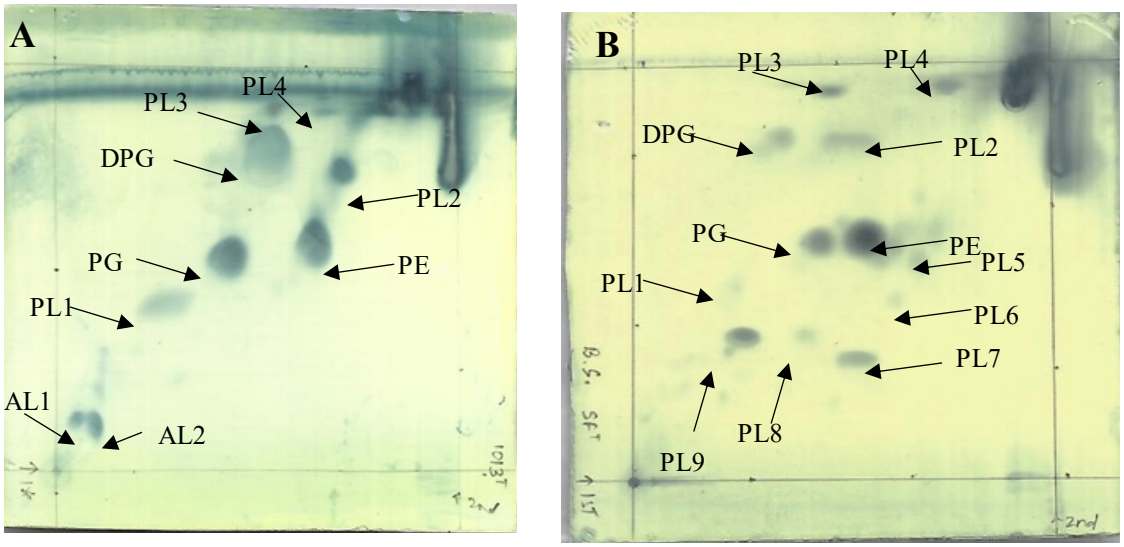


Fig. 33. Two dimensional chromatograms showing polar lipids of
(A) Strain JC1013^T
(B) *M. selenatarsenatis* MTCC 9470^T
PG, phosphatidylglycerol; DPG, diphosphotidylglycerol; PE, phosphatidylethanolamine; AL1,2, unidentified aminolipids; PL1-9, unidentified phospholipids

3.3.3.8. Cell wall amino acid and indole metabolism

m-diaminopimelic acid was the diagnostic cell wall amino acid of strain JC1013^T (Fig. 34). Indole production from L-tryptophan was negative in *M. selenatarsenatis* MTCC 9470^T. Based on UV absorption spectra, five indole derivatives were discovered from strain JC1013^T (Fig. 35), with peak 2 corresponding to indole 3-acetic acid and peak 5 corresponding to indole, and the other three unidentified indoles. The genomic characterisation, as mentioned above, helped in corroborating the HPLC result of indoles.

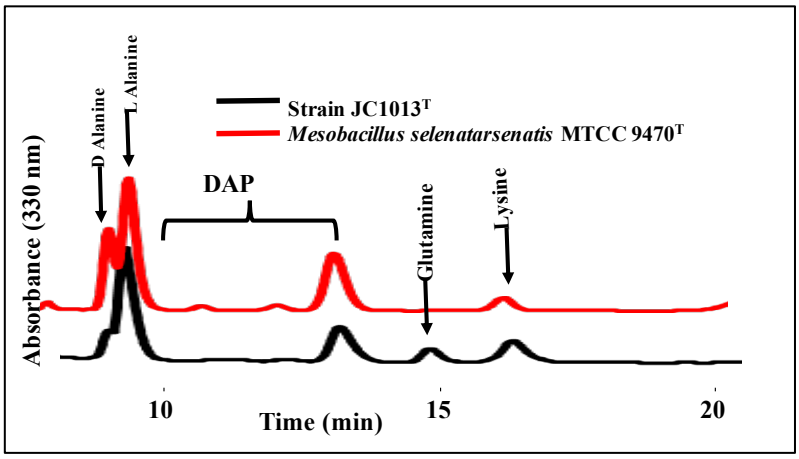


Fig. 34. HPLC chromatogram showing cell wall amino acid peaks of strain JC1013^T and *M. selenatarsenatis* MTCC 9470^T (Strain 1013^T had D-Alanine and L-Lysine)

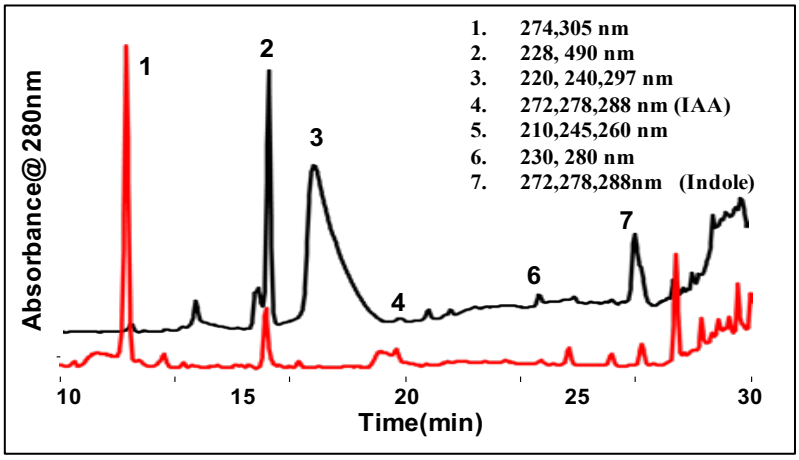


Fig. 35. HPLC chromatogram showing L-tryptophan catabolites produced by strain JC1013^T (black line); control red line). Peaks which have not shown absorption spectra are not labelled. 1, 2, 3, 5, unidentified metabolites; 2, indole-3 acetic acid; 5. indole

3.3.3.9. Carotenoid production and pathway analysis

The carotenoid production was checked in three media of nutrient broth (NB), marine broth (ZM), and Biebl and Pfennig broth (BP) (Fig. 36A). Strain JC1013^T had a short generation/ doubling time of 3.6 hrs in NB compared to when grown in the other two media. Along with the generation time, an increase in fresh weight (12 gL^{-1}) and dry weight (2.5 gL^{1}) was observed in NB (Fig. 36B). Therefore, strain JC1013^T was grown in nutrient broth for further experiments.

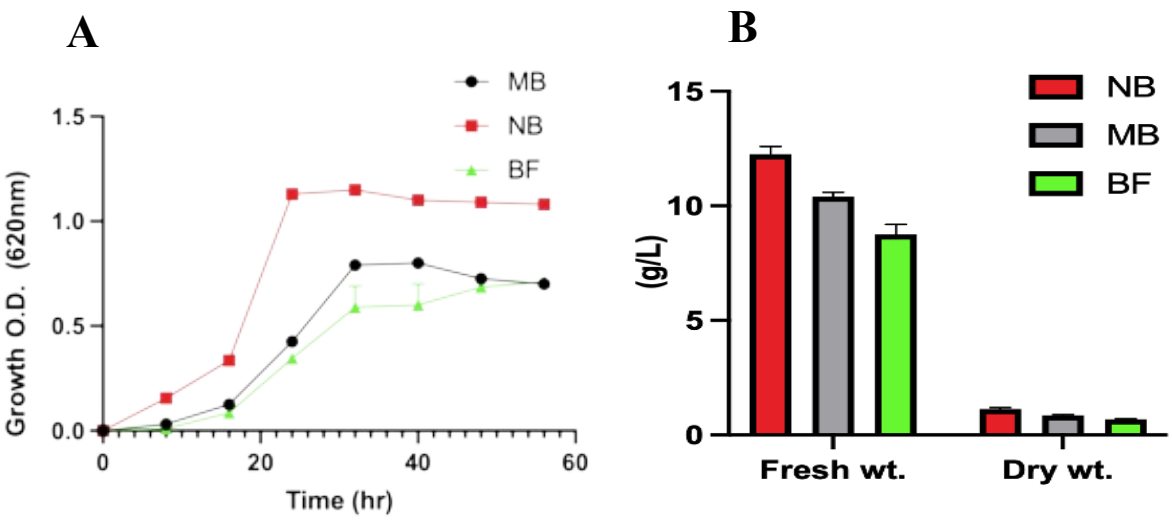


Fig. 36. (A) Growth curve of strain JC1013^T (Generation/doubling time= 3.6 hrs in NB) (B) Biomass (Fresh and dry) of culture of strain JC1013^T grown aerobically under different media at 32℃, rpm 121 for 72 hrs

3.3.3.9.1 Analysis of the crude carotenoid extract

For the crude pigments (secondary metabolites) extracted as mentioned (Section 2.4.4.8.2), the absorption spectrum was measured (300-1100 nm) and showed that the absorption maxima in methanol were observed in 254, 266, 426, 488, 508, and 522 nm (Fig. 37A). Further, the HPLC chromatogram for carotenoids of strain JC1013^T showed seven peaks. Peaks 5 and 6 matched zeaxanthin and lycopene (based on absorption spectrum), respectively, with the remaining five carotenoids remaining undetermined (Fig. 37B). The mass spectrometry analysis of crude extract showed three major peaks, and seven minor peaks of pigments (Fig. 38A). Peaks 1, 2, 3 and 4 could be putatively identified (Fig. 38B, C, D, E). The pigment profile is putatively identified as follows (B) 4-Hydroxyzeaxnthin, (C) 7,8-Didehydroastaxanthin, (D) 7,8- Dihydrozeaxanthin, (E) OH-Neurosporene (Table 15).

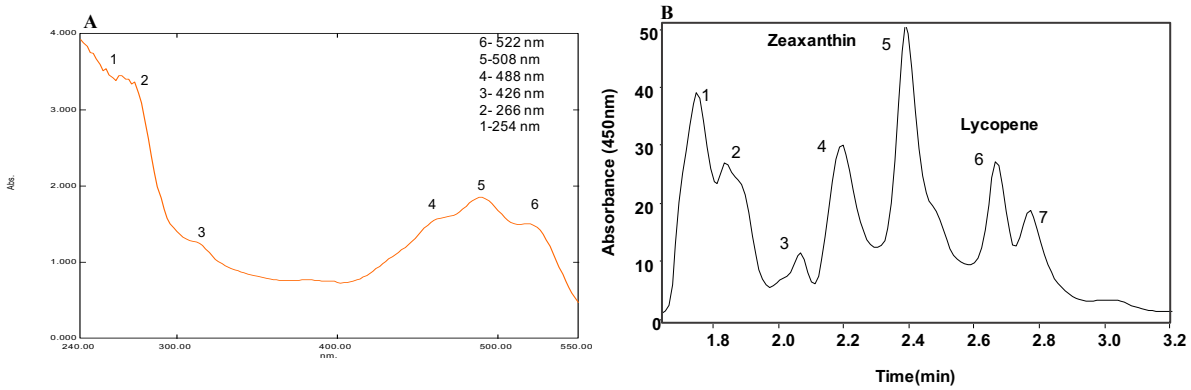


Fig. 37. (A) Absorption spectra of extracted pigments of strain JC1013^T (250-600nm) (B) HPLC chromatogram showing carotenoid peaks of strain JC1013^T ;1-4,7 unidentified carotenoids; 5, Zeaxanthin; 6, Lycopene

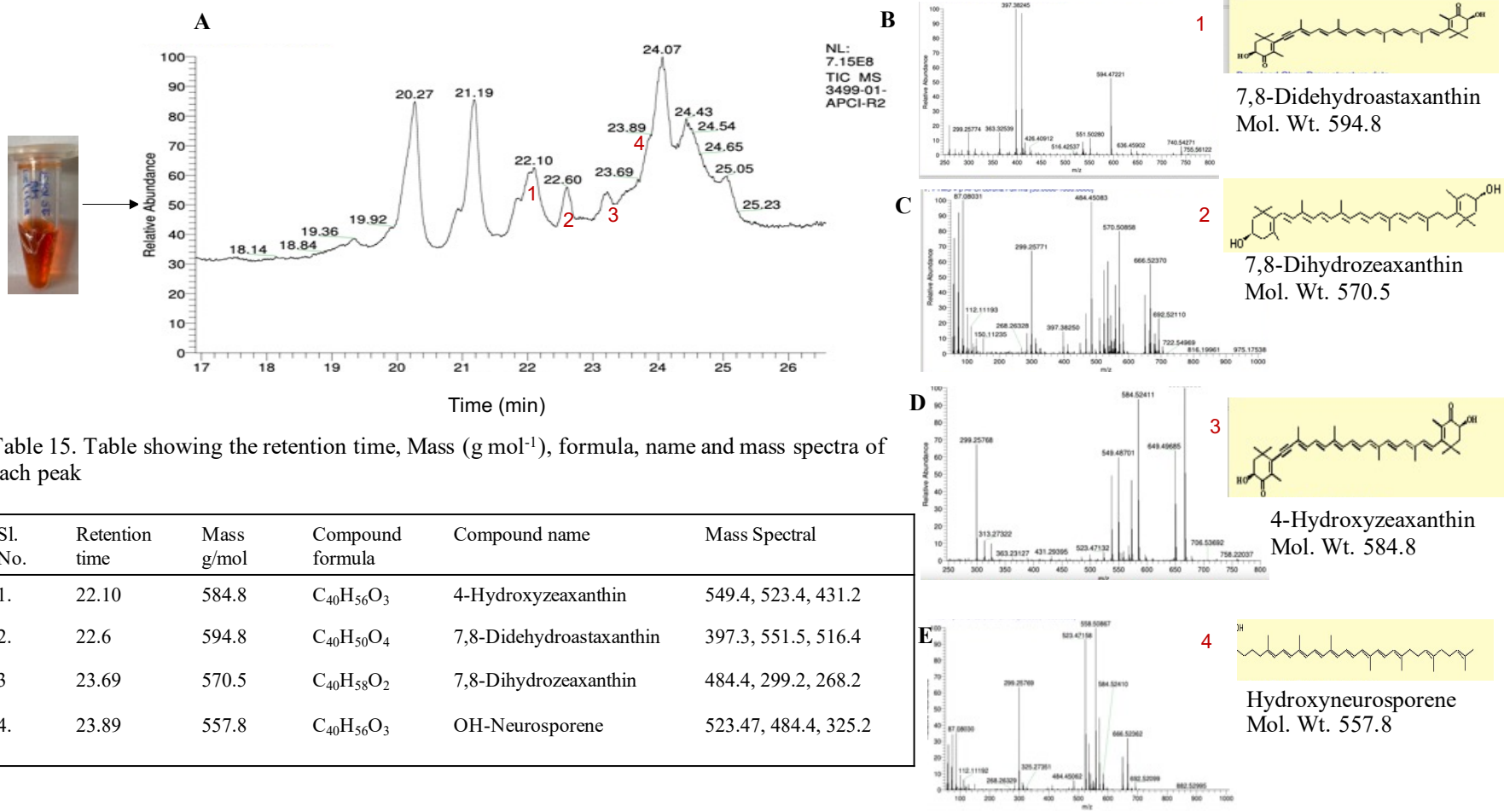


Table 15. Table showing the retention time, Mass (g mol^{-1}), formula, name and mass spectra of each peak

Sl. No.	Retention time	Mass g/mol	Compound formula	Compound name	Mass Spectral
1.	22.10	584.8	$C_{40}H_{56}O_3$	4-Hydroxyzeaxanthin	549.4, 523.4, 431.2
2.	22.6	594.8	$C_{40}H_{50}O_4$	7,8-Didehydroastaxanthin	397.3, 551.5, 516.4
3	23.69	570.5	$C_{40}H_{58}O_2$	7,8-Dihydrozeaxanthin	484.4, 299.2, 268.2
4.	23.89	557.8	$C_{40}H_{56}O_3$	OH-Neurosporene	523.47, 484.4, 325.2

Fig. 38. (A) LC-MS chromatogram of crude carotenoid extract of strain JC1013^T
(B) Mass spectral data of compound corresponding to peak 1 (7,8-Dihydroastaxanthin)
(C) Mass spectral data of compound corresponding to peak 2 (7, 8-Dihydrozeaxanthin)
(D) Mass spectral data of compound corresponding to peak 3 (4-Hydroxyzeaxanthin)
(E) Mass spectral data of compound corresponding to peak 4 (Hydroxyneurosporene)

3.3.3.9.2. Analysis of purified carotenoid fractions

The crude pigments (carotenoid) extracted were purified into 3 fractions by thin layer chromatography (Fig. 39). In fraction 1 (Fig. 39A), three compounds like phytoene, antheraxanthin, phytofluene were identified; in fraction 2 (Fig. 39B), four compounds like 2, 2-Dihydroxyastaxanthin, OH-neurosporene, and compound II were identified. In fraction 3 (Fig. 39C), eschsholtzanthin and compound I was identified. Other information related to putatively identified compounds is depicted in Table 17.

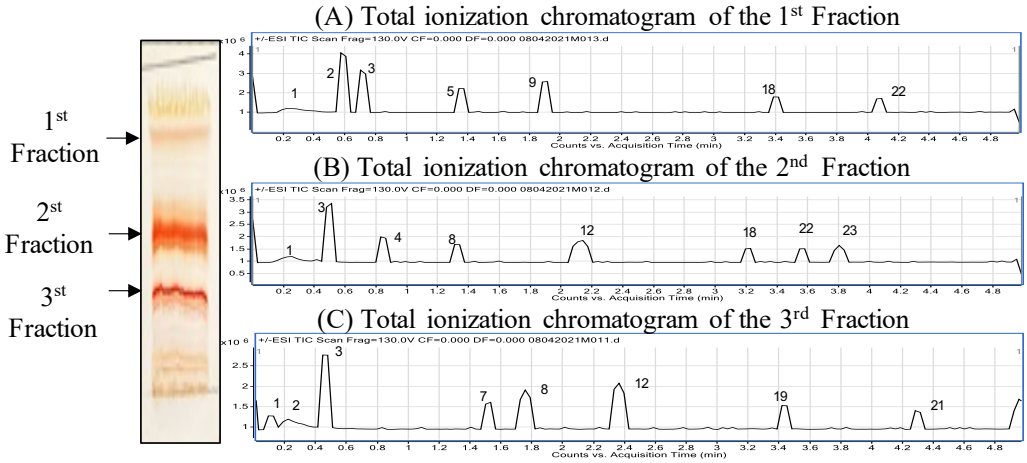


Fig. 39. Total Ionization chromatogram (TIC) of purified three fractions of the carotenoid extract of strain JC1013^T
(A) Fraction 1 (B) Fraction 2 (C) Fraction 3

Table 16. Table showing the retention time, mass (g mol^{-1}), formula, name and mass spectra of each peak in the respective fractions

Frac. No.	Peak No.	Mass g/mol	Compound formula	Compound name	Mass Spectral
1	2	544	$C_{40}H_{64}$	Phytoene	459.6, 333.1, 297.4
1	4	584	$C_{46}H_{56}O_3$	Antheraxanthin	459.2, 386.4, 297.5
1	21	542.9	$C_{40}H_{62}$	Phytofluene	379.2, 311.1, 227.2
2	2	564	$C_{40}H_{52}O_2$	Compound II	453.0, 312.7, 248.7
2	3	628.9	$C_{40}H_{52}O_6$	2,2'-Dihydroxyastaxanthin	521.3, 249.0, 161.0
2	4	557.8	$C_{40}H_{60}O$	OH-neurosporene	450.9, 297.4, 249.4
2	12	570.8	$C_{40}H_{58}O_2$	7,8- Dihydrozeaxanthin	476.1, 311.1, 249.3
3	1	590.5	$C_{42}H_{55}O_2$	Eschscholtzanthin	496.3, 390.5, 291.3
3	21	566.6	$C_{40}H_{54}O_2$	Compound I	465.3, 420.3, 324.4

3.3.3.9.3. Prediction of carotenoid pathway

The result given by the anti-SMASH 5.0 server showed that strain JC1013^T had 33 % of gene identity to the carotenoid biosynthesis gene clusters (total of 10 clusters) with the pre-existing organism's in the database. The identification of carotenoid gene clusters, similarity identity, accession number and the organisms name is given in Table 18. Therefore, combinatorial studies based on analysis of genomic and metabolomic (HPLC, LC-MS) data have resulted in predicting the carotenoid biosynthetic pathway in strain JC1013^T, as illustrated in Fig. 40.

Table 17. Gene identity (%) of strain JC1013^T carotenoid biosynthesis pathway predicted by anti-SMASH5.0

Cluster No.	Gene	% Identity (BlastP)	Organism	Accession no.
1	Squalene/phytoene synthase family protein	93.4	*Bacillus* sp.	KAA0549531
2	Squalene-hopane cyclase	88.7	*Bacillus piezotolerans*	WP_115452006
3	SDR family oxidoreductase	89.9	*Mesobacillus. jeotgali*	WP_079508828
4	Phytoene desaturase	93.2	*Bacillus salsus*	SDP04815
5	Phytoene desaturase	92.9	*Bacillus luteolus*	WP_193539229
6	Alpha/beta hydrolase fold domain containing protein	91.9	*Bacillus* sp. HD4P25	WP_193468623
7	Phytoene desaturase	94.4	*Bacillus* sp. HD4P25	WP_193468621
8	Aldehyde dehydrogenase family protein	90.1	*Bacillus* sp. HD4P25	WP_193468619
9	Phytoene desaturase	94.9	*Bacillus* sp. HD4P25	WP_193468617
10	Glycosyltransferase	93.2	*Bacillus luteolus*	WP_193539214

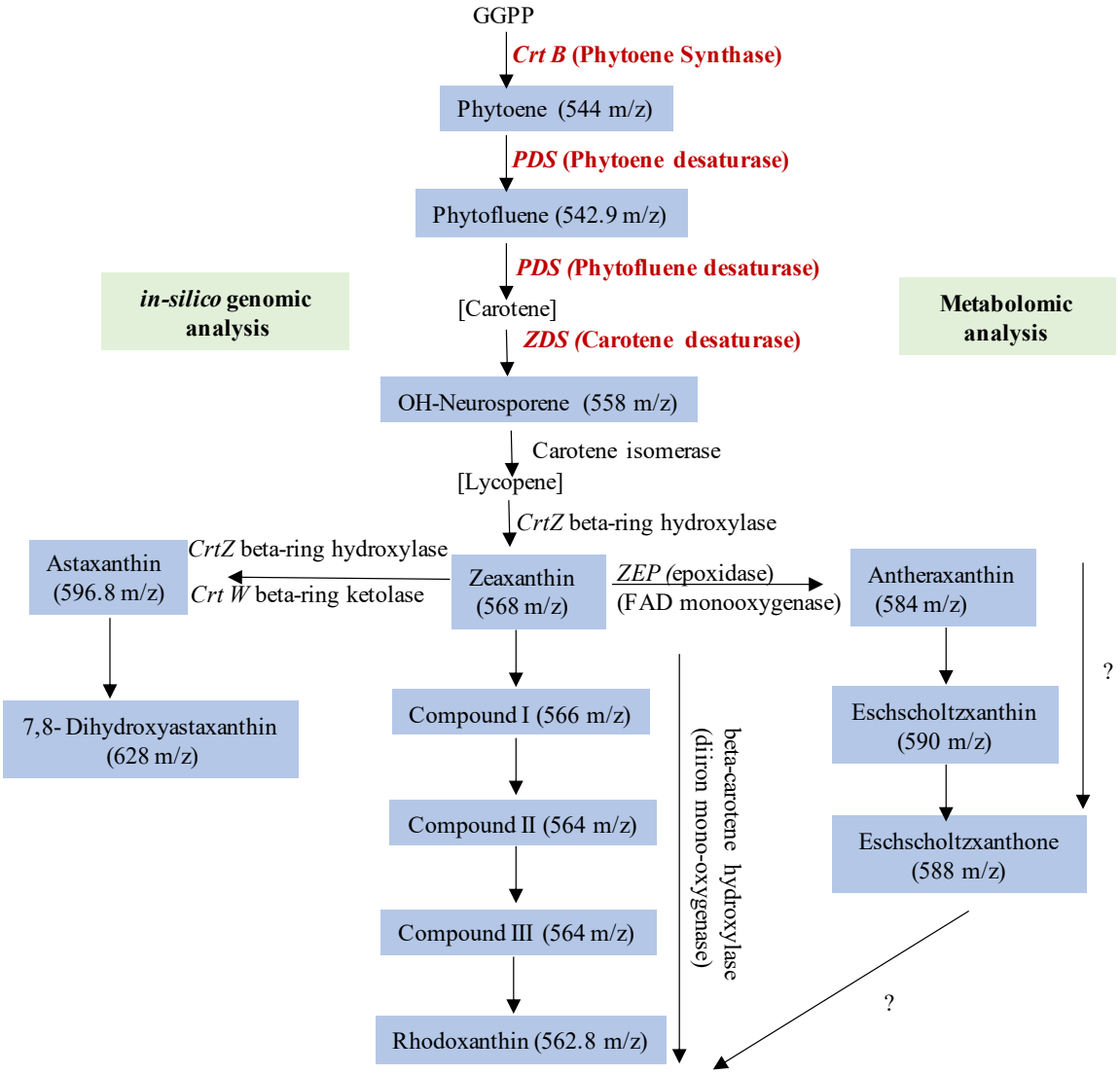


Fig. 40. Predicted pathway of carotenoid biosynthesis of strain JC1013^T derived from genomic and metabolomic data; written in red colour, found in genome sequence; black, not found in genome; compound within blue rectangles, mass (m/z) found in metabolomic data

3.3.4. Characterization of strain JC162^T (= KCTC 32190^T)

3.3.4.1. Home habitat, colony, and cell morphology

Strain JC162^T was isolated from a beach sand sample collected near Pamban bridge, Rameshwaram, Tamil Nadu, India (GPS positioning of the sample collection site is 9°16'31.61" N, 79°13'34.37" E) on 14 November 2007. The vertical section of the beach sand showed stratified colored layers (Fig. 41A). These were broadly categorized into three layers as; purple color-layer 1, orange to brown color-layer 2, and pale cream color-layer 3. From layer 1, *Marichromatioum litoris* (Shivali et al. 2011) was isolated; the color of the sand might be attributed to the purple color of okenone carotenoid pigments produced by this bacterium. The pink pigmented colonies were purified by subsequent streaking on nutrient agar (NA) (HiMedia, M001). Strain JC162^T deposited with KCTC with accession number KCTC 32190^T was procured and characterised along with the type strains of *Roseomonas oryzicola* KCTC 22478^T, which was also further used for comparative taxonomic studies. Both strains were tested for growth in nutrient agar (HiMedia M001), tryptic soy agar (HiMedia M011), and Biebl and Pfennig (BP) media (Lakshmi et al. 2011). Both could grow well on nutrient agar (HiMedia M001), tryptic soy agar (HiMedia M011), and BP media. Strain KCTC 32190^T colonies were 1.0-1.5 mm in diameter, pink in colour, and convex with irregular margins. Cells of strain KCTC 32190^T were 0.6-1 µm in diameter and 1.5-2.0 µm in length with elongated oval to spherical shapes (Fig. 41B, C). Gram-stain-negative and non-motile.

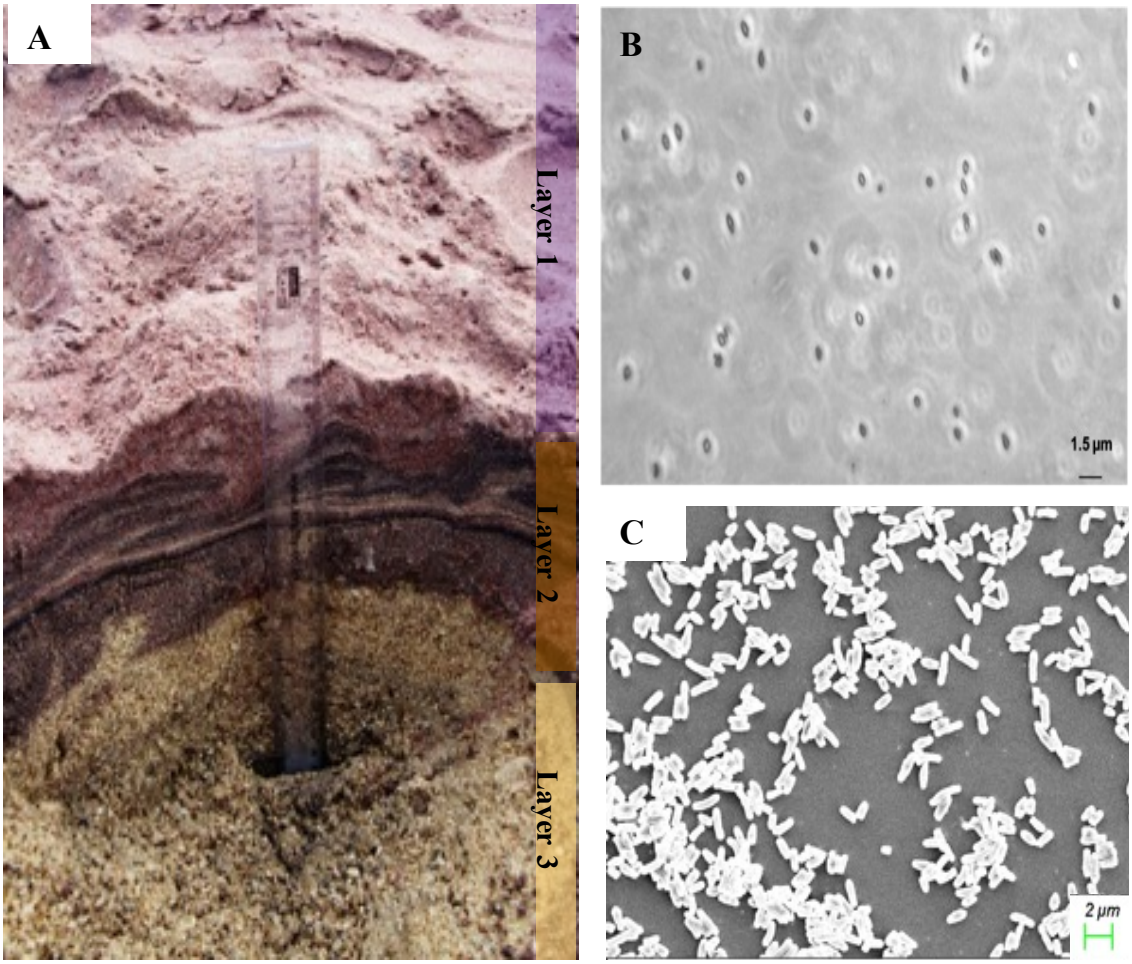


Fig. 41. (A) A beach sand sample collected from Pamban bridge, Rameshwaram, Tamil Nadu, India. (GPS Positioning 9°16'31.61" N, 79°13'34.37" E) Layer 1, purple sand; Layer 2, orange to brown sand; Layer 3, pale cream sand;
(B) Phase contrast microscopic image strain of KCTC 32190^T
(C) FESEM image of strain of KCTC 32190^T

3.3.4.2. Phylogeny based on 16S rRNA and genome sequence

EZBioCloud BLAST results indicated sequence identity (%) of strain KC TC 32190^T with the type strain of *R. oryzicola* YC6724^T (99.78 %), *R. sediminicola* FW-3^T (98.45 %), *R. soli* 5N26^T (98.23 %) and other members of the genus *Roseomonas* (<97.9 %). The phylogenetic analysis of the 16S rRNA gene sequence also revealed that strain KCTC 32190^T cladded with *R. oryzicola* YC6724^T (Fig. 42). However, the 16S rRNA gene based phylogenetic tree indicated that the genus *Roseomonas* was highly heterogenous and was segregating into 7 major clades; Group I *sensu-stricto* (n=4), Group II (n=8), Group III (n=9), Group IV (n=2), Group V (n=2), Group VI (n=7; including strain KCTC 32190^T) and Group VII (n=16) (Fig. 42). Strain KCTC 32190^T was affiliated with Group VI. The members of a phylogenetic tree-defined group shared at least 93.5–99.2 % of the 16S rRNA gene identity (Table 18). The phylogenomic tree constructed based on 92 core gene of 197 genomes by the UBCG tool confirmed the heterogeneousness of the genus *Roseomonas,* as same cladding pattern between members was observed (Group I-VII) (Fig. 43). Genome sequences of Group IV members (*R. aquatica*, *R. fluminis*) were not available (not considered here). *Roseomonas aeriglobus* and genomospecies of *Roseomonas*, clustered away from the members of the family *Acetobacteraceae*, therefore not considered for further analysis (Fig. 43). Based on the 16S rRNA and 92 core genes based phylogenetic tree, the standing of the members of the genus *Roseomonas* was analysed in detail.

Table 18. 16S rRNA gene identity (%) between the members of genus *Roseomonas*

	Group I	**Group II**	**Group III**	**Group IV**	**Group V**	**Group VI**	**Group VII**
I	96.3-98.9						
II	93.5-95.7	94.9-97.8					
III	92.3-94.6	92.3-94.5	93.4-97.9				
IV	94.3-94.7	94-96.4	92.9-94.9	95.6			
V	92.6-94.4	92.9-93.9	92.7-93.9	94.4-95.1	97.1		
VI	92.1-94.3	91.9-93.9	91.3-94.3	92.5-94.6	94.6-98.4	94.4-98.6	
VII	92.4-95.1	91.5-94.7	91.3-04.6	93.4-95.6	92.9-94	92.1-94.3	94.7-99.6

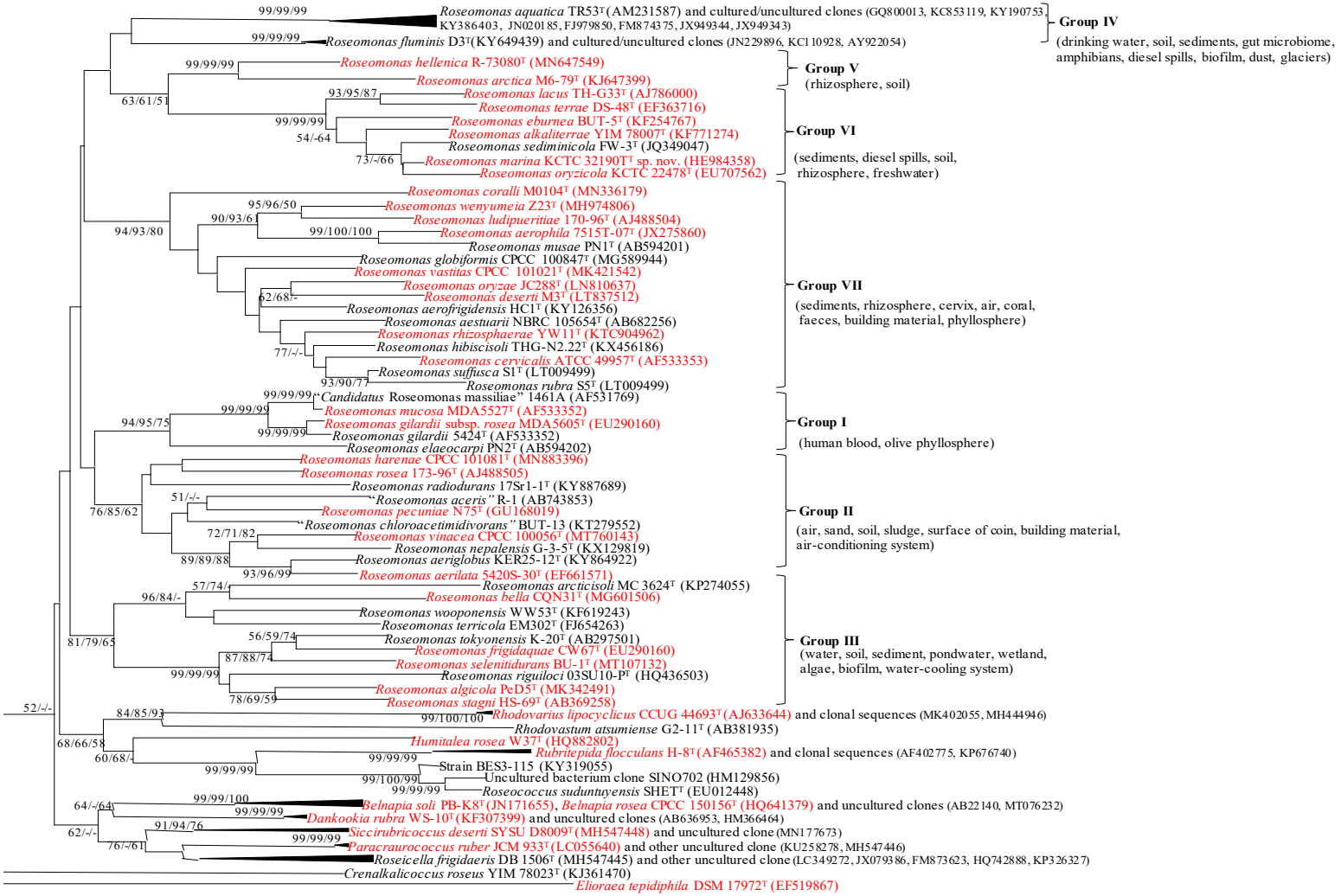


Fig. 42. NJ phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic relationship between the members of genus *Roseomonas* and other closely related members of family *Acetobacteraceae*. *Elioraea tepidiphila* DSM 17972^T (EF519867) was used as an outgroup. Numbers at nodes represent bootstrap values Bar, 0.01 accumulated changes per nucleotide substitutions. (Red color indicates the availability

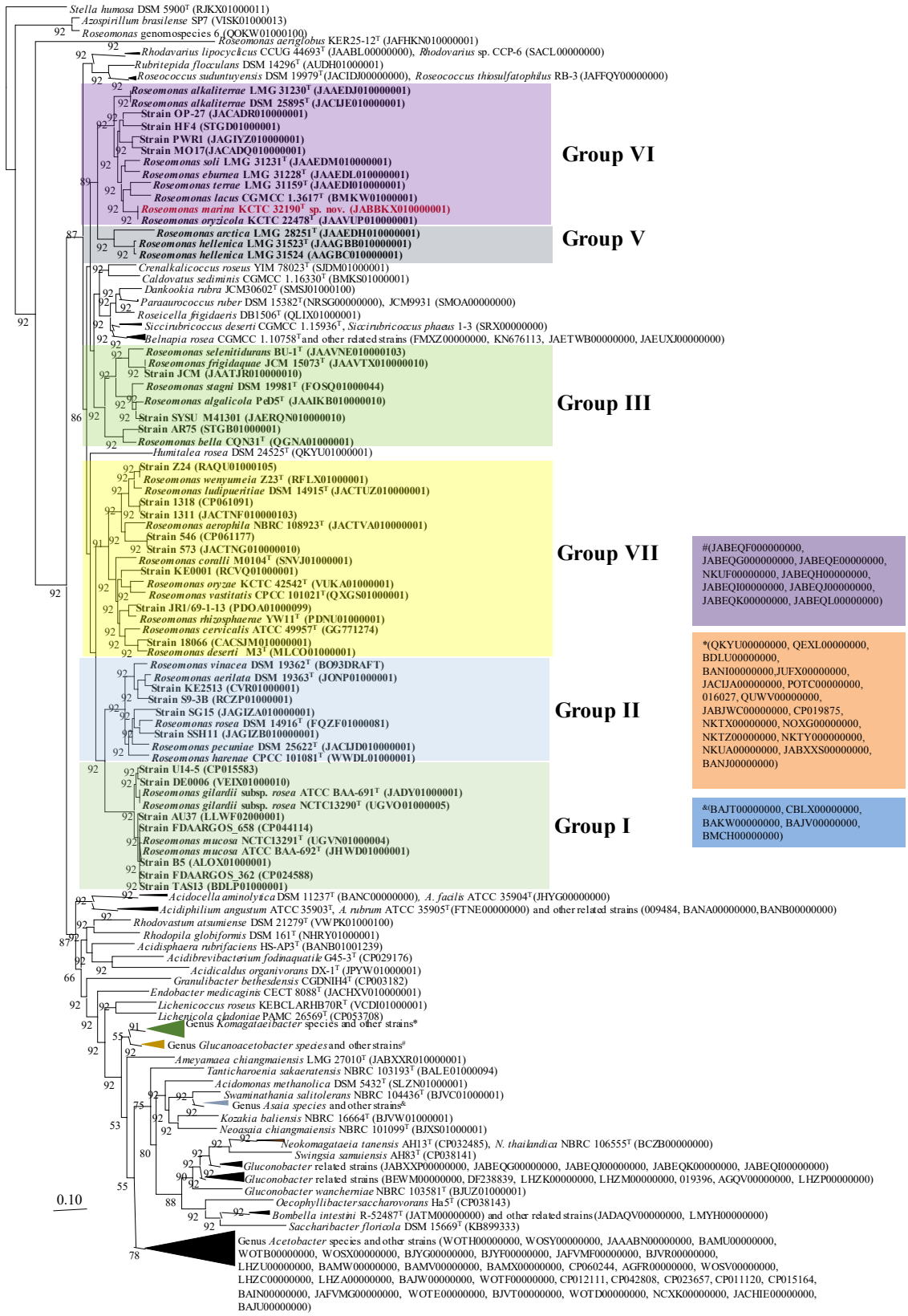
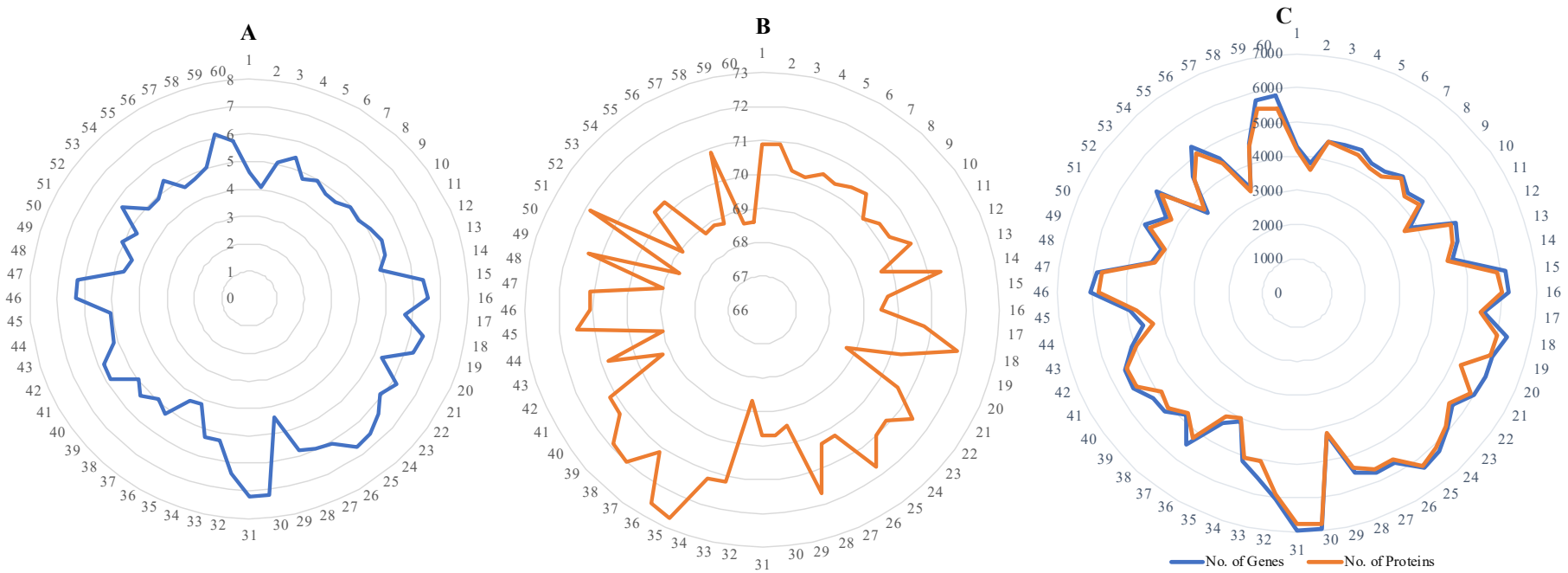


Fig. 43. Phylogenomic tree of the members of genus *Roseomonas* constructed using 92 core genes based on Up-to-date Bacterial Core Gene (UBCG) (Na et al. 2018) by NJ. *Stella humosa* DSM 5900^T (RJKX00000000) as an outgroup

3.3.4.3. Genomic features of *Roseomonas* spp.

Paired-end sequencing of (150×2 bp) fragment library of KCTC 32190^T and *R. oryzicola* KCTC 22478^T resulted in about 5.18 % and 6.5 % quality filtered reads, respectively. The genome size of strain KCTC 32190^T was 5.2 Mbp with 56 scaffolds and an N_{50} value of 626749 bp, whereas the genome size of *R. oryzicola* KCTC 22478^T was 5.3 Mbp with 57 scaffolds and with N_{50} value of 685653 bp. Genomic G+C content of DNA of strains KCTC 32190^T and *R. oryzicola* KCTC 22478^T was calculated as 71.1 % and 71.4 mol %, respectively. The NCBI accession numbers of genome sequences of strain KCTC 32190^T and *R. oryzicola* KCTC 22478^T were JABBKX000000000 and JAAVUP000000000, respectively. Table 2 lists the genomic information for the type strains employed in this study. The genome sizes of Groups I (n=11), II (n=9), III (n=8), V (n=3), VI (n = 12), and VII (n = 17) are 4.1-5.4 Mbp, 4.9-6.5 Mbp, 5.8-6.7 Mbp, 4.4-7.2 Mbp, 4.2-6.4 Mbp, and 4.5-6.3 Mbp, respectively (Fig. 44A). Groups I (n = 11), II (n = 9), III (n = 8), V (n = 3), VI (n = 12), and VII (n = 17) had 70.1-70.9 %, 68.7-71.9 %, 70.3-71.7%, 69.5-69.7 %, 68.7-72.7 %, and 68.6–71.9 % as genomic G+C content (%), respectively (Fig. 44B). Further, the no. of genes and proteins for members are given in Fig. 44C. Group I had the smallest genome size of 4.8 Mbp, whereas Group V had the largest genome size of 6.3 Mbp. Similarly, Group V taxa had the lowest average genomic G+C content (69.6 mol %), while Group VI taxa had the greatest (71.2 mol %). When the genome sizes of the members of the genus *Roseomonas* were compared to those of other genera in the *Acetobacteraceae* family, *Dankookia* (n = 1) had the largest genome size of 7.8 Mbp. The genus *Rubritepida* (n = 1), on the other hand, had the smallest genome size of 3.8 Mbp and the highest genomic G+C content (73.4 mol %) (Table 19).



1.*R. gilardii* subsp. *rosea* ATCC BAA-691^{T}, 2. *R. gilardii* subsp. *rosea* NCTC 13290^{T}, 3. Strain DE0006, 4. Strain U14-5, 5. *R. mucosa* ATCC BAA-692^{T}, 6. *R. mucosa* NCTC 13291^{T}, 7. Strain AU37, 8. Strain B5, 9. Strain TAS13, 10. Strain FDAARGOS_658, 11. Strain FDAARGOS_362, 12. *R. rosea* DSM 14916^{T}, 13. Strain SSH11, 14. *R. pecuniae* N75^{T}, 15. *R. aerilata* DSM 19363^{T}, 16. Strain KE2513, 17. Strain SG15, 18. Strain S9.3B, 19. *R. vinacea* DSM 19362^{T}, 20. *R. harenae* CPCC 101081^{T}, 21. *R. stagni* DSM 19981^{T}, 22. *R. bella* CQN31^{T}, 23. Strain AR75, 24. *R. algicola* PeD5^{T}, 25. Strain SYSU M41301, 26. *R. frigidaquae* JCM 15073^{T}, 27. Strain JCM, 28. *R. selenitidurans* BU-1^{T}, 29. *R. arctica* LMG 28251^{T}, 30. *R. hellenica* LMG 31523^{T}, 31. Strain LMG 31524, 32. *R. lacus* CGMCC 1.3617^{T}, 33. *R. oryzicola* KCTC 22478^{T}, 34. Strain KCTC 32190^{T} =JC162^{T}, 35. *R. alkaliterrae* DSM 25895^{T}, 36. *R. alkaliterrae* LMG 31230^{T}, 37. Strain OP-27, 38. Strain PWR1, 39. Strain HF4, 40. Strain MO17, 41. *R. eburnea* LMG 31228^{T}, 42. *R. terrae* LMG 31159^{T}, 43. *R. soli* LMG 31523^{T}, 44. *R. cervicalis* ATCC 49957^{T}, 45. Strain JR1/69-1-13, 46. *R. deserti* M3^{T}, 47. Strain 18066, 48. *R. oryzae* KCTC 42542^{T}, 49. Strain KE0001, 50. *R. vastitatis* CPCC 101021^{T}, 51. *R. rhizosphaerae* YW11^{T}, 52. *R. aerophila* NBRC 108923^{T}, 53. Strain 546, 54. Strain 573, 55. *R. ludipueritiae* DSM 14915^{T}, 56. Strain 1311, 57. Strain 1318, 58. *R. coralli* M0104^{T}, 59. *R. wenyumeiae* Z23^{T}, 60. Strain Z24

Fig. 44. Radar plot showing (A) genome size (B). G+C% (C). Number of genes and protein of 60 species of genus *Roseomonas*

Table 19. Genome characteristics of the members of the genus *Roseomonas* and closely related strains in the family *Acetobacteraceae*

Organism	Strains	Size (Mbp)	N50 (bp)	L50	DDBJ/EMBL/Gen Bank
Group I (n=11)					
R. gilardii subsp. *rosea*	ATCC BAA-691^T	4.6	257798	6	JADY00000000
	NCTC 13290^T	4.1	2982497	1	UGVO00000000
	Strain DE0006	5.1	366571	6	VEIX00000000
	Strain U14-5	5.4	4328147	1	CP015583
R. mucosa	ATCC BAA-692^T	4.8	238615	8	JHWD00000000
	NCTC 13291^T	5.0	4237410	1	UGVN00000000
	Strain AU37	4.7	48204	30	LLWF00000000
	Strain B5	4.7	51105	24	ALOX00000000
	Strain TAS13	5.0	11441	130	BDLP00000000
	Strain FDAARGOS_658	4.9	4244047	1	CP044114
	Strain FDAARGOS_362	5.1	4180106	1	CP024588
Group II (n=9)					
R. rosea	DSM 14916^T	5.3	136395	14	FQZF00000000
	Strain SSH11	5.2	181234	10	JAGIZB00000000
R. pecuniae	N75^T	4.9	200862	9	JACIJD000000000
R. aerilata	DSM 19363^T	6.4	202657	11	JONP00000000
	Strain KE2513	6.5	149208	14	RCVR00000000
	Strain SG15	5.7	233037	7	JAGIZA00000000
	Strain S9.3B	6.5	123179	14	RCZP00000000
R. vinacea	DSM 19362^T	6.3	192520	9	Go0013226*
R. harenae	CPCC 101081^T	5.3	110178	15	WWDL00000000
Group III (n=8)					
R. stagni	DSM 19981^T	6.2	263857	7	FOSQ00000000
R. bella	CQN31^T	5.9	906722	3	QGNA00000000
	Strain AR75	6.3	1093183	3	STGB00000000
R. algicola	PeD5^T	6.6	452050	5	JAAIKB00000000
	Strain SYSU M41301	6.7	561150	4	JAERQN00000000
R. frigidaquae	JCM 15073^T	6.1	749226	4	JAAVTX00000000
	Strain JCM	6.0	749226	4	JAATJR00000000
R. selenitidurans	BU-1^T	5.8	96152	19	JAAVNE00000000
Group V (n=3)					
R. arctica	LMG 28251^T	4.4	157805	10	JAAEDH00000000
'R. hellenica'	LMG 31523^T	7.2	127414	20	JAAGBB00000000
	Strain LMG 31524	7.2	79618	24	AAGBC00000000
Group VI (n=12)					
R. lacus	CGMCC 1.3617^T	6.4	379484	7	BMKW00000000
R. oryzicola	KCTC 22478^T	5.3	595371	3	JAAVUP00000000
	KCTC 32190^T =JC162^T	5.2	685653	57	JABBKX00000000
R. alkaliterrae	DSM 25895^T	4.2	301642	5	JACIJE00000000
	LMG 31230^T	4.3	16191	69	JAAEDJ000000000
	Strain OP-27	5.2	15335	94	JACADR00000000
	Strain PWR1	4.9	172464	9	JAGIYZ00000000
	Strain HF4	5.3	274811	7	STGD00000000
	Strain MO17	5.0	11683	107	JACADQ00000000
R. eburnea	LMG 31228^T	5.8	161991	12	JAAEDL00000000
R. terrae	LMG 31159^T	5.8	180609	13	JAAEDI00000000
R. soli	LMG 31523^T	5.2	48334	30	JAAGBB00000000

Organism	Strains	Size (Mbp)	N50 (bp)	L50	DDBJ/EMBL/GenBank
Group VII (n=17)					
R. cervicalis	ATCC 49957^T	5.1	8459	145	ADVL00000000
	Strain JR1/69-1-13	5.1	246298	6	PDOA00000000
R. deserti	M3^T	6.3	22965	79	MLCO00000000
	Strain 18066	6.3	33264	62	CACSJM00000000
R. oryzae	KCTC 42542^T	4.7	285650	4	VUKA00000000
	Strain KE0001	4.5	417498	4	RCVQ00000000
R. vastitatis	CPCC 101021^T	5.1	164366	7	QXGS00000000
R. rhizosphaerae	YW11^T	4.7	124656	12	PDNU00000000
R. aerophila	NBRC 108923^T	5.7	86979	21	JACTVA00000000
	Strain 546	4.9	3777457	1	CP061177
	Strain 573	4.9	287631	5	JACTNG00000000
R. ludipueritiae	DSM 14915^T	5.3	26901	49	JACTUZ0000000
	Strain 1311	4.7	184795	9	JACTNF00000000
	Strain 1318	4.8	3565232	1	CP061091
R. coralli	M0104^T	5.0	226823	8	SNVJ000000000
R. wenyumeiae	Z23^T	6.1	179822	10	RFLX00000000
	Strain Z24	5.8	21377	74	RAQU00000000
Other taxa					
Rhodovarius lipocyclicus	CCUG 44693^T	4.6	83545	12	JAAABL00000000
Rubritepida flocculans	DSM 14296^T	3.8	134898	8	AUDH00000000

*Gold IDs in IMG Database

3.3.4.4. Genomic metrics- *d*DDH, ANI, AAI and POCP

ANI score between strain KCTC 32190^T and *R. oryzicola* KCTC 22478^T is 92.4 %, whereas *d*DDH values between the two strains are 48.8 %. Based on the hybridization and ANI score, strain KCTC 32190^T is distinctly related to *R. oryzicola* KCTC 22478^T, as the values are within the recommended standards to delineate a bacterial species (Rosselló-Móra and Amann, 2015; Richter and Rosselló-Móra, 2009). At the intragroup level, members of a phylotaxogenomics defined group (type strains) shared *d*DDH and ANI values of at least 21–100 % and 78.0–99.9 %, respectively (Table 20). Members of various groups had *d*DDH and ANI values above 20.8 % and 76.8 %, respectively, at the inter-group level. Similarly, POCP and AAI values were estimated amongst group members (intra-group) as well as across groups (inter-group) (Table 21) because they are important and accurate for genus determination. Group I (n = 11), Group II (n = 9), Group III (n = 8), Group V (n = 3), Group VI (n = 12), and Group VII (n = 17) taxa had POCP values of 80–98.5 %, 65.2–71.7 %, 66.9–98.7 %, 65.8–99.3 %, 66.4–99.2 %, and 65.2–99.4 %, respectively. The AAI values for the taxa of the respective groups were 91.9–99.0 %, 76.7–84.8 %, 71.8–99.9 %, 73.8–99.9 %, 77.0–99.9 %, and 70.2–99.9 %, respectively. The heatmaps of POCP (Fig. 45A) and AAI (Fig. 45B) support the six-group categorization. The AAI scores of *Roseomonas fauriae* (later heterotypic synonym of *Azospirillum brasilense*) and Group I members range from 51.2 to 51.5 %, indicating that they are separate species belonging to different genera. The values (ANI, AAI, POCP) between the members of groups are presented in Fig. 46.

Table 20. *d*DDH and ANI values between the members of the genus *Roseomonas*

ANI/dDDH	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	*	100	45.3	45.3	20.2	20.3	20.6	20.1	20.1	19.1	19.4	20	19.3	19.4	19.4	20	19.8	19.9	20.3	20.5	20.7	19.5	20.2	20.2	19.8	20.5	19.9	20.3	20.2	20.6	19.9	20.8	19.8	20
2	99.9	*	45.4	45.5	20.2	20.4	20.7	20.2	20.3	19.3	19.6	20.2	19.5	19.2	19.5	19.9	19.7	19.7	20.1	20.7	20.5	19.7	20.1	20.1	19.9	20.7	20.1	20.5	20.4	20.8	20.1	20.9	19.9	20.2
3	91.9	92.2	*	100	20.2	20.6	20.3	20.1	20.2	19	19.7	19.8	19.7	19.6	19.1	19.8	19.6	20.1	20	20.2	20.3	19.6	19.8	20	19.9	20.2	20	20.5	20.4	20.7	20.2	18.5	19.9	20.3
4	91.9	92.1	99.9	*	20.1	20.4	20.2	19.9	20	19.2	19.6	19.4	19.6	19.4	19.2	19.8	19.6	20	20.1	20.1	20.3	19.5	19.6	20.1	19.9	20.1	20.1	20.4	20.3	20.6	20.1	18.4	19.8	20.2
5	75.4	75.3	75.4	75.3	*	24.3	23.8	27.3	23.4	19	18.9	19.1	18.9	19.1	18.6	19.6	19.6	19.4	19.3	19.4	19.6	19	19	19.2	19.5	19.8	19.4	19.9	19	19	19.8	19.6	17.3	17.5
6	75.9	76.1	76.1	76	81.6	*	26.8	24.2	24.3	18.9	19.2	19.3	19.3	19.3	19	19.8	19.5	19.4	19.2	19.7	19.6	19.4	19.3	19.4	19.5	20	19.9	20.7	20.2	20.1	20.1	20.1	19.8	20.1
7	75.9	75.6	75.7	75.6	80.9	83.7	*	24	24.7	18.9	19.1	19.4	18.8	19.2	18.9	19.8	19.9	19.2	19.1	19.4	19.4	19.5	19.3	19.2	19.3	19.8	19.9	20.8	20.9	20.2	19.9	23.6	20.7	20.2
8	75.5	75.4	75.4	75.3	83.7	81.2	81.3	*	23.5	18.8	18.1	19	18.6	18.7	18.7	19.4	19.3	19.2	19.3	19.3	19.4	19.2	19.1	19	19.3	20	19.8	20.7	20.2	20	20.1	20.8	19.8	19.9
9	75.2	75.2	75.3	75.4	80.4	81.5	82.1	80.5	*	18.8	18.9	19.4	18.9	19	18.7	19.3	19.3	19.2	19.3	19.3	19.6	19.6	19.4	19	19.5	15.6	19.7	20.4	19.9	20	19.4	19.9	19.5	20
10	73.8	73.7	73.9	73.8	73.7	73.7	73.8	73.6	73.4	*	33.8	21	21.9	22.3	19.2	19.8	19.7	19.7	19.8	19.9	19.8	19.5	20.1	20	20.4	19.6	19.5	20.3	20.1	19.8	20.1	19.9	19.8	19.9
11	74.6	74.5	74.1	74.2	73.7	74.1	73.8	73.6	73.5	87.9	*	21.3	22	22.2	19.4	20	20	19.8	19.9	19.7	20.1	19.6	20.2	20.1	20.4	19.3	19.2	19.1	19.5	19.5	19.4	19.7	19.2	19.1
12	74.6	74.5	74.4	74.3	73.8	74.3	74.2	73.6	73.7	78.2	78.6	*	21.6	21.2	20.1	20.3	20.4	20.4	20.5	20.1	20.4	20.2	20.2	20.4	20.1	19.7	19.5	20.2	19.9	19.9	19.7	20.2	19.5	19.5
13	74.2	74.2	74.1	74	73.3	73.7	73.5	73.2	73.3	79.2	79.3	78	*	27.2	19.9	19.9	19.9	19.7	19.5	19.4	19.3	19.5	20.3	20.1	19.9	19.6	19.2	20.3	19.7	19.6	19.5	19.7	19.4	19.3
14	74.5	74.4	74.2	74.1	73.7	74.1	73.7	73.7	73.5	79.8	79.6	78.1	83.9	*	19.5	19.9	20.1	19.9	20	19.9	19.8	20	19.3	19.8	20.2	19.7	19.5	20.2	19.9	19.9	19.7	20.2	19.5	19.5
15	73.1	73.1	73.5	73.5	73.6	73.4	73.1	73.1	72.8	73.8	74.1	74.3	74	74.3	*	21.6	21.7	19.6	19.7	19.6	20	19.7	19.9	19.8	19.8	19.6	19.4	20.9	20	19.5	19.5	20.5	19.4	18.8
16	74.4	74.2	74.5	74.3	74	74.3	74.4	73.9	74.3	74.9	75.1	75.5	74.9	75.2	78.1	*	79.3	20.9	20.9	20.7	20.9	20.3	20.1	20.2	20.5	20.1	19.7	20.7	20	19.8	19.7	20.1	19.5	19.5
17	74.6	74.6	74.5	74.5	74.2	74.3	74.4	73.9	74.3	74.9	75	75.4	74.9	75.2	78.5	97.6	*	20.8	20.8	20.3	20.7	20.3	20.1	20.6	20.6	20.2	19.6	20.8	20.1	19.4	19.8	20	19.4	19.3
18	74.4	74.3	74.3	74.2	73.9	74.3	73.9	73.9	73.4	75.2	71	76.3	74.8	75.1	75.1	76.7	76.5	*	48.8	25.6	26.3	24.7	27	24.8	27.2	19.6	19.3	19.8	19.6	19.4	19.5	19.7	19.1	19.2
19	74.4	74.5	74.3	74.2	73.9	74.3	73.9	73.4	73.2	75.2	72.1	76.4	74.3	75.1	75.2	76.4	76.5	92.4	*	26.4	26.5	24.3	27.1	25	27.3	19.7	19.5	20.5	19.8	19.7	19.9	20.1	19.3	19.5
20	73.9	73.8	73.9	73.8	73.7	74.5	73.3	73.8	73.5	74.6	74.9	75.2	74	75	75.3	76.7	76.5	81.7	83.2	*	99.3	23.8	26.4	24	26.4	19.3	19.5	20.5	19.7	19.6	19.6	19.8	19.5	19.5
21	75.5	75.4	75.1	75.4	73.7	74.7	73.3	73.9	73.9	74.5	74	75.2	74.8	75.1	76.6	76.6	76.5	81.2	82.1	99.9	*	24.3	27	24.5	27	19.9	19.5	20.1	19.6	19.7	19.8	20.5	19.6	19.3
22	75.9	75.8	74.7	74.6	74.2	74.8	73.4	73.2	73.4	75.2	75.3	76.8	75.2	71.6	76.4	76.4	76.5	82.6	81.9	81.1	81.4	*	25.3	25.7	25	19.7	19.5	19.7	19.7	19.6	19.8	19.9	19.5	19.4
23	74.8	74.7	74.7	74.5	73.8	74.3	74.1	73.9	73.4	75.4	75.5	75.6	75.5	75.7	76	75.7	76.7	85.1	83.6	83.2	83.4	83.2	*	25.3	30.9	19.8	19.2	20.2	19.4	19.6	19.5	20	19.5	19.3
24	74.3	74.5	73.8	74	73.2	73.7	73.4	73.3	73.3	74.5	74.8	75.3	74.3	74.7	76.1	76.7	76.5	82.9	81.8	82.6	81.4	82.9	82.9	*	24.9	19.8	19.6	20.6	19.7	19.3	19.6	20.1	19.2	19.4
25	74.7	74.3	74.4	74.5	73.9	74.6	74	74.1	73.5	75.6	75.9	76.6	75.1	75.6	76.6	76.4	76.5	85.1	84.3	84.7	83.4	82.8	86.8	82.6	*	19.8	19.7	20.7	20	19.6	19.8	20.5	19.4	19.5
26	75.8	75.7	76.3	76.2	76.1	74.8	74.9	74.6	74.4	74.1	74.4	74.5	74.2	74.6	73.6	74.2	74.2	74.4	74.4	74.1	74.1	74.8	74.6	73.9	74.6	*	22	25.9	23	22.1	22.4	24.1	22.2	21
27	75.3	75.2	75.1	75.2	75.6	74.7	74.6	74.5	74.3	74.1	74.3	73.8	74.1	73.4	73.2	74.3	74.3	74	74	73.8	73.8	74.3	74.4	73.6	74.3	78.1	*	23.2	24.6	24.7	22.5	23.1	21.6	25
28	76.8	76.7	76.9	76.8	75.7	76.3	76.3	75.7	75.5	75.4	72	76.2	75.7	76.3	74.9	76.1	76.1	75.9	75.9	75.6	75.6	76.5	76.3	75.3	76.1	85.2	78	*	24.9	24.9	22.5	27.4	36.8	27.9
29	76	76.1	75.9	75.8	75.2	75.5	75.4	75.2	74.9	74.5	74.8	75.1	74.7	74.9	74.1	75.3	75.3	74.9	74.9	74.3	74.3	75.4	75.4	74.4	75	78	81.2	81.6	*	22.2	23.1	23.1	24.4	23.3
30	76	75.9	76.1	76	74.8	75.5	75.3	75.3	75	74.3	74.3	74.3	73.9	74.5	73.2	74.4	74.4	74.1	74.1	73.9	73.9	74.3	74.5	73.6	74.6	78.1	81.2	79.6	78.9	*	22.2	23.1	21.6	23.7
31	75.9	75.8	75.5	75.4	74.2	74.6	74.9	74.5	74.8	74.2	74.2	74.3	74.2	74.6	73.4	74.7	74.7	74.3	74.3	73.9	73.9	74.4	74.4	73.7	74.4	78.7	78	80.8	80.1	78.9	*	23.5	27.9	22.8
32	76.5	76.4	76.7	76.6	75.2	76	76.9	75.5	74.9	74.8	75.1	75.4	75	75.6	74.4	75.3	75.3	75.2	75.2	74.8	74.8	75.9	76.5	74.8	75.7	80.9	79.1	83.6	81.5	79.5	80.5	*	23.1	23.6
33	75.2	75.3	75.3	75.2	74.2	74.9	74.7	74.7	74.3	73.7	73.8	74.1	74.1	74.5	73	74.4	74.4	74.2	74.2	73.6	73.6	74.4	74.2	73.5	74.4	78.9	78	80.4	79.7	78.2	84.7	80.1	*	23.1
34	75.8	75.9	75.8	75.7	74.8	75.4	75.2	75.3	74.9	74.7	74.3	74.3	74.1	74.9	73.3	74.5	74.5	73.9	73.9	73.5	73.5	73.9	74.6	73.6	74.5	78	81.4	79.6	78.7	91.3	78.9	79.2	78.1	*

1.*R. gilardi* subsp. *rosea* ATCC BAA-691^T, 2. *R. gilardi* subsp. *rosea* NCTC 13290^T, 3. *R. mucosa* NCTC 13291^T, 4. *R. mucosa* ATCC BAA-692^T, 5. *R. aerilata* DSM 19363^T, 6. *R. pecuniae* N75^T, 7. *R. rosea* DSM 14916^T, 8. *R. vinacea* DSM 19362^T, 9. *R. harenae* CP101081^T, 10. *R. stagni* DSM 19981^T, 11. *R. algicola* PeD5^T, 12. *R. bella* CQN31^T, 13. *R. frigidaquae* JCM 15073^T, 14. *R. selenitidurans* BU-1^T, 15. *R.arctica* LMG 28251^T, 16. *R. hellenica* LMG 31523^T, 17. *R. hellenica* LMG 31524, 18. *R. oryzicola* KCTC 22478^T, 19. Strain KCTC 32190^T, 20. *R. alkaliterrae* DSM 25895^T, 21. *R. alkaliterrae* LMG 31230^T, 22. *R. lacus* CGMCC 1.3617^T, 23. *R. eburnea* LMG 31228^T, 24. *R. terrae* LMG 31159^T, 25. *R. soli* LMG 31231^T, 26. *R. deserti* M3^T, 27. *R. aerophila* NBRC 108923^T, 28. *R. cervicalis* ATCC 49957^T, 29. *R. coralli* M0104^T, 30. *R. ludipueritiae* DSM 14915^T, 31. *R. oryzae* KCTC 42542^T, 32. *R. rhizosphaerae* YW11^T, 33. *R. vastitatis* CPCC 101021^T, 34. *R. wenyumeiae* Z23^T

Table 21. POCP and AAI values between the members of genus *Roseomonas*

POCP/ANI	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	*	99.9	93.3	92.9	65.7	66.5	66.3	65.2	65.3	61.9	62.7	61.4	62.7	62.2	61.5	62.6	62.6	63.9	61.8	62.8	62.2	62.2	63.1	61.8	62.7	65.1	64.7	66.5	65.7	65.8	65.3	66.2	64.8	65.3
2	94.1	*	93.2	93.1	65.5	66.4	66.3	65.4	65.7	62.0	62.8	61.5	62.6	62.4	61.7	62.9	62.9	64.0	64.3	62.2	62.7	62.7	63.3	62.3	63.0	65.1	64.8	66.6	65.6	65.9	65.3	66.1	64.9	65.5
3	84.6	81.0	*	99.9	65.2	66.4	65.9	64.8	65.4	61.4	61.6	61.0	62.0	61.5	61.4	62.7	62.7	62.7	63.0	62.1	61.4	62.1	62.5	61.3	61.8	65.1	64.5	66.2	65.2	65.8	64.9	66.0	64.9	65.4
4	82.6	81.7	98.5	*	65.2	66.3	65.7	64.8	65.5	61.2	61.6	61.2	61.8	61.4	61.3	62.7	62.7	62.8	63.0	61.3	62.1	62.6	62.5	61.3	62.0	65.1	64.5	66.2	65.1	65.8	65.0	66.0	64.8	65.2
5	55.9	52.5	56.2	53.1	*	78.8	78.8	83.5	78.0	62.7	62.5	62.1	62.8	62.3	62.6	63.6	63.6	63.6	63.4	61.5	61.6	62.8	61.9	61.7	62.6	63.7	63.7	65.6	64.7	64.9	64.5	64.8	65.2	65.2
6	62.1	59.4	62.7	59.8	66.4	*	82.8	78.6	78.4	63.5	63.4	63.0	63.5	63.3	63.6	64.0	64.0	64.4	64.7	63.7	62.9	64.0	63.5	62.8	63.8	65.1	64.6	66.5	65.7	66.1	65.5	66.1	66.1	65.7
7	59.5	57.7	60.2	57.9	66.6	70.7	*	78.6	81.1	63.2	63.2	63.0	63.4	63.0	62.9	65.0	65.0	64.5	64.5	63.5	62.3	63.6	63.3	62.3	63.6	64.7	64.3	66.7	65.4	66.1	65.7	67.2	65.5	65.5
8	54.1	52.8	54.0	53.9	73.7	67.1	65.8	*	78.9	62.9	63.4	62.7	63.5	62.7	62.9	63.5	63.5	63.7	63.7	61.8	62.8	63.0	62.2	64.4	63.2	63.9	63.6	65.6	64.5	65.4	64.6	65.0	65.9	65.5
9	56.8	56.1	56.3	56.2	65.7	67.9	70.9	65.8	*	63.0	63.3	63.2	63.3	63.0	63.1	65.4	65.4	63.7	63.9	62.8	62.7	62.9	63.3	62.7	63.7	64.9	64.5	65.9	65.6	65.6	64.5	65.8	65.4	65.2
10	52.0	50.9	49.7	49.9	55.5	56.3	58.0	56.0	56.8	*	73.2	75.2	75.6	90.2	64.6	65.3	65.3	67.2	67.4	65.6	65.5	66.8	66.2	65.2	66.5	62.5	62.3	65.5	63.3	62.9	62.5	64.0	62.7	62.3
11	53.4	52.4	48.8	49.1	52.1	53.7	55.7	52.8	53.9	67.9	*	72.0	72.7	73.5	64.3	64.5	64.5	68.8	69.0	66.8	66.0	67.1	67.2	65.6	67.6	62.1	61.8	65.3	63.6	62.7	62.7	64.4	62.4	62.2
12	53.2	51.8	50.8	51.1	55.4	55.9	60.4	56.5	58.2	72.8	66.9	*	83.1	75.3	64.5	65.1	65.1	67.0	67.0	66.0	66.1	66.3	66.1	64.9	65.9	62.5	62.6	65.0	63.5	62.7	62.5	64.1	62.5	62.5
13	53.9	52.4	49.2	49.6	54.4	54.9	57.7	55.7	56.7	73.6	67.8	78.5	*	75.9	64.8	65.8	65.8	67.9	67.9	65.9	65.9	66.7	66.6	65.4	66.5	62.8	62.9	66.0	64.0	63.2	63.0	64.8	63.0	62.7
14	52.8	51.9	49.2	49.5	53.9	55.0	56.9	54.9	56.3	86.7	68.1	71.0	75.7	*	64.6	65.4	65.4	67.3	67.5	65.4	65.5	66.6	66.3	65.1	66.4	62.6	62.3	65.4	63.3	62.8	62.7	64.3	62.4	62.3
15	52.9	53.6	50.3	50.7	52.6	55.4	56.8	53.6	57.4	53.1	57.4	60.5	60.3	60.2	*	*73.8*	*73.8*	67.5	67.7	66.4	66.8	66.9	66.9	66.0	66.9	62.2	62.0	64.1	63.1	61.9	62.8	63.4	62.6	61.6
16	50.1	49.1	49.5	49.8	53.1	53.5	57.9	53.2	58.4	57.4	53.6	58.8	56.9	57.0	*65.8*	*	*99.9*	68.9	68.8	67.3	67.8	67.5	68.0	66.6	68.2	63.4	63.2	65.2	64.7	63.6	63.8	65.1	63.5	62.2
17	50.1	49.1	49.5	49.8	53.1	53.5	57.9	53.2	58.4	57.4	53.6	58.8	56.9	57.0	*65.8*	*99.3*	*	68.9	68.8	67.3	67.8	68.1	68.0	67.2	68.2	63.4	63.2	65.2	64.7	63.6	63.8	65.1	63.5	62.8
18	51.5	53.7	47.1	48.6	49.5	53.2	54.9	50.4	54.5	60.0	64.4	61.4	63.2	62.7	63.5	55.4	55.4	*	94.5	80.8	79.3	81.3	81.9	79.1	81.0	63.8	63.2	66.6	64.5	63.4	63.9	66.0	63.7	63.4
19	52.9	52.2	48.2	47.5	48.5	51.9	53.7	49.4	53.1	58.7	63.9	60.0	61.6	61.3	61.6	54.3	54.3	94.0	*	79.5	81.3	81.2	81.8	79.1	81.1	63.7	63.2	66.2	64.6	63.5	64.1	66.3	63.7	63.6
20	51.8	50.0	47.5	48.6	51.4	52.1	54.2	50.9	54.0	60.4	63.0	61.0	61.1	60.5	60.7	59.8	59.8	69.0	68.3	*	79.7	79.7	79.8	81.5	80.8	63.2	62.7	65.5	64.4	63.1	63.3	65.1	61.9	61.9
21	52.0	52.2	49.9	50.2	52.4	54.8	56.4	53.2	57.0	63.3	63.7	63.0	62.3	63.2	63.3	61.1	61.1	74.9	73.0	75.4	*	79.9	82.1	79.0	81.0	62.4	61.8	64.5	63.6	62.3	63.4	64.0	62.1	64.4
22	51.7	52.0	49.8	50.0	52.1	54.4	56.2	53.0	56.7	62.9	63.5	62.6	61.9	63.5	63.1	60.9	60.9	74.6	73.1	75.1	99.2	*	82.2	79.2	82.4	63.2	62.7	65.6	64.5	63.0	63.4	65.0	63.2	62.8
23	51.2	51.3	49.3	49.7	49.1	51.2	54.4	49.3	53.3	59.1	61.6	60.6	60.3	60.9	60.3	58.9	58.9	73.9	72.6	71.8	74.9	74.7	*	79.3	86.1	63.2	62.4	65.6	64.3	63.0	63.3	66.7	63.0	62.8
24	50.5	50.6	47.5	47.9	50.1	51.7	54.8	50.1	54.6	58.3	58.3	59.2	58.9	58.5	60.4	59.3	59.3	68.9	67.9	72.3	70.6	70.4	69.0	*	79.0	62.5	61.5	64.0	63.5	62.2	62.7	64.2	62.4	61.9
25	51.1	51.3	47.6	48.0	49.5	52.2	54.3	49.8	54.3	59.4	62.8	59.8	60.3	60.3	61.5	57.7	57.7	73.9	72.7	72.9	76.8	76.9	76.3	70.2	*	63.1	62.4	65.8	64.4	63.5	63.8	65.5	63.2	63.2
26	57.8	57.1	56.7	57.0	53.8	56.9	58.9	54.1	57.9	54.0	51.1	57.7	54.5	54.0	53.4	56.1	56.1	51.0	49.8	53.0	55.2	54.9	53.3	54.1	52.7	*	70.8	78.5	73.7	70.9	73.0	76.1	72.7	70.1
27	59.6	59.0	57.2	57.4	53.4	56.0	58.7	53.6	57.2	53.9	51.9	56.7	54.2	53.4	53.1	54.2	54.2	51.9	51.0	52.7	54.9	54.6	52.1	53.0	51.9	67.9	*	72.6	71.1	77.7	71.2	71.8	71.0	77.3
28	56.7	57.2	54.2	54.5	50.0	53.4	55.9	50.4	53.9	55.3	51.9	54.5	54.4	53.1	51.4	49.5	49.5	53.9	53.1	50.3	57.4	57.4	52.6	50.5	52.2	65.1	66.8	*	76.8	72.4	75.6	88.9	75.2	72.1
29	61.1	61.3	59.3	59.7	55.0	60.7	59.7	55.2	59.8	54.5	54.6	57.2	56.0	54.0	55.7	54.9	54.9	54.5	52.9	54.0	58.8	58.8	54.1	55.0	54.8	65.2	66.3	68.6	*	73.4	75.0	79.7	74.9	72.8
30	59.8	59.0	58.9	59.1	54.9	58.4	60.4	55.5	58.0	52.6	52.1	55.9	53.4	52.2	52.0	53.5	53.5	51.2	50.3	50.7	54.3	54.3	51.4	52.3	51.5	65.7	69.6	66.8	66.3	*	75.5	76.7	75.9	72.0
31	60.0	60.0	58.4	58.7	54.2	59.7	59.4	54.3	60.1	52.4	53.0	54.5	52.9	52.5	55.0	53.8	53.8	53.7	52.2	51.4	57.4	57.4	52.4	54.1	53.9	75.0	65.3	66.8	70.3	66.4	*	73.0	72.1	92.3
32	59.1	60.1	59.0	58.9	52.6	59.4	60.8	53.3	57.8	54.1	55.3	57.0	55.8	55.1	56.8	52.8	52.8	57.2	55.8	53.2	55.8	56.1	57.5	54.1	56.5	66.4	66.0	66.0	68.9	65.7	69.3	*	84.7	72.5
33	59.2	58.8	58.5	58.5	56.0	60.7	58.1	55.1	58.8	52.4	50.9	53.8	52.1	50.9	52.2	53.0	53.0	50.1	48.8	50.0	51.9	51.5	49.6	50.7	50.1	65.4	65.8	67.6	69.1	65.7	75.0	66.4	*	72.4
34	57.9	56.8	57.5	57.3	57.7	59.0	59.2	57.5	59.4	53.7	51.3	55.5	52.8	52.1	52.0	53.6	53.6	50.3	49.3	51.4	53.3	53.0	50.4	51.3	51.2	65.1	68.4	68.8	65.2	79.4	65.9	66.9	66.4	*

1.*R. gilardi* subsp. *rosea* ATCC BAA-691^T, 2. *R. gilardi* subsp. *rosea* NCTC 13290^T, 3. *R. mucosa* NCTC 13291^T, 4. *R. mucosa* ATCC BAA-692^T, 5. *R. aerilata* DSM 19363^T, 6. *R. pecuniae* N75^T, 7. *R. rosea* DSM 14916^T, 8. *R. vinacea* DSM 19362^T, 9. *R. harenae* CP101081^T, 10. *R. stagni* DSM 19981^T, 11. *R. algicola* PeD5^T, 12. *R. bella* CQN31^T, 13. *R. frigidaquae* JCM 15073^T, 14. *R. selenitidurans* BU-1^T, 15. *R.arctica* LMG 28251^T, 16. *R. hellenica* LMG 31523^T, 17. *R. hellenica* LMG 31524, 18. *R. oryzicola* KCTC 22478^T, 19. Strain KCTC 32190^T, 20. *R. alkaliterrae* DSM 25895^T, 21. *R. alkaliterrae* LMG 31230^T, 22. *R. lacus* CGMCC 1.3617^T, 23. *R. eburnea* LMG 31228^T, 24. *R. terrae* LMG 31159^T, 25. *R. soli* LMG 31231^T, 26. *R. deserti* M3^T, 27. *R. aerophila* NBRC 108923^T, 28. *R. cervicalis* ATCC 49957^T, 29. *R. coralli* M0104^T, 30. *R. ludipueritiae* DSM 14915^T, 31. *R. oryzae* KCTC 42542^T, 32. *R. rhizosphaerae* YW11^T, 33. *R. vastitatis* CPCC 101021^T, 34. *R. wenyumeiae* Z23^T

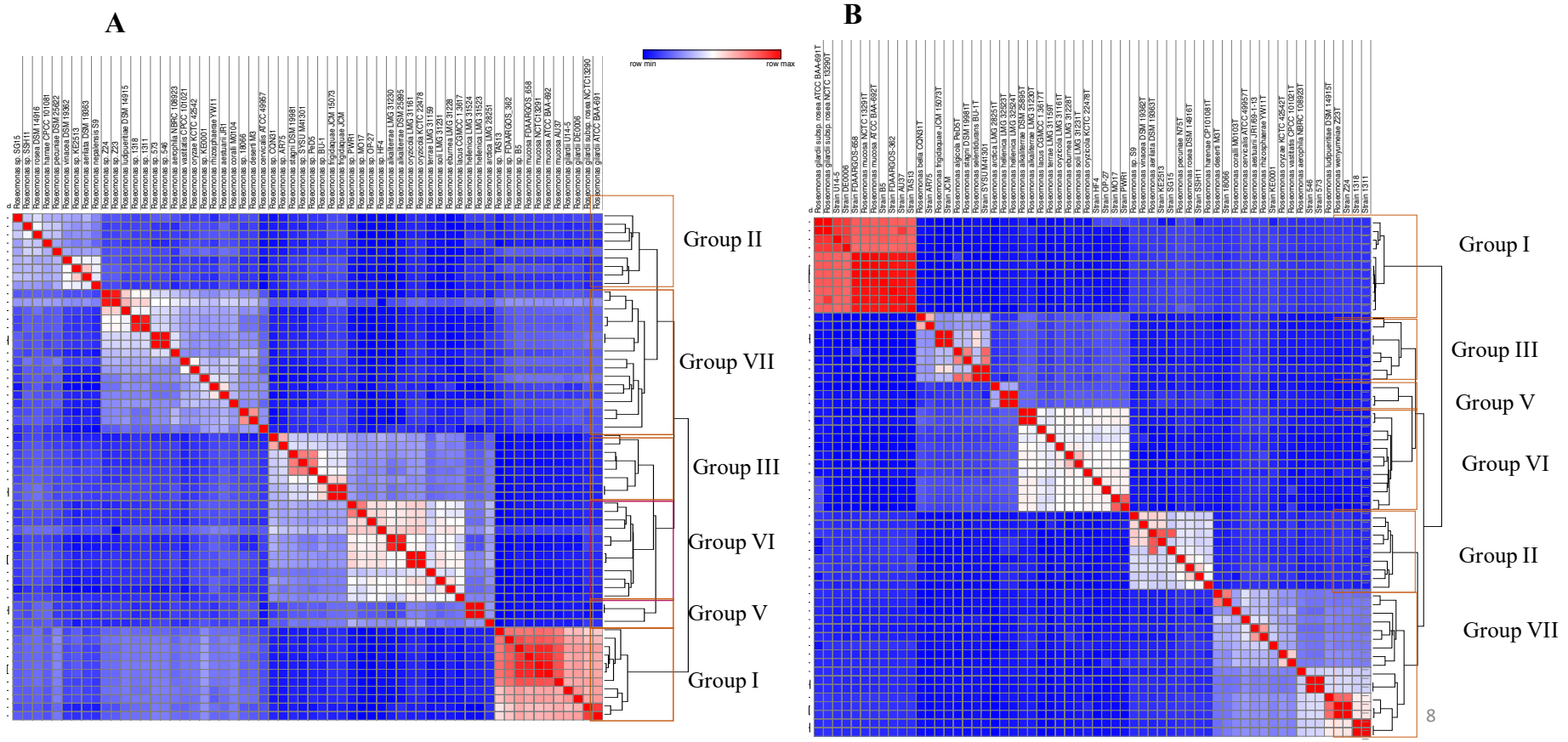


Fig. 45. Heatmap representing (A) POCP (B) AAI values of the members of genus *Roseomonas* clustered with the Euclidean distance method

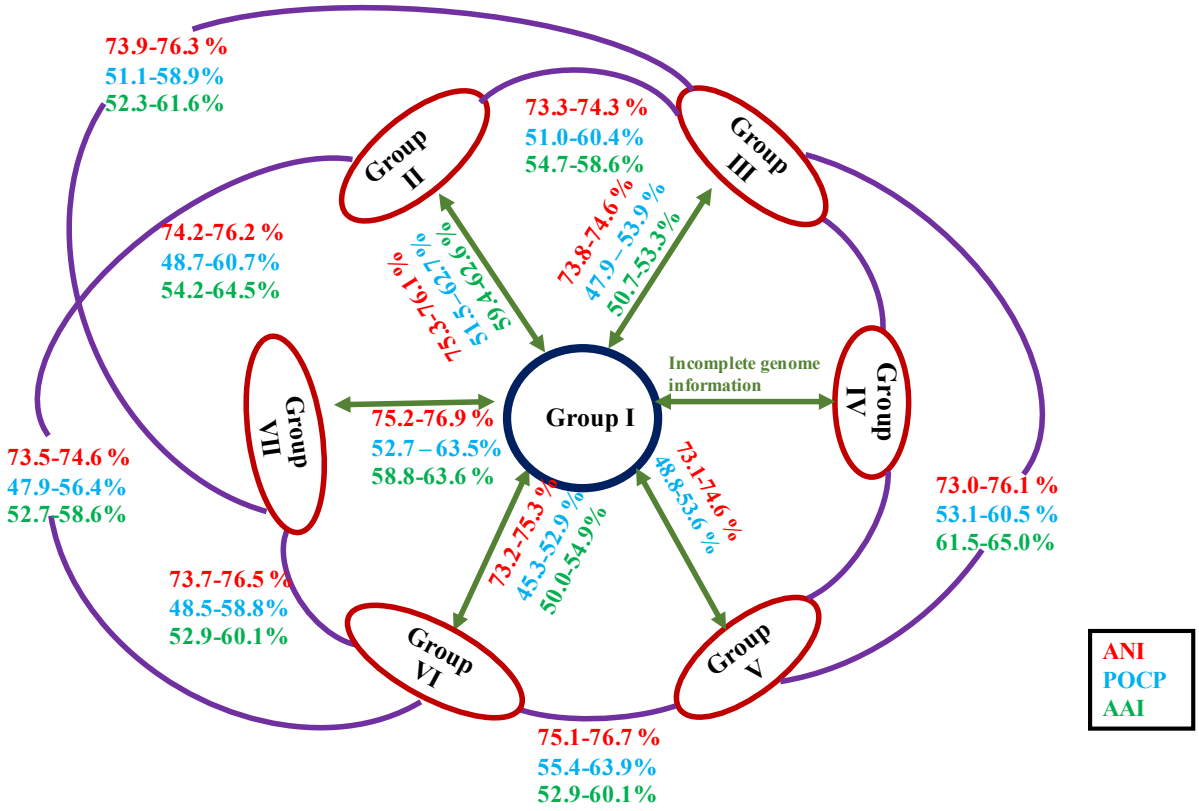
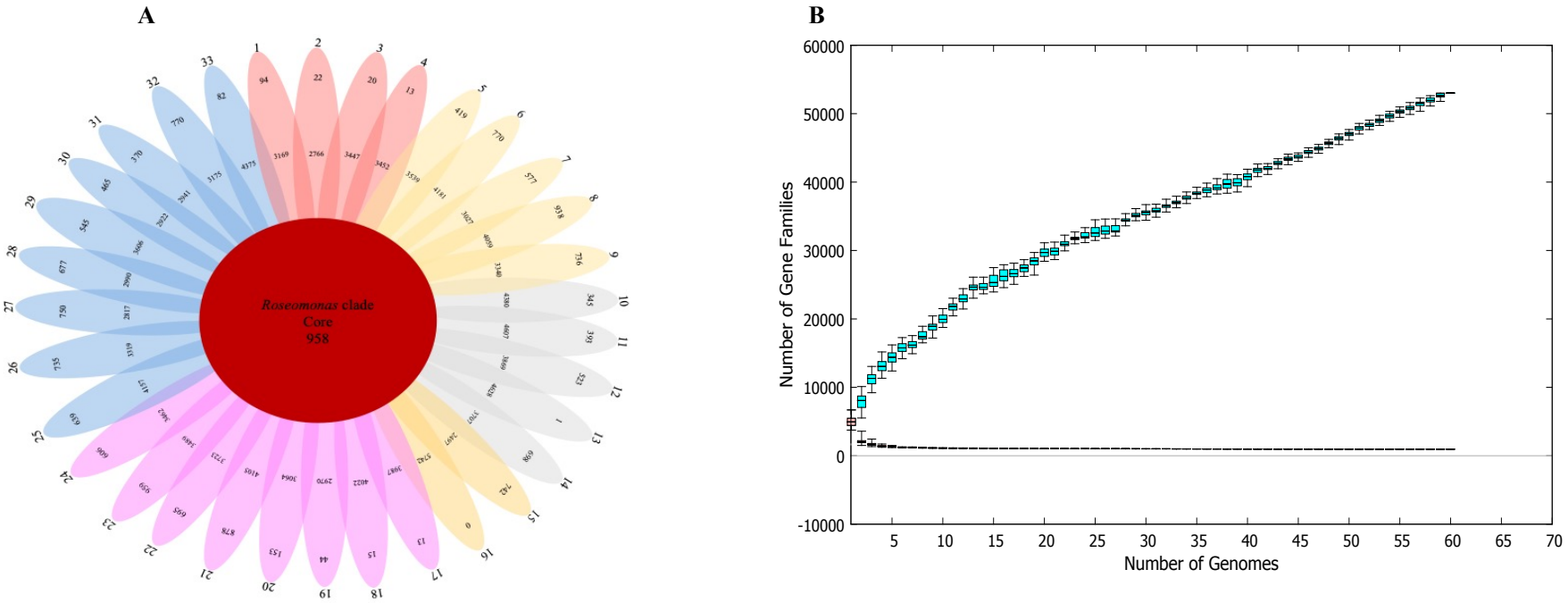


Fig. 46. Diagrammatic representation of the ANI, AAI, and POCP values in between the members of each group of the genus *Roseomonas*

3.3.4.5. Analysis of core and pan-genome

Table 22 comprises the distribution of genes for the genus *Roseomonas* as given by BPGA. Members of the genus *Roseomonas* share 958 core genes (19.1 %), 219,753 accessory genes (72.8 %), and 24,320 strain-specific genes (8.1 %) (Fig. 47A). Only the type strains were selected for the diagrammatic depiction of the pan-genome for the genus *Roseomonas*. The core-pan plot (Fig. 47B) revealed an open pan-genome for the genus *Roseomonas* since it did not plateau and expanded as the number of genomes increased. As the plot levelled out, the core genome was conserved at the genus level. The core genes of members of Groups I, II, III, V, VI, and VII ranged from 1,573 to 3,152. (Table 22).



(1. *R. gilardi* subsp. *rosea* ATCC BAA-691^{T}; 2. *R. gilardi* subsp. *rosea* NCTC 13290^{T}; 3. *R. mucosa* ATCC BAA-692^{T}; 4. *R. mucosa* NCTC 13291^{T}; 5. *R. rosea* DSM 14916 ^{T}; 6. *R. aerilata* DSM 19363^{T}; 7. *R. pecuniae* N75^{T}; 8. *R.* sp. *vinacea* DSM 19362^{T}; 9. *R. harenae* CPCC 101081^{T}; 10. *R. stagni* DSM 19981^{T}; 11. *R. algicola* PeD5 ^{T}; 12. *R. bella* CQN31^{T}; 13. *R. frigidaquae* JCM 15073^{T}; 14. *R. selenitidurans* BU-1^{T}; 15. *R. arctica* LMG 28251^{T}; 16. *'R hellenica'* LMG 31523^{T}; 17. *R. oryzicola* KCTC 22478^{T}; 18. strain KCTC 31290^{T}; 19. *R. alkaliterrae* DSM 25895 ^{T}; 20. *R. alkaliterrae* LMG 31230^{T}; 21. *R. lacus* CGMCC 1.3617^{T}; 22. *R. eburnea* LMG 31228^{T}; 23. *R. terrae* LMG 31159^{T}; 24. *R. soli* LMG 31523^{T}; 25. *R. deserti* M3^{T}; 26. *R. aerophila* NBRC 108923^{T}; 27. *R. cervicalis* ATCC 49957 ^{T}; 28. *R. coralli* M0104^{T}; 29. *R. ludipueritiae* DSM 14915^{T}; 30. *R. oryzae* KCTC 42542 ^{T}; 31. *R. rhizosphaerae* YW11^{T}; 32. *R. vastitatis* CPCC 101021^{T} ; 33. *R. wenyumeiae* Z23^{T})

Fig. 47. Analysis of pan and core genomes using the BPGA pipeline of the 32 type strains and strain KCTC 32190^{T} of genus *Roseomonas*
(A) Flower-pot diagram representing core, accessory and unique genes of the genomes of all strains. It consists of 958 core genes, summing up to 19.1 % of the total pan-genome
(B) Core-pan-plot for all the genomes and shows that the is pan genome is open

Table 22. BPGA-based pan and core genome analysis of the genus *Roseomonas* and its closely related members of the family *Acetobacteraceae*

Microorganisms	No. of core genes	No. of accessory genes	No. of unique genes	No. of exclusively absent genes
(n=60)				
R. gilardii subsp. *rosea* ATCC BAA-691^T	958	3169	94	0
R. gilardii subsp. *rosea* NCTC 13290^T	958	2766	22	3
R. mucosa ATCC BAA-692^T	958	3447	20	0
R. mucosa NCTC 13291^T	958	3452	13	1
Strain DE0006	958	3352	315	0
Strain U14-5	958	3504	408	0
Strain AU37	958	3269	143	0
Strain B5	958	3357	17	0
Strain TAS13	958	3578	308	3
Strain FDAARGOS_658	958	3443	62	0
Strain FDAARGOS_362	958	3382	142	0
R. rosea DSM 14916^T	958	3539	419	1
R. aerilata DSM 19363^T	958	4181	770	0
R. pecuniae N75^T	958	3027	577	1
R. vinacea DSM 19362^T	958	4059	938	0
R. harenae CPCC 101081^T	958	3340	736	1
Strain SSH11	958	3275	574	0
Strain KE2513	958	4239	690	1
Strain S9.3B	958	4012	1220	1
Strain SG15	958	3568	831	4
R. stagni DSM 19981^T	958	4380	345	2
R. algicola PeD5^T	958	4607	393	0
R. bella CQN31^T	958	3869	523	2
R. frigidaquae JCM 15073^T	958	4628	1	0
R. selenitidurans BU-1^T	958	3707	698	1
Strain AR75	958	4068	741	0
Strain SYSU M4 1301	958	4632	443	0
Strain JCM	958	4629	0	0
R. arctica LMG 28251^T	958	2497	742	15
'R. hellenica' LMG 31523^T	958	5742	0	0
Strain LMG 31524	958	5743	0	0
R. oryzicola KCTC 22478^T	958	3987	13	0
Strain KCTC 32190^T	958	4022	53	0
R. alkaliterrae DSM 25895^T	958	2970	44	0
R. alkaliterrae LMG 31230^T	958	3064	153	3
R. lacus CGMCC 1.3617^T	958	4105	878	0
R. eburnea LMG 31228^T	958	3723	695	0
R. terrae LMG 31159^T	958	3489	959	2
R. soli LMG 31523^T	958	3462	606	1
Strain OP-27	958	3608	620	0
Strain PWR1	958	3431	314	0
Strain HF4	958	3537	505	1
Strain MO17	958	3557	438	1
R. deserti M3^T	958	4157	639	2
R. aerophila NBRC 108923^T	958	3319	735	2
R. cervicalis ATCC 49957^T	958	2817	750	20
R. coralli M0104^T	958	2990	677	2
R. ludipueritiae DSM 14915^T	958	3606	545	2
R. oryzae KCTC 42542^T	958	2922	465	1
R. rhizosphaerae YW11^T	958	2941	370	4
R. vastitatis CPCC 101021^T	958	3175	770	0
R. wenyumeia Z23^T	958	4375	82	0

Microorganism	No. of core genes	No. of accessory genes	No. of unique genes	No. of exclusively absent genes
Strain JR1/69-1-13	958	3353	435	0
Strain 18066	958	4117	581	2
Strain KE0001	958	2648	460	3
Strain 546	958	3453	24	0
Strain 573	958	3457	26	0
Strain 1311	958	3344	70	0
Strain 1318	958	3304	112	0
Strain Z24	958	4359	116	0
Group I clade (n=11)				
R. gilardii subsp. *rosea* ATCC BAA-691^T	3152	883	153	3
R. gilardii subsp. *rosea* NCTC 13290^T	3152	558	20	147
R. mucosa NCTC 13291^T	3152	1236	9	1
R. mucosa ATCC BAA-692^T	3152	1230	21	0
Strain DE0006	3152	986	465	45
Strain U14-5	3152	1128	564	11
Strain AU37	3152	1025	164	5
Strain B5	3152	1211	388	24
Strain TAS13	3152	1201	83	3
Strain FDAARGOS_658	3152	1120	174	8
Strain FDAARGOS_362	3152	883	153	3
Group II clade (n=9)				
R. rosea DSM 14916^T	2194	2042	652	25
R. aerilata DSM 19363^T	2194	2774	879	4
R. pecuniae N75^T	2194	1585	742	130
R. vinacea DSM 19362^T	2194	2662	1062	11
R. harenae CPCC 101081^T	2194	1750	1044	88
Strain SSH11	2194	1869	712	37
Strain KE2513	2194	2818	816	19
Strain S9.3B	2194	2557	1380	19
Strain SG15	2194	2019	1116	76
Group III clade (n=8)				
R. stagni DSM 19981^T	1573	3643	450	8
R. algicola PeD5^T	1573	3861	495	0
R. bella CQN31^T	1573	2242	1512	56
R. frigidaquae JCM 15073^T	1573	3979	1	0
R. selenitidurans BU-1^T	1573	2865	898	22
Strain AR75	1573	584	2199	1099
Strain SYSU M4 1301	1573	3873	565	2
Strain JCM	1573	3979	0	0
Group V clade (n=3)				
R. arctica LMG 28251^T	2927	1	1242	3700
R. hellenica LMG 31523^T	2927	3701	0	0
Strain LMG 31524	2927	3700	0	1
Group VI clade (n=12)				
R. lacus CGMCC 1.3617^T	2448	2247	1199	15
R. oryzicola KCTC 22478^T	2448	2464	15	1
Strain KCTC 32190^T=JC162^T	2448	2478	53	2
R. alkaliterrae DSM 25895^T	2448	1454	53	2
R. alkaliterrae LMG 31230	2448	1497	148	4
R. eburnea LMG 31228^T	2448	1967	922	14
R. terrae LMG 31159^T	2448	1655	1262	116
R. soli LMG 31523^T	2448	1836	702	23
StrainOP-27	2448	1953	682	31
Strain PWR1	2448	1853	385	9
Strain HF4	2448	1902	616	29

Microorganism	No. of core genes	No. of accessory genes	No. of unique genes	No. of exclusively absent genes
Group VII clade (n=17)				
R. cervicalis ATCC 49957^T	1591	1982	870	78
R. deserti M3^T	1591	3420	691	5
R. aerophila NBRC 108923^T	1591	2471	928	15
R. coralli M0104^T	1591	2159	858	25
R. ludipueritiae DSM 14915^T	1591	2833	647	6
R. oryzae KCTC 42542^T	1591	2190	548	17
R. rhizosphaerae YW11^T	1591	2139	525	26
R. vastitatis CPCC 101021^T	1591	2400	890	4
R. wenyumeia Z23^T	1591	3699	94	0
Strain JR1/69-1-13	1591	2610	516	3
Strain 18066	1591	3353	672	4
Strain KE0001	1591	1892	572	27
Strain 546	1591	2804	24	0
Strain 573	1591	2808	25	0
Strain 1311	1591	2686	74	0
Strain 1318	1591	2650	116	0
Strain Z24	1591	3668	118	2

3.3.4.6. Functional and metabolic annotations

The alignment of LCB of the two species differed differently. The majority of the proteins of strain KCTC 32190^T share 80-90 % identity with those of *R. oryzicola* KCTC 22478^T, while few other proteins shared 10-30 % similarity. Thus, suggesting a clear-cut dissimilarity of strain KCTC 32190^T from *R. oryzicola* KCTC 22478^T. With other members of the genus *Roseomonas,* the protein sequence identity match was below 80 %. The annotation of orthologous gene clusters between the members of the genus *Roseomonas* showed that strain KCTC 32190^T has 4651 gene clusters, whereas other members have between 3628 to 4902 clusters. For the metabolic characterization (genome based), only members of Group VI have been considered. All the members contain the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP) pathway. All the genes for neurosporene (carotenoid) biosynthesis (and thus the entire pathway) were observed in members of strain KCTC 32190^T and *R. oryzicola* KCTC 22478^T. In contrast other members lacked the complete set of genes for the pathway (Fig. 48A). Similarly, *in-silico* analysis of metabolic pathways like sulphur metabolism, tryptophan metabolism, and

nitrogen metabolism showed that strain KCTC 32190^T differed significantly from other members in gene composition (Fig. 48B, C, D). The presence of a putative gene coding for squalene synthase (EC 2.5.1.21) in strain KCTC 32190^T indicates its probable capability to biosynthesize hopanoids or sterols.

Based on the anti-SMASH 5.0 web server's BGC prediction, the average *Roseomonas* genome had 7 BCGs (420 BCGs were found in 60 genomes). Members of Group I (n = 11) had 65 BCGs (divided into 11 types of BGC families); Group II (n = 9) had 72 (19 BCG families); Group III (n = 8) had 59 (13 BGC families); Group V (n = 3) with 35 (9 BCG families); Group VI (n = 12) with 68 (14 BGC families); and Group VII (n = 17) with 124 BGC families (24 BGC families) (Table 23). Terpenes were the most common (2 per genome on average), but each genome also had one type 1 polyketide synthase (TPKS). Non-ribosomal peptide synthetases (NRPS), thioamitides, and beta-lactone (fengymycin) were found solely in Group I members (asukamycin). Both Group I and II had BGC as phosphonate/terpene (phosphinothricin tripeptide). Group II members had the redox cofactor BGC, and Groups II and III members had type 3 polyketide synthases in common (T3PKS). Group V members possessed NRPS/T1PKS, and arylpolene BGCs. In Group VI, NRPS were found in eight of the 17 genomes, while arylpolene (xanthomonadin) BGCs were found in five genomes. In Group VII, four genomes had just RiPPs BGC, which were ribosomally generated and post-translationally modified peptides (RiPPs). The annotation of genomes revealed that Groups I, II, and III each have 226, 58, and 82 gene clusters, respectively. Members of Group V, VI, VII had distinct 67, 74, and 16 gene clusters, respectively. The genomes of members of the genus *Roseomonas* did not contain any genes relevant to virulence and pathogenicity, as predicted by the IslandViewer 4 service and VirulenceFinder 2.0.

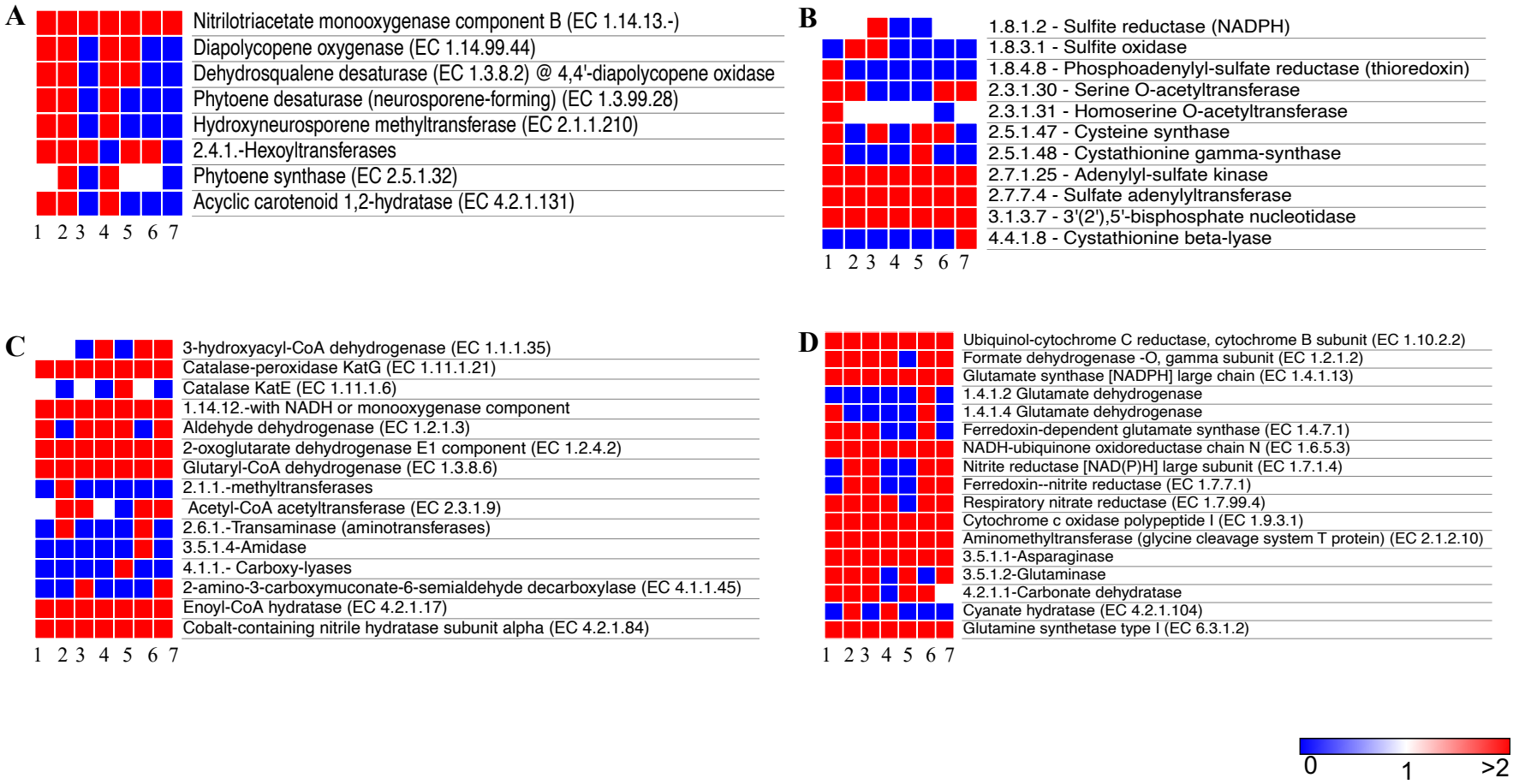


Fig. 48. Heatmap representing no. of genes involved in metabolic processes for strain KCTC 32190^T and Group VI members of genus *Roseomonas*.
(A). Carotenoid biosynthesis; (B). Sulphur metabolism; (C). Tryptophan metabolism; (D). Nitrogen metabolism.
1. Strain KCTC 32190^T; 2, *R. oryzicola* KCTC 22478^T; 3. *R. lacus* CGMCC 1.3617^T; 4. *R. alkaliterrae* DSM 25695^T; 5. *R. terrae* LMG 31159^T; 6. *R. soli* LMG 31231^T; *7. R. eburnea* LMG 31225^T

Table 23. Biosynthetic gene cluster detected by anti-SMASH of the genus *Roseomonas* members (Species list same as given in Table 20)

	A1	A2	A3	A4	A5	A6	A7	A8	A9	B1	B2	B3	B4	B5	B6	B7	B8	B9	C1	C2	C3	C4	C5	C6	C7	C8	C9	D1	D2	D3	D4	D5	D6	D7	D8
1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
6	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
7	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
8	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	2	0	0	0
9	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
10	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	1	0	0	0	1	0	0	0	0	0
12	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0
13	0	0	0	0	0	0	2	1	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0
14	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
16	0	0	0	0	0	0	3	4	1	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0
17	0	0	0	0	0	0	1	6	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	1	0	0	1	0	0	0	1
18	0	0	0	0	0	0	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
19	0	0	0	0	0	0	2	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0
20	0	0	0	0	0	0	3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0
23	0	0	0	0	0	0	1	0	2	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0
25	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
26	0	0	0	1	0	0	2	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0
27	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	1	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	2	1	0
31	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0
32	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
33	0	0	0	0	0	0	2	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2	1	0	0	1	0	0	1	0	0	0	0
34	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	3	0	0	0	1	1	0	0	1	0	0	0	0	1	1	0

A1, NRPS-Heterobactin; A2, NRPS-Paenibactin; A3, NRPS-Myxochelin; A4, NRPS- Griseoviridin/fijimycin; A5, NRPS-streptobactin; A6, NRPS qinichelins; A7, terpene; A8, NRPS; A9, NRPS-like; B1, RiPP-like; B2, NRPS/T1PKS-cystothiazole A; B3, arylpolene-enterobactin; B4, arylpolene-Xanthomonadin; B5, Arylpolyene- Glycinocin A; B6, Arylpolyene-Lipopolysaccharide; B7, Arylpolene-APE-Vf; B8, Arylpolene; B9, NAPAA; C1, Betalactone- Fengycin; C2, Betalactone; C3, Phosphonate/terpene- Phosphinothricintripeptide; C4, Thioamitides-Asukamycin; C5, cyanobactin; C6, TIPKS; C7, T3PKS; C8, T1PKS/NRPS; C9, phosphonate; D1, phenazine; D2, ranthipeptide; D3, Lassopeptide; D4, redox-cofactor; D5, hserlactone; D6, terpene-carotenoid; D7, terpene-Malleobactin; D8, Terpene-hydroxyastaxanthin

3.3.4.7. Physiological and chemotaxonomic characteristics

Optimum growth occurs at 35-37 °C, with a growth range of 15-40 °C. NaCl was not essential for growth but could tolerate up to 1 % (w/v). Strain KCTC 32190^T grows at a pH range of 5.5-8.0 with an optimum at 7.0. Strain KCTC 32190^T grew with D-glucose, D-fructose, D-galactose, L-rhamnose, L-asparagine, L-fucose, sucrose, yeast extract, peptone, pyruvate, D-mannitol, casamino acids, and L-serine. No growth was observed with fumarate, starch, ethanol, glycerol, succinate, benzoate, acetate, D-glutamic acid, L-ascorbic acid, glycine, L-tyrosine, tween 80, maleic acid, L-phenylalanine. The growth of both strains was independent of yeast extract or added vitamins. Strain KCTC 32190^T could utilize only ammonium chloride, whereas *R. oryzicola* KCTC 22478^T could use ammonium chloride and tryptophan as the nitrogen source. Both strains showed negative reactions for nitrate reduction, cellulose, starch, and gelatine hydrolysis. The differences are put up in Table 24. Strain KCTC 32190^T showed indole test negative (L-tryptophan), whereas indole was produced from L-tryptophan by *R. oryzicola* KCTC 22478^T. Strain KCTC 32190^T was sensitive to kanamycin and rifampicin but resistant to streptomycin, nalidixic acid, vancomycin, penicillin, and chloramphenicol.

Whole-cell fatty acid analysis of strain KCTC 32190^T revealed that $C_{18:1}\omega7c/C_{18:1}\omega6c$, *cyclo*-$C_{19:0}\omega8c$, $C_{18:0}$2-OH, and $C_{16:0}$ were the predominant cellular fatty acids and significant proportions of $C_{18:0}$3-OH, $C_{16:1}\omega7c/C_{16:1}\omega6c$, $C_{16:0}$2-OH, and $C_{16:1}\omega5c$ were also detected (Table 25). The differences in the fatty acid profile of strain KCTC 32190^T in comparison to *R. oryzicola* KCTC 22478^T are put up in Table 24. Polar lipids of the strain KCTC 32190^T include phosphatidylglycerol, diphosphotidylglycerol, phosphatidylethanolamine, phosphatidylcholine, an unidentified amino lipid (AL), and three unidentified lipids (L1-3; Fig 49A) whereas for *R. oryzicola* KCTC 22478^T are shown in Fig. 49B. Both the strains had ubiquinone-10 as primary quinone system.

Table 24. Differentiating characteristics of 1. Strain KCTC 32190^T; 2. *R. oryzicola* KCTC 22878^T; 3. *R. sediminicola* FW-3^T (He et al 2014). ; 4. *R. soli* 5N26^T (Kim et al. 2014)

Characteristics	***1**	***2**	**3**	**4**
Isolation source	Sand	Rhizosphere of rice	Fresh water	Cabbage rhizosphere
Colony color			Pale red	White
Cell Size [W x L; µm]	0.6-1 x 1.5-2.0	0.7-0.8 x 1.2-1.6	0.3-0.5 x 1.5-2.2	0.7-1.0 x 1.9-2.4
Cell shape	Coccoid to oval shaped	Coccobacilli	Rod	Rod
Temperature optima (range) ºC	35(15-40)	30 (10-40)	(10-37)	30 (15-40)
pH optimum (range)	7 (5.5-8)	7-8 (5-10)	7 (5.5-10)	6.5 (5.0-8.5)
NaCl tolerance (%)	1	0	1	0
Oxidase/Catalase	+/-	+/-	+/+	-/-
Hydrolysis of				
Casein	+	-	-	-
Urea	+	+	+	-
Indole from L-tryptophan	-	+	-	Na
Utilization of carbon substrate				
Arabinose	+	-	-	-
Glucose	+	-	-	-
D-Galactose	+	-	Na	Na
Fructose	+	-	Na	Na
D-Mannitol	+	-	-	-
Rhamnose	+	-	-	-
Ribose	+	-	+	
Sorbitol	+	-	-	-
Sucrose	+	w	-	-
Cellular FA				
$C_{16:0}$	12.7	12.5	8.9	7.5
$C_{16:0}$2OH	1.0	0.5	-	-
$C_{16:1}\omega5c$	0.5	0.5	0.9	2.4
$C_{16:1}\omega7c/C_{16:1}\omega6c$	1.4	1.3	-	4.1
$C_{18:1}\omega7c/C_{18:1}\omega6c$	50.6	66.7	55.4	63.2
$C_{18:0}$	-	1.1	-	-
$C_{18:1}$2OH	13.0	3.7	29.8	13.4
$C_{18:0}$3OH	1.5	0.7	-	-
$C_{19:0}$ cyclo ω8c	15.3	35.5	-	4.4

*All the tests were performed at author's laboratory under identical conditions.

For both strains common: All the strains are Gram-stain-negative, non-motile and non-spore forming; negative for gelatin, starch hydrolysis;; +, positive or present; -, negative or absent, w: weakly positive; Na, not available

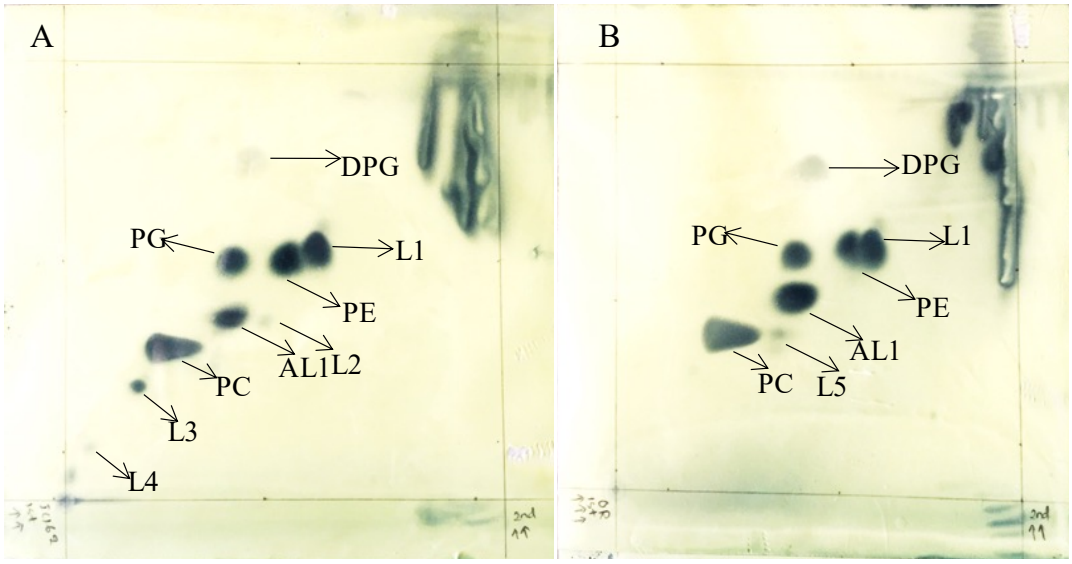


Fig. 49. Two-dimensional TLC of polar lipids. For polar lipids, chloroform-methanol-water (75:32:4 [v/v]) was used as the solvent in the first direction, and chloroform-acetic acid-methanol-water (86: 16: 15:4 [v/v]) was used in the second direction.
(A) strain KCTC 32190^T
(B) *R. oryzicola* KCTC 22478^T
DPG, diphosphatidylglycerol; PE, phosphatidylethanolamine; PC, phosphatidylcholine; PG, phosphatidylglycerol; unidentified lipids (L1-5), unidentified aminolipid (AL1)

For taxonomic demarcation, there should be a correlation between genetic and phenotypic features. The non-motile character of Groups II and VI members distinguish themselves from Group V members. Variable catalase activity distinguished those of Groups II and III from members of Groups I, IV, V, VI, and VII. Groups I and II members have varying oxidase activity, while those in the other groups are either oxidase negative (Groups IV, V) or positive (Groups III, VI, VII). Members of Groups VI and VII have a higher NaCl tolerance than the other groups. Groups II and III members were variable for starch hydrolysis, while Groups IV, V, VI, and VII members were negative. Group VI members tested positive for casein hydrolysis in varying degrees, while the remainder of the groups tested negative. Urease positive for Groups I, IV, and VI members, but Groups II, III, and VII had varying reactions. Likewise, Groups II, III, V, and VII members had a negative reaction to nitrate reduction, whereas the others had a varied reaction. Similarly, Groups II and III differ from the other groups by exhibiting varied gelatin hydrolysis activity. The favourable utilisation of D-glucose distinguishes Group VII from Groups II, III, and IV. Members of Groups V, VI, and VII were able to use L-arabinose for growth, whereas members of Groups II, III, and IV were unable to do so. Sucrose could be used by members of Groups III, VI, and VII, but not by members of Groups IV and V. Group VI members had Q-9 as their unique respiratory quinone, as opposed to individuals in the other groups who had Q-10. Groups III and VII members are distinguished from Groups I, II, IV, V, and VI by the presence of a glycolipid. Groups V and VII are also characterised by the presence of an unidentified lipid as one of the polar lipids. Polar lipid analysis is a useful chemotaxonomic tool frequently employed as a diagnostic tool for reclassification. The phenotypic traits that distinguish the genus *Roseomonas* from other closely related genera in the *Acetobacteraceae* family are listed in Table 2

Table 25. Characteristics differentiating the *Roseomonas* groups from other closely related genera in the family *Acetobacteraceae*

Characteristics	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	14
Motility	v	-	v	v	+	-	v	-	ND	+	-	-	-	-
Oxidase/catalase	v/+	v/v	+/v	-/+	-/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
Optimum temp. °C(range)	35(12-42)	25-30 (5-45)	30-37 (4-45)	28-35 (15-40)	4-45 (15-30)	40-50 (5-55)	28-35 (4-45)	30 (1-30)	25 (4-37)	50 (25-50)	28 (20-40)	30-32 (20-42)	30 (20-37)	40-50 (20-60)
Optimum pH (range)	6-7 (5-8)	6.5-9 (6-10)	7-8 (5-10)	7 (5-9)	4-11(6-8)	7-10 (5.5-11)	7-7.5 (5-8.5)	7 (8-9)	7 (5-8)	7.5 (7.5-8)	7 (6-8)	ND (6.6-6.8)	7 (5.5-8.5)	8 (6-10)
NaCl (%) toleration	0-3	0-1.02	0-2	0-2	0-2	0-6.5	0-6	0	0-3	1-2	0-2	0-4	0-0.5	0-2.5
Hydrolysis of:														
Starch	ND	v	v	-	-	-	-	ND	-	-	-	+	-	-
Casein	ND	-	-	-	-	v	-	ND	-	-	-	-	-	+
Gelatin	-	v	v	-	-	-	-	-	-	+	-	-	+	+
Urease	+	v	v	+	+	+	v	+	v	+	-	+	-	+
Tween 80	ND	v	-	-	v	-	-	+	ND	-	+	-	-	-
Nitrate reduction/ H_2S production	v/v	-/v	-/v	v/-	-/-	v/ND	-/-	-/-	+/-	ND	ND	+/+	+/ND	+/-
Utilization of:														
D-Glucose	v	-	-	-	v	v	+	+	-	+	-	+	+	+
D-Fructose	v	-	v	-	ND	v	+	ND	-	-	-	+	ND	-
D-Galactose	v	-	-	-	ND	v	v	ND	-	-	-	+	ND	+
D-Lactose	v	-	-	-	ND	-	v	-	+	ND	-	+	ND	+
Sucrose	v	v	+	-	-	+	+	-	-	+	ND	-	ND	+
D-Mannitol	v	-	-	-	-	v	v	ND	-	ND	-	-	-	-
L-Arabinose	v	-	-	v	+	+	+	+	-	ND	+	+	-	-
Isoprenoid quinones	Q10	Q10	Q10	Q10	Q9	Q10	Q10	ND	Q9	Q9	Q10	Q10	Q10	Q10
Polar lipid	DPG, PE, PC, AL, PL	DPG, PE, PC, AL, PL	DPG, PE, PC, AL, GL,PL	DPG,PC, PE	DPG, PE, PC, AL, L	DPG, PE, Al, L,PL	DPG, PC, PE, GL, PL, AL, L	DPG, PE, PC, AL	DPG, PC	PC, PE, DPG, AL	PE, PL, AL	ND	PE, PL, AL	DPG, PE, PC
Major FA	$C_{16:0}$, $C_{18:1}$ 2OH, $C_{19:0}$*cyclo*ω7*c*	Summed feature 3 $C_{18:1}$ 2-OH, $C_{18:1}$ 3-OH	Summed feature 3 $C_{16:1}$ *ω*5c $C_{16:0}$	Summed feature 3 $C_{16:0}$ $C_{18:1}$ *ω*7c , $C_{18:1}$ 2-OH	Summed feature 3 C18:1 □7c C16:0	Summed feature 3, $C_{16:0}$ $C_{18:1}$ 2 OH,	Summed feature 3 $C_{16:0}$ $C_{18:1}$ 2 OH	Summed feature 3	$C_{16:0}$	$C_{16:0}$	$C_{18:1}$2OH	$C_{18:1}$ $C_{18:1}$2-OH	$C_{18:1}$2OH	$C_{16:0}$, summed feature 4

All genera are Gram-stain-negative; contains PG as Polar lipid; contains $C_{16:1}\omega7c/C_{16:1}\omega6c$ as major fatty acid; *summed features represent two or three fatty acids that cannot be separated by gas chromatography with the MIDI system. Summed feature 3, $C_{16:1}\omega7c$ and/or iso-$C_{15:0}$ 2-OH; summed feature 4, *anteiso*-$C_{17:1}$ B and/or *iso*-$C_{17:1}$ I.
DPG, diphosphatidylglycerol; PG, phosphatidylglycerol, PE, phosphatidylethanolamine; PC, phosphatidylcholine; GL, unidentified glycolipid; PL, unidentified phospholipid; AL, unidentified amino lipids, L, unidentified Lipid; FA; Fatty acid; *, data obtained from author's laboratory; ND, Not determined; v, variable; +, positive, -, negative; v/ND, variable or not determined; -/v, absent/variable; +/+, positive for both tests; v/+, variable/positive; v/v, both tests show variable results; +/+, both testes positive.

1. Group I (*Roseomonas* genus *sensu-stricto*) (data based on *R. gilardii* 5424^T, *R. gilardii* subsp. *gilardii* ATCC49956^T, *R. gilardii* subsp. *rosea* MDA5605^T, *R. mucosa* MDA5527^T) (Rihs et al. 1993; Han et al. 2003); 2.Group II (data based on *R. pecuniae* N75^T, *R. aerilata* 5420S-30^T, *R. nepalensis* G-3-5^T, *R. rosea* DSM 14916^T) (Sánchez-Porro et al. 2009; Yoo et al. 2008; Chaudhary and Kim 2017; Lopes et al. 2011); 3. Group III (*R. stagni* HS-69^T, *R. frigidaquae* CW67^T, *R. bella* CQN31^T, *R. algicola* PeD5^T, *R. selenitidurans* BU-1^T) (Furuhata et al. 2008; Kim et al. 2009; Zhang et al. 2020; Hou et al. 2020; Kim et al. 2020); 4. Group IV (data based on *R. aquatica* TR5^T, *R. fluminis* D3^T) (Gallego et al. 2006); 5. Group V (based on *R. arctica* M6-79^T, *R. hellenica* R-73070^T) (Qiu et al. 2016; Rat et al. 2021); 6. Group VI (based on *R. oryzicola* KCTC 32190^T, *Strain* KC32190^T; *R. lacus* JCM 13283^T, *R. alkaliterrae* YIM 78007^T) (Jiang et al. 2006; Dong et al. 2014); 7. Group VII (data based on *R. deserti* M3^T, *R. cervicalis* KACC 11686^T, *R. oryzae* JC288^T, *R. rhizosphaerae* KACC 17225^T, *R. aestuarii* JC17^T) (Subhash and Lee 2018; Rihs et al. 1993, Ramaprasad et al. 2015, Chen et al. 2014, Venkat et al. 2010); 8. *Humitalea* (data based on *H. rosea* W37^T) (Margesin and Zhang 2013); 9. *Belnapia* (based on *B. moabensis* CP2C^T, *B. rosea* CGMCC1.10758^T, *B. soli* PB-K8^T)(Jin et al. 2013; Jin et al. 2012; Reddy et al. 2006);10. *Rubritepida* (based on *R. flocculans* DSM 14296^T) (Alarico et al. 2002); 11. *Dankookia* (based on *D. rubra* WS-10^T) (Kim et al. 2016b); 12. *Paracraurococcus* (data based on *P. ruber* NS89^T) (Saitoh et al. 1998; Alarico et al. 2002); 13. *Roseicella* (based on *R. frigidaeris* DB1506^T) (Khan et al. 2019); 14. *Crenalkalicoccus* (based on *C. roseus* YIM 78023^T) (Ming et al. 2016)

3.3.5. Phylogenomic study of *Alcanivorax xenomutans* JC109^T

3.3.5.1. Phylogeny based on 16S rRNA and genome sequence

Alcanivorax xenomutans JC109^T is an important species isolated from a shrimp cultivation pond described from our lab. Most abundantly found in marine habitats. Based on the 16S rRNA gene phylogenetic tree of *Alcanivorax xenomutans* JC109^T along with the other members of the genus *Alcanivorax*, the results revealed that the members were highly polyphyletic (Fig. 50). It was divided into three different clades with high bootstrap values (Fig. 50). Clade I (n=6) consisted of *A. borkumensis* SK2^T, *A. hongdengensis* A-11-3^T, *A. sediminis* PA15-N-34^T, *A. profundi* MTEO17^T, *A. nanhaiticus* 19-m-6^T and *A. jadensis* T9^T. In contrast, clade II (n=2) members were *A. pacificus* W11-5^T and *A. indicus* SW127^T. The other species of the genus *Alcanivorax* like *A. venustensis* ISO4^T, *A. gelatiniphagus* MEBiC08158^T, *A. profundimaris* ST75FaO-1^T, *A. mobilis* MT13131^T, *A. balearicus* MACL04^T, *A. xenomutans* JC109^T, *A. dieselolei* CGMCC 1.3690^T and *A. marinus* R-8-12^T fell under clade III (n=8). As *A. borkumensis* SK2^T is the type species of the genus *Alcanivorax,* therefore clade I shall be marked as *Alcanivorax* genus *sensu-stricto.* The 16S rRNA gene pairwise nucleotide identity among the members showed that some species between different clades had 16S rRNA gene identity values of less than 94 % (Table 26). Additionally, the taxonomic standing of the genus *Alcanivorax* was further supported by the genome phylogeny (Fig. 51). The UBCG-based phylogenomic tree also suggested the same cladding pattern (three clades) as the 16S rRNA gene dendrogram with high gene support index (GSI). Thus, concurring with the proposal of reclassifying the genus *Alcanivorax*.

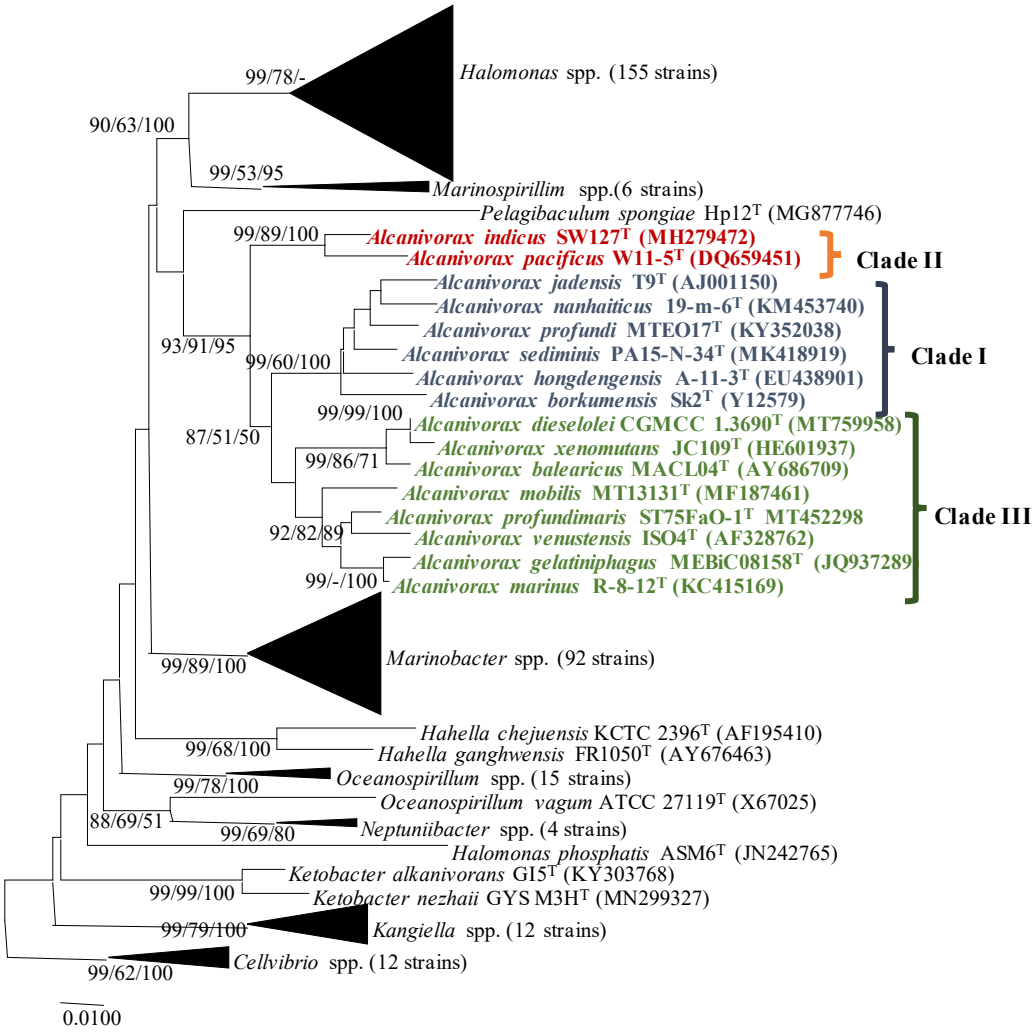


Fig. 50. NJ phylogenetic tree based on the 16S rRNA gene sequences showing the phylogenetic relationship between the members of the genus *Alcanivorax* and other closely related members. Members of genus *Cellvibrio* were used as an outgroup. Numbers at nodes represent bootstrap values (given as percentages of 1,000 replications) of >50 % shown at branch points (NJ/ME/ML). The GenBank accession numbers for the 16S rRNA gene sequences are shown in parentheses. Bar, 0.01 accumulated changes per nucleotide substitutions.

Table 26. 16S rRNA gene identity values between the members of the genus *Alcanivorax*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
1	100																		
2	97.3	100																	
3	97.7	97.9	100																
4	96	96.6	96.4	100															
5	97	97.5	96.9	97.9	100														
6	97	97.2	96.2	96.8	96.9	100													
7	93.9	93.5	93.8	92.9	92.7	93.1	100												
8	93.4	94.4	94.3	92.8	92.8	92.9	96.8	100											
9	94.6	95.4	94.4	94.2	94.5	94.7	93.9	93.6	100										
10	94.9	95.0	94.2	93.4	94.2	94.3	93.6	93.3	99.6	100									
11	93.9	94.2	93.4	93.5	93.7	94.4	94.3	92.9	96.2	96	100								
12	94.1	93.4	94.1	93.3	93.7	94	93.6	93.5	95.4	95	97.7	100							
13	93.6	93.4	92.9	94.6	94.7	94.4	93.8	92.9	95.4	95	98.1	96.9	100						
14	93.9	94.2	94.4	93.2	92.9	93.2	94.5	94.6	95.7	95.7	96.4	96.7	96.1	100					
15	94.5	94.9	93.9	93.7	94.1	94.2	93.4	93.2	99.5	99.1	95.7	94.5	94.7	95.2	100				
16	94.2	93.9	94.0	93.6	93.8	94.1	93.8	93.5	94.9	94.8	97.9	99.2	97.2	96.9	94.5	100			
17	91.6	91.6	91.2	90.7	90.6	91.4	90.2	90.5	91.4	90.7	90.2	90.9	89.6	90.8	90.9	91.6	100		
18	91	91.1	91.2	90.0	91.2	91	89.8	89.8	90.8	90.2	90.1	90.5	90.3	90.2	90.3	91.1	97.7	100	
19	89.6	88.9	88.6	90.2	89.2	89.4	89.1	88.3	90.1	89.7	90.4	89.3	89.8	89.5	89.7	89.6	86.8	86.7	100

1. *A. borkumensis* SK2^T, 2. *A. hongdengensis* A-11-3^T, 3. *A. jadensis* T9^T, 4. *A. nanhaiticus* 19-m-6^T, 5. *A. sediminis* PA15-N-34^T, 6. *A. profundi* MTEO17^T , 7. *A. pacificus* W11-5^T, 8. *A. indicus* SW127^T), 9..*A. venustensis* ISO4^T, 10. *A. dieselolei* B-5^T, 11. *A. xenomutans* JC109^T, 12. *A. balearicus* MACL04^T, 13. *A. mobilis* MT13131^T, 14. *A. profundamaris* ST75FaO-1^T, 15. *A. gelatiniphagus* MEBiC08158^T, 16. *A. marinus* R8-12^T), 17. *Ketobacter nezhaii* GYS_M3H^T, 18. *K. alkanivorans* GI5^T, 19. *Pelagibaculum spongiae* Hp12^T

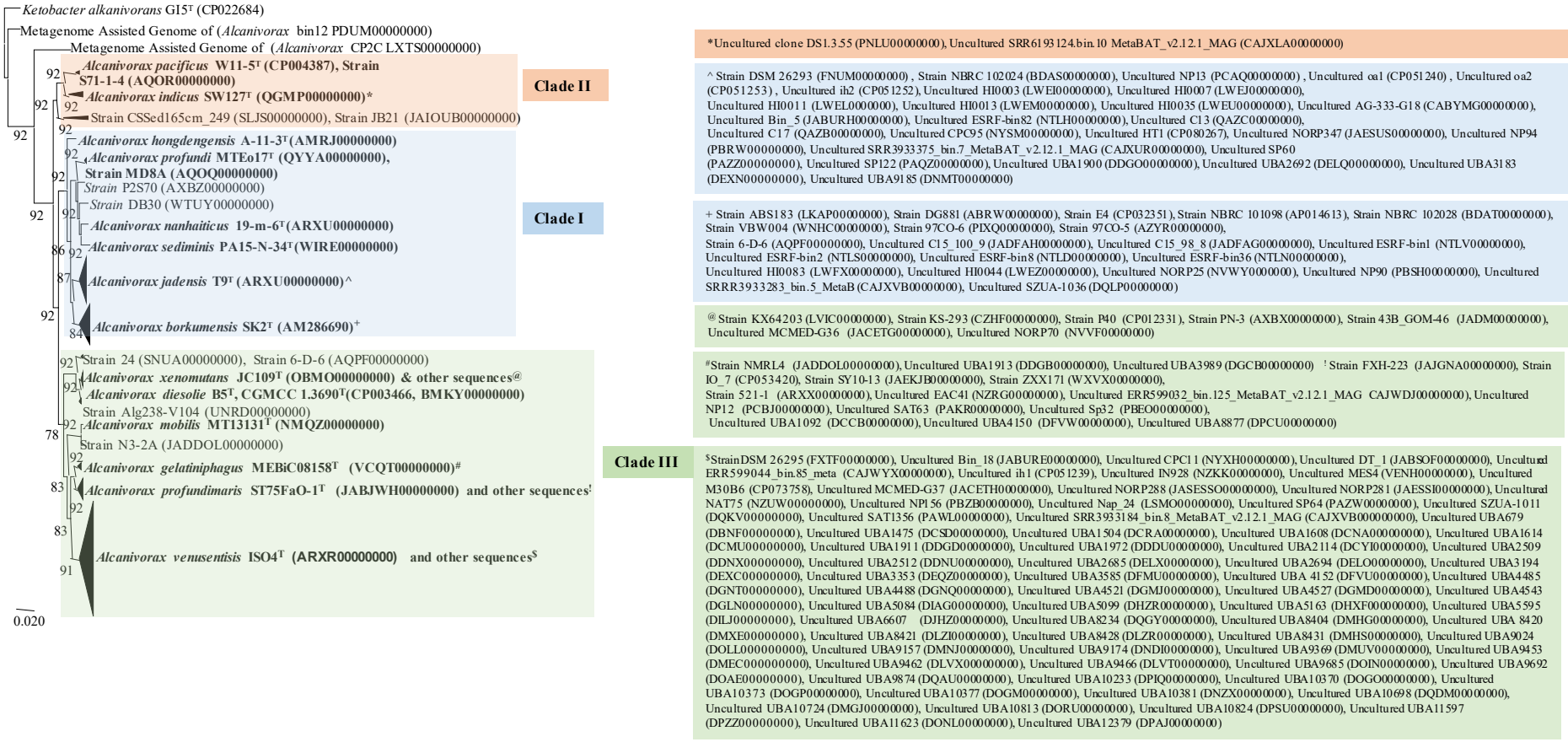


Fig. 51. Phylogenomic tree of members of the genus *Alcanivorax* constructed using 92 core genes tool based on the Up-to-date Bacterial Core Gene (UBCG) (Na et al. 2018). The tree was generated using the MEGAX software (NJ) with *Ketobacter alkanivorans* GI5^T (CP022684) as an outgroup.

3.3.5.2. Genomic features of *Alcanivorax* spp.

The general details of the genomes of the type strains applied in the study are presented in Table 27. The genome size of clade I (n=6) is 3.1-4.1 Mbp, clade II (n=2) is 3.4-4.1 Mbp, and clade III (n=6) is 3.5-5.1 Mbp. The genomic G+C content (mol %) of clade I (n=6), clade II (n=2) and clade III (n=6) were 54.7-60.7 %, 62.6-62.8 %, and 61.3-66.7 %, respectively. On average, clade III members had the highest genome size and G+C content (mol %) of 4.2 Mbp, and 63.8 %, respectively. The clade II and clade III members have CRISPR/Cas, whereas it was present only in the three members of clade I (*A. hongendensis, A. sediminis, A. jadensis*). Additionally, the information related to RNA, pseudogenes, and coding sequences is given in Table 27.

3.3.5.3. Genome based metrics for taxa delineation

Standard genome indices like *d*DDH and ANI were calculated for species delineation, whereas AAI and POCP were calculated for genus delineation. *d*DDH, ANI, AAI, and POCP were calculated at the inter and intra clade level. Within the clade members, as defined by phylogenomics, members had *d*DDH and ANI values of at least 20.1-51.9 % and 75.2-93.2 %, respectively (Table 28). At the inter clade level, the AAI values between clade I and clade II were 61.8-63.2 %; clade I and III were 66.1-69 % and clade II and clade III were 61.8-63.1 % (Table 29). Similarly, POCP values between clade I and II ranged from 60.9-65.5 %; clade I and III ranged between 58.1-68.3 %; whereas clade II and clade III were 53.8-65.8 % (Table 29). The heatmaps of AAI and POCP also suggest the cladding of the genus *Alcanivorax* into three major groups (Fig. 52A, B). For members of other genera from the family *Alcanivoraceae,* the *d*DDH, ANI, AAI, and POCP values are provided in Table 28 and Table 29.

Table 27. Genome characteristics of the members of the genus *Alcanivorax* and closely related strains in the family *Alcanivoraceae*

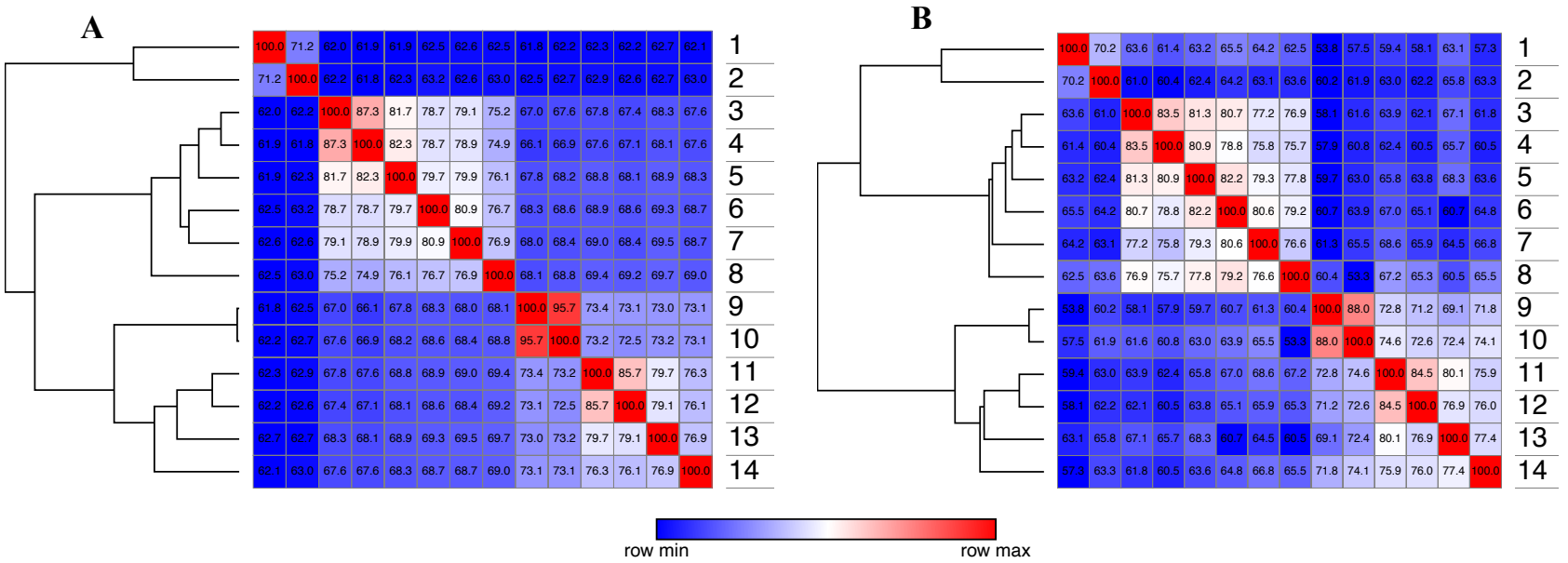
Organisms	Strain	Size (Mb)	G+C (mol%)	Coding sequence	N50 (bp)	L50 (bp)	RNA		Pseudogenes	CRISPR/Cas	DDBJ/EMBL /GenBank
							tRNA	rRNA			
Clade I (n=6)											
A. borkumensis	SK2^T	3.1	54.7	2755	3120143	1	42	6	0	0	AM286690
A. hongendensis	A-11-3^T	3.66	60.7	3416	107364	11	40	4	22	3	AMRJ00000000
A. profundi	MTEO17 ^{T}	3.74	57.5	3459	998536	2	43	3	0	0	QYYA00000000
A. nanhaiticus	19-m-6^T	4.13	56.4	3778	276801	6	42	4	30	0	ARXV00000000
A. sediminis	PA15-N-34^T	3.8	57.2	3390	3189188	1	41	2	18	1	WIRE00000000
A. jadensis	T9^T	3.63	58.4	3266	244829	5	42	3	33	2	ARXU000000000
Clade II (n=2)											
A. pacificus	W11-5^T	4.17	62.6	3759	4168427	1	42	3	72	1	CP004387
A. indicus	SW127^T	3.45	62.8	3095	2638783	1	43	3	0	1	QGMP00000000
Clade III (n=6)											
A xenomutans	JC109^T	4.35	61.7	3918	180371	10	40	6	143	8	OBMO00000000
A. dieselolei	B5^T	5.15	61.3	4417	290175	6	42	6	34	1	UNRD00000000
A. mobilis	MT13131^T	4.1	63.4	3690	109947	12	41	2	0	1	NMQZ00000000
A. gelatiniphagus	MEBiC08158 ^{T}	4.22	65.2	3918	195464	8	43	3	0	7	VCQT00000000
A. profundimaris	ST75FaO-1^T	4.04	66.3	3708	280257	6	42	2	51	3	JABJWH00000000
A. venustensis	ISO4^T	3.54	64.7	3308	94767	11	39	4	0	1	ARXR00000000
Others											
Ketobacter alcanivorans	GI5^T	4.9	51.2	4199	4914503	1	44	6	35	1	CP022684
K. nezhaii	MCCT1A13808^T	5.3	50	4555	433309	4	41	4	69	2	VRKW00000000
Pelagibaculum spongiae	Hp12^T	4.9	43	4306	426703	4	85	12	64	7	QDDL00000000

Table 28. *d*DDH/ANI values between the members of the genus *Alcanivorax*

		CLADE I						CLADE II				CLADE III						
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	15	16	17
1	*Alcanivorax. sediminis* PA15-N-34^T	100	81.8	78.5	75.7	77.7	76.6	71.5	71.6	73.4	73.2	73.9	73.9	73.6	73.1	68	66.7	66
2	*Alcanivorax. nanhaiticus* 19-m-6^T	25.2	100	79	75.5	77.4	76.7	71.3	71.6	73.2	73.4	73.8	73.9	73.9	73.2	68	67.5	65.7
3	*Alcanivorax. profundi* MTEO17^T	23.2	23.8	100	75.9	77.8	76.5	71.2	71.6	73.2	73.3	74.3	74	73.9	73.4	67.9	67	66.1
4	*Alcanivorax. borkurnensis* SK2^T	21.8	21.2	21.4	100	77.2	75.2	70.5	70.6	72.3	72.2	72.5	72.6	72.6	72.2	67.7	67.1	65.7
5	*Alcanivorax. jadensis* T9^T	22.3	22	22.3	22.9	100	78.2	71.2	71.9	74.3	75.2	74.9	74.9	75.2	74.2	68.4	67.5	65.9
6	*Alcanivorax. hongdengensis* A-11-3^T	22.2	22.6	21.8	21	23.2	100	72.4	72.7	74.7	74.7	76.1	76.5	75.9	74.9	68	67.3	65.7
7	*Alcanivorax. indicus* SW127^T	18.8	19.3	19.2	18.6	19.5	19.1	100	75.8	72.3	72.1	73	73	72.9	72.5	68	66.9	65.3
8	*Alcanivorax. pacificus* W11-5^T	18.7	19.5	19.4	20.1	19.1	19.5	20.1	100	72.9	72.7	73.3	73.5	73.2	72.9	67.9	67.3	65.8
9	*Alcanivorax. dieselolei* B-5^T	20.7	20.5	20.3	21.4	20.6	20.9	18.9	19.4	100	93.2	77.9	77.6	77.6	76.7	68.1	67.2	65.6
10	*Alcanivorax. xenomutans* JC109^T	20.4	20.7	20.1	20	20.3	20.6	18.9	19.4	51.9	100	77.3	77.1	77	76.7	68.1	67.2	65.6
11	*Alcanivorax. profundimaris* ST75FaO-1^T	20.4	20.5	20.2	19.8	20.6	20.7	18.8	19.2	21.9	21.4	100	86.1	82.7	79.4	68.2	67.4	65.2
12	*Alcanivorax. gelatiniphagus* MEBiC08158^T	20.4	20.8	20.6	20.7	21	21.1	19	19.2	22.1	21.2	30.3	100	82.1	79.4	68.2	67.2	65.2
13	*Alcanivorax. venustensis* ISO4^T	20	20.4	19.9	19.7	20.9	20.9	18.6	19.4	22	21.3	25.6	25.6	100	79.5	68.3	67.2	63.4
14	*Alcanivorax. mobilis* MT13131^T	20.2	20.5	20.3	20.7	21.1	20.6	18.9	19.5	21.9	20.6	23.2	26.1	23.3	100	68.1	67.1	65.3
15	*Ketobacter alkanivorans* GI5T	18.5	18.8	20.1	24.9	20	20.3	21.4	20.2	21.2	21.9	21.2	19.4	20.7	20.8	100	71.4	66.1
16	*Ketobacter nezhaii* GYS_M3HT	28	22.3	34.3	37.5	27	27.7	29.4	22.9	22.8	21.1	28.1	24.6	20.7	27.7	19.7	100	65.5
17	*Pelagibaculum spongiae* Hp12T	21.9	20.6	21.3	33.3	33.3	36.2	35	35	34.3	37.2	33.3	21.7	32.9	22.6	23.9	20.1	100

Table 29. AAI and POCP values between the members of the genus *Alcanivorax*

		CLADE I						CLADE II				CLADE III						
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	17	18	19
1	*Alcanivorax. sediminis* PA15-N-34^T	100	83.5	81.3	77.2	80.7	76.9	63.6	60.9	58.1	61.6	63.9	62.1	67.1	61.8	43.	40	30
2	*Alcanivorax. nanhaiticus* 19-m-6^T	87.3	100	80.9	75.8	78.8	75.7	61.4	60.4	57.9	60.8	62.4	60.5	65.7	60.5	44.8	42	32.1
3	*Alcanivorax. profundi* MTEO17^T	81.7	82.3	100	79.3	82.2	77.8	63.2	62.4	59.7	62.9	65.8	63.8	68.3	63.6	44.3	43	31.4
4	*Alcanivorax. borkurnensis* SK2^T	78.7	78.7	79.7	100	80.6	76.6	64.2	63.1	61.3	65.5	68.6	65.9	64.5	66.8	43.3	40	29.8
5	*Alcanivorax. jadensis* T9^T	79.1	78.9	79.9	80.9	100	79.2	65.5	64.2	60.7	63.9	67	65.1	60.7	64.8	43.5	45	29.9
6	*Alcanivorax. hongdengensis* A-11-3^T	75.2	74.9	76.1	76.7	76.9	100	62.5	63.6	60.4	53.3	67.2	65.3	60.5	65.5	42.1	42	29.7
7	*Alcanivorax. indicus* SW127^T	62.0	61.9	61.9	62.5	62.6	62.5	100	70.2	53.8	57.5	59.4	58.1	63.1	57.3	44.1	44	31.1
8	*Alcanivorax. pacificus* W11-5^T	62.2	61.8	62.3	63.2	62.6	63.0	71.2	100	60.2	61.9	63	62.2	65.8	63.4	44.2	43	30.2
9	*Alcanivorax. dieselolei* B-5^T	67.0	66.1	67.8	68.3	68	68.1	61.8	62.5	100	88	72.8	71.2	69.1	71.9	37.9	34	27.7
10	*Alcanivorax. xenomutans* JC109^T	67.6	66.9	68.2	68.6	68.4	68.8	62.2	62.7	95.7	100	74.6	72.6	72.4	74.1	39.4	39	29
11	*Alcanivorax. profundimaris* ST75FaO-1^T	67.8	67.6	68.8	68.9	69	69.4	62.3	62.9	73.4	73.2	100	84.5	80.5	75.9	41.2	40	29.9
12	*Alcanivorax. gelatiniphagus* MEBiC08158^T	67.4	67.1	68.1	68.6	68.4	69.2	62.2	62.6	73.1	72.5	85.7	100	76.9	76	39.7	37	28.7
13	*Alcanivorax. venustensis* ISO4^T	68.3	68.1	68.9	69.3	69.5	69.7	62.7	62.7	73	73.2	79.7	79.1	100	77.4	42.5	32.2	30.1
14	*Alcanivorax. mobilis* MT13131^T	67.6	67.6	68.3	68.7	68.7	69	62.1	63.0	73.1	73.1	76.3	76.1	76.9	100	40.1	40.2	28.9
15	*Ketobacter alkanivorans* GI5T	52.1	52.2	52.3	52.7	52.1	52.4	52.1	51.5	51.2	50.7	51.2	51	51.8	51.4	100	41.1	30.9
16	*Ketobacter nezhaii* GYS_M3HT	41.6	41.7	42.4	43.0	41.6	41.4	40.2	40.7	38.2	39.4	40.5	39.8	41.8	39.9	66.4	100	32.1
17	*Pelagibaculum spongiae* Hp12T	51.2	46.9	47.2	47.5	46.9	46.6	46.9	46.5	45.8	46.5	46.5	50.1	46.9	46.3	47.3	47	100



1. *A. indicus* SW127^T, 2. *A. pacificus* W11-5^T, 3. *A. sediminis* PA15-N-34^T, 4. *A. nanhaiticus* 19-m-6^T, 5. *A. profundi* MTEO17^T, 6. *A. jadensis* T9^T, 7. *A. borkumensis* SK2^T, 8. *A. hongdengensis* A-11-3^T, 9. *A. dieselolei* B-5^T, 10. *A. xenomutans* JC109^T,11. *A. profundamaris* ST75FaO-1^T, 12. *A. gelatiniphagus* MeBiC08158^T, 13. *A. venustensis* ISO4^T, 14. *A. mobilis* MT13131^T

Fig. 52. Heatmap representing genomic indexes values for delineation values of the members of genus *Alcanivorax* clustered with the complete h-clust method. (A), AAI; (B), POCP

3.3.5.4. Core and pan-genome analysis

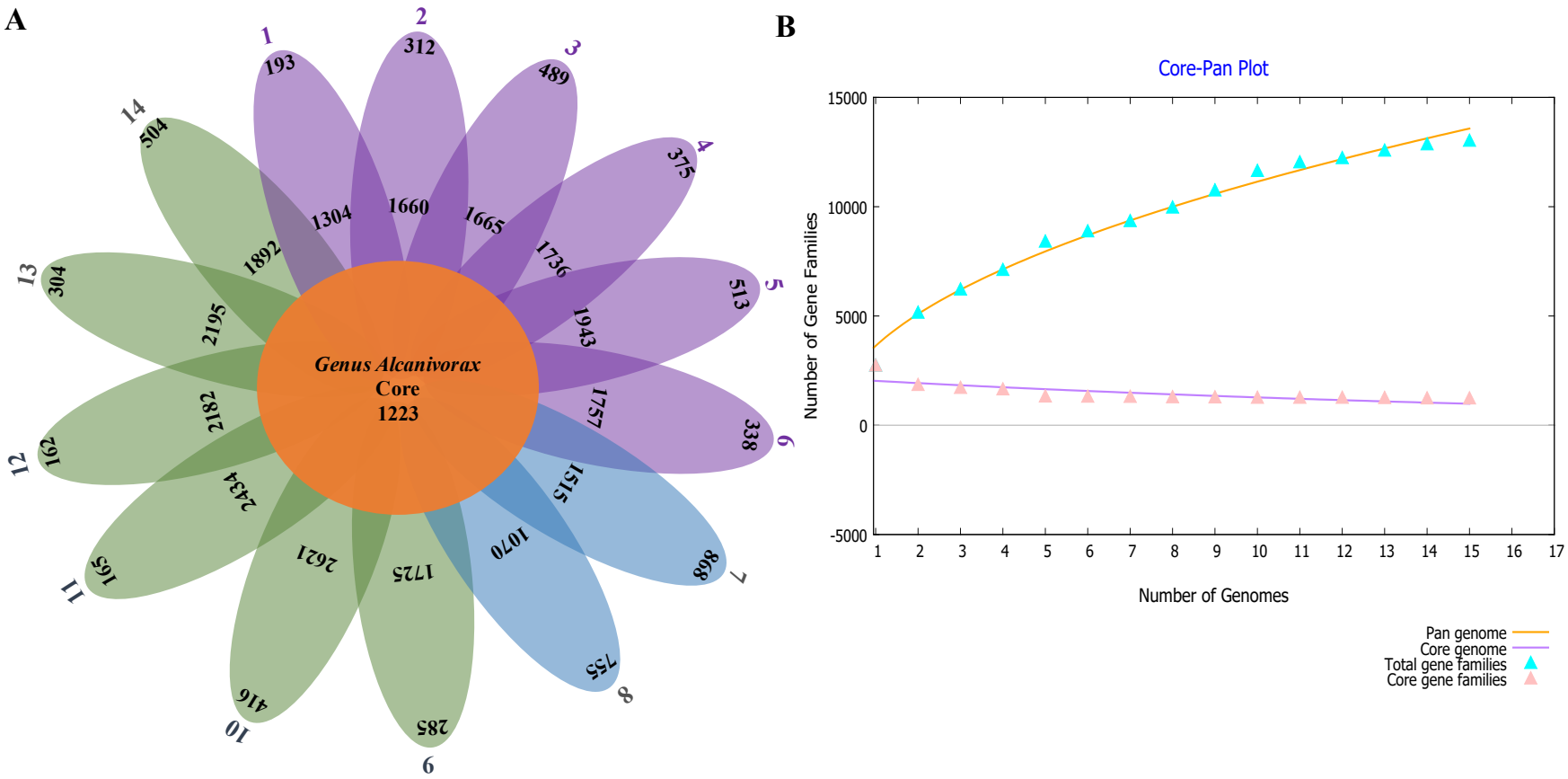
The BPGA software was applied to analyse of the core and pan-genome of the members of the genus *Alcanivorax.* For the members, the number of core genes, accessory genes, and unique genes are presented in Table 30 and Fig. 53. *Alcanivorax* species have 1223 core genes (36.5 %), 24100 accessory genes (51.4 %) and 5679 unique genes (12 %) (Fig. 53A). The core and pan-genome data have been illustrated in Fig. 53. The members of the genus *Alcanivorax* showed an open pan-genome, as was evident from the core and pan-genome plot (Fig. 53B). *Alcanivorax* members (n=14) had core genes ranging between 1994 to 2117 (Table 30). The analysis also showed that clade II (n=2) members did not possess accessory genes, whereas they had the highest average number of unique genes (1208) (Table 30). Similarly, for clade I and III members, core, unique and accessory genes distribution is given in Table 30.

3.3.5.5. Proteome comparison and conserved signature indels

For the proteome comparison wherein *A. borkumensis* SK2^{T} (type species of the genus *Alcanivorax*) was taken as a reference; clade I (n=6), clade II (n=2), clade III (n=6) members showed 70-80 % (Fig. 54A), <70 % (Fig. 54B) and <65 % (Fig. 54C) protein sequence identity, respectively. CSI was analyzed by researching the proteins of *A. borkumensis* SK2^{T}. The result showed that for clade I (n=6), there was one amino acid insertion in Leucyl/phenylanyl-tRNA-protein transferase (EC 2.3.2.6) (Fig. 55A) and one amino acid insertion in DEAD-box ATP-dependent RNA helicase (EC 3.6.4.13) of clade II members (Fig. 55B). Similarly, there was one amino acid deletion for clade III for beta-*N*-acetylglucosaminidase (EC 3.2.1.53) (Fig. 55C).

Table 30. Pan and core genomes of the members of the genus *Alcanivorax*

Organism name	No. of core genes	No. of accessory genes	No. of unique genes
A.borkumensis SK2^T	1223	1304	193
A.jadensis T9^T	1223	1660	312
A.hongdengensis A-11-3^T	1223	1665	489
A.profundi MTEO17 ^{T}	1223	1736	375
A.nanhaiticus 19-m-6^T	1223	1943	513
A.sediminis PA15-N-34^T	1223	1757	338
A.pacificus W11-5^T	1223	1515	868
A.indicus SW127^T	1223	1070	755
A.venustensis ISO4^T	1223	1725	285
A.dieselolie B5^T	1223	2621	416
A.xenomutans JC109^T	1223	2434	165
A.profundimaris ST75FaO-1^T	1223	2182	162
A.gelatiniphagus MEBiC08158^T	1223	2195	304
A.mobilis MT13131^T	1223	1892	504
Clade I			
A.borkumensis SK2^T	1994	502	223
A.jadensis T9^T	1994	819	379
A.hongdengensis A-11-3^T	1994	695	684
A.profundi MTEO17 ^{T}	1994	929	408
A.nanhaiticus 19-m-6^T	1994	1116	564
A.sediminis PA15-N-34^T	1994	946	374
Clade II			
A.indicus SW127^T	2117	0	933
A.pacificus W11-5^T	2117	0	1483
Clade III			
A.dieselolie B5^T	2004	1733	513
A.gelatiniphagus MEBiC08158^T	2004	1288	427
A.mobilis MT13131^T	2004	1029	577
A.profundimaris ST75FaO-1^T	2004	1296	266
A.venustensis ISO4^T	2004	864	360
A.xenomutans JC109^T	2004	1611	201



1. *A. borkumensis* SK2^{T}, 2. *A. jadensis* T9^{T}, 3. *A. hongdengensis* A-11-3^{T}, 4. *A. profundi* MTEO17^{T}, 5. *A. nanhaiticus* 19-m-6^{T}, 6. *A. sediminis* PA15-N-34^{T}, 7. *A. indicus* SW127^{T}, 8. *A. pacificus* W11-5^{T}, 9. *A. dieselolei* B-5^{T}, 10. *A. gelatiniphagus* MeBiC08158^{T}, 11. *A. mobilis* MT13131^{T}, 12. *A. profundamaris* ST75FaO-1^{T}, 13. *A. venustensis* ISO4^{T}, 14. *A. xenomutans* JC109^{T}

Fig. 53. Core and pan-genome plot of the members of genus *Alcanivorax.*
(A) Flower pot illustrating the core genes, unique genes of type strains of genus *Alcanivorax*
(B) Core-pan plot of the members of genus *Alcanivorax*

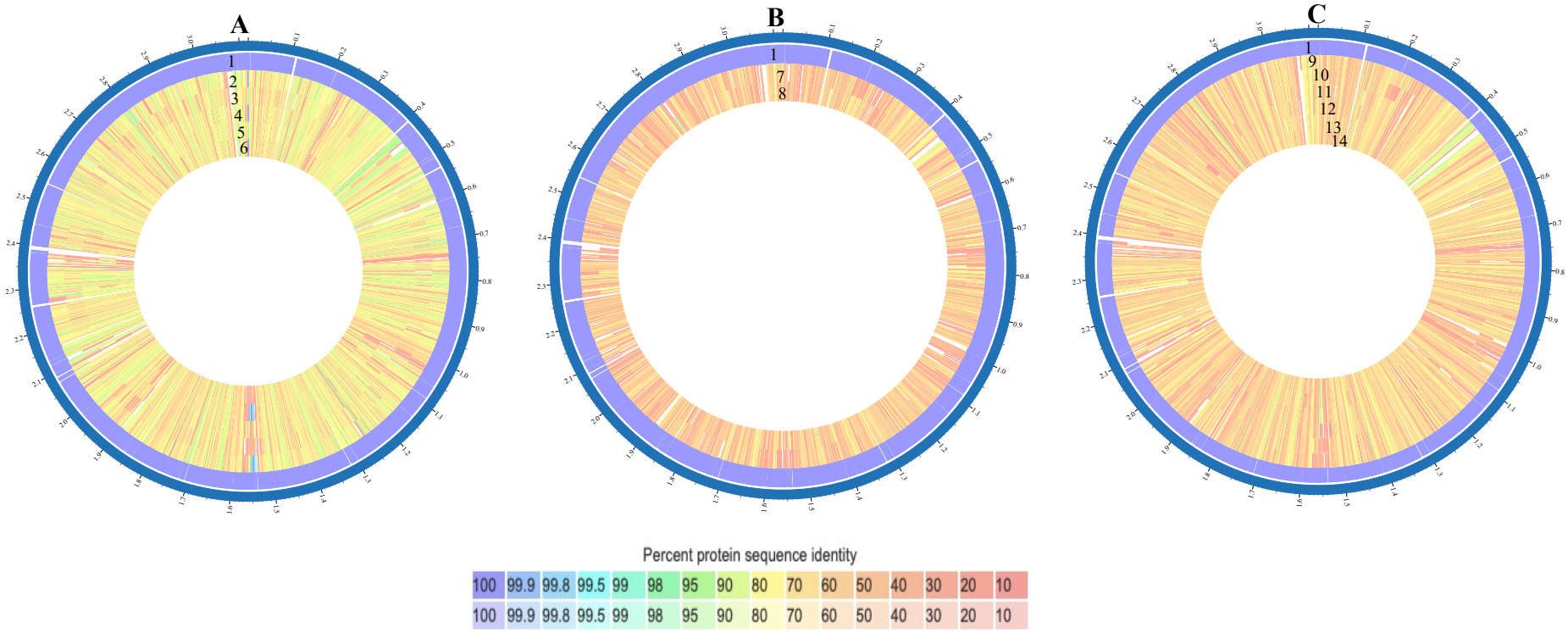


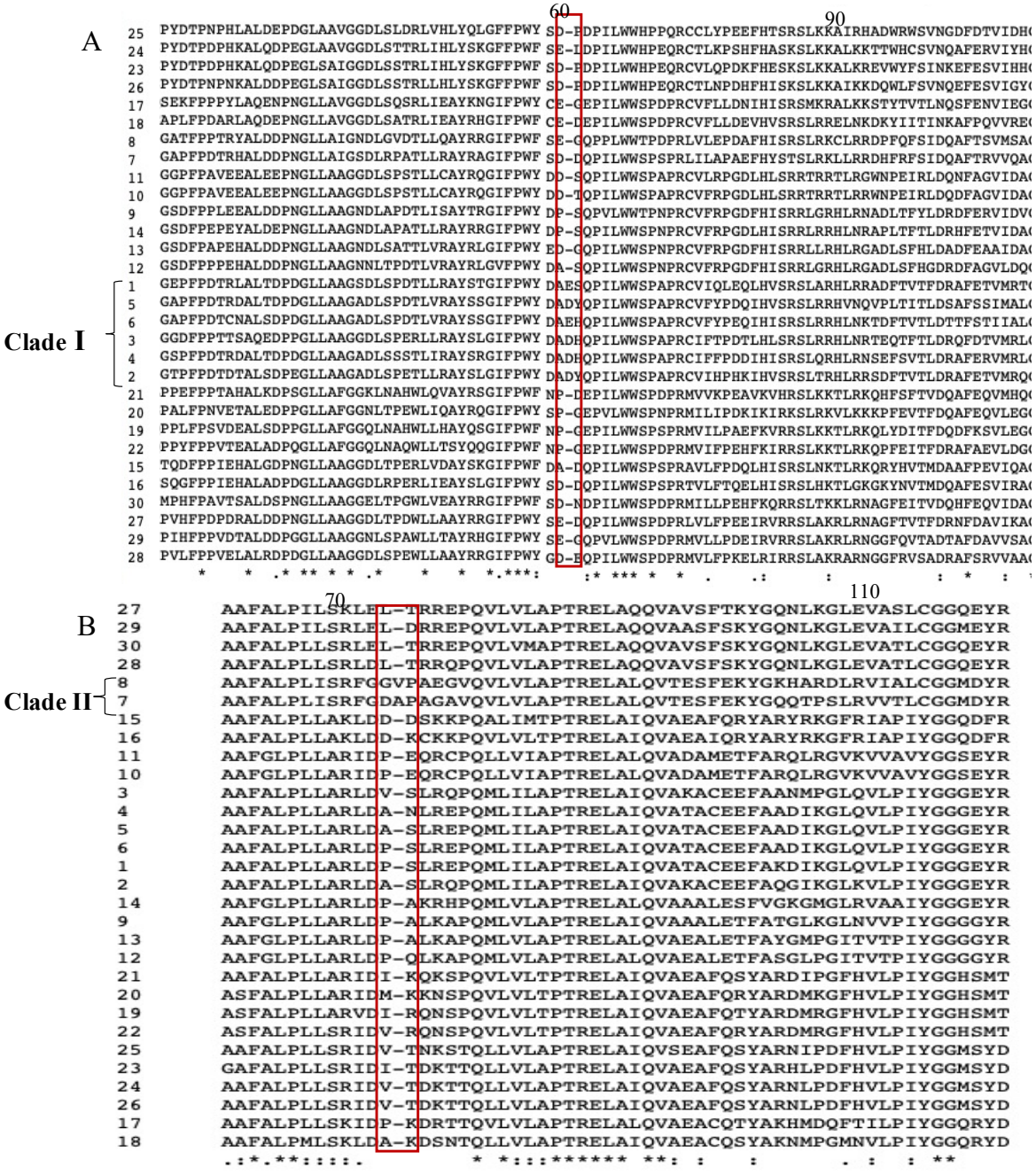
Fig. 54. Proteome comparison of the members of genus *Alcanivorax* with *A. borkumensis* Sk2^T as the reference genome; (numbering outside to inside)

(A) Clade I members with *A. borkumensis* SK2^T as the reference genome

(B) Clade II members with *A. borkumensis* SK2^T as the reference genome

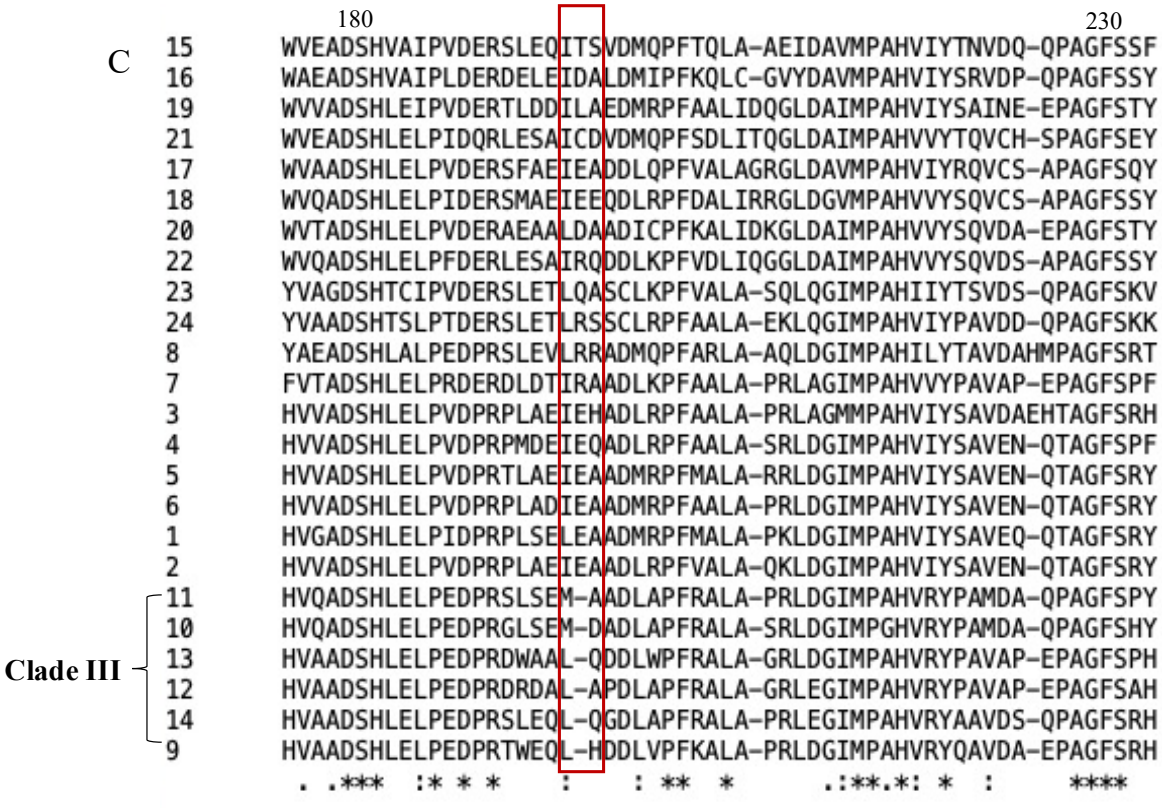
(C) Clade III members with *A. borkumensis* SK2^T as the reference genome

Numbering of the type strains are the same as that of Fig. 53



1. *A. borkumensis* SK2^T, 2. *A. jadensis* T9^T, 3. *A. hongdengensis* A-11-3^T, 4. *A. profundi* MTEO17^T, 5. *A. nanhaiticus* 19-m-6^T, 6. *A. sediminis* PA15-N-34^T, 7. *A. pacificus* W11-5^T, 8. *A. indicus* SW127^T, 9. *A. venustensis* ISO4^T, 10. *A. dieselolei* B-5^T, 11. *A. xenomutans* JC109^T,12. *A. profundamaris* ST75FaO-1^T, 13. *A. gelatiniphagus* MeBiC08158^T, 14. *A. mobilis* MT13131^T, 15. *Ketobacter alkanivorans* GI5^T, 16. *K. nezhaii* MCCC 1A13808^T, 17. *Kangiella taiwanensis* KT1^T, 18. *K. profundi* FT102^T, 19. *Oceanospirillum multiglobuliferum* ATCC 33336^T, 20. *O. linum* ATCC 11336^T, 21. *O. maris* DSM 6286^T, 22. *O. beijerinckii* DSM 7166^T, 23. *Marinomonas ushuaiensis* DSM 15871^T, 24. *M. profundi* M1K-6^T, 25. *M. polaris* DSM 16579^T, 26. *M. mediterranea* MMB-1^T, 27. *Halomonas ilicicola* DSM 19980^T, 28. *H. subterranea* CGMCC 1.6495^T, 29. *H. korlensis* CGMCC 1.6981^T, 30. *H. elongata* DSM 2581^T

Fig. 55. (A) Conserved signature indels (1 amino acid insertion) observed in Leucyl/phenylanyl-tRNA-protein transferase (EC 2.3.2.6) of clade I members
(B) Conserved signature indels (1 amino acid insertion) observed in DEAD-box ATP-dependent RNA helicase of clade II members



1.*A. borkumensis* SK2^{T}, 2. *A. jadensis* T9^{T}, 3. *A. hongdengensis* A-11-3^{T}, 4. *A. profundi* MTEO17^{T}, 5. *A. nanhaiticus* 19-m-6^{T}, 6. *A. sediminis* PA15-N-34^{T}, 7. *A. pacificus* W11-5^{T}, 8. *A. indicus* SW127^{T}, 9. *A. venustensis* ISO4^{T} , 10. *A. dieselolei* B-5^{T}, 11. *A. xenomutans* JC109^{T},12. *A. profundamaris* ST75FaO-1^{T}, 13. *A. gelatiniphagus* MeBiC08158^{T}, 14. *A. mobilis* MT13131^{T}, 15. *Ketobacter alkanivorans* GI5^{T}, 16. *K. nezhaii* MCCC 1A13808^{T}, 17. *Marinobacterium georgiense* DSM 11526^{T} 18. *M. lutimaris* DSM 22012^{T} 19. *M. stanieri* DSM 7027^{T} 20. *Nitrincola lacisaponensis* 4CA^{T} 21. *N. alkalilacustris* ZV-19^{T} 22. *N. tibetensis* xg18^{T} 23. *Marinospirillum minutulum* DSM 6287^{T} 24. *M. celere* DSM 18438^{T}

Fig. 56. (C) Conserved signature indels (1 amino acid deletion) observed in beta-N-acetylglucosaminidase (EC 3.2.1.53) of clade III members

3.3.5.6. Genome annotation - functional and metabolite pathways

The genomes of all the species of the genus *Alcanivorax* (n=14) were annotated with the help of the RAST and PATRIC databases. The metabolic pathways like phenylalanine, tyrosine, and tryptophan biosynthesis; metabolism of methane, propanoate, xenobiotics by cytochrome P450, sulfur, nitrogen, and carotenoid; type II polyketide

biosynthesis (Table 31) were analyzed. It was observed that most of the genes involved in these pathways were found in most of the members. Some genes like enoyl-CoA C-aceyltransferase (EC 4.2.1.17) (propanoate metabolism), glutathione transferase (EC 2.5.1.18) of phenylalanine, tyrosine, and tryptophan biosynthesis; cytochrome-c-oxidase (EC 1.9.3.1) of nitrogen metabolism were present in high copy numbers in all members. However, prephenate dehydrogenase (EC 1.3.1.12) and arogenate dehydrogenase (EC 1.4.1.43) genes were present solely in the members of clade I (n=6). As for the alkane degradation pathway annotated by the RAST software, all the members had a set of genes responsible for the alkane degradation (terminal oxidation) except for *A. indicus* SW127^T and *A. xenomutans* JC109^T. In these two organisms, alcohol dehydrogenase (*alk*J) and cytochrome P450 ω-hydroxylase (*cyp*) genes were absent (Table 31).

Further, for all the members of the genus *Alcanivorax*, the secondary metabolite biosynthetic gene clusters (BCGs) were evaluated. It predicted that on average, species of the genus *Alcanivorax* (n=14) have three BCGs per genome. Clade I (n=6) members had 21 BCGs (grouped into 5 BCG families), clade II (n=2) had 5 BCGs (grouped into 3 families), whereas clade III (n=6) had 19 BCGs (clustered into 6 BCGs families) (Table 33). All the members had at least one gene cluster related to ribosomally synthesized and post-translationally modified peptides (RiPPs) and ectoine. In contrast non-ribosomal peptide (NRPS) was present only in a few (Table 32). Other secondary gene clusters like redox-cofactor, aryl-polyene, and beta-lactone were present only in a few members of clade III and lanthipeptide-class II in clade II members.

Table 31. Genome annotation of the members of the genus *Alcanivorax* (1. *A. sediminis* PA15-N-34^T, 2. *A. nanhaiticus* 19-m-6^T, 3. *A. profundi* MTEO17^T 4. *A. borkurnensis* SK2^T, 5. *A. jadensis* T9^T, 6. *A. hongdengensis* A-11-3^T, 7. *A. indicus* SW127^T, 8. *A. pacificus* W11-5^T, 9. *A. dieselolei* B-5^T, 10. *A. xenomutans* JC109^T, 11. *A. profundimaris* ST75FaO-1^T, 12. *A. gelatiniphagus* MEBiC08158^T, 13. *A. venustensis* ISO4^T 14. *A. mobilis* MT13131^T

		CLADE I						CLADE II		CLADE III					
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Phenylalanine, tyrosine and tryptophan biosynthesis	Shikimate dehydrogenase (1.1.1.25)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Quinate dehydrogenase (1.1.5.8)	0	0	0	0	0	0	0	0	1	1	0	0	0	0
	Prephanate dehydrogenase (1.3.1.12)	1	1	2	1	1	1	0	0	0	0	0	0	0	0
	Arogenate dehydrogenase (1.3.1.43)	1	1	2	1	1	1	0	0	0	0	0	0	0	0
	Anthranilate phosphoribosyltransferase (2.4.2.18)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3-phosphoshikimate 1-carbovinyltransferase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3-deoxy-7-phosphoheptulonate synthase (2.5.1.54)	3	2	2	2	2	2	2	2	1	2	2	2	2	2
	Aspartate transaminase (2.6.1.1)	2	2	2	2	2	3	2	2	3	3	3	3	3	3
	Aromatic-amino-acid-transaminase (2.6.1.517)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Shikimate kinase (2.7.1.71)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Indole-3-glycerol-phosphate synthase (4.1.1.48)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Anthranilate synthase (4.1.3.27)	2	2	2	2	2	2	2	2	3	2	2	1	2	2
	Tryptophan synthase (4.2.1.20)	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Chorismate synthase (4.2.3.5)	1	1	1	1	1	1	1	3	1	1	1	1	1	1
	Chorismate mutase (5.4.99.5)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Methane metabolism	glutathione dehydrogenase (1.1.1.284)	1	1	1	1	1	1	1	3	1	1	1	1	1	1
	glycerate dehydrogenase (1.1.1.29)	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	Malate dehydrogenase (1.1.1.37)	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Phosphoglycerate dehydrogenase (1.1.1.95)	2	1	3	1	2	2	3	3	3	3	3	3	3	3
	Catalase peroxidase (1.11.1.21)	0	1	1	1	2	1	2	0	0	0	2	1	2	1
	Peroxidase (1.11.1.7)	0	1	0	0	1	1	0	0	0	0	0	0	0	0
	catalase (1.11.1.7)	0	2	0	1	3	1	0	3	1	1	1	1	1	1
	Glycerate 2-kinase (2.7.1.165)	0	0	0	0	0	0	0	0	1	1	1	1	0	1
Sulfur metabolism	1.8.1.2 - Sulfite reductase (NADPH)	1	1	1	2	1	3	1	1	3	3	4	1	3	3
	1.8.4.8 - Phosphoadenylyl-sulfate reductase (thioredoxin)	1	1	1	1	1	1	0	0	1	1	1	1	1	1
	1.8.99.2 - Adenylyl-sulfate reductase	0	0	0	0	0	0	0	0	2	0	0	0	0	2
	2.3.1.30 - Serine O-acetyltransferase	2	1	1	1	1	1	1	1	1	1	1	1	1	1
	2.5.1.47 - Cysteine synthase	2	2	2	2	2	2	1	1	1	1	1	1	1	1
	2.5.1.48 - Cystathionine gamma-synthase	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	2.7.1.25 - Adenylyl-sulfate kinase	1	1	1	1	1	1	0	1	1	1	1	1	1	1
	2.7.7.4 - Sulfate adenylyltransferase	1	1	1	1	1	1	2	2	1	1	1	1	1	1
	3.1.3.7 - 3'(2'),5'-bisphosphate nucleotidase	1	1	1	1	1	1	1	1	1	1	1	1	1	1

		CLADE I					CLADE II				CLADE III				
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Nitrogen metabolism	1.10.2.2 - Ubiquinol--cytochrome-c reductase	2	4	2	2	2	2	2	2	2	2	2	2	2	2
	1.13.12.16 - Nitronate monooxygenase	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	1.2.1.2 - Formate dehydrogenase	1	1	1	1	1	1	1	1	0	0	0	0	1	0
	1.4.1.13 - Glutamate synthase (NADPH)	3	4	3	3	4	3	2	12	2	2	2	2	2	2
	1.4.1.2 - Glutamate dehydrogenase	1	1	1	1	2	1	0	0	0	0	0	0	0	0
	1.4.1.4 - Glutamate dehydrogenase (NADP(+))	0	0	0	2	1	0	3	1	1	1	2	2	3	2
	1.4.7.1 - Glutamate synthase (ferredoxin)	1	2	1	1	0	1	2	1	1	1	2	2	2	2
	1.6.5.3 - NADH:ubiquinone reductase (H(+)-translocating)	0	0	0	0	0	6	7	0	0	0	0	7	7	13
	1.7.1.4 - Nitrite reductase (NAD(P)H)	4	4	4	5	3	3	3	2	3	3	2	0	3	3
	1.7.2.1 - Nitrite reductase (NO-forming)	0	0	0	0	0	0	1	0	1	1	1	0	1	0
	1.7.99.4 - Nitrate reductase	1	1	1	5	1	1	1	1	5	5	1	0	1	1
	1.9.3.1 - Cytochrome-c oxidase	6	6	7	7	7	3	11	8	6	6	9	6	11	6
	2.1.2.10 - Aminomethyltransferase	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	3.5.1.2 - Glutaminase	1	1	1	0	0	1	0	0	0	0	0	0	0	0
	3.5.1.49 - Formamidase	1	1	1	1	0	1	1	1	2	2	2	2	1	1
	4.2.1.1 - Carbonate dehydratase	1	2	1	2	2	2	1	2	1	1	1	1	1	1
	4.2.1.104 - Cyanase	0	0	0	0	0	0	0	0	2	1	0	0	0	1
	6.3.1.2 - Glutamate--ammonia ligase	1	1	1	1	1	1	2	2	1	1	2	2	2	2
	6.3.5.4 - Asparagine synthase (glutamine-hydrolyzing)	0	0	0	0	1	0	0	0	0	0	0	1	0	0
Carotenoid biosynthesis	1.14.13.- - With NADH or NADPH as one donor, and incorporation of one atom of oxygen.	1	1	2	2	2	3	0	1	2	2	3	1	2	1
	2.3.1.- - Transferring groups other than amino-acyl groups.	1	1	2	1	0	1	2	1	0	0	0	1	1	0
	2.4.1.- - Hexosyltransferases.	2	2	7	7	8	4	0	5	1	1	1	5	1	2
	Glycerate 2-kinase (2.7.1.165)	0	0	0	0	0	0	0	0	1	1	1	1	0	1
Biosynthesis of type II polyketide products	1.-.-.- - Oxidoreductases.	0	2	0	2	2	2	2	2	0	0	0	1	1	0
	1.1.1.- - With NAD(+) or NADP(+) as acceptor.	0	0	0	1	1	1	1	1	0	0	0	0	2	1
	1.14.13.- - With NADH or NADPH as one donor, and incorporation of one atom of oxygen.	1	2	1	2	2	3	1	1	2	2	3	1	2	1
	1.14.14.- - With reduced flavin or flavoprotein as one donor,	1	1	1	1	1	1	1	1	0	0	1	1	1	1
	2.1.1.- - Methyltransferases.	2	3	2	3	4	3	1	1	2	2	2	1	2	1
	2.4.1.- - Hexosyltransferases.	2	7	2	7	8	4	5	5	1	1	1	5	1	2
	4.-.-.- - Lyases.	0	1	0	0	1	0	0	0	0	0	0	0	0	0
	4.2.1.- - Hydro-lyases.	0	1	0	1	1	1	1	1	0	0	0	0	0	1
Alkane degradation	Alkane-1 monooxygenase AlkB (EC 1.14.15.3)	4	4	2	2	4	4	4	4	3	3	3	2	3	2
	Aldehyde dehydrogenase AlkH (EC 1.2.1.3)	3	2	4	1	3	3	4	4	1	4	7	3	3	2
	Alcohol dehydrogenase AlkJ (EC 1.1.99.-)	1	1	1	2	1	1	0	1	1	0	1	1	1	1
	Putative cytochrome P450 hydroxylase	1	1	1	1	2	1	0	1	1	0	1	1	1	1

Table 32. Secondary metabolites of the members of the genus *Alcanivorax*

	RiPP-like	Ectoine	Redox-cofactor	Lanthipeptide-class-II	Siderophore	NRPS	Aryl-polene	Beta-lactone
1	1	1	0	0	0	0	0	0
2	1	1	0	0	0	1	0	0
3	1	1	0	0	0	2	0	0
4	1	1	0	1	0	0	0	0
5	1	1	0	0	0	0	0	0
6	1	1	0	1	0	3	0	0
7	1	1	0	0	0	0	0	0
8	1	1	0	1	1	0	0	0
9	1	1	1	0	0	1	0	1
10	1	1	0	0	0	0	0	0
11	1	1	1	0	0	0	0	0
12	1	1	0	0	0	0	0	0
13	2	1	0	0	0	0	1	0
14	1	2	0	0	0	0	0	0

1. *A. sediminis* PA15-N-34^T, 2. *A. nanhaiticus* 19-m-6^T, 3. *A. profundi* MTEO17^T 4. *A. borkurnensis* SK2^T, 5. *A. jadensis* T9^T, 6. *A. hongdengensis* A-11-3^T, 7. *A. indicus* SW127^T, 8. *A. pacificus* W11-5^T, 9. *A. dieselolei* B-5^T, 10. *A. xenomutans* JC109^T, 11. *A. profundimaris* ST75FaO-1^T, 12. *A. gelatiniphagus* MEBiC08158^T, 13. *A. venustensis* ISO4^T,14. *A. mobilis* MT13131^T

3.3.5.7. Phenotypic and chemotaxonomic differences

The conventional components of polyphasic studies, phenotypic and chemotaxonomic attributes of the members of the genus *Alcanivorax* were studied. Significant differences were observed in characteristic features like motility, flagellar arrangement, polar lipid, and fatty acid composition. Clade I members were non-motile, whereas clade III members were motile with polar/sub-polar flagella. The polar lipid profiling of three clade members varied differently (Table 33). The distinctive differences were observed in fatty acid profiling as $C_{14:0}$ was found only in members of clade I and $C_{10:0}$ present only in the clade III species. The phenotypes distinguishing the members of the genus *Alcanivorax* from the other genera of the *Alcanivoraceae* family are given in Table 33.

Table 33. Characteristics differentiating the clades of *Alcanivorax* genus from other closely related genera in the family *Alcanivoraceae*

Characteristics	**1**	**2**	**3**	**4**	**5**
Cell size (l×b; µm)	0.2-0.8 × 0.2-2.5	1.7-2.3 × 0.3	0.6-2×0.3-2	4-35×1-1.2	4-35×1-1.2
Motility	-	v	+*	+	-
Flagellar arrangement	-	Peritrichous	Polar/sub-polar	nd	nd
Oxidase/catalase	+/v	+/+	+/+	w+/v	+/-
Optimum temp. °C(range)	18-37 (4-45)	25-30 (8-42)	28-40 (10-42)	25-28 (15-40)	20-22 (4-28)
Optimum pH (range)	6-8 (4-12)	7.5 (5.5-9)	7-8 (5-10)	7-8 (6-9)	7-8(6.5-9.5)
NaCl (%) toleration	0-15	0.5-12	0.5-20	1-8	1-0.3
D-Glucose fermentation	-	nd	-	-	-
β-glucosidase	v	+	-	-	-
Nitrate reduction	v	+	v	-	+
Hydrolysis of:					
Starch	-	-	-	nd	-
Casein	-	+	-	nd	-
Gelatin	v	v	v	v	+
Urease	v	v	-	-	-
Tween 80	v	+	+	+	nd
Utilization as carbon sources					
D-Glucose	-	+	-	nd	-
N-acetylglucosamine	-	+	-	-	-
Capric acid	v	-	v	-	-
Adipic acid	-	w+	v	-	-
Malic acid	-	-	v	-	-
Tri-sodium citrate	-	-	v	-	-
Phenylacetic acid	-	-	v	-	-
L-Arabinose	-	+	v	-	-
Polar lipid (present in clade members)	PL(3) L(3) AL(2)	PL (2) L (1)	PL(4) L(3) AL(3)	L(2)	AL(1) L(1) DPG(1)
Fatty acid (%)	$C_{12:0}$ 3-OH (<4.9), $C_{12:0}$ (<4.9), $C_{14:0}$ (<1.1)	$C_{12:0}$ 3-OH (<14.4), $C_{12:0}$ (<9.2)	$C_{12:0}$ (<10), $C_{12:0}$ 3-OH (<10), $C_{10:0}$ (<5.9)	$C_{14:0}$ (<4),	$C_{10:0}$ 3-OH (1.9), $C_{12:0}$ 3-OH (3.6), $C_{12:}$ 3-OH (2.2)

All genera are Gram-stain-negative, rod-shaped; contains PG, PE as Polar lipid; contains $C_{16:0}$ (>10 %), summed feature 3 (>9.4 %),8 (>21 %) as major fatty acid; *summed features represent two or three fatty acids that cannot be separated by gas chromatography with the MIDI system. Summed feature 3, $C_{16:1}\omega7c$/ $C_{16:1}\omega6c$; summed feature 8, $C_{18:1}\omega7c/\omega6c$; summed feature 9, iso-$C_{17:1}$ $\omega9c$ /$C_{16:0}$ 10-methyl.

DPG, phosphatidylglycerol; PG, phosphatidylglycerol, PE, phosphatidylethanolamine; PL, unidentified phospholipid; AL, unidentified amino lipids, L, unidentified Lipid; FA; Fatty acid; nd, Not determined; +*, except for *A. profundamaris* ST75FaO-1^T; v, variable; +, positive, -, negative; +/v, present/variable; +/+, positive for both tests; w+/v; weakly positive/variable; +/-; positive/absent.

1.Clade I (n=6) (*Alcanivorax* genus *sensu stricto*) (data based on *A. borkumensis* SK2^T, *A. hongdengensis* A-11-3^T; *A. jadensis* T9^T; *A. nanhaiticus* 19-m-6^T, *A. sediminis* PA15-N-34^T, *A. profundi* MTEO17^T (Yakimov et al. 1998; Wu et al. 2009; Bruns and Berthe-Corti 1999; Lai et al. 2016; Liao et al. 2020; Liu et al. 2019); 2. Clade II (n=2) (data based on *A. pacificus* W11-5^T, *A. indicus* SW127^T) (Song et al. 2018; Lai et al. 2011); 3. Clade III (n=8) (data based on *A. venustensis* ISO4^T, *A. dieselolei* B-5^T, *A. xenomutans* JC109^T; *A. balearicus* MACL04^T, *A. mobilis* MT13131^T; *A. profundamaris* ST75FaO-1^T; *A. gelatiniphagus* MEBiC08158^T, *A. marinus* R8-12^T) (Fernández-Martínez et al. 2003; Liu and Shao 2005; Rahul et al. 2014; Rivas et al. 2007; Yang et al. 2018; Dong et al. 2021; Kwon et al. 2015; Lai et al. 2013); 4. Genus *Ketobacter* (n=2) (data based on *K. nezhaii* GYS_M3H^T, *K. alkanivorans* GI5^T (Li et al. 2020; Kim et al. 2018); 5. Genus *Pelagibaculum* (n=1) (data based on *P. spongiae* Hp12^T) (Knobloch et al. 2019)

DISCUSSION

4. DISCUSSION

4.1. Root associated bacteria (RAB) of orchids and their ecological relevance

The concept of plants and their associated microbiota forming a 'holobiont' has become increasingly favored and considered (Vandenkoornhuyse et al. 2015). These associated microbiomes (bacteria and fungi) are dynamically involved in significant activities like nutrition, growth and resistance to biotic and abiotic stress (Herrera et al. 2021). Over the decade, progressive research on plant microbiota has shed light on the importance of these associations and colonization, which are complex and numerous (Kurth et al. 2013). At the outset, the interaction seems futile; however, the additional adaptive functions endorsed by such microorganisms are recognized as an essential feature, adding to the plant's ability to survive environmental changes and conditions (Bulgarelli et al. 2013). Studies, especially on bacterial interactions, have shown that they confer plant growth promoters (acclimatizing seedlings), pathogen resistance (by producing antibiotics, lytic enzymes) and uptake of nutrients (Trivedi et al. 2020; Vandenkoornhuyse et al. 2015).

The concepts of holobionts and RABs in plant evolution and ecology are still emerging and expanding; however, their bacterial community composition and functional role are altogether overlooked in one of the largest and evolutionarily advanced Orchidaceae plant families (Kaur and Sharma 2021). Orchid plants can be depicted as an eminent evolutionary model because of their unique habitats, taxonomic diversity, ecological adaptations (pollination, mycorrhizal interactions, nutrient acquisition) and global distribution (Herrera et al. 2021). The distinctive interactions of orchids and mycorrhiza are well studied; however, the distribution of RABs is invariably less considered. To fill in this research gap and understand the unexplored taxonomic diversity, the study was designed to examine RABs by culture-dependent and independent methods.

4.1.1. Diversity of root associated bacteria of orchids

4.1.1.1. Culture-independent analysis of bacterial communities

Using an amplicon-based metagenome approach, the bacterial communities of the root of orchids (RABs) were explored. It targeted the different taxonomic groups of bacteria that populated roots. To the best of our knowledge, it is the first attempt to understand the diversity by metagenome approach in orchids from various parts of India. 6389 OTUs spanning 17 phyla in ten orchids were observed, amongst which 8 phyla existed common among all the orchid roots. The findings revealed differences in bacterial community profiles at the taxonomic level of phylum. Taxonomic assignment showed OTUs belonging to the major phyla, like *Proteobacteria, Acidobacteria* and *Actinobacteria*, as the most abundant phyla, irrespective of the orchid species but dissimilarity in the distribution occurred in each plant (Fig. 7). *Proteobacteria, Actinobacteria* and *Acidobacteria* are effective root colonizers in several plants such as rice, tomato and wheat (Mitter et al. 2017; Xu et al. 2021). The data associated with the representation of these taxa of microorganisms (*Proteobacteria, Firmicutes, Actinobacteria, Bacteroidetes*) are liable only to the information from cultivated organisms and are often dominated by these phyla. The other phyla like *Planctomycetes*, *Verrucomicrobia*, *Armatimonadetes* and *Chloroflexi* were also present in lower abundances but distributed varyingly in the root microbiome of orchids (Fig. 7).

The study revealed that the root microbiome of epiphytic orchids was similar to that of the terrestrial orchids (Mediterranean) in terms of the abundant phyla like *Proteobacteria, Actinobacteria, Acidobacteria* and *Bacteroidetes* (Alibrandi et al. 2020) as in any of the other metagenome studies of soil, human gut and water (Bradley and Pollard 2017; Qin et al. 2016; Mhete et al. 2020). Epiphyte orchids from provinces of China differed from epiphytes of India in having OTUs belonging to phyla like *Lentisphaerae,*

Tenericutes and *Spirochaetes* (Li et al. 2017). However, unclassified OTUs (genus level) of epiphytes from India were more (50 to 89 %) as compared to the epiphytes from provinces of China (0.02 to 33.84 %) in the root microbiome (Li et al. 2017). It was interesting to see OTUs belonging to the *Acidobacteria* subgroup GP1 present in India's microbiome of orchids roots. *Acidobacteria* subgroup GP1 is predominant in soil (Lin et al. 2019), there are no reports of these members constituting the root microbiomes in orchids. 23 genera were commonly distributed between all the samples, forming the core root microbiota (Table 4). NDMS plot showed that the clustering of bacterial communities was region-specific (Fig. 10). Interestingly, plant-associated genera (iterated by literature) like *Burkholderia*, *Bradyrhizobium*, *Rhizobium*, *Sphingomonas*, *Methylobacterium*, *Mucilaginibacter*, *Microbacterium, Solirubrobacter, Beijerinckia, Acetobacter* formed part of the core root microbiome (Fan et al. 2021; Santos et al. 2020). Therefore, the bacterial communities are assembled in distinctive phylogenetic order in different orchid roots.

Phytopathogen representatives of roots like *Ralstonia* sp., *Pectobacterium* sp. and *Dickeya* sp. were absent in the orchid roots. OTUs of "*Candidatus*" phyla like "*Ca.* Sumerlaeota" (BRC1), "*Ca.* Absconditabacteria" (SR1), "*Ca.* Saccharibacteria" (TM7), "*Ca.* Parcubacteria" (OD1) and "*Ca.* Latescibacteria" (WS3) were also detected in the orchid root microbiome (Fig. 8B) (The Irpcm Phytoplasma/Spiroplasma working team). These "*Candidatus*" members are part of the "Candidate Phyla Radiation" (CPR), which is an expansion of the tree of life which has grown owing to the metagenomic and single-cell genomic studies (Danczak et al. 2017). "*Candidatus"* phyla like "*Ca*. Saccharibacteria" is commonly associated with the roots of citrus plants (*Citrus reticulata* Blanco) (Huang et al. 2021); however, its occurrence reported in this study with the orchid roots is for the first time. Unique phylotypes of bacterial communities of "*Ca.* UH24" to "*Ca.* UH59" (Fig. 12)

were detected in the orchid's roots and harbor taxonomically diverse bacteria, whose ecological functions are yet to be understood.

In the current taxonomic scenario, approximately 75 bacterial and 18 archaeal phyla are revealed by the 16S rRNA phylogeny (https://lpsn.dsmz.de/domain); however, their original count could be higher (Solden et al. 2016). The count of bacterial phyla could be predicted to be around 1500, with "yet uncultivated" taxa accounting for ~85 % of the phylogenetic diversity (Nayfach et al. 2021; Yarza et al. 2014). Therefore, there is a need to update the Candidate Phyla (CP) database, which focuses primarily on metagenomic based taxonomic studies, as it would help in expanding the tree of life. This study showed that rhizo-microbiomes of epiphytic orchids are one of the significant reservoirs of a rich diversity of bacteria, among which many are CPs. Further, insights into the rhizo-microbiomes of epiphytic orchids will help understand their role in plant development and provide a novel perspective on taxonomy, metabolism, microbial evolution and in orchids biotechnology.

4.1.1.2. Culture dependent bacterial analysis of orchids

Along with the amplicon-based metagenome analysis, isolation and culturing of bacteria from roots of orchids in the designed media were also carried out (Table 7). Out of the total isolated strains, eighteen were isolated from different orchid roots and majorly belonged to phyla of Actinobacteria, Proteobacteria and Firmicutes. The cultivated phyla composition was comparable to the abundant phyla in the metagenome data (Fig. 7). The cultured bacterial diversity at genera level were also on par with the observed metagenomic data for species like *Cymbidium* sp., *Aeridis* sp., *Dendrobium* spp., *Rhynchostylis* sp., *Vanda* sp. (Fig. 8A, Table 7). However, the cultured diversity of *Bulbophyllum* spp., *Agrostophyllum* sp. and *Thunia* sp. are lacking, as successful culturing of bacteria was not achieved for reasons unknown. However, the present studies illustrated that most of the

cultivated organisms were from the well-studied taxa; this must have occurred because of the media specificity allowing only the chemotrophs to grow. Amongst the isolated strains, strain JC501^T and strain JC701^T were deemed as novel species established on the basis of 16S rRNA gene identity (%) (Table 7) and overall genome related index (OGRI) like ANI and *d*DDH (Fig. 16). In the subsequent study, these two strains were described as novel species from the roots of orchid species. Both strains were potential candidates for PGPR activities, as shown by the metabolic studies of indole activities (Fig. 21, 25B).

4.1.1.2.1. Description of novel species isolated from the roots of orchids

Strain JC501^T and strain JC701^T were isolated from orchid roots of *Aeridis* sp. and *Dendrobium* sp., respectively. Strain JC501^T and JC701^T were closely related to *Paracoccus marinus* KKL-5^T and *Microbacterium testaceum* DSM 20166^T, respectively, based on the 16S rRNA gene, OGRI-like ANI and *d*DDH values. For both, 16S rRNA gene identity, ANI and *d*DDH values were below < 98.8 %, < 86 % and < 30 %, respectively (Table 7); they were described well at the species level (Rosselló-Móra and Amann 2015). In contrast, *Paracoccus marinus* and *Microbacterium testaceum* were isolated from a marine sample and rice grain, respectively (Khan et al. 2008; Komagata and Iizuka 1964). Both strains (JC501^T and strain JC701^T) ability to grow in an artificial medium demonstrate its free-living nature. Strain JC501^T was unable to grow at salt concentrations beyond 1 % (*P. marinus* NBRC 100637^T was able to thrive at greater salt concentrations), indicating that it is not a true member of the *P. marinus* marine niche. *P. marinus* clade members were isolated from water (Kämpfer et al. 2016), sediments (Li et al. 2009; Pan et al. 2014), blood (Kämpfer et al. 2019) and solvents (Lipski et al. 1998), indicating that they are widespread. Similarly, members of the genus *Microbacterium* are cosmopolitan, with some members being isolated from the rice grain (Komagata and Iizuka 1964), hydrothermal vent (Zhang et al. 2014), human ulcer (Funke et al. 1998) and soil (Kook et al. 2014).

The major polar lipids and fatty acids of strain JC501^T and JC701^T were similar to the earlier reported species of the genus *Paracoccus* and *Microbacterium* (Table 10, 12). Both strains JC501^T and JC701^T have the 2-C-methyl-d-erythritol 4-phosphate/1-deoxy-d-xylulose 5-phosphate (MEP) pathway for the synthesis of the 5-carbon isoprene unit and is observed in bacteria, green algae and higher plants (Perecca et al. 2020). For both the strains of JC501^T and JC701^T, PGPR activities like phosphate solubilization and siderophore production were not observed except for the production of indolic compounds. Therefore, indole metabolism was studied in detail and showed that both organisms could produce an array of indole derived molecules (Fig. 21, 25B). Of these arrays of molecules, only indole acetic acid was identified, while others are remaining as unidentified. Henceforth, suggesting that both strains could be considered a unique repository for indole-derived compounds and metabolism. Being isolated from the roots, strain JC501^T and JC701^T acquisition of IAA synthesis capacities could enhance bacteria–plant interactions, giving it an ecological advantage.

Further, secondary metabolic biosynthetic gene clusters like beta-lactone, terpenes, and type 3 polyketide synthase (T3PKS) were predicted in both strain's genomes. Beta lactones and T3PKS are diverse molecules that are shown to have antibiotic and/or anticancer properties (Awolope et al. 2021; Navarro-Muñoz and Collemare 2020). Studies have suggested that the induction of terpene synthesis by bacteria leads to the protecting of plant cells against reactive oxygen species (ROS) (Salomon et al. 2016). Likewise, both strains could assist orchid species in protecting against ROS during stress. Therefore, a general hypothesis can be formed that members of the genus *Paracoccus* and *Microbacterium* have physiological and genomic traits owing to their varying environmental conditions.

4.1.1.2.2. Proposal of *Paracoccus aeridis* sp. nov.

The phylogenetic study based on the 16S rRNA gene sequence showed that strain JC501^T was affiliated to the genus *Paracoccus* with the highest level of sequence identity with *P. marinus* KKL-A5^T (98.9 %) and other members of the genus *Paracoccus* (<97.3 %). Also, the values of MLSA genes (<84.6 %) are lower than the cut-off values recommended for new species delineation, as shown in *Salini vibrio*, *Thioclava* (members of class *Alphaproteobacteria*) and *Bacillus cereus*, which were between 95.6–97.5 % (López-Hermoso et al. 2017; Liu et al. 2017). The phenotypic traits of strain JC501^T, include colony morphology and cell morphology, temperature, salt tolerance, carbon substrate utilization (Table 10), fatty acid composition (Table 10), as well as differences in polar lipids (Fig. 19), carotenoid profile (Fig. 20), and indole production (Fig. 21), are well reflected in the genomic delineation (MLSA, ANI, *d*DDH). As a result, we propose the name *Paracoccus aeridis* sp. nov. for strain JC501^T as a member of the genus *Paracoccus*. After the year 2020 (publication year), eight new species of the genus *Paracoccus* (*P. alkanivorans* 4-2^T, *P. amoyensis* 11-3^T, *P. aurantiacus* TK008^T, *P. liaowanquigi*i 2251^T, *P. luteus* CFH 10530^T, *P. onubensis* 1011MAR3C25^T, *P. simplex* CCUG 71989^T, *P. xiamenensis* 12-3^T) has been validly described whose 16S rRNA identity (%) is in between 94.7-96.4 % (https://lpsn.dsmz.de/genus/paracoccus). Hereby, these species were not included in this study.

4.1.1.2.2.1. Description of *Paracoccus aeridis* sp. nov.

Paracoccus aeridis (*a.e'ri.dis*. N.L. gen. n. *aeridis* of the orchid genus *Aerides*)

Gram-negative cells, obligately aerobic, chemotropic, and non-motile. The cells have a coccoid form with 0.9–1 μm. When grown on half-strength nutrient agar, colonies are round, pale orange, and elevated, measuring 1–2 mm in diameter with uneven edges and spreading across the plate. At 25–35°C, growth is noted, with optimum growth at 30°C. The strain's optimal pH is 7.0; however, it can thrive in a pH range of 6.0–7.5. NaCl is not obligate for growth but can tolerate up to 1.0 % NaCl (w/v). The activity of catalase and oxidase are positive. Hydrolysis does not occur in casein, gelatin, starch, chitin, Tween 20, Tween 40, Tween 80, urea or cellulose. D-mannitol, D-glucose, pyruvate, rhamnose, D-galactose, ribose, maltose, sorbitol, D-mannose, sucrose, inositol, L-alanine, and histidine all contribute to growth. Propionate and fructose are examples of non-growth substrates. L-tryptophan produces indole 3-acetic acid and a few unidentified indole-derived molecules. Vitamins aren't required for growth. The only isoprenoid quinone is Q10. Polar lipids include phosphatidylcholine, phosphatidylethanolamine, phosphatidylglycerol, two unknown amino lipids, four unknown phospholipids and two unknown lipids. $C_{16:0}$, $C_{17:0}$, summed feature 8 ($C_{18:1}$7c and/or $C_{18:1}$6c), and anteiso-$C_{11:0}$ are the major fatty acids. The type strain has a genomic G+C content of 69.4 mol %.

The type strain JC501^T (=LMG 30532^T=NBRC 113644^T) was isolated from a root of an epiphytic orchid species (*Aerides maculosa*) sample taken in Amboli, Western Ghats, Karnataka, India. The 16S rRNA gene sequence has the entry number LT799401 in GenBank, while the draft genome has the accession number VJYZ000000000 in GenBank.

4.1.1.2.3. Proposal of *Microbacterium indolicus* sp. nov.

The phylogenetic analyses based on the 16S rRNA gene sequence showed that strain JC701^T was affiliated to the genus *Microbacterium* with the highest level of sequence identity with *M. testaceum* NBRC 12675^T (99.3 %) and other members of the genus *Microbacterium* (<98.7 %). The phenotypic traits of strain JC701^T include colony and cell morphology, temperature, salt tolerance, carbon substrate utilization (Table 12), fatty acid composition (Table 12), polar lipids (Fig. 25A), and indole production (Fig. 25B), are well reflected in the genomic delineation (ANI, *d*DDH) and gene composition of strain JC701^T from its closest strain. As a result, the proposal of the name *Microbacterium indolicus* sp. nov. for strain JC701^T belonging to the genus *Microbacterium*.

4.1.1.2.3.1. Description of *Microbacterium indolicus* sp. nov.

Microbacterium indolicus (*in.dol'i.cus*. N.L. neut. n. *indolum*, indol; L. masc. adj. suff. -*icus*, suffix used with the sense of pertaining to; N.L. masc. adj. *indolicus*, pertaining to indole, referring to the ability of the organism to produce indole)

Gram-stain-positive, aerobic, non-motile, non-flagellated, non-endospore-forming, rod-shaped bacterium (0.8–1 µm). Colonies are shiny, yellow, convex, opaque, with smooth and entire margins (0.5–1 mm in diameter). Growth occurs at 14 to 37 °C (optimum 37 °C) and at pH 5.0–8.0 (optimum pH 7.0). No requirement of NaCl for growth but tolerated up to 6 % (w/v). Oxidase positive and catalase negative. Positive growth in the presence of a number of carbon sources like D-glucose, ribose, sorbitol, sucrose, inositol, pyruvate, arabinose, lactose, fructose, galactose, maltose and rhamnose. No growth was observed with carbon substrates like mannose. No growth when peptone and ammonium sulphate was given as nitrogen sources. No requirement of additional supplements of vitamins for growth; however, growth yields are enhanced with the addition of yeast extract (0.01 %). Negative for hydrolysis of gelatin and starch, H_2S production and nitrate

reduction. Positive for indole test. The major polyamine is spermidine. Major polar lipids were phosphatidylglycerol, diphosphosphotidylglycerol, unidentified phospholipid, and three unidentified glycolipids (GL1, 2, 3). The predominant fatty acids are anteiso-$C_{15:0}$, anteiso-$C_{17:0}$, iso-$C_{16:0}$ and $C_{16:0}$. The predominant isoprenoid menaquinones are MK-11 and MK-10. The genomic DNA G+C content of the type strain is 70.5 mol %.

The type strain, JC701^T (=KCTC 49385^T=MCC 4427^T=NBRC 114338^T), was isolated from the root of an orchid in India. The accession number of GenBank for the 16S rRNA gene sequence and the draft genome is LR656263 and JAJNCV000000000, respectively.

Therefore, the findings and discussions based on the existing knowledge on orchid RABs show them as a potential source of biotic interactions for the general functioning of the orchid/host plants. A general hypothesis is that the creating an open-access and indexed bacterial culture collection that includes all microbiota members characterized by culture-independent (16S rRNA ribotype sequences) accompanied by isolating techniques will be a crucial tool in the future. This will allow the systematic testing of synthetic communities in the lab for plant growth and health functions (defined nutrient conditions). The study of plant microbiome entails tying microbial ecology to the general functioning of the host as well as prospecting microbes as a store of novel genes for the functioning of the host (Kaur and Sharma 2021). Further, it will assist in the developing communities that will resolve ecosystem networking, allow identification of syntrophic interactions and ultimately define principles driving niche composition and competition (Bulgarelli et al. 2013).

4.2. Proposal and descriptions of novel taxa

In addition to the orchids, a diversity of bacteria from key ecological habitats like saltpans, acid mines, coastal sand and hot springs were also investigated (Table 7). The methodological constraints to isolate organisms from such habitats are no longer there. It is now possible to fully sequence a bacterial genome from such unique habitats to analyze their specific genes and genetic information (Romalde et al. 2019). In the current scenario, bacterial genomics is assisting in reinforcing boundaries of genera and a higher hierarchy of taxonomy with constant revising of the denotation of the term "species". Furthermore, whole-genome sequencing has aided in the resolution of complicated taxonomic positions in recent years (Hördt et al. 2020; Orata et al. 2018). Furthermore, a better understanding of the phylogeny of such organisms from unique habitats allows scientists to predict microbes' genetic capability for biotechnological uses and adaptation to environmental changes. Therefore, given below are the description and reclassification of a few bacterial taxa isolated from ecologically important habitats (Table 7). The strains are delineated per the current genomic related indexes and the differentiating polyphasic attributes.

4.2.1. *Mesobacillus* sp. JC1013^T

4.2.1.1. Ecology

strain JC1013^T and *M. selenatarsenatis* MTCC 9470^T were isolated from an orange-colored pond of a solar saltern and the sediment of an effluent drain in Japan (Yamamura et al. 2007), respectively. *Mesobacillus* spp. that are closely related were isolated either from freshwater (Kanso et al. 2002), soil (Ahmed et al. 2007) or fermented foods (Yoon et al. 2017). In laboratory circumstances, strain JC1013^T develops on artificial media, confirming its free-living nature. As illustrated in Fig. 40, strain JC1013^T is predicted to have a unique pathway for the production of carotenoid pigments. Carotenoids are thought to play a protective role, as in most cases (Sandmann 2019). Since zeaxanthin and lycopene

are effective UV-protectants, this could be especially advantageous for strain JC1013^T, which was cultured from a solar saltern wherein temperatures can vary up to 45 °C, also with a lot of UV light (Khaneja et al. 2010). Although strain JC1013^T was shown to grow in a salt range of 1 to 7 %, it was not required mandatorily for its growth. This could be due to its ability to tolerate high salinity on a facultative basis, making it resilient to saline and non-saline environment. Members of the family *Bacillaceae* have a diverse range of species, many of which are isolated from various habitats. Pigmentation, other physiological and genetic traits may have contributed to strain JC1013^T adaptation to saline and hot environments.

4.2.1.2. Proposal of *Mesobacillus aurantius* sp. nov.

The phylogenomic based species delineation of strain JC1013^T from *M. selenatarsenatis* MTCC 9470^T is well supported by the differences in phenotypic characters such as colony and cell morphology, temperature, salt tolerance, organic carbon substrate utilisation (Table 14), fatty acid composition (Table 14), as well as differences in polar lipids (Fig. 33), indole production (Fig. 35) and carotenoid profile (Fig. 37). Strain JC1013^T is characterised as a type strain representing a distinct species of *Mesobacillus* and proposed as *Mesobacillus aurantius* sp. nov.

4.2.1.3. Description of *Mesobacillus aurantius* sp. nov.

Mesobacillus aurantius (au.ran.ti.us. N.L. adj. *aurantius* orange, related to colony color; N.L. adj. n. *aurantius* orange-colored colony)

Cells are obligatory aerobes that are non-motile. Slender rods (2–3 × 0.1–0.2 µm) with terminal endospores. Colonies on nutrient agar are uneven, flat, dry, and orange-pink, with a diameter of 1–3 mm. The ideal temperature for growth is 30 °C (range 25–40 °C) and, the ideal pH for growth is 7 (range 6.0–8.5). For growth NaCl is not essential; however, can tolerate up to 7 % (w/v). Positive for catalase and negative for oxidase. Casein, gelatin,

chitin, CMC, and tween 20/40/80 are not hydrolyzed. Lipase and urease activity are both negative. Indole tested positive. Lactose, inositol, mannose, sorbitol, and pyruvate promote good growth, while arabinose, galactose and glucose support weak growth. Fructose is not utilized. The nitrogen sources, including ammonium chloride and, L-tryptophan were utilized. Vitamins were not necessary; however, the addition of yeast extract promoted growth. Menaquinone 7 is a component of the quinone system. The principal polar lipids are diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, four unidentified phospholipids and two unidentified amino lipids. Iso-$C_{15:0}$, anteiso-$C_{15:0}$, $C_{16:0}$3-OH, iso-$C_{17:0}$10c, and summed feature 4 (iso-$C_{17:1}$I/ anteiso B) are the most common fatty acids. The diagnostic amino acids of the cell wall are *m*-diaminopimelic acid.

A pond near a solar salt saltern yielded the type strain JC1013^T (= NBRC 114146^T = KACC 21451^T). The 16S rRNA gene sequence and draft genome of *Mesobacillus aurantius* JC1013^T have been deposited in GenBank under the accession number LS998022 and JAAVUM000000000, respectively.

4.2.2. Phylogenomic based reappraisal of the genus *Roseomonas*

The genus *Roseomonas* has been extensively researched because its members have severe environmental and clinical implications (Rihs et al. 1993; Romano-Bertrand et al. 2016; Subhash et al. 2016). In 2020–2021, six new species names have been validly published. Due to the emergence of the genomic era, whole-genome sequencing data gives precise and more decisive information for taxonomic delineation than the 16S rRNA gene marker alone. The genus *Roseomonas* species has traditionally been delineated using a polyphasic method. According to Romano-Bertrand et al. (2016), the genus *Roseomonas* was divided into 7 distinct clades based on the 16S rRNA gene phylogeny and ecology of the members. Similarly, a study conducted by Hördt et al. (2020) revealed that the genus *Roseomonas* was heterogeneous and polyphyletic, requiring substantial reclassification. Taxogenomics is thought to be more exact and superior for delineating taxa since it improves data repeatability and reliability (Lalucat et al. 2022). A complete taxogenomic study was carried out to evaluate the absolute phylogenetic standing between the existing members of the genus *Roseomonas*.

The phylogenetic tree based on the 16S rRNA gene revealed that the genus *Roseomonas* was polyphyletic, with 7 major clades: Groups I to VII (Fig. 42). In terms of species composition, the cladding patterns of Groups I, II, III, VI, and VII members in this study are similar to the cladding pattern found by Romano-Bertrand et al. (2016). *R. aquatica* (Group IV) clustered outside the specified clade discovered by Romano-Bertrand et al. (2016), but *R. fluminis*, another member of this group, was later described by Ko et al. (2018). Further, *R. arctica* (Group V) was related to the *R. stagni* (Group III) group. The phylogenomic tree based on 92 concatenated genes for *Roseomonas*, on the other hand, revealed a similar cladding pattern (Fig. 43; except for Group IV members for which the genome sequences are unavailable).

Genome size variation has been proven to be a reliable taxonomic marker by Hördt et al. (2020) and Nouioui et al. (2018). They showed that the genome size of members of the class *Alphaproteobacteria* and phylum *Actinobacteria* appeared to be taxa-specific. Both the investigation showed that the genome size and G+C content (mol %) were phylogenetically conserved. As a result, the genome sizes have been incorporated into the descriptions formally. In this case, when compared to the other groups, Group I had the smallest genome size. Using the BPGA system to analyse pan and core genomes, showed that the *Roseomonas* pan-genome is open and contains 958 core genes, accounting for 19.1 % of the overall pan-genome (Table 22). The only *Roseomonas* groups that were capable of producing specific secondary metabolites (BCGs) were Group I (NRPS, asukamycin, fengymycin,); Group II (redox-cofactor); and Group V (arylpolene, NRPS/T1PKS) (Table 23). The genome annotation also showed that the gene make-up of the clusters was unique to each group, with Group I having the most and Group VII having the least number of gene clusters.

For species delineation, Chun et al. (2018) recommended minimum standards based on an overall genome-related index (OGRI) such as ANI and *d*DDH. For species delineation, Wayne et al. (1984) recommended a 70 % *d*DDH value cut-off. In the current scenario, ≥ANI values of 95 % indicate that the 2 strains belong to the same species, whereas the values of <95 % indicate that the 2 strains belong to separate species (75 % for different genera) (Richter and Rosselló-Móra 2009; Rodriguez and Konstantinidis 2014; Rosselló-Móra and Amann 2015). Both indices of *d*DDH and ANI are compatible with the suggested standards at the intra-group level, indicating that all *Roseomonas* members are well-defined at the level of species (Table 20). For genus delineation, a cut-off of 50 % was suggested for the POCP values (Qin et al. 2014). In between the groups, the POCP values were <65 % and ranged between 40 to 65 % (Table 21, Fig. 45A) and < 70 % (20-60 %)

with other members of the family *Acetobacteraceae*. As a result, the computed values of POCP for genus delineation for *Roseomonas* species did not coincide with those observed by Qin et al. (2014). Although the recommended genus border of the POCP value for bacterial lineages was set at 50 %, several research later revealed exceptions when comparing inter-genera comparisons. Further, the general cut-off of 50 % is regarded conservative because it is only an index/mark of relatedness (Liu et al. 2020), as evidenced by its inadequate in distinguishing different genera of the *Bacillaceae* (Aliyu et al. 2016), *Burkholderiaceae* (Lopes-Santos et al. 2017), Methylococcaceae (Orata et al., 2018), and *Rhodobacter* (Wirth and Whitman, 2018). AAI values of related but different genera ranged between 60 to 80 %, according to Luo et al. (2014). AAI values were lower than 70 % in this investigation between different groups of *Roseomonas* (Table 21, Fig. 45B) and other genera in the *Acetobacteraceae* family, consistent with Luo et al. (2014).

Hördt et al. (2020) showed that certain *Roseomonas* species, such as *R. stagni* cladded with *H. rosea* whereas *R. lacus* with *Rubritepida flocculans*. The *d*DDH, ANI, AAI, and POCP values clarified the differentiation between the stated species pairs. Between *R. stagni* and *H. rosea*; *R. lacus* and *R. flocculans*, the *d*DDH, ANI, AAI, and POCP values were 19.4, 74.1, 64.2, and 64.2 % and 19.8, 74.8, 66.6, and 57.1 %, respectively. The same analysis revealed *Rhodovarius lipocyclicus* was nested under the genus *Roseomonas* (Hördt et al. 2020). However, *R. lipocyclicus* is a well-distinguished member of the genus *Rhodovarius* as evident from the AAI and POCP values of 50.8–66.8 % and 53.2–64.7 %, respectively; which is below the proposed cut-off for the genus determination. The following indices distinguishes *Roseomonas* species from *Rubritepida*, *Humitalea*, and *Rhodovarius* genera. According to the findings of Romano-Bertrand et al. (2016) and Hördt et al. (2020), the current studies support the formation of 7 distinct groups within the genus *Roseomonas*.

4.2.2.1. Ecology of members of the genus *Roseomonas*

The genus *Roseomonas* is ubiquitously present in the environment. Group I members were the only members to be isolated from human blood (*R. gilardii, R. mucosa*, *R. gilardii* subsp. *rosea*) (Rihs et al. 1993; Han et al. 2003), except for *R. elaeocarpi*, which was otherwise isolated from olive phyllosphere (Damtab et al. 2016). *Roseomonas* strains have been commonly reported in association with human illnesses and isolated from clinical samples such as sputum, wounds, and genitourinary sites (Rihs et al. 1993; Wallace et al. 1990). It was also linked to immunocompromised patients (Marin et al. 2001). Other cases of *Roseomonas* related bacteremia include spinal osteomyelitis (Nahass et al. 1995), endocarditis (Shao et al. 2019), infected teeth (Diesendorf et al. 2017), and peritonitis (Malini et al. 2016). The genomic islands are rich in pathogenic genes and virulence factors. As a result of horizontal gene transfer, they become part of the genome; further, these genes impart virulence and pathogenicity to the bacteria (da Silva Filho et al. 2018). However, the genomic studies of the genus *Roseomonas* members do not suggest genes responsible for their virulence and pathogenicity, therefore cannot corroborate claims mentioned above. As a result, *Roseomonas* strains may be considered incidental rather than causative of pathogenicity. *Roseomonas* spp. have been isolated from a variety of sources, like sediment (He et al. 2014), soil (Kim and Ka 2014), polluted soil (Subhash and Lee 2018), freshwater (Baik et al. 2012), and human sources such as blood (Rihs et al. 1993).

The presence of accessory genes in *Roseomonas* members reveals their divergence due to environmental adaptations/remodelling in an ecological aspect. The presence of accessory genes indicates that genes were acquired in response to the selective pressure (Brito et al. 2015) or to colonise new biological areas (McInerney et al. 2017). The members global distribution might be ascribed to the acquisition of accessory genes for survival, which could be related to their simple varying organisation, genomic sizes, or

horizontal gene transfer rates (Aherfi et al. 2018). This is also evidenced by the different genetic patterns discovered during the pan-genome analysis for each group, which could have come from environmental changes.

4.2.2.2. Proposal to reclassify the genus *Roseomonas* into six novel genera

The results of the large taxogenomics study support the split of several groupings of the genus *Roseomonas* into distinct genera. Because *Roseomonas* is a highly divergent genus, members of it can no longer be grouped into a single phylogenetic genus, allowing the formation of six new genera. A revised description and reclassification are provided for each of the novel genera. The creation of seven distinct clades as independent genera is supported by phylogenetic trees based on the 16S rRNA gene (Fig. 42) and whole genome sequences (Fig. 43). Although 16S rRNA gene studies have several disadvantages, they can be nevertheless accountable for the classification of *Roseomonas* members. This analysis would be the first attempt to reclassify the genus *Roseomonas* by forming 6 novel genera: Group II as *Pararoseomonas* gen. nov., Group III as *Falsiroseomonas* gen. nov., Group IV as *Paeniroseomonas* gen. nov., Group V as *Plastoroseomonas* gen. nov., Group VI as *Neoroseomonas* gen. nov and Group VII as *Pseudoroseomonas* gen. nov. In the absence of genome sequences, the genus definition for Group IV is based on 16S rRNA gene analyses and phenotypic features. The genus definition based on 16S rRNA gene sequences and phenotypic features provides enough clarity to separate Group IV members from the other *Roseomonas* groups. The availability of genomic sequences, on the other hand, may help to solidify the reclassification in the upcoming future. The newly formed genera species have been precisely defined according to the required ANI and *d*DDH standards (Table 20). 16S rRNA gene identity, ANI, AAI, and POCP values match the genus delineation standards supported by other phenotypic features for the genus *Roseomonas* (Fig. 46).

4.2.2.2.1. Emended description of the genus *Roseomonas* by Rihs et al. (1998)

Except for a few changes, the description is the similar to the one given by Rihs et al. (1993) and Sánchez-Porro et al. (2009) for the genus *Roseomonas*. H_2S production varies, and the genome-based G+C content is ~70 mol %.

The type species is *Roseomonas gilardii*.

4.2.2.2.2. Description of *Pararoseomonas* gen. nov.

Pararoseomonas (Pa.ra.ro.se.o.mo'nas. Gr. pref. *para*-, next to, resembling; N.L. fem. n. *Roseomonas*, a bacterial genus; N.L. fem. n. *Pararoseomonas*, next to the genus *Roseomonas*).

Aerobic and Gram-negative. Coccoid/short rods and non-motile members. Both catalase and oxidase variable. Diphosphatidylglycerol, phosphatidylcholine, phosphatidylglycerol, phosphatidylethanolamine, unidentified amino lipids and unidentified phospholipids are major polar lipids. The primary fatty acids are $C_{18:1}$2-OH, $C_{18:1}$3-OH, $C_{16:0}$, $C_{18:0}$, and summed feature 3. The genomic size of the members is 4.9–6.4 Mbp, with a G+C content of 69.7–71.4 mol %. The 16S rRNA gene, 92 core genes (phylogenomics), AAI indices, POCP values, and morphological and genomic traits are used to define/delineate the genus.

The type species is *Pararoseomonas rosea.*

4.2.2.2.2.1. Description of *Pararoseomonas rosea* comb. nov.

Pararoseomonas rosea (ro'se.a. L. fem. adj. *rosea* rose-colored or pink).

Basonym: *Roseomonas rosea* (Kämpfer et al., 2003; Sánchez-Porro et al., 2009)

The description is similar which is determined by Sánchez-Porro et al. (2009). AJ488505 and FQZF00000000 is the accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is 173/96^T (=CIP 107419^T = DSM 14916^T).

4.2.2.2.2.2. Description of *Pararoseomonas vinacea* comb. nov.

Pararoseomonas vinacea (vi.na'ce.a. L. fem. adj. *vinacea*, of /belonging to wine/ the grape, alluding to the colony color).

Basonym: *Roseomonas vinacea* (Zhang et al., 2008)

The description is similar as determined by Zhang et al. 2008. MT760143 and BO93DRAFT are the accession numbers for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is CPCC 100056^T (=KCTC 22045^T = CCM 7468^T).

4.2.2.2.2.3. Description of *Pararoseomonas nepalensis* comb. nov.

Pararoseomonas nepalensis (ne.pal.en'sis. N.L. fem. adj. *nepalensis*, related to Nepal, the country where soil samples were collected).

Basonym: *Roseomonas nepalensis* (Chaudhary and Kim, 2017)

The description is similar to that given by Chaudhary and Kim in 2017. KX129819 is the accession number for the 16S rRNA gene sequence. The type strain is G-3-5^T (=JCM 31470^T = KACC 18908^T).

4.2.2.2.2.4. Description of *Pararoseomonas aeriglobus* comb. nov.

Pararoseomonas aeriglobus (a.e.ri.glo'bus Gr. masc. n. *aêr*, air; L. masc. n. *globus*, a sphere; N.L. masc. n. *aeriglobus*, a sphere living in the air).

Basonym: *Roseomonas aeriglobus* (Lee and Jeon, 2018)

The description is similar as determined by Lee and Jeon in 2018. KY864922 is the accession number for the 16S rRNA gene sequence. The type strain is KER25-12^T (=KACC 19282^T = JCM 32049^T).

4.2.2.2.2.5. Description of *Pararoseomonas aerilata* comb. nov.

Pararoseomonas aerilata (ae.ri.la'ta. Lat. masc. n. *aêr*, air; L. fem. perf. part. *lata*, carried by; N.L. fem. part. adj. *aerilata*, airborne).

Basonym: *Roseomonas aerilata* (Yoo et al., 2008)

The description is similar to *Roseomonas aerilata* as determined by Yoo et al. (2008). EF661571 and JONP00000000 are the accession numbers of the 16S rRNA gene and genome sequences, respectively. The type strain is 5420S-30^T (=KACC 12521^T= DSM 19363^T).

4.2.2.2.2.6. Description of *Pararoseomonas radiodurans* comb. nov.

Pararoseomonas radiodurans (ra.di.o.du'rans. L. masc. n. *radius*, a beam or ray; N.L. pref. *radio-*, pertaining to radiation; L. pres. part. *durans*, enduring; N.L. part. adj. *radiodurans*, enduring radiation).

Basonym: *Roseomonas radiodurans* (Kim et al., 2018)

The description is similar as determined by Kim et al. (2018) for *Roseomonas radiodurans*. The accession number for the 16S rRNA gene sequence is KY887689. The type strain is 17Sr1-1^T (=KCTC 52899^T = NBRC 112872^T).

4.2.2.2.2.7. Description of *Pararoseomonas pecuniae* comb. nov.

Pararoseomonas pecuniae (pe.cu'ni.ae. L. gen. n. *pecuniae*, of or from money or a coin, pertaining to the source of sample for the isolation of the type strain).

Basonym: *Roseomonas pecuniae* (Lopes et al., 2011)

The description is similar as given for *Roseomonas pecuniae* as determined by Lopes et al. (2011). GU168019 and JACIJD00000000 are the accession number for the 16S rRNA gene and the genome sequence, respectively. The type strain is N75^T (=LMG 25481^T = CIP 110074^T).

4.2.2.2.2.8. Description of *Pararoseomonas harenae* comb. nov.

Pararoseomonas harenae (ha.re'nae. L. gen. fem. n. *harenae*, of sand, pertaining to the isolation of the type strain from the desert sand).

Basonym: *Roseomonas harenae* (Deng et al., 2020)

The description is similar as given for *Roseomonas harenae* as determined by Deng et al. (2020). MN883396 and WWDL00000000 are the accession numbers for the 16S rRNA gene and genome sequences, respectively. The type strain is CPCC 101081^T (=KCTC 62852^T = NBRC 113512^T).

4.2.2.2.3. Description of *Falsiroseomonas* gen. nov.

Falsiroseomonas (Fal'si.ro.se.o.mo'nas. L. adj. *falsus*, false; N.L. fem. n. *Roseomonas*, a name of bacterial genus; N.L. fem. n. *Falsiroseomonas*, false *Roseomonas*).

This genus members are aerobic and Gram-negative with coccoid-rod-shaped cells. Within the genus, motility varies. Catalase is variable, and oxidase is positive. Diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylcholine, glycolipid, unidentified phospholipid, and amino lipid are major polar lipids. The primary fatty acids are $C_{16:1}$5c and $C_{16:0}$ and summed feature 3. The members have a genome size of 5.8–6.6 Mb and a G+C content of 70–72 mol % (genomes). The 16S rRNA gene, 92 core genes (phylogenomics), AAI, POCP values, and morphological and genomic traits are used to delineate the genus.

The type species is *Falsiroseomonas stagni*.

4.2.2.2.3.1. Description of *Falsiroseomonas stagni* comb. nov.

Falsiroseomonas stagni (stag'ni. L. gen. neut. n. *stagni*, of a pond, referring to the site of isolation of the type strain).

Basonym: *Roseomonas stagni* (Furuhata et al., 2008)

The description is similar to *Roseomonas stagni* as determined by Furuhata et al. (2008). AB369258 and FOSQ00000000 is the accession number for the 16S rRNA gene and genome sequences, respectively. The type strain is HS-69^T (=DSM 19981^T = JCM 15034^T = KCTC 22213^T).

4.2.2.2.3.2. Description of *Falsiroseomonas wooponensis* comb. nov.

Falsiroseomonas wooponensis (woo.po.nen'sis. N.L. masc./fem. adj. *wooponensis*, of/belonging to Woopo wetland, South Korea, the geographical origination of the type strain).

Basonym: *Roseomonas wooponensis* (Lee et al., 2015)

The description is similar as determined for *Roseomonas wooponensis* as given by Lee et al. (2015). The accession number for the 16S rRNA gene sequence is KF619243. The type strain is WW53^T (=KCTC 32534^T = JCM 19527^T).

4.2.2.2.3.3. Description of *Falsiroseomonas terricola* comb. nov.

Falsiroseomonas terricola (ter.ri'co.la. L. fem. n. *terra*, earth, soil; L. masc./fem. suff. -*cola*, inhabitant, dweller; from L. masc./fem. n. *incola*, dweller; N.L. masc./fem. n. *terricola*, soil-dweller, alluding to the isolation of the type strain from soil).

Basonym: *Roseomonas terricola* (Kim et al., 2017)

The description is similar as determined for *Roseomonas terricola* by Kim et al. (2017). The accession number for the 16S rRNA gene sequence is FJ654263. The type strain is EM302^T(=KACC 13942^T = KCTC 42906^T = NBRC 111477^T).

4.2.2.2.3.4. Description of *Falsiroseomonas selenitidurans* comb. nov.

Falsiroseomonas selenitidurans (se.le.ni.ti.du'rans. N.L. neut. n. *selenitum*, selenite; L. v. *duro*, withstand; N.L. part. adj. *selenitidurans*, withstanding/resisting selenite).

Basonym: *Roseomonas selenitidurans* (Hou et al., 2020)

The description is similar as determined for *Roseomonas selenitidurans* by Hou et al. (2020). MT107132 and JAAVNE00000000 is the accession number for the 16S rRNA gene and genome sequence. The type strain is BU-1^T (=GDMCC 1.1776^T = KACC 21750^T).

4.2.2.2.3.5. Description of *Falsiroseomonas fridaquae* comb. nov.

Falsiroseomonas frigidaquae (fri.gi.da'quae. L. masc. adj. *frigidus*, cold; L. fem. n. *aqua*, water; N.L. gen. fem. n. *frigidaquae*, from or of cold water, alluding to the isolation of the type strain from a water-cooling system).

Basonym: *Roseomonas frigidaquae* (Kim et al., 2009)

The description is similar as determined by Kim et al. (2009) for *Roseomonas frigidaquae.* EU210160 and JAAVTX00000000 is the accession number for the 16S rRNA gene sequence and genome sequence. The type strain is CW67^T (=JCM 15073^T = KCTC 22211^T).

4.2.2.2.3.6. Description of *Falsiroseomonas tokyonensis* comb. nov.

Falsiroseomonas tokyonensis (to.ky.o.nen'sis. N.L. masc./fem. adj. *tokyonensis*, of Tokyo, source from where the type strain was isolated).

Basonym: *Roseomonas tokyonensis* (Furuhata et al., 2014)

The description is similar as determined for *Roseomonas tokyonensis* by Furuhata et al. (2013). The gene accession number for the 16S rRNA gene sequence is AB297501. The type strain is K-20^T (=JCM 14634^T = KCTC 32152^T).

4.2.2.2.3.7. Description of *Falsiroseomonas riguiloci* comb. nov.

Falsiroseomonas riguiloci (ri.gu.i.lo'ci. L. masc. adj. *riguus*, well-watered; L. masc. n. *locus*, a site; N.L. gen. n. *riguiloci*, type strain isolated from a well-watered place).

Basonym: *Roseomonas riguiloci* (Baik et al., 2012)

The description is similar as determined by Baik et al. (2012) for *Roseomonas riguiloci.* The accession number for the 16S rRNA gene sequence is HQ436503. The type strain is 03SU10-P^T (=KCTC 23339^T = JCM 17520^T = DSM 29515^T).

4.2.2.2.3.8. Description of *Falsiroseomonas algicola* comb. nov.

Falsiroseomonas algicola (al.gi'co.la. L. fem. n. *alga* (*gen. algae*), an alga; L. masc./fem. suff. -*cola*, dweller; from L. masc./fem. n. *incola*, an inhabitant; N.L. masc./fem. n. *algicola*, an inhabitant of algae).

Basonym: *Roseomonas algicola* (Kim et al., 2020)

The description is similar as determined by Kim et al. (2020) for *Roseomonas algicola.* MK342491 and JAAIKB00000000 is the accession number for the 16S rRNA gene and genome sequence, respectively. The type strain is PeD5^T (=JCM 33309^T = KACC 19925^T).

4.2.2.2.3.9. Description of *Falsiroseomonas arcticisoli* comb. nov.

Falsiroseomonas arcticisoli (arc'ti.cus arc.ti.ci.so'li L. masc. adj. *arcticus*, northern; L. neut. n. *solum*, soil; N.L. gen. n. *arcticisoli*, referring to soil from the Arctic).

Basonym: *Roseomonas arcticisoli* (Kim et al., 2016)

The description is similar to that described by Kim et al. (2016) for *Roseomonas arcticisoli.* The accession number for the 16S rRNA gene sequence is KP274055. The type strain is MC 3624^T (=CCTCC AB 2014278^T = LMG 28637^T).

4.2.2.2.3.10. Description of *Falsiroseomonas bella* comb. nov.

Falsiroseomonas bella (bel'la. L. fem. adj. *bella*, pretty/beautiful).

Basonym: *Roseomonas bella* (Zhang et al., 2020)

The description is similar as determined by Zhang et al. (2020) for *Roseomonas bella*. MG601506 and QGNA00000000 is the accession number for the 16S rRNA gene and genome sequence, respectively. The type strain is CQN31^T (=KCTC 62447^T= MCCC 1H00309^T).

4.2.2.2.4. Description of *Paeniroseomonas* gen. nov.

Paeniroseomonas (Pa.e.ni.ro.se.o.mo'nas. L. adv. *paene*, almost; - i-, connecting vowel; N.L. fem. n. *Roseomonas*, name of bacterial genus; N.L. fem. n. *Paeniroseomonas*, almost a *Roseomonas*).

Aerobic and Gram-negative with coccoid-rod shaped bacteria. The motility of cells varies. Catalase is positive, oxidase is negative. The principal polar lipids include diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, and phosphatidylcholine. The primary fatty acids are $C_{18:1}$7c, $C_{18:1}$2-OH, and $C_{16:0}$ and summed feature 3. 68.6–73.1 % G+C content (mol %). The phylogeny of the 16S rRNA gene and phenotypic characteristics are used to define the genus.

The type species is *Paeniroseomonas aquatica.*

4.2.2.2.4.1. Description of *Paeniroseomonas aquatica* comb. nov.

Paeniroseomonas aquatica (a.qua'ti.ca L. fem. adj. *aquatica*, isolated from water, aquatic).

Basonym: *Roseomonas aquatica* (Gallego et al., 2006)

The description is similar as determined by Gallego et al. (2006) for *Roseomonas aquatica.* The accession number for the 16S rRNA gene sequence is AM231587. The type strain is TR53^T (=CECT 7131^T = JCM 13556^T = DSM 19438^T).

4.2.2.2.4.2. Description of *Paeniroseomonas fluminis* comb. nov.

Paeniroseomonas fluminis (flu'mi.nis L. gen. neut. n. *fluminis*, of/from a river).

Basonym: *Roseomonas fluminis* (Ko et al., 2018)

The description is similar as determined for *Roseomonas fluminis* by Ko et al. (2018). The accession number for the 16S rRNA gene sequence is KY649439. The type strain is D3^T (=JCM 31968^T= KACC 19269^T).

4.2.2.2.5. Description of *Plastoroseomonas* gen. nov.

Plastoroseomonas (Plas.to.ros.e.o.mo'nas. Gr. adj. *plastos*, false; N.L. fem. n. *Roseomonas*, a bacterial genus; N.L. fem. n. *Plastoroseomonas*, a false *Roseomonas*).

The genus contains aerobic and Gram-negative with rod-shaped cells. Motility is the property of members. Catalase is positive and oxidase is negative. The principal polar lipids are diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, unidentified amino lipid, and an unidentified lipid. The major fatty acids are summed feature 3, $C_{18:1}$2-OH and $C_{16:0}$. Genomic size ranges from 4.4 to 7.2 Mbp, whereas G+C mol % is 70 mol %. The 16S rRNA gene, 92 core genes (phylogenomic), POCP values, AAI indices, and genomic and phenotypic traits are used to delineate the genus.

The type species is *Plastoroseomonas arctica.*

4.2.2.2.5.1. Description of *Plastoroseomonas arctica* comb. nov.

Plastoroseomonas arctica (arc'ti.ca. L. fem. adj. *arctica*, northern, from the Arctic, alluding to the place where the type strain was isolated).

Basonym: *Roseomonas arctica* (Qiu et al., 2016)

The description is similar to *Roseomonas arctica* as determined by Qiu et al. (2016). KJ647399 and JAAEDH000000000 are the accession numbers for the 16S rRNA gene and genome sequences, respectively. The type strain is M6-79^T (=CCTCC AB 2013101^T = LMG 28251^T).

4.2.2.2.5.2. Description of *Plastoroseomonas hellenica* comb. nov.

Plastoroseomonas hellenica (hel-lé-ni-ka. Gr. adj. *ellenikos*, Greek, L. fem. adj. *hellenica*, Greek, referring to Greece, the country from where the bacterium was firstly isolated).

Basonym: *Roseomonas hellenica* (Rat et al., 2021)

The description is similar to *Roseomonas hellenica* as determined by Rat et al. (2021). MN647549 and JAAGBB000000000 are the accession numbers for the 16S rRNA gene and genome sequences, respectively. The type strain is R-73080^T (=LMG 31523^T = CECT 30032^T).

4.2.2.2.6. Description of *Pseudoroseomonas* gen. nov.

Pseudoroseomonas (Pseu.do.ro.se.o.mo'nas. Gr. adj. *pseudês*, false; N.L. fem. n. *Roseomonas*, a bacterial genus; N.L. fem. n. *Pseudoroseomonas*, false *Roseomonas*).

Members are aerobic and Gram-negative. Cell morphology variable (ranges from coccoid to short rods). Cells are non-motile. Oxidase and catalase positive. The major polar lipids are diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, unidentified amino lipid, unidentified lipid, and unidentified phospholipid. The primary fatty acids are $C_{18:1}$2-OH and $C_{16:0}$, and summed feature 3. Members' genomic sizes range from 4.2 to 6.4 Mbp, with G+C content ranging from 68.7 to 72.7 %. The 16S rRNA gene, 92 core genes (phylogenomic), AAI indices, POCP values, and phenotypic and genomic traits are used to define the genus.

The type species is *Pseudoroseomonas cervicalis*.

4.2.2.2.6.1. Description of *Pseudoroseomonas cervicalis* comb. nov.

Pseudoroseomonas cervicalis [cer.vi.ca'lis. L. fem. n. *cervix* (*gen. cervicis*), neck; L. masc./fem. Adj. suff. -*alis*, suffix denoting pertaining to; N.L. fem. adj. *cervicalis*, referring to cervix, isolated from cervix].

Basonym: *Roseomonas cervicalis* (Rihs et al., 1998)

The description is similar to *Roseomonas cervicalis* determined by Rihs et al. (1993). AF533353 and ADVL00000000 are the accession number for the 16S rRNA gene and genome sequence, respectively. The type strain is E7107^T (=ATCC 49957^T = CIP 104027^T).

4.2.2.2.6.2. Description of *Pseudoroseomonas suffusca* comb. nov.

Pseudoroseomonas suffusca (suf.fus'ca. L. fem. adj. *suffusca*, light brown, alluding to the color of colonies).

Basonym: *Roseomonas suffusca* (Subhash and Lee, 2017)

The description is similar to *Roseomonas suffusca,* as determined by Subhash and Lee (2017). The accession number for the 16S rRNA gene sequence is LT009497. The type strain is S1^T(=KEMB 563-465^T = JCM 31176^T).

4.2.2.2.6.3. Description of *Pseudoroseomonas rubra* comb. nov.

Pseudoroseomonas rubra (ru'bra. L. fem. adj. *rubra*, red/rose-pink).

Basonym: *Roseomonas rubra* (Subhash et al., 2016)

The description is similar to *Roseomonas rubra* as determined by Subhash et al. (2016). The accession number for the 16S rRNA gene sequence is LT009499. The type strain is S5^T (=JCM 31177^T = KEMB 563-468^T).

4.2.2.2.6.4. Description of *Pseudoroseomonas hibiscisoli* comb. nov.

Pseudoroseomonas hibiscisoli (hi.bis.ci.so'li. N.L. masc. n. *Hibiscus*, Mugunghwa/Hibiscus syriacus; L. neut. n. *solum*, soil; N.L. gen. neut. n. *hibiscisoli*, of soil of a *Hibiscus,* the isolation source).

Basonym: *Roseomonas hibiscisoli* (Yan et al., 2017)

The description is similar to *Roseomonas hibiscisoli,* determined by Yan et al. (2017). The accession number for the 16S rRNA gene sequence is LT009499. The type strain is THG-N2.2^T (=KACC 18935^T = CCTCC AB 2016176^T).

4.2.2.2.6.5. Description of *Pseudoroseomonas rhizosphaerae* comb. nov.

Pseudoroseomonas rhizosphaerae (rhi.zo.sphae'rae. Gr. fem. n. *rhiza*, a root; Gr. fem. n. *sphaîra*, a ball, a sphere; N.L. fem. n. *rhizosphaera*, the rhizosphere; N.L. gen. fem. N. *rhizosphaerae*, of/from the rhizosphere).

Basonym: *Roseomonas rhizosphaerae* (Chen et al., 2014)

The description is similar to *Roseomonas rhizosphaerae* as determined by Chen et al. (2014). KC904962 and PDNU00000000 are the accession number for the 16S rRNA gene and genome sequence, respectively. The type strain is YW11^T (= KACC 17225^T = CCTCC AB2013041^T).

4.2.2.2.6.6. Description of *Pseudoroseomonas aestuarii* comb. nov.

Pseudoroseomonas aestuarii (aes.tu.a'ri.i. L. gen. n. *aestuarii*, of an estuary, the place from which the type strain was isolated).

Basonym: *Roseomonas aestuarii* (Venkata Ramana et al., 2010)

The description is similar to *Roseomonas aestuarii* as determined by Venkata Ramana et al. (2010). The accession number for the 16S rRNA gene sequence is AB682256. The type strain is JC17^T (=CCUG 57456^T = KCTC 22692^T = NBRC 105654^T).

4.2.2.2.6.7. Description of *Pseudoroseomonas aerofrigidensis* comb. nov.

Pseudoroseomonas aerofrigidensis (a.e.ro.fri.gi.den'sis. Gr. masc. n. *aêr*, air; L. masc. adj. *frigidus*, cold, cool, chilled; L. masc./fem. adj. suff. -*ensis*, suffixes used in the sense of "belonging to" /"coming from"; N.L. fem. adj. *aerofrigidensis*, referring to cooling air, isolation source from an air conditioner).

Basonym: *Roseomonas aerofrigidensis* (Hyeon and Jeon, 2017)

The description is similar to *Roseomonas aerofrigidensis* as determined by Hyeon and Jeon (2017). The accession number for the 16S rRNA gene sequence is KY126356. The type strain is HC1^T (=JCM 31878^T = KACC 19097^T).

4.2.2.2.6.8. Description of *Pseudoroseomonas oryzae* comb. nov.

Pseudoroseomonas oryzae (o.ry'zae. L. gen. fem. n. *oryzae*, of rice, referring to the source of isolation of the type strain, from rice paddy soil).

Basonym: *Roseomonas oryzae* (Ramaprasad et al., 2015)

The description is similar to *Roseomonas oryzae*, as determined by Ramaprasad et al. (2015). LN810637 and VUKA00000000 is the accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is JC288^T(=KCTC 42542^T= LMG 28711^T).

4.2.2.2.6.9. Description of *Pseudoroseomonas vastitatis* comb. nov.

Pseudoroseomonas vastitatis (vas.ti.ta'tis. L. gen. fem. n. *vastitatis*, of/from a desert, pertaining to the isolation source of the type strain).

Basonym: *Roseomonas vastitatis* (Zhao et al., 2020)

The description is similar to *Roseomonas vastitatis* given by Zhao et al. (2020). MK421542 and QXGS00000000 is the accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is CPCC 101021^T (=J1A743^T = KCTC 62043^T).

4.2.2.2.6.10. Description of *Pseudoroseomonas globiformis* comb. nov.

Pseudoroseomonas globiformis (glo.bi.for'mis. L. masc. n. *globus*, sphere; L. fem. n. *forma*, shape; N.L. masc./fem. adj. *globiformis*, spherical shape).

Basonym: *Roseomonas globiformis* (Fang et al., 2018)

The description is similar to *Roseomonas globiformis* as determined by Fang et al. (2018). The accession number for the 16S rRNA gene sequence is MG589944. The type strain is CPCC 100847^T (=KCTC 52099^T).

4.2.2.2.6.11. Description of *Pseudoroseomonas wenyumeiae* comb. nov.

Pseudoroseomonas wenyumeiae (wen.yu.mei'ae. N.L. gen. fem. n. *wenyumeiae*, of Yumei Wen, a famous microbiologist, for her contribution to the Hepatitis B vaccine and anti-HBs research and her fundamental role in immunology in China).

Basonym: *Roseomonas wenyumeiae* (Tian et al., 2019)

The description is similar to *Roseomonas wenyumeiae* as determined by Tian et al. (2019). MH974806 and RFLX00000000 is the accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is Z23^T (=CGMCC 1.16540^T = DSM 106207^T).

4.2.2.2.6.12. Description of *Pseudoroseomonas ludipueritiae* comb. nov.

Pseudoroseomonas ludipueritiae [lu.di.pu.e.ri'ti.ae. L. masc. n. *ludus*, a place of exercise /practice, a school for elementary instruction; L. masc. n. *puer (gen. pueri)*, a child; N.L. gen. n. *ludipueritiae*, of a playing place of childhood, a kindergarten].

Basonym: *Roseomonas ludipueritiae* (Kämpfer et al., 2003; Sánchez-Porro et al., 2009)

The description is similar to *Roseomonas ludipueritia* as determined by Kämpfer et al. (2003) and Sánchez-Porro et al. (2009). AJ488504 and JACTUZ0000000 is the

accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is 170-96^T (=CIP 107418^T= DSM 14915^T).

4.2.2.2.6.13. Description of *Pseudoroseomonas aerophila* comb. nov.

Pseudoroseomonas aerophila (a.e.ro'phi.la. Gr. masc. n. *aêr*, air; Gr. masc. adj. *philos*, loving; N.L. fem. adj. *aerophila*, air-loving).

Basonym: *Roseomonas aerophila* (Kim et al., 2013)

The description is similar to *Roseomonas aerophila* as determined by Kim et al. (2013). JX275860 and JACTVA00000000 is the accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is 7515T-07^T (=KACC 16529^T = NBRC 108923^T).

4.2.2.2.6.14. Description of *Pseudoroseomonas musae* comb. nov.

Pseudoroseomonas musae (mu'sae. L. gen. fem. n. *musae,* of Musa, isolated from leaf of banana *Musa* sp).

Basonym: *Roseomonas musae* (Nutaratat et al., 2017)

The description is same as determined by Nutaratat et al. (2013). AB594201 is the accession number for the 16S rRNA gene sequence. The type strain is PN1^T (=BCC 44863^T = NBRC 107870^T).

4.2.2.2.6.15. Description of *Pseudoroseomonas coralli* comb. nov.

Pseudoroseomonas coralli (co.ral'li. L. gen. n. *coralli*, of coral, from where the type strain was isolated).

Basonym: *Roseomonas coralli* (Li et al., 2021)

The description is similar to *Roseomonas coralli* as determined by Li et al. (2021). MN336179 is the accession number for the 16S rRNA gene sequence and SNVJ000000000 for the genome sequence. The type strain is M0104^T (=KCTC 62359^T = MCCC 1K03632^T).

4.2.2.2.6.16. Description of *Pseudoroseomonas deserti* comb. nov.

Pseudoroseomonas deserti (de.ser'ti. L. gen. neut. n. *deserti*, of/from a desert).

Basonym: *Roseomonas deserti* (Subhash and Lee, 2018)

The description is similar to *Roseomonas deserti* as determined by Subhash and Lee (2018). LT837512 and MLCO00000000 is the accession number for the 16S rRNA gene sequence and genome sequence, respectively. The type strain is M3^T (=KEMB 2255-459^T = JCM 31275^T).

4.2.2.2.7. Description of *Neoroseomonas* gen. nov.

Neoroseomonas (Ne.o.ro.se.o.mo'nas. Gr. adj. *neos*, new; N.L. fem. n. *Roseomonas*, a bacterial genus; N.L. fem. n. *Neoroseomonas*, referring to a novel group of *Roseomonas*).

Members of the genus are aerobic and Gram-staining negative. Cell shape are coccoid to short rod. Within the members, motility is variable. Positive for oxidase and catalase. The principal polar lipids are diphosphatidylglycerol, phosphatidylcholine, phosphatidylglycerol, phosphatidylethanolamine, unknown glycolipid, unknown amino lipid, unknown lipid, and unknown phospholipid. The primary fatty acids are $C_{18:1}$2-OH and $C_{16:0}$, and summed feature 3. Members' genome sizes range from 4.7 to 6.3 Mbp, with a G+C content of 68.8 to 71.5 %. The 16S rRNA gene, 92 core genes (phylogenomic), AAI indices, POCP values, and phenotypic and genomic traits were used to define the genus.

The type species is *Neoroseomonas lacus.*

4.2.2.2.7.1. Description of *Neoroseomonas lacus* comb. nov.

Neoroseomonas lacus (la'cus. L. gen. n. *lacus*, of a lake, referring to the site of isolation of the type strain).

Basonym: *Roseomonas lacus* (Jiang et al., 2006)

The description is similar to *Roseomonas lacus,* as determined by Jiang et al. (2006). AJ786000 and BMKW00000000 is the accession number for 16S rRNA gene sequence and genome sequence, respectively. The type strain is G33^T (=CGMCC 1.3617^T=JCM 13283^T).

4.2.2.2.7.2. Description of *Neoroseomonas terrae* comb. nov.

Neoroseomonas terrae (ter'rae. L. gen. n. *terrae*, of/from the soil).

Basonym: *Roseomonas terrae* (Yoon et al., 2007)

The description is similar to *Roseomonas terrae* as determined by Yoon et al. (2007). EF363716 is the accession number for the 16S rRNA gene sequence, whereas JAAEDI000000000 is the genome sequence accession number. The type strain is DS-48^T (=KCTC 12874^T=JCM 14592^T).

4.2.2.2.7.3. Description of *Neoroseomonas eburnea* comb. nov.

Neoroseomonas eburnea (e.bur'ne.a. L. fem. adj. *eburnea*, white color as ivory).

Basonym: *Roseomonas eburnea* (Wang et al., 2016)

The description is similar to *Roseomonas eburnea,* given by Wang et al. (2016). KF254767 is the 16S rRNA gene sequence accession, whereas JAAEDL000000000 is the accession number for genome sequence. The type strain is BUT-5^T(=CCTCC AB2013276^T=KACC 17166^T).

4.2.2.2.7.4. Description of *Neoroseomonas alkaliterrae* comb. nov.

Neoroseomonas alkaliterrae (al.ka.li.ter'rae. Arabic n. *al-qaliy*, the ashes of saltwort; L.n. *terra*, soil/earth; N.L. gen. n. *alkaliterrae*, of/from alkaline soil).

Basonym: *Roseomonas alkaliterrae* (Dong et al., 2014)

The description is similar to *Roseomonas alkaliterrae,* as Dong et al. (2014) determined. KF771274 and JACIJE00000000 are the accession number for 16S rRNA gene sequence and genome sequence, respectively. The type strain is YIM 78007^T (=BCRC 80644^T= JCM 19656^T).

4.2.2.2.7.5. Description of *Neoroseomonas oryzicola* comb. nov.

Neoroseomonas oryzicola (o.ry.zi'co.la. L. n. *oryza*, rice; L. masc./fem. suff. *-cola*, an inhabitant; from L. masc./fem. n. *incola*, dweller; N.L. fem. n. *oryzicola*, an inhabitant of rice).

Basonym: *Roseomonas oryzicola* (Chung et al., 2015)

The description is similar to *Roseomonas oryzicola,* as determined by Chung et al. (2015). EU707562 and JAAVUP00000000 are the accession number for 16S rRNA gene sequence and genome sequence, respectively. The type strain is YC6724^T (=KCTC 22478^T=NBRC 109439^T).

4.2.2.2.7.6. Description of *Neoroseomonas soli* comb. nov.

Neoroseomonas soli (so'li. L. gen. n. *soli*, of soil, the isolation source of the type strain).

Basonym: *Roseomonas soli* (Kim and Ka, 2014)

The description is similar to *Roseomonas soli* as determined by Kim and Ka (2014). The accession numbers for 16S rRNA and genome sequence are JN575264 and JAAEDM00000000, respectively. The type strain is 5N26^T(=KACC 16376^T=NBRC 109097^T).

4.2.2.2.7.7. Description of *Neoroseomonas sediminicola* comb. nov.

Neoroseomonas sediminicola (se.di.mi.ni.co'la. L. n. *sedimen -inis*, sediment; L. masc./fem. suff. -cola, inhabitant, dweller; N.L. fem. n. *sediminicola*, sediment-dweller, pertaining to the source of the type strain).

Basonym: *Roseomonas sediminicola* (He et al., 2019)

The description is similar to *Roseomonas sediminicola,* as detemined by He et al. (2014). The accession number for the 16S rRNA gene sequence is JQ349047. The type strain is FW-3^T (=KACC 16616^T= JCM 18210^T).

4.2.2.2.7.8. *Neoroseomonas* sp. JC162^T (=KCTC 32190^T)

4.2.2.2.7.8.1. Proposal of *Neoroseomonas marina* sp. nov

Strain KCTC 32190^T was distinct (Table 24) from its nearest phylogenetic neighbour *N. oryzicola* KCTC 22478^T with respect to organic substrate utilization, hydrolysis of casein, salinity tolerance, fatty acid and polar lipid profiling (Fig. 49) supported by the genotypic distinctiveness (ANI and *d*DDH). Hence, it justifies the description of strain KCTC 32190^T as representative of a novel species, for which the name *Neoroseomonas marina* sp. nov., is proposed.

4.2.2.2.7.8.2. Description of *Neoroseomonas marina* sp. nov.

Neoroseomonas marina (ma.ri'na. L. fem. adj. *marina*, of the sea, marine)

Four days old colonies on tryptone soya agar are small (1.0-1.5 mm diameter), smooth, mucoid, circular, convex and pink coloured. Cells range from coccoid to oval-shaped, 0.6-1.0 µm in width and 1.5-2.0 µm in length. Gram-stain-negative, non-spore forming and non-motility. NaCl is not obligatory and can tolerate up to 1 % (w/v). Mesophile (range 15-40°C). Optimum growth occurs at pH 7.0 (range 5.5-8.0). Casein is hydrolysed but not cellulose, starch or gelatin. Nitrate reduction and indole test negative. Urease, oxidase and catalase positive. Good growth occurs with rhamnose, melibiose, D-

glucose, arabinose, L-fucose, sucrose, D-galactose, D-mannitol, sorbitol, ribose and pyruvate. No growth was observed in citrate, acetate, malate, benzoate, glutamate, methanol, ethanol, aspartate, D-glutamic acid, starch, L-ascorbic acid, glycine, succinate, DL-aspartic acid, L-glutamine, tween 80 and D-cellobiose. Sensitive to kanamycin, tetracycline and rifampicin and resistant to carbenicillin and chloramphenicol. $C_{18:1}\omega7c/C_{18:1}\omega6c$, *cyclo*-$C_{19:0}\omega8c$, $C_{18:0}$ 2-OH, $C_{16:0}$, $C_{18:0}$ 3-OH, $C_{16:1}\omega7c/C_{16:1}\omega6c$, $C_{16:0}$ 2-OH and $C_{16:1}\omega5c$ are the cellular fatty acids. Phosphatidylglycerol, diphosphotidylglycerol, phosphatidylethanolamine, phosphatidylcholine, an unidentified amino lipid (AL) and three unidentified lipids (L1-3) are the polar lipids. Ubiquinone-10 is the major quinone system. Genomic DNA G+C content of the strain KCTC 32190^T is 70.9 mol %.

The type strain KCTC 32190^T (= CGMCC1.12364^T) was isolated from a beach sand sample collected below Pamban bridge, Rameshwaram, Tamil Nadu, India. The GenBank accession number for the 16S rRNA gene sequence of *Roseomonas marina* JC162^T (=KCTC 32190^T) is HE984358. The draft genome has been archived at DDBJ/GenBank/ENA under the accession number JABBKX000000000 for strain KCTC 32190^T. The version described in this study is JABBKX010000000.

4.2.3. Phylogenomic-based reappraisal of the genus *Alcanivorax*

Strain JC109^T is a member of *Alcanivorax xenomutans*, in the family *Alcanivoracaceae* of class *Gammaproteobacteria*. *Alcanivorax* members play an important role in bioremediation processes. Therefore genomic analysis of these hydrocarbonoclastic marine bacteria will allow us to better understand their alkane metabolic capacities. *Alcanivorax xenomutans* JC109^T was isolated from a shrimp cultivation pond described from our lab. This taxon is the most abundantly found species in marine habitats on India's east and west coasts (unpublished data). Over the past few years, the isolation and identification of these members of the genus *Alcanivorax* have been of paramount interest for its biodegradation based applicative purposes. This has simultaneously resulted in the increased description of members of *Alcanivorax* and other genera of the family *Alcanivoraceae*. Therefore, it necessitates congruent phylogenetic analysis and taxonomic description based on the recent and advanced delineation criteria for evading unwanted and inconsistencies in the future. With the advancement of molecular procedures and sequencing techniques, canonical classification of bacterial taxa based on genome sequence data (in contrast to single genes) has been highly recommended (Chun and Rainey 2014; Parks et al. 2018). Therefore, in this study we have meticulously re-evaluated the taxonomic standing of the members of the genus *Alcanivorax* and deduced its ranking based on the modern approaches.

The taxonomic standing of the members of the genus *Alcanivorax* was firstly reviewed by 16S rRNA gene marker (Table 26), overall genome-related index (OGRI) like ANI and *d*DDH (Table 28) for species delineation (Chun et al. 2018). Hence, calculated values of these two-genome metrics showed that all the species of the genus *Alcanivorax* are congruous and well-reported at the species level (cut-off values discussed in Section 4.2.2). However, based on the 16S rRNA gene based dendrogram, these members were

divided into three major clades (I, II and III) (Fig. 50). The 16S rRNA gene identity between the some of members of the genus *Alcanivorax* was <94 %, indicating that they might belong to different genera as <95 % is the standard threshold specified for genus differentiation (Rosselló-Móra and Amann 2015; Yarza et al. 2014). However, the 16S rRNA gene identity should not be the only criterion for taxa delineation as it can be highly ambiguous (Chan et al. 2012; Olm et al. 2020). The phylogenomic tree based on the genomes of the species of *Alcanivorax* also indicated the same cladding order (Fig. 51). In this study, clade III members had the highest genome size and G+C content (mol%) (Table 27). The genome phylogeny should be further supported by whole genome metrics. Therefore, at the protein level, AAI and POCP at higher taxonomic rank have been used as an effective tool for enhancing taxa delineation. This is advantageous in terms of low identity in a limited orthologous gene between genomes of distant members of genera. In our analysis, the AAI and POCP (Table 29, Fig. 52A, B) values between inter-clades of the genus *Alcanivorax* were <69.0 % and <68.3 %, respectively. Hence, the AAI values are in accordance with the to the standards as given by Luo et al. (2014). As for POCP, later several studies showed that the given cut-off of 50 % to be highly conserved (Liu et al. 2020) and has been applied only as relatedness for many genera.

The evaluation of core and pan-genome studies summarized that the pan-genome of the genus *Alcanivorax* is open as the plot did not plateau with the addition of genomes. Altogether, the members had 1223 core genes aggregating up to 36.5 % of the sum-total pan-genome (Table 30, Fig. 53). Interestingly, with the group-wise study of the species, there was a drastic rise in the number of core genes (Table 30). This clearly indicated the divergence between the inter-clade members. The protein sequence identity of clade II and clade III members with *A. borkumensis* SK2^{T} (type species) were lower (65-70 %) as compared to its own clade members (70-80) (Fig. 54). Further, studies on CSIs showed that

there were three CSI signatures specifically present to each of the clades (Fig. 55), therefore separating clades from each other. These CSIs include insertion or deletion into proteins which are remarkable and can be linked to evolutionary linked members (Gupta 2006). Over the years, many taxa like the genus *Bacillus* (Patel and Gupta 2020), *Mycobacterium* (Gupta et al. 2018), and order *Pasteurellales* have been re-ordered with the aid of CSI (Naushad et al. 2014).

The genomic annotation supported aforementioned and revealed that all the members of clades consisted of genes required in the aerobic alkane degradation (Table 31). The annotation results showed that the genes responsible for the larger *n*-alkane degradation were part of the terminal oxidation (Rojo 2009) pathway. Our reclassification has accounted phenotypic characters like motility, flagellar arrangement and fatty acids.

4.2.3.1. Ecology of the genus *Alcanivorax*

The species of the genus *Alcanivorax* are ubiquitous, indispensable and spanning an extensive domain of marine habitats. *Alcanivorax* bacterial groups are capable of degrading hydrocarbon pollutants, thus becoming a major area of research for their bio-remediation applications (Chernikova et al. 2020). From an ecological overview, the species of genus *Alcanivorax* are chiefly isolated from habitats like sea water, oil polluted sea water, saline pond, thermal vents and marine sediments (Table 2). Thus, causing their distribution to be confined to restricted areas and simultaneously allowing them to adapt (loss/gaining of genes) in accordance to its environment.

4.2.3.2. Proposal to reclassify the genus *Alcanivorax* into two novel genera

The detailed phylotaxogenomic studies along with the other phenotypic data show that the genus under study is highly divergent (Table 33). Therefore, members should not be retained under a singular genus and justifies the division of the genus *Alcanivorax* into

three genera as shown earlier (Fig. 50, 51). The emended description based on reclassification for each of the genera are presented. Hereby, clade II has been reclassified as *Isoalcanivorax* gen. nov. and clade III as *Alloalcanivorax* gen. nov., along with the emended description of the genus *Alcanivorax sensu stricto* (Clade I).

4.2.3.2.1. Emended description of the genus *Alcanivorax* by Yakimov et al. 1998

For the members of the genus *Alcanivorax*, the description is similar as given by Yakimov et al. (1998) except for a few changes. The cell size of the members ranged from 0.2-0.8×0.2-2.5 μm. Oxidase positive and catalase variable. Nitrate reduction is variable. Major polar lipids are phosphatidylethanolamine, phosphatidylglycerol, unknown polar lipid, unknown phospholipid and unknown aminolipid. Summed feature 3, $C_{16:0}$, $C_{12:0}$ 3-OH, $C_{12:0}$, $C_{14:0}$ are the fatty acids.

The type species is *Alcanivorax borkumensis*.

4.2.3.2.2. Description of *Isoalcanivorax* gen. nov.

Isoalcanivorax (Iso.al.ca.ni.vo'rax. Gr. adj. *isos*, equal; N.L. masc. n. *Alcanivorax*, a bacterial genus; N.L. masc. n. *Isoalcanivorax*, like the genus *Alcanivorax*).

Members are aerobic, Gram-stain-negative and rod-shaped. Variably motile. Oxidase and catalase positive. Major polar lipids consisted of phosphatidylethanolamine and an unidentified phospholipid. The major fatty acids are summed features 8, 3 and $C_{16:0}$. Members consists of a genome size of around 3.5-4.2 Mbp and G+C content of 62.6-62.8 %. The genus delineation is based on the 16S rRNA gene, 92 core genes (phylogenomics), AAI indices, POCP values, phenotypic and other genomic features.

The type species is *Isoalcanivorax pacificus*

4.2.3.2.2.1. Description of *Isoalcanivorax pacificus* comb. nov.

Isoalcanivorax pacificus (pa.ci'fic.us. L. masc. adj. *pacificus*, pacific, related to the Pacific Ocean).

Basonym: *Alcanivorax pacificus* (Lai et al., 2011)

The description is similar as that of *Alcanivorax pacificus* given by Lai et al. (2011). DQ659451 is the accession number for the 16S rRNA gene sequence and CP004387 for the genome sequence. The type strain is W11-5^T (=MCCC 1A00474^T = CCTCC AB 208236^T = LMG 25514^T).

4.2.3.2.2.2. Description of *Isoalcanivorax indicus* comb. nov.

Isoalcanivorax indicus (in'di.cus. L. masc. adj. *indicus*, related to the Indian Ocean, wherein type strain was isolated).

Basonym: *Alcanivorax indicus* (Song et al., 2018)

The description is similar to the one as determined by Song et al. (2018) for *Alcanivorax indicus.* The accession numbers of 16S rRNA gene sequence and genome sequence are MH279472 and QGMP00000000, respectively. The type strain is SW127^T (=CGMCC 1.16233^T =KCTC 62652^T).

4.2.3.2.3. Description of *Alloalcanivorax* gen. nov.

Alloalcanivorax (Al.lo.al.ca.ni.vo'rax. Gr. masc. adj. *allos*, other; N.L. masc. n. *Alcanivorax*, a bacterial genus; N.L. masc. n. *Alloalcanivorax,* like the genus *Alcanivorax*).

Members are aerobic, Gram-stain-negative, rod-shaped and motile. Oxidase and catalase positive. Polar lipids consisted of phosphatidylglycerol, unidentified aminolipid and unidentified phospholipid. The major fatty acids are summed features 8, 3 and $C_{16:0}$. Members have a genomic size of around 3.5-5.2 Mb and a genomic G+C content of 61.3-66.4 %. The genus delineation is based on the 16S rRNA gene, 92 core genes (phylogenomics), AAI indices, POCP values, phenotypic and other genomic features.

The type species is *Alcanivorax venustensis.*

4.2.3.2.3.1. Description of *Alloalcanivorax venustensis* comb. nov.

Alloalcanivorax venustensis (ve.nust.en'sis. L. masc. adj. *venustensis*, from 'Portus Venustus', Elegant Port, is the ancient Latin names of Santa Pola in Roman times, a coastal town south of Alicante, the closest town to the sites from which the two strains of the species were isolated. It might also allude to the elegant/slender aspect of the rods).

Basonym: *Alcanivorax venustensis* (Fernández-Martínez et al., 2003)

The description of the species is similar as determined for *Alcanivorax venustensis* by Fernández-Martínez et al. (2003). The accession number of 16S rRNA gene sequence and genome sequence are AF32876 and ARXR00000000, respectively. The type strain is ISO4^T (=CECT 5388^T =DSM 13974^T).

4.2.3.2.3.2. Description of *Alloalcanivorax dieselolei* comb. nov.

Alloalcanivorax dieselolei (die.sel.o'le.i. N.L. masc. n. *dieselius*, latinized family name of Rudolf Diesel; L. gen. n. *olei*, of oil; N.L. gen. n. *dieselolei*, of Diesel oil, the carbon substrate used in the isolation process of the two strains).

Basonym: *Alcanivorax dieselolei* (Liu and Shao, 2005)

The description is the same as given for *Alcanivorax dieselolei* by Liu and Shao (2005). MT759958 and CP003466 are the accession numbers for 16S rRNA gene and genome sequence, respectively. The type strain is B-5^T (=CGMCC 1.3690 ^{T} =DSM 16502 ^{T}).

4.2.3.2.3.3. Description of *Alloalcanivorax xenomutans* comb. nov.

Alloalcanivorax xenomutans (xe.no.mu'tans. Gr. masc. n. *xenos*, foreign; L. pres. part. *mutans*, transforming, converting; N.L. part. adj. *xenomutans*, transforming foreign or xenobiotic compounds).

Basonym: *Alcanivorax xenomutans* (Rahul et al., 2014)

The description for the species is similar as determined by Rahul et al. (2014) for *Alcanivorax xenomutans.* The 16S rRNA and genome sequence accession numbers are HE601937 and OBMO00000000, respectively. The type strain is JC109^T (=KCTC 23752^T =NBRC 108843^T).

4.2.3.2.3.4. Description of *Alloalcanivorax balearicus* comb. nov.

Alloalcanivorax balearicus (ba.le.a'ri.cus. N.L. masc. adj. *balearicus*, of the Balearic Islands, source of isolation).

Basonym: *Alcanivorax balearicus* (Rivas et al., 2007)

The description for the species is similar as determined by Rivas et al. (2007) for *Alcanivorax balearicus.* AY686709 is the 16S rRNA gene sequence accession number. The type strain is MACL04^T (=CECT 5683^T =DSM 23776^T =LMG 22508^T).

4.2.3.2.3.5. Description of *Alloalcanivorax mobilis* comb. nov.

Alloalcanivorax mobilis (mo'bi.lis. L. masc. adj. *mobilis*, motile).

Basonym: *Alcanivorax mobilis* (Yang et al., 2018)

The description is the same as based on *Alcanivorax mobilis* given by Yang et al. (2018). The accession number for 16S rRNA gene sequence is MF187461 whereas for genome sequence is NMQZ00000000. The type strain is MT13131^T (=KCTC 52985^T =MCCC 1A11581^T).

4.2.3.2.3.6. Description of *Alloalcanivorax profundamaris* comb. nov.

Alloalcanivorax profundimaris (pro.fun.di.ma'ris. L. masc. adj. *profundus*, deep; L. gen. n. *maris*, from/of the sea; N.L. gen. n. *profundimaris*, from/of the deep-sea, source from where the type strain was isolated).

Basonym: *Alcanivorax profundimaris* (Dong et al., 2021)

The description is similar as determined for *Alcanivorax profundimaris* by Dong et al. (2021). The accession number for the 16S rRNA gene sequence is MT452298 whereas for genome sequence is JABJWH00000000. The type strain is ST75FaO-1^T (=KCTC 82142^T =MCCC 1A17714^T).

4.2.3.2.3.7. Description of *Alloalcanivorax gelatiniphagus* comb. nov.

Alloalcanivorax gelatiniphagus (ge.la.ti.ni.pha'gus. N.L. neut. n. *gelatinum*, gelatin; Gr. masc./suff.-*phagos*, glutton; N.L. masc. adj. *gelatiniphagus*, gelatin eater).

Basonym: *Alcanivorax gelatiniphagus* (Kwon et al., 2015)

The description is similar as given for *Alcanivorax gelatiniphagus* by Kwon et al. (2015). JQ937289 is the accession number for the 16S rRNA gene sequence and VCQT00000000 for the genome sequence. The type strain is MEBiC08158^T (=JCM 18425^T =KCCM 42990^T).

4.2.3.2.3.8. Description of *Alloalcanivorax marinus* comb. nov.

Alloalcanivorax marinus (ma.ri'nus. L. masc. adj. *marinus*, of the sea/marine).

Basonym: *Alcanivorax marinus* (Lai et al., 2013)

The description of the species is similar as determined for *Alcanivorax marinus* by Lai et al. (2013). The accession number of 16S rRNA gene sequence is KC415169. The type strain is R8-12^T (=CCTCC AB 208234^T =LMG 24621^T =MCCC 1A00382^T).

Note: A new notification by Oren and Garrity (2022) has been released wherein new names have been assigned for the phylum of *Actinobacteria, Acidobacteria, Armatimonadetes, Bacteroidetes, Chloflexi, Gemmatimonadetes, Planctomycetes* and *Verrucomicrobia* as *Actinomycetota, Acidobacteriota, Armatimonadota, Bacteroidota, Chloroflexota, Gemmatimonadota, Planctomycetota* and *Verrucomicrobiota*, respectively. However, these changes have not been incorporated as this development has happened after the completion of thesis work.

4.2.4. Notes on bacterial taxonomy

We understand that standard approaches such as 16S rRNA gene sequence and DDH alone are not superior for determining phylogenetic relationships. The polyphasic data supported by the phylogenomic studies is considered ideal for taxa delineation in the present-day. This improved precision and inclusivity of genome annotations (from genotypes to phenotypes) is allowing researchers to gain more detailed insights into the biology of species, communities, and individuals. Furthermore, the genomes reflect the organism's biology as well as its evolutionary lineage pattern, preventing the creation of redundant species. Therefore, genome-based taxonomy becomes increasingly important for separating closely related species and revealing species-specific characteristics. Such a system of taxonomy is integrated taxonomy. For effective identification, new technologies and study systems are being developed and tested. As a potential prospect for the future, the following ambitious systems/concepts have been visioned. It is a supposition which will allow gaps and systemic limitations to be overcome in taxonomic studies. The future of the taxonomy can be summarized as follows:

i) Systems taxonomy: Despite the fact that integrated taxonomy provides the genetic framework for delineation, however system-based taxonomy is the most recommended and prudent classification system. It is a long-term and ambitious aim is to conduct comprehensive taxonomy investigations that include genome, proteome, and metabolomics, as well as other components. In fact, it would take into account all of the elements that influence the bacteria's ability to survive. It is not too far away, owing to significant technological advancements in the development of system-based taxonomy, which includes a multidisciplinary area for a complete yet exact functional taxonomy (Fig. 56). Furthermore, systems-based taxonomy encompasses every consolidated component of systems biology that has an impact on the organism's survival.

ii) Virtual taxa: It is widely accepted that just 1 % of the total microbial wealth is known to be under cultivation, with the rest remaining unexplored (Nayfach et al. 2021; Sutcliffe et al. 2021). Metagenomics has also bolstered scientific belief in the existence of untapped microbial wealth. As a result, creating ranks using virtual taxon would be the pressing need of the hour. The identification and classification of bacteria based on their genome by annotating their cellular characteristics can lead in the creation of a virtual taxa. It is a holistic approach of integrating genomics to create virtual taxa, which would in return represent a modelled species with all its phenotypic, metabolic and physiological characteristics. It will also strengthen/enhance the taxonomy of the uncultured or "*Candidatus*" species status in particular, allowing virtual taxon to be strongly connected. It's also on par with "*Candidatus*" species rank because cultivating them is difficult. Further, the metagenomic assisted genome (MAG) and single genome assisted genome (Bowers et al. 2017) can also be genome sequenced to frame the organism's virtual taxa. In addition, the SeqCode developed by Brian et al. (2020) (under review) applies genomic sequence data as a common method for typifying cultivated and uncultivated microorganisms, and it adheres to the ICNP's precepts by following comparable priority principles

In general, identifying and assigning taxa to the microorganisms under study would be a holistic approach. Despite the fact that microbial transcriptomics, proteomics and metabolomics are varied and dynamic in nature under various environmental situations, they nonetheless play an important part in the study of the bacterial physiological activities. All of these aspects have so far been ignored when delineating taxa, and as a result, we have been unable to comprehend key evolutionary convergences and divergences that lead to speciation.

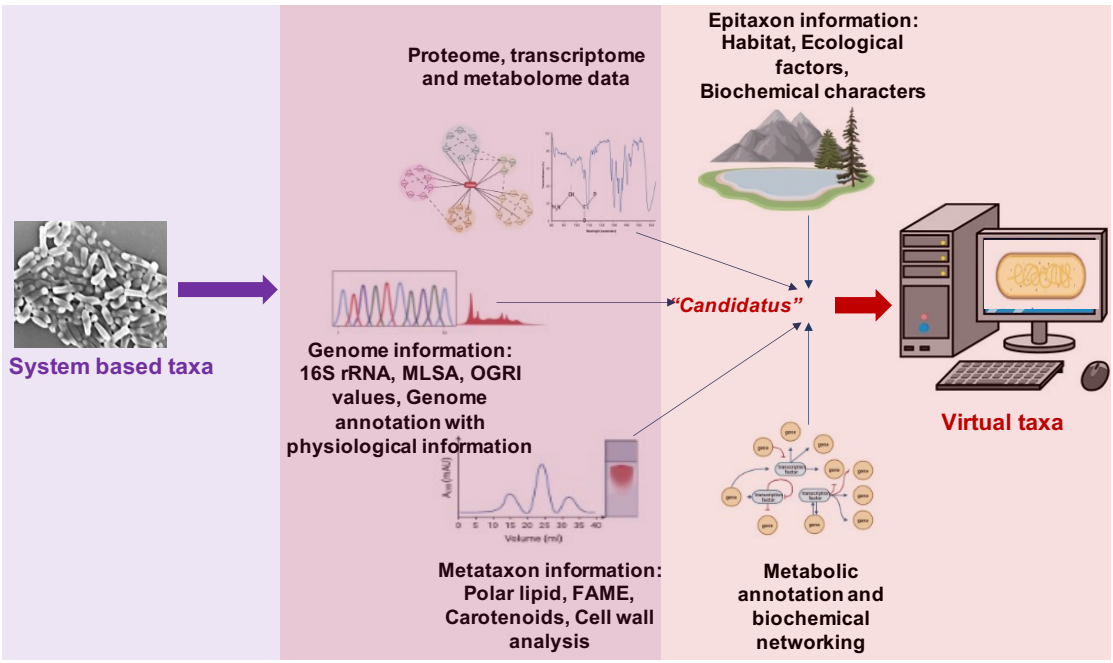


Fig. 56. Various components of "omics" based studies integrated to form systems-based taxa and virtual taxa

As a result, such concepts like the cell's architecture (shape/size), biochemical activities, membrane components, meta-taxon information (polar lipids/quinones, etc.), and other physiological activities (growth mode/ respiration/ reproduction/ energy metabolism) would all be taken into account, as well as its genomic fingerprints for effective application of taxonomy. Therefore, taxonomic studies are based on assets rather than liabilities. The system-based taxonomy and virtual taxa can be widely adapted in the future due to its functionality/applicability. It may be a slow shift, but it certainly aids in obtaining a larger picture by concatenating its DNA with its current state of being (phenotype) as well as numerous environmental circumstances. On the other hand, creating virtual taxa related to taxonomy is a liability that is only employed to help taxonomists grasp the intrinsic quality of the uncultivated state. It's a tool for visualization. In spite of the fact that many ways have been established and are continuously being developed, there is an unending demand for a brisk and efficient means of identifying bacteria from various sources.

MAJOR FINDINGS

5. MAJOR FINDINGS

- This study has enriched the existing knowledge on the number of bacterial phyla, with the addition of value to the previously guesstimated numbers. These "*Candidatus*" phyla's "probably" might be endemic to epiphytic orchids owing to their ubique habitat however, it needs further validation.
- Cultivation studies of bacteria from orchid root microbiome revealed that isolated members belonged primarily to *Proteobacteria* and *Actinobacteria*, among which, *Paracoccus aeridis* and *Microbacterium indolicus* are described as new species.
- Among the array of indole derivatives produced by the two novel species, only a few are identified while a large number of these remain still unidentified.
- *Mesobacillus aurantius* sp. nov. is identified as a novel strain for producing zeaxanthin and its derivatives.
- The phylotaxogenomic studies helped in the reappraisal of the members of genus *Roseomonas* and *Alcanivorax* into six and two novel genera, respectively.
- The following reclassifications were proposed:
  - Genus *Roseomonas* is reclassified into six novel valid genera names: *Pararoseomonas* gen. nov., *Falsiroseomonas* gen. nov., *Paeniroseomonas* gen. nov., *Plastoroseomonas* gen. nov., *Pseudoroseomonas* gen. nov., and *Neoroseomonas* gen. nov. and emended description of *Roseomonas sensu stricto. "Neoroseomonas marina"* JC162^{T} sp. nov is described as a novel member of the genus *Neoroseomonas.*
  - Genus *Alcanivorax* is reclassified into two novel genera: "*Isoalcanivorax"* gen. nov., and "*Alloalcanivorax"* gen. nov. with emended description of *Alcanivorax sensu stricto.*
- Based on the experience gained and owing to the rapid development in the field of bacterial taxonomy, a "note" is proposed on the future vision on bacterial taxonomy.

REFERENCES

REFERENCES

Abram K, Udaondo Z, Bleker C, Wanchai V, Wassenaar TM, Robeson MS et al (2021) Mash-based analyses of *Escherichia coli* genomes reveal 14 distinct phylogroups. Commun Biol 26: 117. doi: 10.1038/s42003-020-01626-5

Abraham BS, Caglayan D, Carrillo NV, Chapman MC, Hagan CT, Hansen ST et al (2020) Shotgun metagenomic analysis of microbial communities from the Loxahatchee nature preserve in the Florida Everglades. Environ Microbiome 15:2. doi: 10.1186/s40793-019-0352-4

Afzal I, Shinwari ZK, Sikandar S, Shahzad S (2019) Plant beneficial endophytic bacteria: mechanisms, diversity, host range and genetic determinants. Microbiol Res 221:36–49. 10.1016/j.micres.2019.02.001

Aherfi S, Andreani J, Baptiste E, Oumessoum A, Dornas FP, Andrade ACDSP et al (2018) A large open pangenome and a small core genome for giant pandoraviruses. Front Microbiol 9:1486. doi: 10.3389/fmicb.2018. 01486

Ahmed I, Yokota A, Fujiwara T (2007) A novel highly boron tolerant bacterium, *Bacillus boroniphilus* sp. nov., isolated from soil, that requires boron for its growth. Extremophiles 11:217-24. doi: 10.1007/s00792-006-0027-0

Alarico S, Rainey FA, Empadinhas N, Schumann P, Nobre MF, da Costa MS et al (2002) *Rubritepida flocculans* gen. nov., sp. nov., a new slightly thermophilic member of the alpha-1 sub-class of the *Proteobacteria*. Syst Appl Microbiol 25:198–206. doi: 10.1078/0723-2020-00116

Allali I, Arnold JW, Roach J, Cadenas MB, Butz N, Hassan HM et al (2017) A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. BMC Microbiol 17:194. doi:org/10.1186/s12866-017-1101-8

Alibrandi P, Schnell S, Perotto S, Cardinale M (2020) Diversity and structure of the endophytic bacterial communities associated with three terrestrial orchid species as revealed by 16S rRNA gene metabarcoding. Front Microbiol 11:604964. doi: 10.3389/fmicb.2020.604964

Aliyu H, Lebre P, Blom J, Cowan D, De Maayer P (2016) Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. Syst Appl Microbiol. 39:527-533. doi: 10.1016/j.syapm.2016.09.004

Almpanis A, Swain M, Gatherer D, McEwan N (2018) Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. Microb Genom 4: e000168. doi:10.1099/mgen.0.000168

Anil Kumar P, Srinivas TN, Takaichi S, Maoka T, Sasikala Ch, Ramana ChV (2009) *Phaeospirillum chandramohanii* sp. nov., a phototrophic alphaproteobacterium with carotenoid glycosides. Int J Syst Evol Microbiol 59:2089-93. doi: 10.1099/ijs.0.009019-0

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. doi:10.1016/S0022-2836(05)80360-2

Ankenbrand MJ, Keller A (2016) bcgTree: automatized phylogenetic tree building from bacterial core genomes. Genome 59:783-791. doi: 10.1139/gen-2015-0175

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-9. doi: 10.1038/75556

Ashford RT, Muchowski J, Koylass M, Botstein D, Butler H, Cherry JM et al (2020) Application of whole genome sequencing and pan-family multi-locus sequence analysis to characterize relationships within the family *Brucellaceae*. Front Microbiol 14:11-1329. doi: 10.3389/fmicb.2020.01329

Auch AF, von Jan M, Klenk HP, Göker M (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci 2:117-134. doi:10.4056/sigs.531120

Aurass P, Karste S, Trost E, Glaeser SP, Kämpfer P, Flieger A (2017) Genome sequence of *Paracoccus contaminans* LMG 29738^T, isolated from a water microcosm. Genome Announc 5:e00487-17. doi: 10.1128/genomeA.00487-17

Awolope OK, O'Driscoll NH, Di Salvo A, Glaeser SP, Kämpfer P, Flieger A (2021) De novo genome assembly and analysis unveil biosynthetic and metabolic potentials of *Pseudomonas fragi* A13BB. BMC Genom Data 15. doi:org/10.1186/s12863-021-00969-0

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA et al (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. doi: 10.1186/1471-2164-9-75

Bajerski F, Wagner D, Mangelsdorf K (2017) Cell membrane fatty acid composition of *Chryseobacterium frigidisoli* pb4t, isolated from Antarctic glacier forefield soils, in response to changing temperature and ph conditions. Front Microbiol 19:677. doi: 10.3389/fmicb.2017.00677

Baik KS, Park SC, Choe HN, Kim SN, Moon JH, Seong CN (2012) *Roseomonas riguiloci* sp. nov., isolated from wetland freshwater. Int J Syst Evol Microbiol 62:3024–3029. doi: 10.1099/ijs.0.036186-0

Bale NJ, Sorokin DY, Hopmans EC, Koenen M, Rijpstra WIC, Villanueva L et al (2019) New insights into the polar lipid composition of extremely halo(alkali)philic euryarchaea from hypersaline lakes. Front Microbiol 12:10-377. doi: 10.3389/fmicb.2019.00377

Belin BJ, Busset N, Giraud E, Molinaro A, Silipo A, Newman DK (2018) Hopanoid lipids: from membranes to plant-bacteria interactions. Nat Rev Microbiol 16:304-315. doi: 10.1038/nrmicro.2017.173

Bennett JW (2011) Microbiology in the 21st Century. In: National Research Council (US) Board on Research Data and Information; Uhlir PF, editor. Designing the Microbial Research Commons: Proceedings of an International Symposium. Washington (DC): National Academies Press (US)

Bertelli C, Laird MR, Williams KP (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45:W30–W35. doi: 10.1093/nar/gkx343

Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin Microbiol Infect 24:335-341. doi: 10.1016/j.cmi.2017.10.013

Bharti R, Grimm DG (2021) Current challenges and best-practice protocols for microbiome analysis. Brief Bioinform 22:178-193. doi: 10.1093/bib/bbz155.

Biebl H, Pfennig N (1981) Isolation of members of the family *Rhodospirillaceae*. In: Starr MP, Stolp H, Truper HG, Balows A, Schlegel HG (editors). The Prokaryotes: a Handbook on Habitats, Isolation and Identification of Bacteria, 1:pp. 267–273.

Bird E (2000) Coastal geomorphology: An Introduction, Chichester, England, John Wiley & Sons, Ltd, 322 pp.

Bishop DH, Pandya PK, King HK (1962) Ubiquinone and vitamin K in bacteria. Biochem J 83:606-14. doi: 10.1042/bj0830606

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20. doi: 10.1093/bioinformatics/btu170

Bowers R, Kyrpides N, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK et al (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. doi.org/10.1038/nbt.3893

Blevins SM, Bronze MS (2010) Robert Koch and the 'golden age' of bacteriology. Int J Infect Dis 14:e744-51. doi: 10.1016/j.ijid.2009.12.003

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY et al (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81-W87. doi: 10.1093/nar/gkz310

Bradley PH, Pollard KS (2017) Proteobacteria explain significant functional variability in the human gut microbiome. Microbiome 23:36. doi: 10.1186/s40168-017-0244-z

Brian PH, Maria C, Hugenholtz P, Konstantinos T, Murray A, Palmer M, et al. (2020) SeqCode, a nomenclatural code for prokaryotes described from sequence data. (Under review)

Breslauer KJ, Frank R, Blöcker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci USA 83:3746-50. doi: 10.1073/pnas.83.11.3746

Brito AF, Braconi CT, Weidmann M, Dilcher M, Alves JM, Gruber A, et al (2015) The pangenome of the *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV). Genome Biol Evol 27:94–108. doi: 10.1093/ gbe/evv231

Brooksbank C, Bergman MT, Apweiler R, Birney E, Thornton J (2014) The European Bioinformatics Institute's data resources. Nucleic Acids Res 42: D18-25. doi: 10.1093/nar/gkt1206

Bruns A, Berthe-Corti L (1999) *Fundibacter jadensis* gen. nov., sp. nov., a new slightly halophilic bacterium, isolated from intertidal sediment. Int J Syst Bacteriol 49:441-8. doi: 10.1099/00207713-49-2-441

Buddhi S, G S, Gupta D, Sasikala Ch, Ramana ChV (2020) *Afifella aestuarii* sp. nov., a phototrophic bacterium. Int J Syst Evol Microbiol 70:327-333. doi: 10.1099/ijsem.0.003756

Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P (2013) Structure and functions of the bacterial microbiota of plants. Annu Rev Plant Biol 64:807-38. doi: 10.1146/annurev-arplant-050312-12010

Busse J, Auling G (1988) Polyamine pattern as a chemotaxonomic marker within the Proteobacteria. Syst Appl Microbiol 11:1-8.

Callahan BJ, Sankaran K, Fukuyama JA, McMurdie PJ, Holmes SP (2016) Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. F1000Res 5:1492. doi: 10.12688/f1000research.8986.2

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-6. doi: 10.1038/nmeth.f.303

Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC et al. (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. Nature 419:512-9. doi: 10.1038/nature01099

Carriche GM, Almeida L, Stüve P, Velasquez L, Dhillon-LaBrooy A, Roy U et al (2021) Regulating T-cell differentiation through the polyamine spermidine. J Allergy Clin Immunol 147:335-348.e11. doi: 10.1016/j.jaci.2020.04.037

Chakraborty J, Sapkale V, Rajput V, Shah M, Kamble S, Dharne M et al (2020) Shotgun metagenome guided exploration of anthropogenically driven resistomic hotspots within Lonar soda lake of India. Ecotoxicol Environ Saf 194:110443. doi: 10.1016/j.ecoenv.2020.110443

Chan JZM, Halachev MR, Loman NJ, Constantinidou C, Pallen MJ (2012) Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*. BMC Microbiol 12:302. doi: 10.1186/1471-2180-12-302

Chaudhari NM, Gupta VK, Dutta C (2016) BPGA- an ultra-fast pan-genome analysis pipeline. Sci Rep 13:24373. doi: 10.1038/srep24373

Chaudhary DK, Kim J (2017) *Roseomonas nepalensis* sp. nov., isolated from oil-contaminated soil. Int J Syst Evol Microbiol 67:981–987. doi: 10. 1099/ijsem.0.001727

Chen B, Yu T, Xie S, Du K, Liang X, Lan Y et al (2018) Comparative shotgun metagenomic data of the silkworm Bombyx mori gut microbiome. Sci Data 5:180285. doi: 10.1038/sdata.2018.285

Chen LX, Huang LN, Méndez-García C, et al (2016) Microbial communities, processes and functions in acid mine drainage ecosystems. Curr Opin Biotechnol 38:150-8. doi: 10.1016/j.copbio.2016.01.013

Chen Q, Sun LN, Zhang XX, He J, Kwon SW, Zhang J et al (2014) *Roseomonas rhizosphaerae* sp. nov., a triazophos-degrading bacterium isolated from soil. Int J Syst Evol Microbiol 64:1127–1133. doi: 10.1099/ijs.0.057000-0

Chen W, Sheng XF, He LY, Huang Z (2015) *Rhizobium yantingense* sp. nov., a mineral-weathering bacterium. Int J Syst Evol Microbiol 65:412-417. doi: 10.1099/ijs.0.064428-0

Chen X, Zhang Y, Zhang Z, Zhao Y, Sun C, Yang M et al (2018) PGAweb: A web server for bacterial pan-genome analysis. Front Microbiol 21:1910. doi: 10.3389/fmicb.2018.01910

Chernikova TN, Bargiela R, Toshchakov SV, Shivaraman V, Lunev EA, Yakimov MM et al (2020) Hydrocarbon-degrading bacteria *Alcanivorax* and *Marinobacter* associated with microalgae *Pavlova lutheri* and *Nannochloropsis oculata*. Front Microbiol 28:572931. doi: 10.3389/fmicb.2020.572931

Chevreul ME (1823) Sur plusieurs corps gras, et particulièrement sur leurs combinaisons avec les alcalis. Annales de Chimie 88:225-261.

Chris AH, Floyd WS, Rebecca S, Smith A, Ware J, Zaspel J (2021) The future for a prominent taxonomy. Ins Syst Div 5:2. doi: org/10.1093/isd/ixaa020

Chun J, Rainey FA (2014) Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int J Syst Evol Microbiol 64, 316-324. doi: 10.1099/ijs.0.054171-0

Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS et al (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. doi:10.1099/ijsem.0.002516

Colwell RR (1970) Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus* and related *Vibrio* species. J Bacteriol 104:410–433.

Collins MD, Jones D (1981) Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. Microbiol Rev 45:316-54. doi: 10.1128/mr.45.2.316-354.1981

Costa SS, Guimarães LC, Silva A, Soares SC, Baraúna RA (2020) First steps in the analysis of prokaryotic pan-genomes. Bioinform Biol Insights 14:1177932220938064. doi: 10.1177/1177932220938064

Crane FL (1965) Distribution of quinones. 183-206. In R. A. Morton (ed), Biochemistry of quinones. Academic Press, Inc, London

Christenhusz MJM, Byng JW (2016) The number of known plants species in the world and its annual increase. Phytotaxa 3: 201-217.

Cvejic JH, Putra SR, El-Beltagy A, Hattori R, Hattori T, Rohmer M (2000) Bacterial triterpenoids of the hopane series as biomarkers for the chemotaxonomy of *Burkholderia, Pseudomonas and Ralstonia* spp. FEMS Microbiol Lett 15:295-9. doi: 10.1111/j.1574-6968.2000.tb08974.x

da Silva Filho AC, Raittz RT, Guizelini D, De Pierri CR, Augusto DW, Dos Santos-Weiss ICR et al (2018) Comparative analysis of genomic Island prediction tools. Front Genet 12:619. doi: 10.3389/fgene.2018.00619

Damtab J, Nutaratat P, Boontham W, Srisuk N, Duangmal K, Yurimoto H et al (2016) *Roseomonas elaeocarpi* sp. nov., isolated from olive (*Elaeocarpus hygrophilus* Kurz.) phyllosphere. Int J Syst Evol Microbiol 66:474–480. doi: 10.1099/ijsem.0.000748

Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ (2017) Members of the Candidate Phyla Radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. Microbiome 5:112. doi:org/10.1186/s40168-017-0331-1

Dar GH, Khuroo AA (2020). [Topics in Biodiversity and Conservation] Biodiversity of the Himalaya: Jammu and Kashmir State 18. doi:10.1007/978-981-32-9174-4

Das D, Kalra I, Mani K, Salgoankar B, Braganca J (2019) Characterization of extremely halophilic archaeal isolates from Indian salt pans and their screening for production of hydrolytic enzymes. Environmental Sustainability 2:227–239. doi:org/10.1007/s42398-019-00077-x

Dastager SG, Deepa CK, Li WJ, Tang SK, Pandey A (2011) *Paracoccus niistensis* sp. nov., isolated from forest soil, India. Antonie Van Leeuwenhoek 99:501-6. doi: 10.1007/s10482-010-9515

Datta S, Rajnish KN, Samuel MS, Pugazlendhi, Selvarajan E (2020) Metagenomic applications in microbial diversity, bioremediation, pollution monitoring, enzyme and drug discovery. A review. Environ Chem Lett 18:1229–1241. doi: org/10.1007/s10311-020-01010-z

de Carvalho CCCR, Caramujo MJ (2018) The various roles of fatty acids. Molecules 23:2583. doi: 10.3390/molecules23102583

Deepshikha G, Mohammed M, Mekala LP, Sasiakal Ch, Ramana ChV (2019) iTRAQ-based quantitative proteomics reveals insights into metabolic and molecular responses of glucose-grown cells of *Rubrivivax benzoatilyticus* JA2. J Proteomics. 194:49-59. doi: 10.1016/j.jprot.2018.12.027

Deng T, Chen X, Zhang Q, Zhong Y, Guo J, Sun G et al (2017) *Ciceribacter thiooxidans* sp. nov., a novel nitrate-reducing thiosulfate-oxidizing bacterium isolated from sulfide-rich anoxic sediment. Int J Syst Evol Microbiol 67:4710-4715. doi: 10.1099/ijsem.0.002367

Deng Y, Sun Y, Wang H, Yu LY, Zhang YQ (2020) *Roseomonas harenae* sp. nov., from desert gravel soil. Int J Syst Evol Microbiol 70:5711–5716. doi: 10.1099/ijsem.0.004467

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069-72. doi: 10.1128/AEM.03006-05

Dhir B (2018) Biotechnological tools for remediation of acid mine drainage (removal of metals from wastewater and leachate) //Bio-Geotechnologies for Mine Site Rehabilitation. Elsevier, 67–82.

Diesendorf N, Köhler S, Geißdörfer W, Grobecker-Karl T, Karl M, Burkovski A (2017) Characterisation of *Roseomonas mucosa* isolated from the root canal of an infected tooth. BMC Res Notes 10:212. doi: 10.1186/s13104- 017- 2538- 4

Divyasree B, Suresh G, Sasikala C, Ramana ChV (2018) *Chryseobacterium salipaludis* sp. nov., isolated at a wild ass sanctuary. Int J Syst Evol Microbiol 68:542-546. doi: 10.1099/ijsem.0.002536

Dobritsa AP, Samadpour M (2019) Reclassification of *Burkholderia insecticola* as *Caballeronia insecticola* comb. nov. and reliability of conserved signature indels as molecular synapomorphies. Int J Syst Evol Microbiol 69:2057-2063. doi: 10.1099/ijsem.0.003431

Dong L, Ming H, Yin YR, Duan YY, Zhou EM, Nie GX et al (2014) *Roseomonas alkaliterrae* sp. nov., isolated from an alkali geothermal soil sample in Tengchong. Yunnan, south-west China. Antonie van Leeuwenhoek 105: 899–905. doi: 10.1007/s10482-014-0144-1

Dong C, Lai Q, Liu X, Gu L, Zhang Y, Xie Z, et al (2021) *Alcanivorax profundimaris* sp. nov., a novel marine hydrocarbonoclastic bacterium isolated from seawater and deep-sea sediment. Curr Microbiol 78:1053-1060. doi: 10.1007/s00284-020-02322-7

Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M et al (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359:eaar4120. doi:10.1126/science.aar4120

Drancourt M, Raoult D (1994) Taxonomic position of the *Rickettsiae*: current knowledge. FEMS Microbiol Rev. 13:13-24. doi: 10.1111/j.1574-6976.1994.tb00032.x.

Drews G (2000) The roots of microbiology and the influence of Ferdinand Cohn on microbiology of the 19th century. FEMS Microbiol Rev 24:225-49. doi: 10.1111/j.1574-6976.2000.tb00540.x

Du MZ, Zhang C, Wang H, Liu S, Wei W, Guo FB (2018) The GC content as a main factor shaping the amino acid usage during bacterial evolution process. Front Microbiol 9:2948. doi: 10.3389/fmicb.2018.02948

Dugail I, Kayser BD, Lhomme M (2017) Specific roles of phosphatidylglycerols in hosts and microbes. Biochimie 141:47-53. doi: 10.1016/j.biochi.2017.05.005

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-7. doi: 10.1093/nar/gkh340

Emmanuel NE, Suleman E, Du PM, Deepak B, Raj TG (2021) Bioinformatics tools for gene and genome annotation analysis of microbes for synthetic biology and cancer biology applications. In: Singh V., Kumar A. (eds) Advances in Bioinformatics. Springer, Singapore. doi:org/10.1007/978-981-33-6191-1_16

Enebe MC, Babalola OO (2021) the influence of soil fertilization on the distribution and diversity of phosphorus cycling genes and microbes community of maize rhizosphere using shotgun metagenomics. Genes (Basel) 12:1022. doi: 10.3390/genes12071022

Ejigu GF, Jung J (2020) Review on the computational genome annotation of sequences obtained by next-generation sequencing. biology (Basel). 9:295. doi: 10.3390/biology9090295

El-Najjar N, Gali-Muhtasib H, Ketola RA, Vuorela P, Urtti A, Vuorela H (2011) The chemical and biological activities of quinones: overview and implications in analytical detection. Phytochem Rev 10: 353. doi:org/10.1007/s11101-011-9209-1

Eze MO, Hose GC, George SC, Daniel R (2021) Diversity and metagenome analysis of a hydrocarbon-degrading bacterial consortium from asphalt lakes located in Wietze, Germany. AMB Express 11:89. doi: 10.1186/s13568-021-01250-4

Fan D, Smith DL (2021) Characterization of Selected Plant Growth-Promoting Rhizobacteria and Their Non-Host Growth Promotion Effects. Microbiol Spectr 9: e00279-21. doi: org/10.1128/Spectrum.00279-21

Fang XM, Bai JL, Zhang DW, Su J, Zhao LL, Liu HY et al (2018) *Roseomonas globiformis* sp. nov., an airborne bacteria isolated from an urban area of Beijing. Int J Syst Evol Microbiol 68:3301–3306. doi: 10.1099/ijsem.0. 002985

Favre-Godal Q, Gourguillon L, Lordel-Madeleine S, Gindro K, Choisy P (2020) Orchids and their mycorrhizal fungi: an insufficiently explored relationship. Mycorrhiza 30:5–22. doi: org/10.1007/s00572-020-00934-2

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. doi: 10.1111/j.1558-5646

Fernández-Martínez J, Pujalte MJ, García-Martínez J, Mata M, Garay E, Rodríguez-Valera F (2003) Description of *Alcanivorax venustensis* sp. nov. and reclassification of *Fundibacter jadensis* DSM 12178^T (Bruns and Berthe-Corti 1999) as *Alcanivorax jadensis* comb. nov., members of the emended genus *Alcanivorax*. Int J Syst Evol Microbiol 53:331-338. doi: 10.1099/ijs.0.01923-0

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenza* RD. *Science* 269:496–512. doi: 10.1126/science.7542800

Flores-Félix JD, Menéndez E, Peix A, García-Fraile P, Velázquez E (2020) History and current taxonomic status of genus *Agrobacterium*. Syst Appl Microbiol. 43:126046. doi: 10.1016/j.syapm.2019.126046

Funke G, Lawson PA, Nolte FS, Weiss N, Collins MD (1998) *Aureobacterium resistens* sp. nov., exhibiting vancomycin resistance and teicoplanin susceptibility. FEMS Microbiol Lett 158:89-93. doi: 10.1111/j.1574-6968.1998.tb12805.x.

Furuhata K, Ishizaki N, Edagawa A, Fukuyama M (2013) *Roseomonas tokyonensis* sp. nov. isolated from a biofilm sample obtained from a cooling tower in Tokyo. Japan Biocontrol Sci 18:205–209. doi: 10.4265/bio.18.205

Furuhata K, Miyamamoto H, Goto K, Kato Y, Hara M, Fukuyama M (2008) *Roseomonas stagni* sp. nov., isolated from pond water in Japan. J Gen Appl Microbiol 54:167–171. doi: 10.2323/jgam.54.167

Gallego V, Sanchez-Porro C, Garcia MT, Ventosa A (2006) *Roseomonas aquatica* sp. nov., isolated from drinking water. Int J Syst Evol Microbiol 56: 2291–2295. doi: 10.1099/ijs.0.64379-0

Galperin MY, Makarova KS, Wolf YI, Koonin EV (2015) Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res 43:D261-9. doi: 10.1093/nar/gku1223

Garabito MJ, Marquez MC, Ventosa A (1998) Halotolerant *Bacillus* diversity in hypersaline environments. Can J Microbiol 44:95–102.

García-Martínez J, Acinas SG, Antón AI, Rodríguez-Valera F (1999) Use of the 16S--23S ribosomal genes spacer region in studies of prokaryotic diversity. J Microbiol Methods 36:55-64. doi: 10.1016/s0167-7012(99)00011-1

Garrity GM (2016) A new genomics-driven taxonomy of bacteria and archaea: are we there yet? J Clin Microbiol 54:1956-63. doi: 10.1128/JCM.00200-16

Gest H (2004) The discovery of microorganisms by Robert Hooke and Antoni *Van* Leeuwenhoek, fellows of the Royal Society. Notes Rec R Soc Lond 58:187-201. doi: 10.1098/rsnr.2004.0055

Gonzalez JM, Saiz-Jimenez C (2005) A simple fluorimetric method for the estimation of DNA-DNA relatedness between closely related microorganisms by thermal denaturation temperatures. Extremophiles 9:75-9. doi: 10.1007/s00792-004-0417-0

Gonzales-Siles L, Karlsson R, Schmidt P, Salvà-Serra F, Jaén-Luchoro D, Skovbjerg S et al (2020) Approach for discerning species-unique gene markers for identifications of *Streptococcus pneumoniae* and *Streptococcus pseudopneumoniae*. Front Cell Infect Microbiol 19:10-222. doi: 10.3389/fcimb.2020.00222

Goris J, Konstantinidis KT, Klappenbach JA, Salvà-Serra F, Jaén-Luchoro D, Skovbjerg S, et al (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. doi:10.1099/ijs.0.64483-0

Gross K, Sun M, Schiestl FP (2016) Why Do Floral Perfumes Become Different? Region-Specific Selection on Floral Scent in a Terrestrial Orchid. PLoS One 11:e0147975. doi: 10.1371/journal.pone.0147975

Gu W, Miller S, Chiu CY (2019) Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu Rev Pathol 14:319-338. doi: 10.1146/annurev-pathmechdis-012418-012751

Guan F, Han Y, Yan K, Zhang Y, Zhang Z, Wu N, Tian J (2020) Highly efficient production of chitooligosaccharides by enzymes mined directly from the marine metagenome. Carbohydr Polym 234:115909. doi: 10.1016/j.carbpol.2020.11590

Gupta RS, Lo B, Son J (2018) Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. Front Microbiol 13: 9-67. doi: 10.3389/fmicb.2018.00067

Gupta RS (1998) Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol Mol Biol Rev 62:1435-91. doi: 10.1128/MMBR.62.4.1435-1491.1998

Gupta RS (2006) Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (*Campylobacterales*). BMC Genomics 7:167. doi: org/10.1186/1471-2164-7-167

Gupta RS, Patel S, Saini N, Chen S (2020) Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the *Subtilis* and *Cereus* clades of species. Int J Syst Evol Microbiol 70:5753-5798. doi: 10.1099/ijsem.0.004475

Gunde-Cimerman N, Plemenitaš A, Oren A (2018) Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. FEMS Microbiol Rev 42:353-375. doi: 10.1093/femsre/fuy009

Gzyl A, Augustynowicz E, Mosiej E, Zawadka M, Gniadek G, Nowaczek A, et al (2005) Amplified fragment length polymorphism (AFLP) versus randomly amplified polymorphic DNA (RAPD) as new tools for inter- and intra-species differentiation within *Bordetella*. J Med Microbiol 54:333-346. doi: 10.1099/jmm.0.45690-0

Hamana K, Matsuzaki S (1992) Polyamines as a chemotaxonomic marker in bacterial systematics. Crit Rev Microbiol 18:261-83. doi: 10.3109/10408419209113518

Han XY, Pham AS, Tarrand JJ, Rolston KV, Helsel LO, Levett PN (2003) Bacteriologic characterization of 36 strains of *Roseomonas* species and proposal of *Roseomonas mucosa* sp. nov. and *Roseomonas gilardii* subsp *rosea* subsp. nov. Am J Clin Pathol 120:256–264. doi: 10.1309/731V- VGVC- KK35- 1Y4J

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245-9. doi: 10.1016/s1074-5521(98)90108-9

He D, Kim JK, Jiang XY, Park HY, Sun C, Yu HS et al (2014) *Roseomonas sediminicola* sp. nov., isolated from fresh water. Antonie Van Leeuwenhoek. 105:191-197. doi: 10.1007/s10482-013-0065

Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. Genomics 107:1-8. doi: 10.1016/j.ygeno.2015.11.003

Herbst EJ, Dion AS (1970) Polyamine changes during development of *Drosophila melanogaster*. Fed Proc 29:1563-7.

Herrera H, Fuentes A, Soto J, Valadares R, Arriagada (2021) Orchid-associated bacteria and their plant growth promotion capabilities. In: Merillon JM., Kodja H. (eds) Orchids Phytochemistry, Biology and Horticulture. Reference Series in Phytochemistry. Springer, Cham.

Hirashi A, Hoshino, Kitamura H (1984) Isoprenoid quinone composition in the classification of *Rhodospirllaceae.* J Gen Appl Microbiol 30:197-210.

Horath T, Bachofen R (2009) Molecular characterization of an endolithic microbial community in dolomite rock in the central alps (Switzerland). Microb Ecol 58:292-306. doi: 10.1007/s00248-008-9483-7

Hördt A, López MG, Meier-Kolthoff JP, Schleuning M, Weinhold LM, Tindall BJ et al (2020) Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of Alphaproteobacteria. Front Microbiol 11:468. doi: 10.3389/fmicb.2020.00468.

Hou X, Liu H, Wei S, Ding Z, Sang F, Zhao Y et al (2020) *Roseomonas selenitidurans* sp. nov., isolated from urban soil, and emended description of *Roseomonas frigidaquae*. Int J Syst Evol Microbiol 70:5937–5942. doi: 10. 1099/ijsem.0.004496

Hu D, Cha G, Gao B (2018) A phylogenomic and molecular markers-based analysis of the class *Acidimicrobiia*. Front Microbiol 15:9:987. doi: 10.3389/fmicb.2018.0098

Huang XX, Xu L, Sun JQ (2021) *Gracilibacillus suaedae* sp. nov., an indole acetic acid-producing endophyte isolated from a root of Suaeda salsa. Int J Syst Evol Microbiol 71. doi: 10.1099/ijsem.0.005140

Huang Z, Wang P, Pu Z, Lu L, Chen G, Hu X et al (2021) Effects of mancozeb on citrus rhizosphere bacterial community. Microb Pathog 154:104845. doi: 10.1016/j.micpath.2021.104845

Hutchings P (1986) Biological destruction of coral reefs. Coral Reefs. 4:239–52.

Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S et al (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol 12:e1004957. doi: 10.1371/journal.pcbi.1004957

Hyeon JW, Jeon CO (2017) *Roseomonas aerofrigidensis* sp. nov., isolated from an air conditioner. Int J Syst Evol Microbiol 67:4039–4044. doi: 10.1099/ ijsem.0.002246

Igarashi K, Kashiwagi K (2010) Modulation of cellular function by polyamines. Int J Biochem Cell Biol 42:39-51. doi: 10.1016/j.biocel.2009.07.009

The Irpcm Phytoplasma/Spiroplasma Working Team-Phytoplasma Taxonomy Group. '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects. Int J Syst Evol Microbiol 54:1243-1255. doi: 10.1099/ijs.0.02854-0

Imhoff JF, Caumette P (2004) Recommended standards for the description of new species of anoxygenic phototrophic bacteria. Int J Syst Evol Microbiol 54:1415-1421. doi: 10.1099/ijs.0.03002-0

Janda JM, Abbott SL (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol 45:2761-2764. doi:10.1128/JCM.01228-07

Jay ZJ, Inskeep WP (2015) The distribution, diversity, and importance of 16S rRNA gene introns in the order *Thermoproteales*. Biol Direct 10:35. doi: 10.1186/s13062-015-0065-6

Jiang CY, Dai X, Wang BJ, Zhou YG, Liu SJ (2006) *Roseomonas lacus* sp. nov., isolated from freshwater lake sediment. Int J Syst Evol Microbiol 56: 25–28. doi: 10.1099/ijs.0.63938-0

Jin L, Lee HG, No KJ, Ko SR, Kim HS, Ahn CY et al (2013) *Belnapia soli* sp. nov., a proteobacterium isolated from grass soil. Int J Syst Evol Microbiol 63:1955–1959. doi: 10.1099/ijs.0.045302-0

Jin R, Su J, Liu HY, Li QP, Zhang YQ, Yu LY (2012) Description of *Belnapia rosea* sp. nov. and emended description of the genus *Belnapia* Reddy et al., 2006. Int J Syst Evol Microbiol 62:705–709. doi: 10.1099/ijs.0.031 021-0

Kämpfer P, Aurass P, Karste S, Flieger A, Glaeser SP (2016) *Paracoccus contaminans* sp. nov., isolated from a contaminated water microcosm. Int J Syst Evol Microbiol 66:5101-5105. doi: 10.1099/ijsem.0.001478

Kämpfer P, Andersson MA, Jäckel U, Salkinoja-Salonen M (2003) *Teichococcus ludipueritiae* gen. nov. sp. nov., and *Muricocccus roseus* gen. nov. sp. nov. representing two new genera of the alpha-1 subclass of the Proteobacteria. Syst Appl Microbiol 26:23–29. doi: 10.1078/ 072320203322337272

Kämpfer P, Busse HJ, Galatis H, Criscuolo A, Clermont D, Bizet C et al (2019) *Paracoccus haematequi* sp. nov., isolated from horse blood. Int J Syst Evol Microbiol 69:1682-1688. doi: 10.1099/ijsem.0.003376

Kämpfer P, Andersson MA, Jäckel U, Salkinoja-Salonen M (2003) *Teichococcus ludipueritiae* gen. nov. sp. nov., and *Muricocccus roseus* gen. nov. sp. nov. representing two new genera of the alpha-1 subclass of the Proteobacteria. Syst Appl Microbiol 26:23–29. doi: 10.1078/ 072320203322337272

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27-30. doi: 10.1093/nar/28.1.27

Kang D, Huang Y, Nesme J, Herschend J, Jacquiod S, Kot W et al (2021) Metagenomic analysis of a keratin-degrading bacterial consortium provides insight into the keratinolytic mechanisms. Sci Total Environ 761:143281. doi: 10.1016/j.scitotenv.2020.143281

Kanso S, Greene AC, Patel BKC (2002) *Bacillus subterraneus* sp. nov., an iron- and manganese-reducing bacterium from a deep subsurface Australian thermal aquifer. Int J Syst Evol Microbiol 52:869-874. doi: 10.1099/00207713-52-3-869

Karray F, Ben Abdallah M, Kallel N, Hamza M, Fakhfakh M, Sayadi S (2018) Extracellular hydrolytic enzymes produced by halophilic bacteria and archaea isolated from hypersaline lake. Mol Biol Rep 45:1297-1309. doi: 10.1007/s11033-018-4286-5

Kates M (1986) Lipidology Tof. Isolation, analysis and identification of lipids. Laboratory Techniques in Biochemistry and Molecular Biology, 2. Amsterdam: Burdon RH and van Knippenberg PH, Elsevier 100–112.

Kaur J, Sharma J (2021) Orchid root associated bacteria: Linchpins or Accessories? Front Plant Sci 12:661966. doi: 10.3389/fpls.2021.661966

Khan SA, Jeong SE, Jung HS, Quan ZX, Jeon CO (2019) *Roseicella frigidaeris* gen. nov., sp. nov., isolated from an air-conditioning system. Int J Syst Evol Microbiol 69:1384–1389. doi: 10.1099/ijsem.0.00 3322

Khan ST, Takaichi S, Harayama S (2008) *Paracoccus marinus* sp. nov., an adonixanthin diglucoside-producing bacterium isolated from coastal seawater in Tokyo Bay. Int J Syst Evol Microbiol 58:383–386. doi: 10.1099/ijs.0.65103-0

Khaneja R, Perez-Fons L, Fakhry S, Baccigalupi L, Steiger S, To E et al (2010) Carotenoids found in Bacillus. J Appl Microbiol 108:1889–1902. doi: 10.1111/j.1365-2672.2009.04590.x.

Kim DU, Lee H, Kim SG, Ka JO (2017) *Roseomonas terricola* sp. nov., isolated from agricultural soil. Int J Syst Evol Microbiol 67:4836–4841. doi: 10.1099/ijsem.0.002389

Kim DU, Ka JO (2014) *Roseomonas soli* sp. nov., isolated from an agricultural soil cultivated with Chinese cabbage (*Brassica campestris*). Int J Syst Evol Microbiol 64:1024-1029. doi: 10.1099/ijs.0.053827-0

Kim JY, Kim DU, Kang MS, Jang JH, Kim SJ, Kim MJ et al (2018) *Roseomonas radiodurans* sp. nov., a gamma-radiation-resistant bacterium isolated from gamma ray-irradiated soil. Int J Syst Evol Microbiol 68:2443–2447. doi: 10.1099/ijsem.0.002852

Kim MS, Baik KS, Park SC, Rhee MS, Oh HM, Seong CN (2009) *Roseomonas frigidaquae* sp. nov., isolated from a water-cooling system. Int J Syst Evol Microbiol 59:1630–1634 doi: 10.1099/ijs.0.004812-0

Kim HM, Khan SA, Han DM, Chun BH, Jeon CO (2020) *Roseomonas algicola* sp. nov., isolated from a green alga *Pediastrum duplex*. Int J Syst Evol Microbiol 11:5634–5639. doi: 10.1099/ijsem.0.004454

Kim MC, Rim S, Pak S, Ren L, Zhang Y, Chang X et al (2016) *Roseomonas arcticisoli* sp. nov., isolated from Arctic tundra soil. Int J Syst Evol Microbiol 66:4057–4064. doi: 10.1099/ijsem.0.001310

Kim WH, Kim DH, Kang K, Ahn TY (2016) *Dankookia rubra* gen. nov., sp. nov., an alphaproteobacterium isolated from sediment of a shallow stream. J Microbiol 54:420–425. doi: 10.1007/s12275-016-60 54-3

Kim SH, Kim JG, Jung MY, Kim SJ, Gwak JH, Yu WJ et al (2018) *Ketobacter alkanivorans* gen. nov., sp. nov., an n-alkane-degrading bacterium isolated from seawater. Int J Syst Evol Microbiol 68:2258-2264. doi: 10.1099/ijsem.0.002823

Kim SJ, Weon HY, Ahn JH, Hong SB, Seok SJ, Whang KS et al (2013) *Roseomonas aerophila* sp. nov., isolated from air. Int J Syst Evol Microbiol 63:2334–2337. doi: 10.1099/ijs.0.046482-0

Kim I, Chhetri G, Kim J, Kang M, Seo T (2020) *Lewinella aurantiaca* sp. nov., a carotenoid pigment-producing bacterium isolated from surface seawater. Int J Syst Evol Microbiol 70:6180-6187. doi: 10.1099/ijsem.0.004515

Ko Y, Yim J, Hwang WM, Kang K, Ahn TY (2018) *Roseomonas fluminis* sp. nov. isolated from sediment of a shallow stream. Int J Syst Evol Microbiol 68:782–787. doi: 10.1099/ijsem.0.002578

Kodama Y, Mashima J, Kosuge T, Kaminuma E, Ogasawara O, Okubo K et al (2018) DNA Data Bank of Japan: 30th anniversary. Nucleic Acids Res 46:D30-D35. doi: 10.1093/nar/gkx926

Kook M, Son HM, Yi TH (2014) *Microbacterium kyungheense* sp. nov. and *Microbacterium jejuense* sp. nov., isolated from salty soil. Int J Syst Evol Microbiol 64:2267-2273. doi: 10.1099/ijs.0.054973-0

Komagata K, Iizuka H (1964) New species of *Brevibacterium* isolated from rice. (Studies on the microorganisms of cereal grains, Part VII). Jour Agri Chem Society of Japan 38:496-502.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120. doi: 10.1007/BF01731581

Konstantinidis KT, Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 187:6258-64. doi: 10.1128/JB.187.18.6258-6264.2005

Knobloch S, Daussin A, Jóhannsson R, Marteinsson V (2019) *Pelagibaculum spongiae* gen. nov., sp. nov., isolated from a marine sponge in south-west Iceland. Int J Syst Evol Microbiol 69:2129-2134. doi: 10.1099/ijsem.0.003448

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-4. doi: 10.1093/molbev/msw054

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547-1549. doi: 10.1093/molbev/msy096

Kumari A, Singh RK (2019) Medicinal chemistry of indole derivatives: Current to future therapeutic prospectives. Bioorg Chem 89:103021. doi: 10.1016/j.bioorg.2019.103021

Kurth, F, Zeitler K, Feldhahn L, Neu TR, Weber T, Krištůfek V et al (2013) Detection and quantification of a mycorrhization helper bacterium and a mycorrhizal fungus in plant-soil microcosms at different levels of complexity. BMC Microbiol 13:205. doi: 10.1186/1471-2180-13-205

Kunitsky C, Osterhout G, Sasser M (2006) Identification of microorganisms using fatty acid methyl ester (FAME) analysis and the MIDI Sherlock Microbial Identification System. Encyclopedia of rapid microbiological methods. 3: 1-18.

Kwon KK, Hye OJ, Yang SH, Seo HS, Lee JH (2015) *Alcanivorax gelatiniphagus* sp. nov., a marine bacterium isolated from tidal flat sediments enriched with crude oil. Int J Syst Evol Microbiol 65:2204-2208. doi: 10.1099/ijs.0.000244

Lai Q, Wang J, Gu L, Zheng T, Shao Z (2013) *Alcanivorax marinus* sp. nov., isolated from deep-sea water. Int J Syst Evol Microbiol 63: 4428-4432. doi: 10.1099/ijs.0.049957-0

Lai Q, Zhou Z, Li, G, Li G, Shao Z (2016) *Alcanivorax nanhaiticus* sp. nov., isolated from deep sea sediment. Int J Syst Evol Microbiol 66:3651-3655. doi: 10.1099/ijsem.0.001247

Lai Q, Wang L, Liu Y, Fu Y, Zhong H, Wang B et al (2011) *Alcanivorax pacificus* sp. nov., isolated from a deep-sea pyrene-degrading consortium. Int J Syst Evol Microbiol 61:1370-1374. doi: 10.1099/ijs.0.022368-0

Lan R, Reeves PR (2002) *Escherichia coli* in disguise: molecular origins of *Shigella*. Microbes Infect. 4:1125-32. doi: 10.1016/s1286-4579(02)01637-4

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 82:6955-9. doi: 10.1073/pnas.82.20.6955

Lalucat J, Gomila M, Mulet M, Zaruma A, García-Valdés E (2022) Past, present and future of the boundaries of the *Pseudomonas* genus: Proposal of *Stutzerimonas* gen. nov. Syst Appl Microbiol. 45: 126289. doi: 10.1016/j.syapm.2021.126289

Lebreton F, Manson AL, Saavedra JT, (2017) Tracing the enterococci from paleozoic origins to the hospital. Cell. 169: 849-861. doi: 10.1016/j.cell.2017.04.027

Lee I, Kim YO, Park SC, Chun J (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. doi:10.1099/ijsem.0.000760

Lee JH, Kim MS, Baik KS, Kim HM, Lee KH, Seong CN (2015) *Roseomonas wooponensis* sp. nov., isolated from wetland freshwater. Int J Syst Evol Microbiol 65:4049–4054. doi: 10.1099/ijsem.0.000536

Lee JH, Lee J (2010) Indole as an intercellular signal in microbial communities. FEMS Microbiol Rev 34:426-44. doi: 10.1111/j.1574-6976.2009.00204.x.

Lee Y, Jeon CO (2018) *Roseomonas aeriglobus* sp. nov., isolated from an air-conditioning system. Antonie Van Leeuwenhoek 111:343–351. doi: 10.1007/ s10482- 017-0956- x

Lodha TD, Srinivas A, Sasikala C, Ramana CV (2015) Hopanoid inventory of *Rhodoplanes* spp. Arch Microbiol 197:861-7. doi: 10.1007/s00203-015-1112-5

Li F, Huang Y, Hu W, Li Z, Wang Q, Huang S et al (2021) *Roseomonas coralli* sp. nov., a heavy metal resistant bacterium isolated from coral. Int J Syst Evol Microbiol 71:004624. doi: 10.1099/ijsem.0.004624

Li HF, Qu JH, Yang JS, Li ZJ, Yuan HL (2009) *Paracoccus chinensis* sp. nov., isolated from sediment of a reservoir. Int J Syst Evol Microbiol 59:2670-4. doi: 10.1099/ijs.0.004705-0

Li J, Qi M, Lai Q, Liu X, Shao Z (2020) *Ketobacter nezhaii* sp. nov., a marine bacterium isolated from coastal sediment. Int J Syst Evol Microbiol 70:4960-4965. doi: 10.1099/ijsem.0.004365

Li O, Xiao R, Sun L, Guan C, Kong D, Hu X (2017) Bacterial and diazotrophic diversities of endophytes in *Dendrobium catenatum* determined through barcoded pyrosequencing. PLoS One 12:e0184717. doi: 10.1371/journal.pone.0184717

Li T, Yang W, Wu S, Selosse MA, Gao J (2021) Progress and prospects of mycorrhizal fungal diversity in Orchids. Front Plant Sci 12:646325. doi: 10.3389/fpls.2021.646325

Liao X, Lai Q, Yang J, Dong C, Li D, Shao Z (2020) *Alcanivorax sediminis* sp. nov., isolated from deep-sea sediment of the Pacific Ocean. Int J Syst Evol Microbiol 54:4280-4284. doi: 10.1099/ijsem.0.004285

Liesack W, Weyland H, Stackebrandt E (1991) Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. Microb Ecol 21:191-8. doi: 10.1007/BF02539153

Lin YT, Lin YF, Tsai IJ, Chang EH, Jien SH, Lin YJ et al (2019) Structure and diversity of soil bacterial communities in offshore islands. Sci Rep 9:4689. doi:org/10.1038/s41598-019-41170-9

Lipski A, Reichert K, Reuter B, Spröer C, Altendorf K (1998) Identification of bacterial isolates from biofilters as *Paracoccus alkenifer* sp. nov. and *Paracoccus solventivorans* with emended description of *Paracoccus solventivorans*. Int J Syst Bacteriol 48 2:529-36. doi: 10.1099/00207713-48-2-529

Liu J, Ren Q, Zhang Y, Li Y, Tian X, Wu Y et al (2019) *Alcanivorax profundi* sp. nov., isolated from deep seawater of the Mariana Trench. Int J Syst Evol Microbiol 69:371-376. doi: 10.1099/ijsem.0.003145

Liu C, Shao Z (2005) *Alcanivorax dieselolei* sp. nov., a novel alkane-degrading bacterium isolated from sea water and deep-sea sediment. Int J Syst Evol Microbiol 55:1181-1186. doi: 10.1099/ijs.0.63443-0

Liu A, Zhang YJ, Cheng P, Peng YJ, Blom J, Xue QJ (2020) Whole genome analysis calls for a taxonomic rearrangement of the genus *Colwellia*. Antonie Van Leeuwenhoek 113: 919–931. doi: 10.1007/s10482-020-01405-6

Liu Y, Lai Q, Göker M, Meier-Kolthoff JP, Wang M, Sun Y et al (2015) Genomic insights into the taxonomic status of the *Bacillus cereus* group. Sci Rep 16:14082. doi: 10.1038/srep14082

Liu Y, Lai Q, Shao Z (2017) A multilocus sequence analysis scheme for phylogeny of *Thioclava* bacteria and proposal of two novel species. Front Microbiol 8:1321. doi: 10.3389/fmicb.2017.01321

Liu YX, Qin Y, Chen T, Lu M, Qian X, Guo X, et al (2021) A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell 12:315-330. doi: 10.1007/s13238-020-00724-8

Livingstone PG, Morphew RM, Whitworth DE (2018) Genome sequencing and pan-genome analysis of 23 *Corallococcus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets. Front Microbiol. 19:3187. doi: 10.3389/fmicb.2018.03187

Logan NA, Berge O, Bishop AH, Busse HJ, De Vos P, Fritze D et al (2009) Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. Int J Syst Evol Microbiol 59:2114-21. doi: 10.1099/ijs.0.013649-0

López-Hermoso C, de la Haba RR, Sánchez-Porro C, Papke RT, Ventosa A (2017) Assessment of multilocus sequence analysis as a valuable tool for the classification of the genus *Salinivibrio*. Front Microbiol 8:1107. doi: 10.3389/fmicb.2017.01107

Lopes A, Espirito SC, Grass G, Chung AP, Morais PV (2011) *Roseomonas pecuniae* sp. nov., isolated from the surface of a copper-alloy coin. Int J Syst Evol Microbiol 61: 610–615. doi: 10.1099/ijs.0.020966-0

Lopes-Santos L, Castro DBA, Ferreira-Tonin M, Chung AP, Morais PV (2017) Reassessment of the taxonomic position of *Burkholderia andropogonis* and description of *Robbsia andropogonis* gen. nov., comb. nov. Antonie Van Leeuwenhoek 110:727–736. doi: 10.1007/s10482-017-0842-6

Luo C, Rodriguez-R LM, Konstantinidis KT (2014). MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences. Nucleic Acids Res 42:e73. doi: 10.1093/nar/gku169

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Giga Sci 1:18. doi:org/10.1186/2047-217X-1-18

Mahajan GB, Balachandran L (2017) Sources of antibiotics: Hot springs. Biochem Pharmacol 134:35-41. doi: 10.1016/j.bcp.2016.11.021.

Maoka T (2020) Carotenoids as natural functional pigments. J Nat Med 74:1–16. doi:org/10.1007/s11418-019-01364-x

Matsuda K (2017) PCR-based detection methods for single-nucleotide polymorphism or mutation: real-time pcr and its substantial contribution toward technological refinement. Adv Clin Chem 80:45-72. doi: 10.1016/bs.acc.2016.11.002

Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R et al (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 17:3140-5. doi: 10.1073/pnas.95.6.3140

Malini S, Goh BL, Lim TS (2016) A rare case of *Roseomonas gilardii* peritonitis in a patient on continuous ambulatory peritoneal dialysis. Perit Dial Int 36:578. doi: 10.3747/pdi.2016.00103

Mathé C, Sagot MF, Schiex T, Rouzé P (2002) Current methods of gene prediction, their strengths and weaknesses. Nucleic Acids Res 30:4103-17. doi: 10.1093/nar/gkf543

Margesin R, Zhang DC (2013) *Humitalea rosea* gen. nov., sp. nov., an aerobic bacteriochlorophyll-containing bacterium of the family *Acetobacteraceae* isolated from soil. Int J Syst Evol Microbiol 63:1411–1416. doi: 10.1099/ijs.0.043018

Marin ME, Marco DJ, Diabar E, Fernandez Caniggia L, Greco G, Flores Y et al (2001) Catheter-related bacteremia caused by *Roseomonas gilardii* in an immunocompromised patient. Int J Infect Dis 5:170–171. doi: 10.1016/ s1201-9712(01)90095- 5

Martens M, Dawyndt P, Coopman R, Gillis M, De Vos P, Willems A (2008) Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). Int J Syst Evol Microbiol 58:200–214. doi: 10.1099/ijs.0.65392-0

Martin FM, Uroz S, Barker DG (2017) Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. Science 356:eaad4501. doi: 10.1126/science.aad4501

McKerrow J, Vagg S, McKinney T, Seviour EM, Maszenan AM, Brooks P et al (2000) A simple HPLC method for analysing diaminopimelic acid diastereomers in cell walls

of Gram-positive bacteria. Lett Appl Microbiol 30:178-82. doi: 10.1046/j.1472-765x.2000.00675.x

McInerney JO, McNally A, O'Connell MJ (2017) Why prokaryotes have pangenomes. Nat Microbiol 28:2–17040. doi: 10.1038/nmicrobiol.2017.40

Meera KB, Khan MA, Khan ST (2019) Next-Generation Sequencing (NGS) Platforms: An Exciting Era of Genome Sequence Analysis. In: Tripathi V, Kumar P, Tripathi P, Kishore A, Kamle M (eds) Microbial Genomics in Sustainable Agroecosystems. Springer, Singapore. doi:org/10.1007/978-981-32-9860-6_6

Meier-Kolthoff JP, Klenk HP, Göker M (2014) Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 64:352-356. doi: 10.1099/ijs.0.056994-0

Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M (2022) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 50:D801-D807. doi: 10.1093/nar/gkab902

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M et al (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9:386. doi: org/10.1186/1471-2105-9-386

Mezzasoma A, Coleine C, Sannino C, Selbmann L (2021) Endolithic bacterial diversity in lichen-dominated communities is shaped by sun exposure in McMurdo dry valleys, Antarctica. Microb Ecol 83:328-339. doi: 10.1007/s00248-021-01769-w

Mhete M, Eze PN, Rahube TO, Akinyemi FO (2020) Soil properties influence bacterial abundance and diversity under different land-use regimes in semi-arid environments. Sci African 7:e00246. doi:10.1016/j.sciaf.2019.e00246

Michon AL, Saumet L, Bourdier A, Haouy S, Sirvent N, Marchandin H (2014) Bacteremia due to imipenem-resistant *Roseomonas mucosa* in a child with acute lymphoblastic leukemia. J Pediatr Hematol Oncol 36: e165–e168.

Ming H, Duan YY, Yin YR, Meng XL, Li S, Zhou EM et al (2016) *Crenalkalicoccus roseus* gen. nov., sp. nov., a thermophilic bacterium isolated from alkaline hot springs. Int J Syst Evol Microbiol 66:2319–2326. doi:10.1099/ijsem.0.001029

Mineta K, Gojobori T (2016) Databases of the marine metagenomics. Gene 1:724-8. doi: 10.1016/j.gene.2015.10.035.

Mishra S, Lin Z, Pang S, Zhang W, Bhatt P, Chen S (2021) Recent advanced technologies for the characterization of xenobiotic-degrading microorganisms and microbial communities. Front Bioeng Biotechnol 9:632059. doi: 10.3389/fbioe.2021.632059

Mitter EK, de Freitas JR, Germida JJ (2017) Bacterial Root Microbiome of Plants Growing in Oil Sands Reclamation Covers. Front Microbiol 8:849. doi: 10.3389/fmicb.2017.00849

Mohanty P, Singh PK, Chakraborty D, et al (2021) Insight into the role of PGPR in sustainable agriculture and environment. Front Microbiol 5:1-12. doi: org/10.3389/fsufs.2021.667150

Mokashe N, Chaudhari B, Patil U (2018) Operative utility of salt-stable proteases of halophilic and halotolerant bacteria in the biotechnology sector. Int J Biol Macromol 117:493-522. doi: 10.1016/j.ijbiomac.2018.05.217

Mujahid M, Sasikala C, Ramana CV (2010) Aniline-induced tryptophan production and identification of indole derivatives from three purple bacteria. Curr Microbiol 61:285–290. doi:org/10.1007/s00284-010-9609-2

Na SI, Kim YO, Yoon SH, Ha SM, Baek I, Chun J (2018) UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. J Microbiol 56:280-285. doi: 10.1007/s12275-018-8014-6

Naushad HS, Lee B, Gupta RS (2014) Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya, Pectobacterium* and *Brenneria*. Int J Syst Evol Microbiol 64:366-383. doi: 10.1099/ijs.0.054213-0

Navarro-Muñoz JC, Collemare J (2020) Evolutionary Histories of Type III Polyketide Synthases in Fungi. Front Microbiol 10:3018. doi: 10.3389/fmicb.2019.03018

Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F et al (2021) A genomic 200 catalog of Earth's microbiomes. Nat Biotechnol 39:499-509. doi: 10.1038/s41587-020-0718-6

Netz N, Opatz T (2015) Marine Indole Alkaloids. Mar Drugs 13:4814-914. doi: 10.3390/md13084814

Nguyen TM, Kim J (2017) A rapid and simple method for identifying bacterial polar lipid components in wet biomass. J Microbiol 55:635-639. doi: 10.1007/s12275-017-7092-1

Ning K, Tong Y (2019) The fast track for microbiome research. Genomics Proteomics Bioinformatics 17:1-3. doi: 10.1016/j.gpb.2019.04.001

Nouioui I, Carro L, Garcia-Lopez M, Meier-Kolthoff JP, Woyke T, Kyrpides NC et al (2018) Genome-based taxonomic classification of the phylum *Actinobacteria*. Front Microbiol 9:2007. doi: 10.3389/fmicb.2018.02007

Nowicka B, Kruk J (2010) Occurrence, biosynthesis and function of isoprenoid quinones. Biochim Biophys Acta 1797:1587-605. doi: 10.1016/j.bbabio.2010.06.007

Nutaratat P, Srisuk N, Duangmal K, Yurimoto H, Sakai Y, Muramatsu Y et al (2013) *Roseomonas musae* sp. nov., a new bacterium isolated from a banana phyllosphere. Antonie Van Leeuwenhoek 103:617–624. doi: 10.1007/ s10482- 012- 9845- 5

Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF (2020) Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems 5: e00731-19. doi: 10.1128/ mSystems.00731- 719

Orata FD, Meier-Kolthoff JP, Sauvageau D, Stein LY (2018) Phylogenomic analysis of the gammaproteobacterial methanotrophs (Order *Methylococcales*) calls for the reclassification of members at the genus and species levels. Front Microbiol 9:3162. doi: 10.3389/fmicb.2018.03162

Oren A (2002) Diversity of halophilic microorganisms; environments, phylogeny, physiology and applications. J Ind Microbial Biotechnol 28:56–63. doi: 10.1038/sj/jim/7000176

Oren A (2015) 70th anniversary collection for the microbiology society: international journal of systematic and evolutionary microbiology. Int J Syst Bacteriol 65:4291–4293. doi: 10.1099/ijsem.0.000531

Oren A, Garrity GM (2022) Notification that new names of prokaryotes, new combinations, and new taxonomic opinions have appeared in volume 71, part 10 of the IJSEM. Int J Syst Evol Microbiol 72. doi: 10.1099/ijsem.0.005165.

Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G et al (2015) Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. Bioinform Biol Insights 9:75-88. doi: 10.4137/BBI.S12462

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M et al (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res 33:5691-702. doi: 10.1093/nar/gki866

Parker CT, Tindall BJ, Garrity GM (2019) International Code of Nomenclature of Prokaryotes. Int J Syst Evol Microbiol 69:S1-S111. doi: 10.1099/ijsem.0.000778

Pan J, Sun C, Zhang XQ, Huo YY, Zhu XF, Wu M (2014) *Paracoccus sediminis* sp. nov., isolated from Pacific Ocean marine sediment. Int J Syst Evol Microbiol 64:2512-2516. doi: 10.1099/ijs.0.051318-0

Panda MK, Sahu MK, Tayung K (2013) Isolation and characterization of a thermophilic *Bacillus* sp. with protease activity isolated from hot spring of Tarabalo, Odisha. India. Iran J Microbiol 5:159–165.

Parag B, Sasikala C, Ramana CV (2015) *Barrientosiimonas endolithica* sp. nov., isolated from pebbles, reclassification of the only species of the genus *Tamlicoccus, Tamlicoccus marinus* Lee 2013, as *Barrientosiimonas* marina comb. nov. and emended description of the genus *Barrientosiimonas*. Int J Syst Evol Microbiol 65:3031-3036. doi: 10.1099/ijs.0.000374

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36: 996–1004. doi: 10.1038/nbt.4229

Patel NY, Baria DM, Yagnik SM, Rajput KN, Panchal RR, Raval VH (2021) Bio-prospecting the future in perspective of amidohydrolase L-glutaminase from marine habitats. Appl Microbiol Biotechnol 105:5325–5340. doi:org/10.1007/s00253-021-11416-6

Patel S, Gupta RS (2020) A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. Int J Syst Evol Microbiol 70: 406–438. doi:10.1099/ijsem.0.003775

Perreca E, Rohwer J, González-Cabanelas D, Loreto F, Schmidt A, Gershenzon J et al (2020) Effect of drought on the methylerythritol 4-phosphate (mep) pathway in the isoprene emitting conifer *Picea glauca*. Front Plant Sci 11:546295. doi: 10.3389/fpls.2020.546295

Phillips RD, Reiter N, Peakall R (2020) Orchid conservation: from theory to practice. Ann Bot 3:345-362. doi:10.1093/aob/mcaa093

Poddar A, Das SK (2018) Microbiological studies of hot springs in India: a review. Arch Microbiol 200:1–18. doi:org/10.1007/s00203-017-1429-3

Qin QL, Xie BB, Zhang XY, Chen XL, Zhou BC, Zhou J et al (2014) A proposed genus boundary for the prokaryotes based on genomic insights. J Bacteriol 196:2210-5. doi: 10.1128/JB.01688-14

Qiu X, Qu Z, Jiang F, Lin Y, Zhang Y, Chang X et al (2016) *Roseomonas arctica* sp. nov., isolated from arctic glacial foreland soil. Int J Syst Evol Microbiol 66:1218–1223. doi: 10.1099/ijsem.0.000857

Oren A (2002) Diversity of halophilic microorganisms: Environments, phylogeny, physiology, and applications. J Ind Microbiol Biotech 28:56–63. doi.org/10.1038/sj/jim/7000176

Qin Y, Hou J, Deng M, Liu Q, Wu C, Ji Y et al (2016) Bacterial abundance and diversity in pond water supplied with different feeds. Sci Rep 6:35232. doi: 10.1038/srep35232

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:590-6. doi: 10.1093/nar/gks1219

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N (2017) Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35:833-844. doi: 10.1038/nbt.3935

Ramaprasad EVV, Sasikala C, Ramana CV (2015) *Roseomonas oryzae* sp. nov., isolated from paddy rhizosphere soil. Int J Syst Evol Microbiol 65:3535–3540. doi: 10.1099/ijsem.0.000449

Rahul K, Sasikala C, Tushar L, Debadrita R, Ramana CV (2014) *Alcanivorax xenomutans* sp. nov., a hydrocarbonoclastic bacterium isolated from a shrimp cultivation pond. Int J Syst Evol Microbiol 64:3553-3558. doi: 10.1099/ijs.0.061168-0

Rat A, Naranjo HD, Lebbe L, Cnockaert M, Krigas N, Grigoriadou K et al (2021) *Roseomonas hellenica* sp. nov., isolated from roots of wild-growing *Alkanna tinctoria*. Syst Appl Microbiol 44:126206. doi: 10.1016/j.syapm.2021. 126206

Reddy GS, Nagy M, Garcia-Pichel F (2006) *Belnapia moabensis* gen. nov., sp. nov., an alphaproteobacterium from biological soil crusts in the Colorado Plateau, USA. Int J Syst Evol Microbiol 56:51–58. doi: 10.1099/ijs.0.63764-0

Reid R, Visscher P, Decho A, Stolz JF, Bebout BM, Dupraz C et al (2000) The role of microbes in accretion, lamination and early lithification of modern marine stromatolites. Nature 406:989–992. doi: 10.1038/35023158

Rosselló-Móra R, Amann, R (2015) Past and future species definitions for bacteria and Archaea. Syst Appl Microbiol 38:209–202. doi: 10.1016/j.syapm. 2015.02.001

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126-31. doi: 10.1073/pnas.0906412106

Rihs JD, Brenner DJ, Weaver RE, Steigerwalt AG, Hollis DG, Yu VL (1993) *Roseomonas*, a new genus associated with bacteremia and other human infections. J Clin Microbiol 31: 3275–3283. doi: 10.1128/JCM.31.12. 3275- 3283.1993

Rivas R, García-Fraile P, Mateos PF, Martínez-Molina E, Velázquez E (2007) *Alcanivorax balearicus* sp. nov., isolated from Lake Martel. Int J Syst Evol Microbiol 57:1331-1335. doi: 10.1099/ijs.0.64912-0

Rodriguez LM and Konstantinidis KT (2014). Bypassing cultivation to identify bacterial species. Microbe 9:111–117.

Rogers HJ, Perkins HR, Ward JB (1980) Microbial Cell Walls and Membranes. Chapman and Hall, London.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open-source tool for metagenomics. PeerJ 4:e2584. doi: 10.7717/peerj.2584

Rojo F (2009) Degradation of alkanes by bacteria. Environ Microbiol 11:2477-90. doi: 10.1111/j.1462-2920.2009. 01948.x

Romalde JL, Balboa S, Ventosa A (2019) Editorial: microbial taxonomy, phylogeny and biodiversity. Front Microbiol 10:1324. doi: 10.3389/fmicb.2019.01324

Rosselló-Móra R, Amann R (2001) The species concept for prokaryotes. FEMS Microbiol Rev 25:39–67. doi: 10.1111/j.1574-6976.2001.tb00571.x

Romano-Bertrand S, Bourdier A, Aujoulat F, Aujoulat F, Michon AL, Masnou A et al (2016) Skin microbiota is the main reservoir of *Roseomonas mucosa*, an emerging opportunistic pathogen so far assumed to be environmental. Clin Microbiol Infect 22:737.e1-7. doi: 10.1016/j.cmi.2016.05.024

Rosselló-Móra R, Amann R (2015) Past and future species definitions for Bacteria and Archaea. Syst Appl Microbiol 38:209-16. doi: 10.1016/j.syapm.2015.02.001

Rosselló-Mora R (2005) Updating prokaryotic taxonomy. J Bacteriol 187:6255-7. doi: 10.1128/JB.187.18.6255-6257.2005

Rosselló-Móra R, Trujillo ME, Sutcliffe IC (2017) Introducing a digital protologue: a timely move towards a database-driven systematics of archaea and bacteria. Antonie Van Leeuwenhoek 110:455–456. doi: 10.1007/s10482-017-0841-7

Salomon MV, Purpora R, Bottini R, Piccoli P (2016) Rhizosphere associated bacteria trigger accumulation of terpenes in leaves of *Vitis vinifera* L. cv. Malbec that protect cells against reactive oxygen species. Plant Physiol Biochem 106:295-304. doi: 10.1016/j.plaphy.2016.05.007

Sánchez-Porro C, Gallego V, Busse HJ, Kämpfer P, Ventosa A (2009) Transfer of *Teichococcus ludipueritiae* and *Muricoccus roseus* to the genus *Roseomonas*, as *Roseomonas ludipueritiae* comb. nov. and *Roseomonas* rosea comb. nov., respectively, and emended description of the genus *Roseomonas*. Int J Syst Evol Microbiol 59:1193–1198. doi: 10.1099/ijs.0.004 820-0

Sandmann G (2019) antioxidant protection from uv- and light-stress related to carotenoid structures. Antioxidants (Basel) 8:219. doi: 10.3390/antiox8070219

Santos RM, Diaz PAE, Lobo LLB, Rigobelo EC (2020) Use of plant growth-promoting rhizobacteria in maize and sugarcane: characteristics and applications. Front Sustain Food Syst 4:136. doi: 10.3389/fsufs.2020.00136

Saitoh S, Suzuki T, Nishimura Y (1998) Proposal of *Craurococcus roseus* gen. nov., sp. nov. and *Paracraurococcus ruber* gen. nov., sp. nov., novel aerobic bacteriochlorophyll a-containing bacteria from soil. Int J Syst Evol Microbiol 48:1043–1047. doi: 10.1099/00207713-48-3-1043

Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I (2019) GenBank. Nucleic Acids Res 47:D94-D99. doi: 10.1093/nar/gky989

Scalzitti N, Jeannin-Girardon A, Collet P, Poch O, Thompson JD (2020) A benchmark study of *ab initio* gene prediction methods in diverse eukaryotic organisms. BMC Genomics 21:293. doi:org/10.1186/s12864-020-6707-9

Schlechter RO, Miebach M, Remus-Emsermann MNP (2019) Driving factors of epiphytic bacterial communities: A review. J Adv Res 19:57-65. doi: 10.1016/j.jare.2019.03.003

Schleifer KH (1985) Analysis of the chemical composition and primary structure of murein. Meth Microbiol 18:123-156.

Schleifer KH (2009) Classification of Bacteria and Archaea: past, present and future. Syst Appl Microbiol 32:533-42. doi: 10.1016/j.syapm.2009.09.002

Schleifer KH, Kandler O (1972) Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriol Rev 36:407-77. doi: 10.1128/br.36.4.407-477.1972

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537-41. doi: 10.1128/AEM.01541-09

Schildkraut CL, Marmur J, Doty P (1961) The formation of hybrid DNA molecules and their use in studies of DNA homologies. J Mol Biol 3:595–617. doi: 10.1016/S0022-2836(61)80024-7

Sentausa E, Fournier PE (2013) Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 19: 790-5. doi: 10.1111/1469-0691.12181

Setubal JC, Almeida NF, Wattam AR (2018) Comparative Genomics for Prokaryotes. Methods Mol Biol 1704:55-78. doi: 10.1007/978-1-4939-7463-4_3

Shah S, Chand K, Rekadwad B, Shouche YS, Sharma J, Pant B (2021) A prospectus of plant growth promoting endophytic bacterium from orchid (*Vanda cristata*). BMC Biotechnol 21:16. doi: org/10.1186/s12896-021-00676-9

Shao S, Guo X, Guo P, Cui Y, Chen Y (2019) *Roseomonas mucosa* infective endocarditis in patient with systemic lupus erythematosus: case report and review of literature. BMC Infect Dis 19:140. doi: 10.1186/s12879-019- 3774- 0

Sharma A, Pandey A, Shouche YS, Kumar B, Kulkarni G (2009) Characterization and identification of *Geobacillus* spp. isolated from Soldhar hot spring site of Garhwal Himalaya, India. J Basic Microbiol 49:187–194. doi:org/10.1002/jobm.200800194

Sharmili AS, Ramasamy P (2016) Fatty Acid Methyl Ester (FAME) analysis of moderately thermophilic bacteria isolated from the coramandal coast, Chennai, Tamil Nadu. Eur J Exp Biol 6:1-7.

Sharpton TJ (2014) An introduction to the analysis of shotgun metagenomic data. Front Plant Sci 5:209. doi: 10.3389/fpls.2014.00209

Shivali K, Sasikala Ch, Ramana ChV (2012) MLSA barcoding of *Marichromatium* spp. and reclassification of *Marichromatium fluminis* (Sucharita et al., 2010) as

Phaeochromatium fluminis gen. nov. comb. nov. Syst Appl Microbiol 35:221-5. doi: 10.1016/j.syapm.2012.03.002

Shivali K, Venkata Ramana V, Ramaprasad EVV, Sasikala Ch, Ramana ChV (2011) *Marichromatium litoris* sp. nov. and *Marichromatium chrysaorae* sp. nov. isolated from beach sand and from a jelly fish (*Chrysaoracolorata*). Syst Appl Microbiol 34:600-605. doi: 10.1016/j.syapm.2011.08.002

Siller H, Rainey FA, Stackebrandt E, Winter J (1996) Isolation and characterization of a new gram-negative, acetone-degrading, nitrate-reducing bacterium from soil, *Paracoccus solventivorans* sp. nov. Int J Syst Bacteriol 46:1125-30. doi: 10.1099/00207713-46-4-1125

Singer SJ (1972) A fluid lipid-globular protein mosaic model of membrane structure. Ann NY Acad Sci 195:16–23 doi: 10.1111/j.1749-6632.1972.tb54780.x

Singh P, Dash SS (eds) (2018) Plant discoveries 2017. BSI, Kolkata.

Skerman VBD, McGowan V, Sneath PHA (1980) Approved lists of bacterial names. Int J Syst Bacteriol 30:225–420.

Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E et al (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. Proc Natl Acad Sci USA 111:E4096-102. doi: 10.1073/pnas.1321152111

Smibert RM, Krieg NR (1981) General characterization. In: Gerhardt P, Murray RGE, Costilow RN, Nester EW, Wood WA et al. (editors). Manual of Methods for General Bacteriology. Am Society Microbiol 24:409–43.

Solden L, Lloyd K, Wrighton K (2016) The bright side of microbial dark matter: lessons learned from the uncultivated majority. Curr Opin Microbiol 31:217-226. doi: 10.1016/j.mib.2016.04.020

Song L, Liu H, Cai S, Huang Y, Dai X, Zhou Y (2018) *Alcanivorax indicus* sp. nov., isolated from seawater. Int J Syst Evol Microbiol 68:3785-3789. doi: 10.1099/ijsem.0.003058

Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ et al (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Sys Evol Microbiol 52:1043–1047. doi:10.1099/00207713-52-3-1043

Stackebrandt E, Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 33:152–155.

Subhash Y, Lee SS (2017) *Roseomonas suffusca* sp. nov., isolated from lagoon sediments. Int J Syst Evol Microbiol 67:2390–2396. doi: 10.1099/ijsem. 0.001966

Subhash Y, Lee SS (2018) *Roseomonas deserti* sp. nov., isolated from crude oil contaminated desert sand. Int J Syst Evol Microbiol 68:675–680. doi: 10.1099/ijsem.0.002565

Subhash Y, Bang JJ, You TH, Lee SS (2016) *Roseomonas rubra* sp. nov., isolated from lagoon sediments. Int J Syst Evol Microbiol 66:3821–3827. doi: 10.1099/ijsem.0.001271

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592. doi: 10.1073/pnas.48.4.582

Szarvas J, Ahrenfeldt J, Cisneros JLB, Thomsen MCF, Aarestrup FM, Lund O (2020) Large scale automated phylogenomic analysis of bacterial isolates and the evergreen online platform. Commun Biol 3: 137. doi: org/10.1038/s42003-020-0869-5

Sutcliffe IC, Rosselló-Móra R, Trujillo ME (2021) Addressing the sublime scale of the microbial world: reconciling an appreciation of microbial diversity with the need to describe species. New Microbes New Infect 43:100931. doi: 10.1016/j.nmni.2021.100931

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL et al (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 27:13950-5. doi: 10.1073/pnas.0506758102

Tian Z, Lu S, Jin D, Yang J, Pu J, Lai XH et al (2019) *Roseomonas wenyumeiae* sp. nov., isolated from faeces of Tibetan antelopes (*Pantholops hodgsonii*) on the Qinghai-Tibet Plateau. Int J Syst Evol Microbiol 69:2979– 2986. doi: 10.1099/ijsem.0.003479

Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 60:249-266. doi: 10.1099/ijs.0.016949-0

Tiwari S, Sandhya M, Gurnule WB, Bhagat DS, Gunjal A (2022) Chapter 2 - Habitat-specific microbial community associated with the biodiversity hotspot. In: Gunjal A, Shinde S (eds) Microbial Diversity in Hotspots. Academic Press.

Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK (2020) Plant-microbiome interactions: from community assembly to plant health. Nat Rev Microbiol 18:607-621. doi: 10.1038/s41579-020-0412-1

Trüper HG, Pfennig N (1981) Characterization and identification of the anoxygenic phototrophic bacteria. In: Starr MP, Stolp H, Trüper HG, Balows A, Schlegel HG (eds) The Prokaryotes. Springer, Berlin, Heidelberg. doi: org/10.1007/978-3-662-13187-9_18

Tsavkelova EA, Egorova MA, Leontieva MR, Malakho SG, Kolomeitseva GL, Netrusov AI (2016) *Dendrobium nobile* Lindl. seed germination in co-cultures with diverse

associated bacteria. Plant Growth Regul 80:79–91. doi: 10.1007/s10725-016-0155-1

Urdiain M, López-López A, Gonzalo C, Busse HJ, Langer S, Kämpfer P et al (2008) Reclassification of *Rhodobium marinum* and *Rhodobium pfennigii* as *Afifella marina* gen. nov. comb. nov. and *Afifella pfennigii* comb. nov., a new genus of photoheterotrophic *Alphaproteobacteria* and emended descriptions of *Rhodobium, Rhodobium orientis* and *Rhodobium gokarnense*. Syst Appl Microbiol 31:339-51. doi: 10.1016/j.syapm.2008.07.002

Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A (2015) The importance of the microbiome of the plant holobiont. New Phytol. 206:1196-206. doi: 10.1111/nph.13312

Venkata RV, Sasikala CH, Takaichi S, Ramana ChV (2010) *Roseomonas aestuarii* sp. nov., a bacteriochlorophyll-a containing alphaproteobacterium isolated from an estuarine habitat of India. Syst Appl Microbiol 33:198–203. doi: 10.1016/j.syapm.2009.09.004

Walker JJ, Spear JR, Pace NR (2005) Geobiology of a microbial endolithic community in the Yellowstone geothermal environment. Nature 434:1011-1014. doi: 10.1038/nature03447

Wallace PL, Hollis DG, Weaver RE, Moss CW (1990) Biochemical and characterization of pink-pigmented oxidative bacteria. J Clin Microbiol 28:689–693.

Wang C, Deng S, Liu X, Yao L, Shi C, Jiang J et al (2016) *Roseomonas eburnea* sp. nov., isolated from activated sludge. Int J Syst Evol Microbiol 66:385–390. doi: 10.1099/ijsem.0.000728

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261-7. doi: 10.1128/AEM.00062-07

Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, et al (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol 37:463–464. doi: org/10.1099/00207713-37-4-463

Wierzchos J, DiRuggiero J, Vítek P, Artieda O, Souza-Egipsy V, Škaloud P et al (2015) Adaptation strategies of endolithic chlorophototrophs to survive the hyper arid and extreme solar radiation environment of the Atacama Desert. Front Microbiol 6:934. doi: 10.3389/fmicb.2015.00934

Wirth JS, Whitman WB (2018) Phylogenomic analyses of a clade within the *Roseobacter* group suggest taxonomic reassignments of species of the genera *Aestuariivita*, *Citreicella, Loktanella, Nautella, Pelagibaca, Ruegeria, Thalassobius, Thiobacimonas* and *Tropicibacter*, and the proposal of six novel genera. Int J Syst Evol Microbiol 68: 2393–2411. doi: 10.1099/ijsem.0.002833

Woese CR (2004) A new biology for a new century. Microbiol Mol Biol Rev 68:173-86. doi: 10.1128/MMBR.68.2.173-186.2004

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088-90. doi: 10.1073/pnas.74.11.5088

Wu L, Ma J (2019) The Global Catalogue of Microorganisms (GCM) 10K type strain sequencing project: providing services to taxonomists for standard genome sequencing and annotation. Int J Syst Evol Microbiol 69:895-898. doi: 10.1099/ijsem.0.003276

Wu Y, Lai Q, Zhou Z, Liu C, Shao Z (2009) *Alcanivorax hongdengensis* sp. nov., an alkane-degrading bacterium isolated from surface seawater of the straits of Malacca and Singapore, producing a lipopeptide as its biosurfactant. Int J Syst Evol Microbiol 59:1474-9. doi: 10.1099/ijs.0.001552-0

Xie CH, Yokota A (2003) Phylogenetic analyses of *Lampropedia hyali*na based on the 16S rRNA gene sequence. J Gen Appl Microbiol 49:345-9. doi: 10.2323/jgam.49.345

Xu C, Liu Q, Diao B, Kan B, Song X, Li D (2009) Optimization of pulse-field gel electrophoresis for *Bartonella* subtyping. J Microbiol Methods 76:6-11. doi: 10.1016/j.mimet.2008.09.001

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H et al (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 47: W52-W58. doi:10.1093/nar/gkz333

Xu L, Dong Z, Chiniquy D, Pierroz G, Deng S, Gao C et al (2021) Genome-resolved metagenomics reveals role of iron metabolism in drought-induced rhizosphere microbiome dynamics. Nat Commun 12: 3209. doi:org/10.1038/s41467-021-23553-7

Yakimov MM, Golyshin PN, Lang S, Moore ER, Abraham WR, Lünsdorf H et al (1998) *Alcanivorax borkumensis* gen. nov., sp. nov., a new, hydrocarbon-degrading and surfactant-producing marine bacterium. Int J Syst Bacteriol 48:339-48. doi: 10.1099/00207713-48-2-339

Yamamoto S, Hamanaka K, Suemoto Y, Ono B, Shinoda S (1986) Evidence for the presence of a novel biosynthetic pathway for norspermidine in *Vibrio*. Can J Microbiol 32:99-103. doi: 10.1139/m86-021

Yamamura S, Yamashita M, Fujimoto N, Kuroda M, Kashiwa M, Sei K et al (2007) *Bacillus selenatarsenatis* sp. nov., a selenate and arsenate-reducing bacterium isolated from the effluent drain of a glass-manufacturing plant. Int J Syst Evol Microbiol 57:1060-1064. doi: 10.1099/ijs.0.64667-0

Yan ZF, Lin P, Li CT, Kook M, Wang QJ, Yi TH (2017) *Roseomonas hibiscisoli* sp. nov., isolated from the rhizosphere of Mugunghwa (*Hibiscus syriacus*). Int J Syst Evol Microbiol 67:2873–2878. doi: 10.1099/ ijsem.0.002036

Yang S, Li M, Lai Q, Li G, Shao Z (2018) *Alcanivorax mobilis* sp. nov., a new hydrocarbon-degrading bacterium isolated from deep-sea sediment. Int J Syst Evol Microbiol 68:1639-1643. doi: 10.1099/ijsem.0.002612

Yang SH, Tandon K, Lu CY, Wada N, Shih CJ, Hsiao SS et al (2019) Metagenomic, phylogenetic, and functional characterization of predominant endolithic green sulfur bacteria in the coral *Isopora palifera*. Microbiome 7:3. doi: 10.1186/s40168-018-0616-z

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH et al (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12: 635–645. doi:org/10.1038/nrmicro3330

Ye SH, Siddle KJ, Park DJ, Sabeti PC (2019) Benchmarking metagenomics tools for taxonomic classification. Cell 178:779-794. doi: 10.1016/j.cell.2019.07.010

Yoo SH, Weon HY, Noh HJ, Hong SB, Lee CM, Kim BY et al (2008) *Roseomonas aerilata* sp. nov., isolated from an air sample. Int J Syst Evol Microbiol 58:482–485. doi: 10.1099/ijs.0.65385-0

Yoon JH, Kang SJ, Park S, Oh TK (2005) Reclassification of the three species of the genus *Krokinobacter* into the genus *Dokdonia* as *Dokdonia genika* comb. nov., *Dokdonia diaphoros* comb. nov. and *Dokdonia eikasta* comb. nov., and emended description of the genus *Dokdonia* Yoon et al. 2005. Int J Syst Evol Microbiol 62:1896-1901. doi: 10.1099/ijs.0.035253-0

Yoon JH, Kang SS, Lee KC, Kho YH, Choi SH, Kang KH et al (2001) *Bacillus jeotgali* sp. nov., isolated from jeotgal, Korean traditional fermented seafood. Int J Syst Evol Microbiol 51:1087-1092. doi: 10.1099/00207713-51-3-1087

Yoon JH, Kang SJ, Oh HW, Oh TK (2007) *Roseomonas terrae* sp. nov. Int J Syst Evol Microbiol 57:2485–2488. doi: 10.1099/ijs.0.65113-0

Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H et al (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613-1617. doi: 10.1099/ijsem.0.001755

Yoon SH, Ha Sm, Lim J, Kwon S, Chun JA (2017) A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek 110:1281–1286. doi:org/10.1007/s10482-017-0844-4

Zeng LB, Wang D, Hu NY, Zhu Q, Chen K, Dong K et al (2017) A novel pan-genome reverse vaccinology approach employing a negative-selection strategy for screening surface-exposed antigens against leptospirosis. Front Microbiol 8:396. doi:10.3389/fmicb.2017.00396

Zeyad MT, Rajawat MVS, Kumar M, Malik A, Anas M, Ansari WA et al (2021) Role of microorganisms in plant adaptation towards climate change for sustainable

agriculture. In: Lone S.A., Malik A. (eds) Microbiomes and the Global Climate Change. Springer, Singapore. doi: org/10.1007/978-981-33-4508-9_14

Zhao LL, Deng Y, Sun Y, Liu HY, Yu LY, Zhang YQ (2020) *Roseomonas vastitatis* sp. nov. isolated from Badain Jaran desert in China. Int J Syst Evol Microbiol 70:1186–1191. doi: 10.1099/ijsem.0.003898

Zhao Y, Sun C, Zhao D, Zhang Y, You Y, Jia X et al (2018) PGAP-X: extension on pan-genome analysis pipeline. BMC Genomics 19:36. doi: 10.1186/s12864-017-4337-7

Zhang YQ, Yu LY, Wang D, Liu HY, Sun CH, Jiang W et al (2008) *Roseomonas vinacea* sp. nov., a Gram-negative coccobacillus isolated from a soil sample. Int J Syst Evol Microbiol 58:2070–2074. doi: 10.1099/ijs.0.65789-0

Zhang B, Han D, Korostelev Y, Yan Z, Shao N, Khrameeva E et al (2016) Changes in snoRNA and snRNA abundance in the human, chimpanzee, macaque, and mouse brain. Genome Biol Evol 8:840-50. doi: 10.1093/gbe/evw038

Zhang JY, Jiang XB, Zhu D, Wang XM, Du ZJ, Mu DS (2020) *Roseomonas bella* sp. nov., isolated from lake sediment. Int J Syst Evol Microbiol 70:5473–5478. doi: 10.1099/ijsem.0.004436

Zhang L, Jiao Y, Ling L, Wang H, Song W, Zhao T et al (2021) *Microbacterium stercoris* sp. nov., an indole acetic acid-producing actinobacterium isolated from cow dung. Int J Syst Evol Microbiol 71. doi: 10.1099/ijsem.0.005099

Zheng L, Fleith M, Giuffrida F, O'Neill BV, Schneider N (2019) Dietary polar lipids and cognitive development: a narrative review. Adv Nutr 10:1163-1176. doi: 10.1093/advances/nmz051

Zhang S, Yang Y, Li J, Qin J, Zhang W, Huang W et al (2018) Physiological diversity of orchids. Plant Divers 40:196-208. doi: 10.1016/j.pld.2018.06.003

Zhang Y, Ren H, Zhang G (2014) *Microbacterium hydrothermale* sp. nov., an actinobacterium isolated from hydrothermal sediment. Int J Syst Evol Microbiol 64:3508-3512. doi: 10.1099/ijs.0.061697-0

Zhu D, Adebisi WA, Ahmad F, Sethupathy S, Danso B, Sun J (2020) Recent development of extremophilic bacteria and their application in biorefinery. Front Bioeng Biotechnol 8:483. doi: 10.3389/fbioe.2020.00483

Zuo G, Hao B (2015) CVTree3 Web Server for whole-genome-based and alignment-free prokaryotic phylogeny and taxonomy. Genomics Proteomics Bioinformatics. 13:321-31. doi: 10.1016/j.gpb.2015.08.004

PUBLICATIONS

7. PUBLICATIONS

1. **Rai A**, Indu, Smita N et al (2019) Emerging Concepts in Bacterial Taxonomy. In: Satyanarayana T, Johri B, Das S. (eds) Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications. Springer, Singapore.
2. **Rai A**, Smita N, Suresh G, et al (2019) *Paracoccus aeridis* sp. nov., an indole producing bacterium isolated from the rhizosphere of an orchid, *Vanda* sp. Int J Syst Evol Microbiol 70:1720-1728.
3. **Rai A**, Uppada J, Sasikala Ch, et al (2020) Genus *Paraclostridium*. In: Bergey's Manual of Systematics of Archaea and Bacteria. Wiley, USA.
4. **Rai A**, Smita N, Shabbir A et al (2021) *Mesobacillus aurantius* sp. nov., isolated from an orange-colored pond near a solar saltern. Arch Microbiol 203:1499-1507.
5. **Rai A,** Jagadeshwari U, Gupta D, et al (2021) Phylotaxogenomics for the reappraisal of the genus *Roseomonas* with the creation of six new genera. Front Microbiol 12:1787.
6. Jagadeshwari U, **Rai A,** Sasikala Ch, et al (2021) Genus *Marispirochaetia*. In: Bergey's Manual of Systematics of Archaea and Bacteria. Wiley, USA.
7. **Rai A,** Jagadeshwari U, Sasikala Ch, et al (2021) Genus *Paeniclostridium* In: Bergey's Manual of Systematics of Archaea and Bacteria. Wiley, USA. (Under review)
8. **Rai A,** Suresh G, Biswas R, et al (2021) Genome revisitation justifies the classification of the genus *Alcanivorax* into *Alloalcanivorax* gen. nov. and *Isoalcanivorax* gen. nov. Int J Syst Evol Microbiol. (Under review)
9. **Rai A**, Jagadeeshwari U, Deepshika G, et al (2021) *Neoroseomonas marina* sp. nov., isolated from a beach sand. Curr Microbiol. (Under review)

CONFERENCES AND SEMINARS

1. Presented a poster at the "8th International conference on Photosynthesis and Hydrogen Energy Research for Sustainability" during 30th Oct to 3rd Nov, 2017 at UoH, Hyderabad.
2. Presented poster at the "59th Annual Conference of Association of Microbiologists of India & International Symposium on host-pathogen interaction" during 9th to 12th December, 2018 at UoH, Hyderabad.
3. Presented an oral presentation at "National Conference on Frontiers in Plant Biology" during 31st Jan to 1st Feb, 2019 at UoH, Hyderabad.

Emerging Concepts in Bacterial Taxonomy

1

Anusha Rai, Indu, N. Smita, G. Deepshikha, K. Gaurav, K. Dhanesh, G. Suresh, Ch. Sasikala, and Ch. V. Ramana

Abstract

Bacterial taxonomy has progressed over the years by virtue of the brisk and competent scientific developments. Ground-breaking molecular techniques have added an edge in the phylogenetic studies, resulting in the quality description of the taxa under studies. New avenues are rapidly developing whose validation has always been embraced and included, which will assist in resolution. It began with the simple application of objective procedures for classification, and now we have arrived at the genome-based taxonomy. This pedantic step has led to the meticulous examination and served to reconcile certain conflicts of the status of the taxa. This field is dynamic and is exploring more options like proteomics and metabolomics in gaining more insights into the lineal heritage. Even though there has been a significant change and addition, there is an ever-growing need for a comprehensive study, which would thread all the attributes together into one functional unit of classification. In this review, we examine the paradigm shift from traditional taxonomy to integrated taxonomy useful in the characterisation of bacteria which in addition aids in the identity of biotechnological targets.

Keywords

Bacterial taxonomy · Polyphasic · Phylogenomics · Integrated taxonomy · Average nucleotide sequence index (ANI)

A. Rai · Indu · N. Smita · G. Deepshikha · K. Gaurav · K. Dhanesh · G. Suresh · C. V. Ramana (✉)
Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana, India

C. Sasikala (✉)
Centre for Environment, Institute of Science & Technology, JNT University Hyderabad, Hyderabad, Telangana, India

T. Satyanarayana et al. (eds.), *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications*, https://doi.org/10.1007/978-981-13-8315-1_1

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION
Rai *et al.*, *Int. J. Syst. Evol. Microbiol.* 2020;70:1720–1728
DOI 10.1099/ijsem.0.003962

MICROBIOLOGY SOCIETY

Paracoccus aeridis sp. nov., an indole-producing bacterium isolated from the rhizosphere of an orchid, *Aerides maculosa*

Anusha Rai[1], Smita N[1], Suresh G[1], Shabbir A[1], Deepshikha G[1], Sasikala Ch[2,*] and Ramana Ch.V[1,*]

Abstract

A Gram-stain-negative, non-motile, coccoid-shaped, catalase- and oxidase-positive, non-denitrifying, neutrophilic bacterium designated as strain JC501^T was isolated from an epiphytic rhizosphere of an orchid, *Aerides maculosa*, growing in the Western Ghats of India. Phylogenetic analyses based on the 16S rRNA gene sequence indicated that strain JC501^T belonged to the genus *Paracoccus* and had the highest levels of sequence identity with *Paracoccus marinus* KKL-A5^T (98.9 %), *Paracoccus contaminans* WPAn02^T (97.3%) and other members of the genus *Paracoccus* (<97.3 %). Strain JC501^T produced indole-3 acetic acid and other indole derivatives from tryptophan. The dominant respiratory quinone was Q-10 and the major fatty acid was $C_{18:1}\omega7c/C_{18:1}\omega6c$, with significant quantities of $C_{18:1}\omega9c$, $C_{17:0}$ and $C_{16:0}$. The polar lipids of strain JC501^T comprised phosphatidylglycerol, phosphatidylcholine, diphosphatidylglycerol, an unidentified glycolipid, two unidentified aminolipids, two unidentified lipids and four unidentified phospholipids. The genome of strain JC501^T was 3.3 Mbp with G+C content of 69.4 mol%. For the resolution of the phylogenetic congruence of the novel strain, the phylogeny was also reconstructed with the sequences of eight housekeeping genes. Based on the results of phylogenetic analyses, low (<85.9%) average nucleotide identity, digital DNA–DNA hybridization (<29.8%), chemotaxonomic analysis and physiological properties, strain JC501^T could not be classified into any of the recognized species of the genus *Paracoccus*. Strain JC501^T represents a novel species, for which the name *Paracoccus aeridis* sp. nov. is proposed. The type strain is JC501^T (=LMG 30532^T=NBRC 113644^T).

Bacteria residing in the rhizosphere of the plants play a vital role in the holistic development of the plant system [1–3]. Epiphytic orchid-associated bacteria have functional and ecological roles in the development of their host plant [4]. Epiphytes do not interact directly with the soil or its microbiota and thus constitute a unique system of ecology. Therefore, epiphytes have their own distinctive structural system for their sustenance, wherein they take up the nutrients and moisture from the atmosphere on the surface of the host plant aided by the microbial association [5, 6]. While investigating this unique diversity and its subsequent role in the development of the orchids, we have isolated strain JC501^T from the rhizosphere of an epiphytic orchid (*Aerides maculosa*). This strain belongs to the genus *Paracoccus* based on 16S rRNA gene sequence analysis. The genus *Paracoccus* was first described by Davis and his co-workers in 1969 [7] and belongs to the family '*Rhodobacteraceae*' of the class *Alphaproteobacteria* in the phylum *Proteobacteria*. There are more than 50 species of *Paracoccus* with validly published names (www.bacterio.net). Members have been isolated from environmental samples such as soil [8, 9], sediment [10, 11], water [12, 13], sludges [14, 15], foodstuffs [16], clinical specimens [17] and insects [18]. *Paracoccus halotolerans* [19], *Paracoccus salipaludis* [20], *Paracoccus fontiphilus* [13], *Paracoccus alimentarius* [16], *Paracoccus endophyticus* [21], *Paracoccus haematequi* [22] and *Paracoccus nototheniae* [23] are the valid names published during the year 2018–2019, while *Paracoccus jeotgali* [24] and *Paracoccus indicus* [25] are effective publications. Members of this genus are Gram-stain-negative, mostly non-motile and chemoorganotrophs [26]. Their major fatty acid is $C_{18:1}\omega7c$ and they are metabolically versatile [27]. The members of the genus *Paracoccus* have a genome size ranging from 2.9 to 5.6

Author affiliations: [1]Department of Plant Sciences, School of Life Sciences, University of Hyderabad P.O. Central University, Hyderabad 500046, India; [2]Bacterial Discovery Laboratory, Centre for Environment, Institute of Science and Technology, J. N. T. University Hyderabad, Kukatpally, Hyderabad 500085, India.
***Correspondence:** Sasikala Ch, sasi449@yahoo.ie; Ramana Ch.V, cvr449@gmail.com
Keywords: *Paracoccus*; epiphytic orchid; Proteobacteria; indole.
Abbreviations: AIC, Akaike information criterion; ANI, average nucleotide identity; dDDH, digital DNA–DNA hybridization; IAA, indole-3-acetic acid; IAM, indole-3-acetamide; LCB, local collinear block; ML, maximum-likelihood; MLSA, multilocus sequence analysis; NA, nutrient agar.
The GenBank/EMBL/DDBJ accession number for the 16S rRNA gene sequence of strain JC501^T is LT799401. The Whole Genome Shotgun project is SELD00000000. The genome sequence of *P. marinus* NBRC 100637^T is VJYZ00000000.
Five supplementary tables and ten supplementary figures are available with the online version of this article.

Archives of Microbiology
https://doi.org/10.1007/s00203-020-02146-w

ORIGINAL PAPER

Mesobacillus aurantius sp. nov., isolated from an orange-colored pond near a solar saltern

Anusha Rai[1] · N. Smita[1] · A. Shabbir[1] · U. Jagadeeshwari[2] · T. Keertana[1] · Ch. Sasikala[2] · Ch. V. Ramana[1]

Received: 23 June 2020 / Revised: 3 November 2020 / Accepted: 3 December 2020

Abstract

An endospore producing, strict aerobic, Gram-stain-positive, orange-colored colony forming bacterium designated as strain JC1013^T was isolated from an orange pond near a solar saltern of Tamil Nadu, India. Phylogenetic analysis of the 16S rRNA gene sequences indicated that strain was affiliated to the family *Bacillaceae* of the phylum *Firmicutes*. Strain showed highest 16S rRNA gene sequence identity of 98.7% with *Mesobacillus selenatarsenatis* SF-1 ^{T} and below 98.3% with other members of the genus *Mesobacillus*. Strain JC1013^T produced carotenoid pigments and indole compounds. Major cellular fatty acids of strain JC1013^T were iso-$C_{15:0,}$ anteiso-$C_{15:0}$, $C_{16:0}$ 3-OH, iso-$C_{17:0}\omega10c$ and summed feature 4 (iso-$C_{17:1}$ I/ anteisoB). Polar lipids were diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, two unidentified aminolipids and four unidentified phospholipids. Strain JC1013^T constituted *m*-diaminopimelic acid as diagnostic cell wall amino acids. MK-7 is the predominant menaquinone of strain JC1013^T. The genome size of strain JC1013^T was 4.6 Mbp and its G + C content was 42.7 mol%. For the affirmation of strain's taxonomic status, a detailed phylogenomic study was done. Based on the phylogenetic analyses, low ANI (84.6%), AAI (88.5%) values, in-silico DDH (< 29%) value, morphological, physiological and chemo-taxonomical characteristics, strain JC1013^T was clearly distinguished from the nearest phylogenetic neighbor, *Mesobacillus selenatarsenatis* SF-1^T to conclude that it is a new species of the genus *Mesobacillus*. We propose the name as *Mesobacillus aurantius* with type strain JC1013^T (= NBRC 114146^T = KACC 21451 ^{T}).

Keywords *Mesobacillus* · Sp. nov. · Endospore · Salt tolerant

Communicated by Erko Stackebrandt.

The GenBank/EMBL/DDBJ accession number for 16S rRNA gene sequence of strain JC1013^T is LS998022**.** The GenBank/EMBL/DDBJ accession for the whole genome shotgun sequence is JAAVUM000000000.

Supplementary information The online version contains supplementary material at (doi:https://doi.org/10.1007/s00203-020-02146-w).

✉ Ch. Sasikala
sasikala.ch@gmail.com

✉ Ch. V. Ramana
cvr449@gmail.com

[1] Department of Plant Sciences, School of Life Sciences, University of Hyderabad, P.O. Central University, Hyderabad 500 046, India

[2] Bacterial Discovery Laboratory, Centre for Environment, Institute of Science and Technology, J. N. T. University H, Kukatpally, Hyderabad 500 085, India

Abbreviations

NCBI	National centre for biotechnology information
GCM	The global catalogue of microorganisms
ANI	Average nucleotide identity
AAI	Average amino identity
*d*DDH	Digital DNA–DNA hybridization
BLAST	Basic local alignment search tool
MUSCLE	Multiple sequence comparison by log-expectation
TLC	Thin-layer chromatography
HPLC	High-pressure liquid chromatography
NBRC	Biological resource center
NITE	KACC, Korean agricultural culture collection
CMC	Carboxymethyl cellulose

Introduction

Hyper-saline environments like salt pan lakes are hubs for unique ecological layouts spanning diverse growth conditions (pH, alkalinity, salinity and temperature) and

Published online: 04 January 2021

frontiers in Microbiology

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fmicb.2021.677842

Phylotaxogenomics for the Reappraisal of the Genus *Roseomonas* With the Creation of Six New Genera

Anusha Rai[1†], Uppada Jagadeeshwari[2†], Gupta Deepshikha[1], Nandardhane Smita[1], Chintalapati Sasikala[2] and Chintalapati Venkata Ramana[1*]*

[1] Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, India, [2] Bacterial Discovery Laboratory, Centre for Environment, Institute of Science and Technology (IST), Jawaharlal Nehru Technological (JNT) University Hyderabad, Hyderabad, India

OPEN ACCESS

Edited by:
Iain Sutcliffe,
Northumbria University,
United Kingdom

Reviewed by:
Wen-Jun Li,
Sun Yat-sen University, China
Helene Marchandin,
Centre Hospitalier Universitaire de Nîmes, France

***Correspondence:**
Chintalapati Sasikala
sasi449@yahoo.ie;
sasikala.ch@gmail.com
Chintalapati Venkata Ramana
cvr449@gmail.com

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 08 March 2021
Accepted: 08 June 2021
Published: 13 August 2021

Citation:
Rai A, Jagadeeshwari U, Deepshikha G, Smita N, Sasikala C and Ramana CV (2021) Phylotaxogenomics for the Reappraisal of the Genus Roseomonas With the Creation of Six New Genera. Front. Microbiol. 12:677842. doi: 10.3389/fmicb.2021.677842

The genus *Roseomonas* is a significant group of bacteria which is invariably of great clinical and ecological importance. Previous studies have shown that the genus *Roseomonas* is polyphyletic in nature. Our present study focused on generating a lucid understanding of the phylogenetic framework for the re-evaluation and reclassification of the genus *Roseomonas.* Phylogenetic studies based on the 16S rRNA gene and 92 concatenated genes suggested that the genus is heterogeneous, forming seven major groups. Existing *Roseomonas* species were subjected to an array of genomic, phenotypic, and chemotaxonomic analyses in order to resolve the heterogeneity. Genomic similarity indices (*d*DDH and ANI) indicated that the members were well-defined at the species level. The Percentage of Conserved Proteins (POCP) and the average Amino Acid Identity (AAI) values between the groups of the genus *Roseomonas* and other interspersing members of the family *Acetobacteraceae* were below 65 and 70%, respectively. The pan-genome evaluation depicted that the pan-genome was an open type and the members shared 958 core genes. This claim of reclassification was equally supported by the phenotypic and chemotaxonomic differences between the groups. Thus, in this study, we propose to re-evaluate and reclassify the genus *Roseomonas* and propose six novel genera as *Pararoseomonas* gen. nov., *Falsiroseomonas* gen. nov., *Paeniroseomonas* gen. nov., *Plastoroseomonas* gen. nov., *Neoroseomonas* gen. nov., and *Pseudoroseomonas* gen. nov.

Keywords: phylotaxogenomics, average amino acid Identity (AAI), percentage of conserved proteins (POCP), reclassification, *Roseomonas*

Abbreviations: NCBI, National Center for Biotechnology Information; ANI, Average Nucleotide Identity; AAI, Average Amino acid Identity; POCP, Percentage of Conserved Proteins; *d*DDH, digital DNA-DNA Hybridization; BLAST, Basic Local Alignment Search Tool; MUSCLE, MUltiple Sequence Comparison by Log-Expectation; KCTC, Korean Collection for Type Cultures; CGMCC, China General Microbiological Culture Collection Centre; ATCC, American Type Culture Collection; CIP, Institute Pasteur Collection; DSM, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH; JCM, Japan Collection of Microorganisms-RIKEN BioResource Center; KACC, Korean Agricultural Culture Collection; NBRC, NITE Biological Resource Center; LMG, Laboratorium voor Microbiologie; MCCC, Marine Culture Collection of China; GDMCC, Guangdong Microbial Culture Collection Center; CECT Colección Española de Cultivos Tipo; KEMB, Korea Environmental Microorganisms Bank; CCUG, Culture Collection University of Göteborg; BCC, BIOTEC Culture Collection; UBCG, Up-to-date Bacterial Core Gene set.

Genus

Firmicutes/Clostridia/Clostridiales/Peptostreptococcaceae/

Paraclostridium

Sasi Jyothsna et al. 2016

Anusha Rai and Ch V. Ramana, *Department of Plant Sciences, University of Hyderabad, P.O. Central University, Hyderabad, India*

Jagadeeshwari Uppada and Ch Sasikala, *Centre for Environment, Institute of Science and Technology, J. N. T. University Hyderabad, Hyderabad, India*

Edited by: Fred A. Rainey, *University of Alaska Anchorage, Anchorage, AK, USA*

Pa.ra.clos.tri'di.um. Gr. prep. *para* next to, resembling; N.L. neut. n. *Clostridium* a bacterial genus; N.L. neut. n. *Paraclostridium* next to *Clostridium*.

Cells are Gram-stain-positive and rod shaped. Obligate anaerobes thriving in the mesophilic conditions belonging to the class *Clostridia* and family *Peptostreptococcaceae*. Cells are motile and reproduce by binary fission. Members produce endospores. Catalase- and oxidase-negative. Metabolic activities such as indole and H_2S production, starch and gelatin hydrolysis, and nitrate reduction may vary within species. Cells grow on a number of organic substrates even without the supplementation of growth factors and NaCl. $C_{16:0}$ is the major fatty acid, with $C_{18:0}$, $C_{18:1}$ ω7*c*, $C_{17:0}$, $C_{16:1}$ ω9*c*, and *iso*-$C_{16:0}$ being the minor fatty acids. The DNA G+C content is 28–29.3 mol%. Cells contain diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine, and unidentified amino lipids majorly as the polar lipids. Two species have been validly published under this genus namely, *Paraclostridium bifermentans* and *Paraclostridium benzoelyticum*. Mostly isolated from soil samples, marine habitats, polluted waters, clinical specimen such as wounds, blood, and ulcers and occasionally from humans intestinal microbiota. *P. bifermentans* is a rare cause of infection in humans which can be fatal. Antimicrobial compounds are produced.

DNA G+C content (mol%): 28–29.3.

Type species: **Paraclostridium bifermentans** (Weinberg and Séguin 1918) Sasi Jyothsna et al. 2016 (basonym: **Clostridium bifermentans** Weinberg and Séguin 1918 (Approved Lists 1980).

Members of the genus *Paraclostridium* are **Gram-stain-positive** and **rod shaped**. Cells are **motile** with either **peritrichous** flagella **or pseudo** filaments, reproduce by **binary fission**, and produce **endospores**. They are **obligate anaerobes** belonging to the phylum ***Firmicutes*, class *Clostridia***, and **family *Peptostreptococcaceae***. Members thrive optimally in **mesophilic** conditions. **Catalase-** and **oxidase-negative**. Metabolic activities such as **indole** and **H_2S production**, **starch** and **gelatin hydrolysis**, and **nitrate reduction** vary within the species. Cells grow on a number of organic substrates even without the supplementation of growth factors and NaCl. **$C_{16:0}$** is the major fatty acid, with **$C_{18:0}$**, **$C_{18:1}$ ω7*c*, $C_{17:0}$**, **$C_{16:1}$ ω9*c***, and ***iso*-$C_{16:0}$** being the minor fatty acids. The DNA G+C content is 28–29.3 mol%. **Diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine,** and **unidentified amino lipids** are majorly seen to constitute

DOI: 10.1002/9781118960608.gbm01635. Published by John Wiley & Sons, Inc., in association with Bergey's Manual Trust.

Genus

Spirochaetes/Spirochaetia/Spirochaetales/Spirochaetaceae/

Marispirochaeta

Shivani et al. 2017[VP]

Jagadeeshwari Uppada, *J. N. T. University Hyderabad, Hyderabad, India*
Anusha Rai, *Department of Plant Sciences, University of Hyderabad, P.O. Central University, Hyderabad, India*
Ch. Sasikala, *J. N. T. University Hyderabad, Hyderabad, India*
Ch. V Ramana, *Department of Plant Sciences, University of Hyderabad, P.O. Central University, Hyderabad, India*

Edited by: Xue-Wei Xu, *Second Institute of Oceanography, MNR, Hangzhou, China*

Marispirochaeta (Ma.ri.spi.ro.chaéta. L. neut. n. *mare*, the sea; N.L. fem. n. *Spirochaeta* a bacterial genus; N.L. fem. n. *Marispirochaeta* a marine *Spirochaeta*).

Cells are helical in structure and Gram-stain-negative, motile with helical movement, and obligate anaerobes. Members of the class Spirochaete, order Spirochaetales, and family *Spirochaetaceae*. Production of the sphaeroplasts occurs during the stationary phase or during an unfavorable growth condition. Catalase- and oxidase-negative. Indole production from L-tryptophan and gelatin hydrolysis is negative. Starch hydrolysis is positive. Exhibits chemoorganoheterotrophic mode of nutrition. Cells grow on a number of organic substrates. Traces of yeast extract are required for its growth. Glucose fermentation is observed in the member. The major fatty acids are $C_{14:0}$, $C_{16:0}$, and iso-$C_{15:0}$. Cells contain diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine as the major polar lipids. The DNA G+C content is 53.6–54.9 mol%. The only validly published species under this genus is *Marispirochaeta aestuarii*. One species of the rank *candidatus* is also described, that is, *Candidatus* Marispirochaeta associata. Both the species were isolated from sediment samples of the marine coasts of Gujarat, India.

DNA G+C content (mol%): 53.6–54.9 (Genome).

Type species: **Marispirochaeta aestuarii** Shivani et al. 2017[VP].

Candidatus species: *Candidatus* Marispirochaeta associata Shivani et al. 2016.

Members of the genus *Marispirochaeta* are **helical** in structure (Figures 1 and 2). In nature, it is **motile** with the helical movement. **Gram-stain-negative** and **nonflagellated**. They are **obligate anaerobes** belonging to the phylum ***Spirochaetes*, class *Spirochaetia*, order *spirochaetales***, and family ***Spirochaetaceae*. Sphaeroplasts** are formed during the stationary phase or under an unfavorable growth condition. Their doubling time period is 2–4 h. Continuous subculturing at the fifth day is required for the viability of the cells. Usually, the cells fail to grow in the solid medium containing agar, agarose, or the gellan gum. They are **catalase-** and **oxidase-negative**. Members are **mesophilic** and **slightly halophilic** in nature. Its growth mode is **chemoorganotrophic**. Metabolic activities such as **indole production** from L-tryptophan and **gelatin hydrolysis** are negative. **Starch hydrolysis** is positive. Cells utilize a number of organic substrates with the supplementation of traces of yeast extract for

DOI: 10.1002/9781118960608.gbm01938. Published by John Wiley & Sons, Inc., in association with Bergey's Manual Trust.

Diversity of root associated bacteria of Orchids and phylotaxogenomics of a few bacteria

by Anusha Rai

Submission date: 30-Mar-2022 11:39AM (UTC+0530)
Submission ID: 1796743631
File name: Anusha_Rai_16LPPH13.pdf (749.98K)
Word count: 37635
Character count: 203259

Diversity of root associated bacteria of Orchids and phylotaxogenomics of a few bacteria

ORIGINALITY REPORT

28%	24%	21%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.frontiersin.org Internet Source	12%
2	Anusha Rai, N. Smita, A. Shabbir, U. Jagadeeshwari, T. Keertana, Ch. Sasikala, Ch. V. Ramana. "Mesobacillus aurantius sp. nov. isolated from an orange-colored pond near a solar saltern", Archives of Microbiology, 2021 Publication	2%
3	www.ncbi.nlm.nih.gov Internet Source	2%
4	link.springer.com Internet Source	1%
5	dokumen.pub Internet Source	1%
6	Anusha Rai, Uppada Jagadeeshwari, Gupta Deepshikha, Nandardhane Smita, Chintalapati Sasikala, Chintalapati Venkata Ramana. "Phylotaxogenomics for the Reappraisal of the Genus Roseomonas With the Creation of	1%

These papers published by the student itself

Ch. Venkata Ramana 12/4/2022

Prof. Ch. Venkata Ramana Ph.D
Department of Plant Sciences School of Life Sciences
University of Hyderabad, P.O. Central University,
Hyderabad - 500 046, India

These papers published by student itself

Ch. Venkata Ramana 12/4/2022

Prof. Ch. Venkata Ramana Ph.D
Department of Plant Sciences School of Life Sciences
University of Hyderabad, P.O. Central University,
Hyderabad - 500 046, India

	Six New Genera", Frontiers in Microbiology, 2021 Publication	
7	Submitted to University of Hyderabad, Hyderabad Student Paper	1 %
8	www.science.gov Internet Source	1 %
9	"Bergey's Manual® of Systematic Bacteriology", Springer Nature, 2012 Publication	<1 %
10	ijs.microbiologyresearch.org Internet Source	<1 %
11	lpsn.dsmz.de Internet Source	<1 %
12	ijs.sgmjournals.org Internet Source	<1 %
13	Submitted to Jawaharlal Nehru Technological University Student Paper	<1 %
14	www.microbiologyresearch.org Internet Source	<1 %
15	v3r.esp.org Internet Source	<1 %

16	"The Prokaryotes", Springer Science and Business Media LLC, 2014 Publication	<1 %
17	fjfsdata01prod.blob.core.windows.net Internet Source	<1 %
18	escholarship.org Internet Source	<1 %
19	ijsb.sgmjournals.org Internet Source	<1 %
20	www.scipress.com Internet Source	<1 %
21	intl-ijs.sgmjournals.org Internet Source	<1 %
22	assets.researchsquare.com Internet Source	<1 %
23	www.mdpi.com Internet Source	<1 %
24	hdl.handle.net Internet Source	<1 %
25	Bergey's Manual® of Systematic Bacteriology, 2005. Publication	<1 %
26	docksci.com Internet Source	<1 %

27	pdffox.com Internet Source	<1 %
28	Noel R. Krieg, Wolfgang Ludwig, Jean Euzéby, William B. Whitman. "Chapter 3 Phylum XIV. Bacteroidetes phyl. nov.", Springer Science and Business Media LLC, 2010 Publication	<1 %
29	new.esp.org Internet Source	<1 %
30	Garrity, George M., Julia A. Bell, and Timothy Lilburn. "Class I. Alphaproteobacteria class. nov.", Bergey's Manual® of Systematic Bacteriology, 2005. Publication	<1 %
31	repository.helmholtz-hzi.de Internet Source	<1 %
32	The Prokaryotes, 2014. Publication	<1 %
33	Submitted to University of Queensland Student Paper	<1 %
34	Karl-Heinz Schleifer. "Phylum XIII. Firmicutes Gibbons and Murray 1978, 5 (Firmacutes [sic] Gibbons and Murray 1978, 5)", Systematic Bacteriology, 2009 Publication	<1 %
	Submitted to National Chung Hsing University	

35	Student Paper	<1%
36	Submitted to Incheon National University Student Paper	<1%
37	ediss.uni-goettingen.de Internet Source	<1%
38	cyberleninka.org Internet Source	<1%
39	ebin.pub Internet Source	<1%
40	academic.oup.com Internet Source	<1%
41	Submitted to Ana's Test Account Student Paper	<1%
42	www.unboundmedicine.com Internet Source	<1%
43	www.biorxiv.Org Internet Source	<1%
44	www.freepatentsonline.com Internet Source	<1%
45	Submitted to Pacific University Student Paper	<1%
46	brcdownloads.vbi.vt.edu Internet Source	<1%

47	mic.sgmjournals.org Internet Source	<1 %
48	zh.wikipedia.org Internet Source	<1 %
49	pub.epsilon.slu.se Internet Source	<1 %
50	www.plantcell.org Internet Source	<1 %
51	Submitted to University of Newcastle upon Tyne Student Paper	<1 %
52	acikbilim.yok.gov.tr Internet Source	<1 %
53	irgu.unigoa.ac.in Internet Source	<1 %
54	scholarworks.bwise.kr Internet Source	<1 %
55	turpaduunari.fi Internet Source	<1 %
56	www.piplinks.org Internet Source	<1 %
57	www.ppjonline.org Internet Source	<1 %

58	Hyo-Jin Lee, Kyung-Sook Whang. "Roseomonas rosulenta sp. nov., isolated from rice paddy soil", Research Square Platform LLC, 2022 Publication	<1 %
59	Kefu Yu, Guanghua Wang, Jianfeng Liu, Yuanjin Li, Jin Li, Jixin Luo, Biao Chen, Zhiheng Liao, Hongfei Su, Jiayuan Liang. "Description of Prasinibacter corallicola gen. nov., sp. nov., a zeaxanthin-producing bacterium isolated from stony coral Porites lutea", Research Square Platform LLC, 2021 Publication	<1 %
60	etd.aau.edu.et Internet Source	<1 %
61	www.biorxiv.org Internet Source	<1 %
62	1library.net Internet Source	<1 %
63	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
64	Submitted to University of Venda Student Paper	<1 %
65	jcoagri.uobaghdad.edu.iq Internet Source	<1 %

66	Joana Bondoso. "Epiphytic Planctomycetes on macroalgae: insights of their morphology, physiology and ecology", Repositório Aberto da Universidade do Porto, 2014. Publication	<1 %
67	Submitted to University of Central Florida Student Paper	<1 %
68	ouci.dntb.gov.ua Internet Source	<1 %
69	Submitted to Pukyong National University Student Paper	<1 %
70	Submitted to University of Pune Student Paper	<1 %
71	priam.prabi.fr Internet Source	<1 %
72	www.faqs.org Internet Source	<1 %
73	Bo Dong, Xiaoqian Lin, Xiaohuan Jing, Tongyuan Hu et al. "A Bacterial Genome And Culture Collection of Gut Microbial in Weanling Piglet", Research Square Platform LLC, 2021 Publication	<1 %
74	Goodfellow, Michael. "Phylum XXVI. Actinobacteria phyl. nov.", Bergey's Manual® of Systematic Bacteriology, 2012.	<1 %

Publication

75	Submitted to KYUNG HEE UNIVERSITY Student Paper	<1 %
76	Ram Hari Dahal, Jungmin Kim, Dhiraj Kumar Chaudhary, Dong-Uk Kim, Hyein Jang, Jaisoo Kim. "Genome Mining Revealed Polyhydroxybutyrate Biosynthesis by Ramlibacter Agri Sp. Nov., Isolated from Agriculture Soil in Korea", Research Square Platform LLC, 2021 Publication	<1 %
77	Submitted to Temple University Student Paper	<1 %
78	coralreefalgaelab.org Internet Source	<1 %
79	file.scirp.org Internet Source	<1 %
80	journals.plos.org Internet Source	<1 %
81	Andrew M. Boddicker, Annika C. Mosier. " Unexpected versatility in the metabolism and ecophysiology of globally relevant nitrite-oxidizing bacteria ", Cold Spring Harbor Laboratory, 2018 Publication	<1 %
82	biosoc.pk Internet Source	

		<1 %
83	pdf-s3.xuebalib.com:1262 Internet Source	<1 %
84	plant.ucr.edu Internet Source	<1 %
85	thesesups.ups-tlse.fr Internet Source	<1 %
86	www.islandscholar.ca Internet Source	<1 %
87	www.nature.com Internet Source	<1 %
88	Lizah T. van der Aart, Imen Nouinoui, Alexander Kloosterman, José Mariano Ingual et al. " Classification of the gifted natural product producer sp. nov. by polyphasic taxonomy ", Cold Spring Harbor Laboratory, 2018 Publication	<1 %
89	Submitted to University of Bristol Student Paper	<1 %
90	publikationen.bibliothek.kit.edu Internet Source	<1 %
91	Danhua Li, Rexiding Abuduaini, Mengxuan Du, Yujing Wang et al. "Alkaliphilus flagellata sp.	<1 %

nov., Butyricicoccus intestinisimiae sp. nov., Clostridium mobile sp. nov., Clostridium simiarum sp. nov., Dysosmobacter acutus sp. nov., Paenibacillus brevis sp. nov., Peptoniphilus ovalis sp. nov., and Tissierella simiarum sp. nov., isolated from monkey feces", Cold Spring Harbor Laboratory, 2021
Publication

92 Submitted to Hankuk University of Foreign Studies <1%
Student Paper

93 Jennah E. Dharamshi, Natalia Gaarslev, Karin Steffen, Tom Martin, Detmer Sipkema, Thijs J. G. Ettema. "Genomic diversity and biosynthetic capabilities of sponge-associated chlamydiae", Cold Spring Harbor Laboratory, 2021 <1%
Publication

94 Khongsai L Lhingjakim, Nandardhane Smita, Gaurav Kumar, Uppada Jagadeeshwari et al. "Paludisphaera rhizosphaereae sp. nov., a new member of the family Isosphaeraceae,isolated from the rhizosphere soil of Erianthus ravennae", Research Square Platform LLC, 2022 <1%
Publication

95 Yu-Sheng Wang, Yu-Yan Yue, Yi-Lin Bai, Xu-Yang Zhang, Shuo Wang, Zong-Jun Du. <1%

"Rhodohalobacter sulfatireducens sp. nov., isolated from a marine solar saltern", Research Square Platform LLC, 2022
Publication

96	epdf.pub Internet Source	<1 %
97	open.bu.edu Internet Source	<1 %
98	www.hrpub.org Internet Source	<1 %
99	Eprints.ncl.ac.uk Internet Source	<1 %
100	Ndl.ethernet.edu.et Internet Source	<1 %
101	Sabrina Naud, Issam Hasni, Sara Bellali, Hoang Thong Kieu et al. "Genomic and phenotypic description of Streptococcus resistens sp. nov., Streptococcus buccae sp. nov. and Streptococcus mediterraneus sp. nov., three new members of the Streptococcus genus isolated from the human oral cavity", Research Square, 2021 Publication	<1 %
102	Samad Ashrafi, Nemanja Kuzmanović, Sascha Patz, Ulrike Lohwasser et al. " Two new species isolated from root nodules of	<1 %

common sainfoin ( ) show different plant colonization strategies ", Cold Spring Harbor Laboratory, 2022
Publication

103	Yingying Jiang, Chaochao Zheng, Tianfei Yu, Jing Li, Jiamin Ai, Maiping Li, Xiaodong Liu, Zhenshan Deng. "Rhodococcus Yananensis Sp. Nov., Isolated From Microbial Fermentation Bed Material From a Pig Farm", Research Square Platform LLC, 2021 Publication	<1 %
104	Yuliya V Zakalyukina, Ilya A Osterman, Jacqueline Wolf, Meina Neumann-Schaal, Imen Nouioui, Mikhail Biryukov. "Amycolatopsis Camponoti Sp. Nov., New Tetracenomycin-Producing Actinomycete Isolated from Carpenter Ant Camponotus Vagus", Research Square Platform LLC, 2021 Publication	<1 %
105	bmcmicrobiol.biomedcentral.com Internet Source	<1 %
106	bmcplantbiol.biomedcentral.com Internet Source	<1 %
107	centaur.reading.ac.uk Internet Source	<1 %
108	etheses.whiterose.ac.uk Internet Source	<1 %

109	oceanrep.geomar.de Internet Source	<1 %
110	onlinelibrary.wiley.com Internet Source	<1 %
111	patents.justia.com Internet Source	<1 %
112	psasir.upm.edu.my Internet Source	<1 %
113	purehost.bath.ac.uk Internet Source	<1 %
114	tel.archives-ouvertes.fr Internet Source	<1 %
115	www.jove.com Internet Source	<1 %
116	www.mysciencework.com Internet Source	<1 %
117	www.rainforest-alliance.org Internet Source	<1 %
118	0354-4664, 2008 Publication	<1 %
119	Dario Arizala, Shefali Dobhal, Anne M. Alvarez, Mohammad Arif. "Elevation of Clavibacter michiganensis subsp. californiensis to species level as Clavibacter californiensis sp. nov.,	<1 %

merging and re-classification of Clavibacter michiganensis subsp. chilensis and Clavibacter michiganensis subsp. phaseoli as Clavibacter chilensis sp. nov. based on complete genome in-silico analyses", Cold Spring Harbor Laboratory, 2022
Publication

---

120 Fuchs, Bernhard M, Harder, Jens, Riedel, Thomas, Spring, Stefan, Spröer, Cathrin and Yan, Shi. "Taxonomy and evolution of bacteriochlorophyll a-containing members of the OM60/NOR5 clade of marine gammaproteobacteria: description of Luminiphilus syltensis gen. nov., sp. nov., reclassification of Haliea rubra as Pseudohaliea rubra gen. nov., comb. nov., and emendation of Chromatocurvus halotolerans.", Helmholtz Zentrum für Infektionsforschung Repository, 2013.
Publication

<1 %

---

121 Hui-Ning Jiang, Shuai-Ting Yun, Bao-Xun Wang, Ming-Jing Zhang, Yu Ma, Yan-Xia Zhou. "Jeotgalibacillus kunyuensis sp. nov., a novel orange-pigmented species with relative genes or gene clusters, isolated from wetland soil in Yantai, China", Research Square Platform LLC, 2022
Publication

<1 %

122	fr.wikipedia.org Internet Source	<1 %
123	"Bergey's Manual® of Systematic Bacteriology", Springer Nature, 2010 Publication	<1 %
124	George M. Garrity, Julia A. Bell, Timothy Lilburn. "Chapter 2 Class II. Betaproteobacteria class. nov.", Springer Science and Business Media LLC, 2005 Publication	<1 %
125	Madhaiyan Munusamy, Saravanan Venkatakrishnan Sivaraj, See-Too Wah-Seng. "Genome based analyses reveals the presence of heterotypic synonyms and subspecies in Bacteria and Archaea", Cold Spring Harbor Laboratory, 2020 Publication	<1 %
126	eprints.whiterose.ac.uk Internet Source	<1 %
127	orcid.org Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		