



OPEN ACCESS

EDITED BY

Javier Pascual,
Darwin Bioprospecting Excellence, Spain

REVIEWED BY

Mohammad Ali Amoozegar,
University of Tehran, Iran
Chirayu Desai,
Gujarat Biotechnology University, India
Jörn Petersen,
German Collection of Microorganisms and Cell
Cultures GmbH (DSMZ), Germany

*CORRESPONDENCE

Paulina Corral
✉ paulina.corral@unina.it

RECEIVED 20 June 2023

ACCEPTED 17 August 2023

PUBLISHED 29 September 2023

CITATION

Gattoni G, Di Costanzo F, de la Haba RR,
Fernández AB, Guerrero-Flores S,
Selem-Mojica N, Ventosa A and Corral P (2023)
Biosynthetic gene profiling and genomic
potential of the novel photosynthetic marine
bacterium *Roseibaca domitiana*.
Front. Microbiol. 14:1238779.
doi: 10.3389/fmicb.2023.1238779

COPYRIGHT

© 2023 Gattoni, Di Costanzo, de la Haba,
Fernández, Guerrero-Flores, Selem-Mojica,
Ventosa and Corral. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Biosynthetic gene profiling and genomic potential of the novel photosynthetic marine bacterium *Roseibaca domitiana*

Giuliano Gattoni¹, Fabiana Di Costanzo¹, Rafael R. de la Haba²,
Ana B. Fernández^{3,4}, Shaday Guerrero-Flores⁵,
Nelly Selem-Mojica⁵, Antonio Ventosa² and Paulina Corral^{1,2*}

¹Department of Biology, University of Naples Federico II, Naples, Italy, ²Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, ³Institute for Multidisciplinary Research in Applied Biology, Public University of Navarre, Pamplona, Spain, ⁴Research & Development Department, Bioinsectis SL, Navarre, Spain, ⁵Centro de Ciencias Matemáticas, Universidad Nacional Autónoma de México UNAM, Morelia, Mexico

Shifting the bioprospecting targets toward underexplored bacterial groups combined with genome mining studies contributes to avoiding the rediscovery of known compounds by revealing novel, promising biosynthetic gene clusters (BGCs). With the aim of determining the biosynthetic potential of a novel marine bacterium, strain V10^T, isolated from the Domitian littoral in Italy, a comparative phylogenomic mining study was performed across related photosynthetic bacterial groups from an evolutionary perspective. Studies on polyphasic and taxogenomics showed that this bacterium constitutes a new species, designated *Roseibaca domitiana* sp. nov. To date, this genus has only one other validly described species, which was isolated from a hypersaline Antarctic lake. The genomic evolutionary study linked to BGC diversity revealed that there is a close relationship between the phylogenetic distance of the members of the photosynthetic genera *Roseibaca*, *Roseinatronobacter*, and *Rhodobaca* and their BGC profiles, whose conservation pattern allows discriminating between these genera. On the contrary, the rest of the species related to *Roseibaca domitiana* exhibited an individual species pattern unrelated to genome size or source of isolation. This study showed that photosynthetic strains possess a streamlined content of BGCs, of which 94.34% of the clusters with biotechnological interest (NRPS, PKS, RRE, and RiPP) are completely new. Among these stand out T1PKS, exclusive of *R. domitiana* V10^T, and RRE, highly conserved only in *R. domitiana* V10^T and *R. ekhonensis*, both categories of BGCs involved in the synthesis of plant growth-promoting compounds and antitumoral compounds, respectively. In all cases, with very low homology with already patented molecules. Our findings reveal the high biosynthetic potential of infrequently cultured bacterial groups, suggesting the need to redirect attention to microbial minorities as a novel and vast source of bioactive compounds still to be exploited.

KEYWORDS

biosynthetic profiling, photosynthetic bacteria, bacteriochlorophyll, genome mining, *Roseibaca*, biosynthetic gene cluster (BGC)

1. Introduction

Microbial diversity and all its metabolic potential are being explored by massive sequencing techniques and the use of genome mining strategies. Metagenomic studies and further identification of specialized metabolite biosynthetic gene clusters (BGCs) have demonstrated that we are obviating most of the prokaryotic diversity and, consequently, all its metabolic potential (Scherlach and Hertweck, 2021; Paoli et al., 2022). Most of what is known to date about microbial diversity is restricted to genomic sequences, which are reconstructed as metagenome-assembled genomes (MAGs). Thus, the vast majority of the bacterial community remains uncultivated. In some environments, such as the Mediterranean Sea, a large proportion of its microbial population is represented by members of the phylum *Pseudomonadota*, which in turn encompasses groups still to be discovered and that count as minorities (Haro-Moreno et al., 2018). It is also well known that culture failure is due to many bacteria thriving solely in their natural environment and community as ecto/endosymbionts (Yang et al., 2022) or as part of complex structures as microbial mats (Chen et al., 2020).

Recognizing the challenges of culturing new taxa, new strategies have been developed and implemented to address the “non-cultivable” fraction and overcome this limitation. Among them, the exploration of extreme and poorly explored environments, such as the seabed (Feling et al., 2003; Eustáquio et al., 2011; Jensen et al., 2015), the employment of *in situ* cultivation techniques through isolation devices (Berdy et al., 2017; Lodhi et al., 2018), and other isolation approaches based on microbial co-culture (Stewart, 2012; Sánchez-Andrea et al., 2018), stand out. All of them have enabled the isolation of new taxa that are currently being explored at a biotechnological level, mainly in the pharmaceutical industry (Bauman et al., 2022). In addition to these methods, we have recently proposed the diel and seasonal sampling as a strategy to increase the probability of isolating new marine microorganisms. These studies considered dynamic phenomena, such as diel vertical migration (DVM) and seasonality with the consequent microbial redistribution, allowing the isolation of the new bacterium *Leeuwenhoekiella parthenopeia*, belonging to the rare biosphere and, interestingly, with the capacity to inhibit the viability of tumor cells *in vitro* (Gattoni et al., 2023).

The primary resource to establish the biosynthetic potential of a bacterium is the genome mining strategy, whose application highly contributes to the prevention of the rediscovery of already-known natural products (Ward and Allenby, 2018; Gavriilidou et al., 2022). The biosynthetic capacity is assessed by predicting the main classes of BGCs, such as polyketide synthase (PKS), non-ribosomal peptide synthase (NRPS), ribosomally synthesized and post-translationally modified peptides (RiPPs), and RiPP recognition element (RRE), all involved in the secondary/specialized metabolic pathways targeted to synthesize bioactive compounds of biotechnological interest (Medema et al., 2015; Kloosterman et al., 2020). A homology search with known BGCs from other organisms encoding the synthesis of chemically deciphered and experimentally characterized compounds is a common approach to estimating the metabolic capacity of the bacterium of interest as a potential source of lead candidates before investing more time and money in further biomedical and biotechnological applications

(Doroghazi et al., 2014; Pan et al., 2017; Parkinson et al., 2018; Belknap et al., 2020).

In the present study, we determined the biosynthetic profile and the genetic potential of a novel photosynthetic marine bacterium, designated as *Roseibaca domitiana*, which was isolated after diel and seasonal sampling from the Tyrrhenian Sea coast, as previously reported for *Leeuwenhoekiella parthenopeia*. This pink-pigmented bacterium is phylogenetically closely related to the only representative species of the genus *Roseibaca*, *Roseibaca ekhonensis*, a bacterium originally isolated from a saline lake in Antarctica (Labrenz et al., 2009). However, the new isolate showed the closest similarity with an uncharacterized cultured species, “*Roseibaca calidilacus*,” which was isolated from a microbial mat of the heliothermal hypersaline lake Hot Lake (Washington, USA) (Romine et al., 2017). Those two related cultivated species of the genus *Roseibaca* are derived from extreme aquatic environments. Thus, *Roseibaca domitiana* constitutes the first representative of this genus isolated from a marine environment. Taxonomically, this new bacterium belongs to the genus *Roseibaca*, classified in the family *Paracoccaceae* (Liang et al., 2021; Göker, 2022), order *Rhodobacterales*, class *Alphaproteobacteria*, and phylum *Pseudomonadota* (Oren and Garrity, 2021). In particular, the members of the family *Paracoccaceae* are widely distributed in the marine environment, representing up to 30% of the bacterial community in some aquatic environments (Lamy et al., 2011; Sato-Takabe et al., 2016; Simon et al., 2017). The ability to synthesize pigments, such as bacteriochlorophyll, is an intrinsic characteristic of some genera of purple photosynthetic bacteria, especially of Aerobic Anoxygenic Phototrophs (AAP) (Madigan and Jung, 2009), whose biosynthetic potential remains unexplored.

The comparative genomic study carried out allowed us to determine the biosynthetic capacity of *Roseibaca domitiana* with respect to closely related members of the family *Paracoccaceae* from an evolutionary perspective and based on the novelty of BGCs. We provide the first insights into the lack of correlation between the phylogenetic distance and the BGC profiles of members of this family. Specifically, we highlight and discuss the novel BGCs identified in *Roseibaca domitiana* as potential synthesizers of compounds for plant improvement and antitumor agents, proposing this species as a promising bioresource in the agricultural and pharmaceutical industries.

2. Materials and methods

2.1. Sampling and isolation

The strain V10^T was isolated from the littoral Domitian Coast, Italy (40°54′50.2″N 14°01′26.2″E), during a diel and seasonal study of the Tyrrhenian Sea. The seawater samples were taken during the fall and winter of 2020–2021 and subsequently processed for 2 h on media prepared with the same filtered natural seawater. Salinity, pH, conductivity, particulate [total dissolved solids (TDSs)], and temperature were measured *in situ* with a refractometer and a digital multiparameter pH/EC/TDS/TEMP (Hanna HI-9812-51). The isolation was performed by extinction dilution technique on solidified seawater oligotrophic medium (SWOM). The plates were aerobically incubated at room

temperature ($\sim 21^{\circ}\text{C}$) following the light day cycle, and the colonies were visible after 6 days of incubation. Following five subculture passages, the strain V10^T was able to grow on the synthetic medium, marine agar (MA) 2216 (Difco). For long-term conservation, the strain V10^T was grown in marine broth (MB) for 72 h and stored at -80°C in cryovials with 40% (v/v) of glycerol.

SWOM composition (g/L of natural seawater): casamino acids, 0.5; yeast extract, 1.0; tryptone, 1.0; and agar, 15.0. Supplements: glycerol, 0.1 M; mineral solution, 500 $\mu\text{L/L}$; and vitamin solution, 500 $\mu\text{L/L}$. All supplement solutions were sterilized by filtration and added after autoclaving. The vitamin solution composition (mg/L) was as follows: biotin, 4.0; folic acid, 4.0; pyridoxine-HCl, 20.0; riboflavin, 10.0; thiamine-HCl 2H₂O, 10.0; nicotinamide, 10.0; D-Ca-pantothenate, 10.0; vitamin B₁₂, 0.20. The volume was brought up to 1,000 mL using distilled water. The solution was stored in the dark at 4°C.

2.2. Molecular identification of the strain V10^T

The strain V10^T was initially identified using 16S rRNA gene sequencing after PCR amplification. Universal primers 27F and 1492R were used in a 40 μL master mix that included a 5 \times reaction buffer, which contains dNTPs, MgCl₂, and enhancers. MyTaq Taq DNA polymerase from Bioline and nuclease-free water were also part of the mix. The PCR procedure involved an initial denaturation at 95°C for 5 min, followed by 25 cycles that included 30 s at 94°C, 30 s at 50°C, and 90 s at 72°C. A final extension was carried out for 10 min at 72°C. Afterward, the PCR product was verified through 1% agarose gel electrophoresis and purified using a GeneJET PCR purification kit from Invitrogen. The sequencing was conducted by Stab Vida in Portugal using the Sanger method. Finally, sequence quality control and assembly were conducted using 4Peaks v.1.8 and ChromasPro v.2.1.10 software, respectively.

The assembled 16S rRNA gene sequence from strain V10^T was compared with the EzBioCloud database (Yoon et al., 2017) using the 16S-based ID service. The sequence identity was also calculated using NCBI BLAST (RRID: SCR_004870) against the NCBI nr/nt nucleotide database (RRID: SCR_004860) restricted to sequences from type material but also including uncultured/environmental sample sequences.

2.3. Genomes sequence retrieval and estimation of Overall Genome Relatedness Indices

For genome sequencing, high-quality DNA was obtained using the DNeasy UltraClean kit (QIAGEN, Hilden, Germany), following the protocol for Gram-negative bacteria. DNA purity and quantification were determined by fluorometry using a Qubit assay kit for double-stranded DNA broad range (dsDNA BR, Invitrogen) and measured with a Qubit 4 fluorometer (Invitrogen, MA, USA).

The genome of strain V10^T was sequenced using the Illumina NovaSeq 6000 system 2 \times 150 PE (Novogene Co., UK). Raw read quality was assessed using FastQC (RRID: SCR_014583) v.0.11.9,

and those containing adapters and low-quality bases ($Q \leq 20$) were removed with Trimmomatic (RRID: SCR_011848) v.0.36 (Bolger et al., 2014). *De novo* assembly of the filtered reads was performed using SPAdes (RRID: SCR_000131) v.3.15.4 (Bankevich et al., 2012) and SOAPdenovo2 (RRID: SCR_005503) (Luo et al., 2012). Assembly quality was estimated with QUAST (RRID: SCR_001228) v.5.1.0rc1 (Gurevich et al., 2013), and its completeness and contamination were assessed using CheckM (RRID: SCR_016646) v.1.0.5 (Parks et al., 2015). The curated genome sequence was annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAP) (RRID: SCR_021329) (Tatusova et al., 2016) and deposited in GenBank/EMBL/DDJB under accession number JALZWP000000000.

For comparative genomic studies, 36 genome sequences of cultivated species closely related to the strain V10^T were obtained from NCBI GenBank (Supplementary Table 1). This genomic comparison was performed between strain V10^T and type strains of validly published species names. Additionally, the study included the genomes of non-type strains of closely related cultivated species: “*Roseibaca calidilacus*” HL-91, *Roseibaca* sp. Y0-43, *Roseinatronobacter* sp. HJB301, and “*Natronohydrobacter thiooxidans*” AH01.

To circumscribe the strain V10^T as a new taxon, several Overall Genome Relatedness Indices (OGRI) (Chun and Rainey, 2014) were estimated. The average nucleotide identity based on BLASTn+ (RRID: SCR_001598) (denoted as ANIb) was calculated using pyANI v.0.2.12 software (Pritchard et al., 2016). The digital DNA–DNA hybridization (dDDH) was calculated using the Genome-to-Genome Distance Calculator (GGDC) v.3.0, with formula d_4 (a.k.a. GGDC formula 2) (Meier-Kolthoff et al., 2013). The results were evaluated based on the cutoff for species boundaries: ANI, 95–96% (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Chun and Rainey, 2014), and dDDH, 70% (Auch et al., 2010).

2.4. Phylogenetic and phylogenomic reconstruction

The 16S rRNA gene-based phylogeny was performed using the online tool Type Strain Genome Server (TYGS) (Meier-Kolthoff and Göker, 2019; Meier-Kolthoff et al., 2022), including, among others, the genome sequences of the strain V10^T and its most closely related species of the genus *Roseibaca*. Given the well-known inaccuracy of determining evolutionary relationships by using single locus phylogenies (Corral et al., 2018; de la Haba et al., 2018; Infante-Domínguez et al., 2020), a more detailed genome-based tree was inferred, as explained below.

Orthologous gene clusters present in all the genomes (core-genome dataset) were determined using all-vs.-all BLASTp+ (RRID: SCR_001010) comparisons among the translated coding sequences (CDS) of the annotated genomes under study, as implemented in the “anvi-pan-genome” script (options “–min-percent-identity 40 –minbit 0.5 –mcl-inflation 5 –use-ncbi-blast”) from Anvi'o suite v.7.1 (Eren et al., 2020). Then, single-copy core protein sequences were individually aligned with Muscle v. 3.8.1551 (Edgar, 2004). Alignment columns for each individual gene with

more than 50% gaps or highly hypervariable were removed by means of the “alignment_pruner.pl” Perl script (<https://github.com/novigit/davinciCode/blob/master/perl>). Pruned alignments were finally concatenated into a super-matrix using the AMAS tool (Borowiec, 2016), which was further analyzed to estimate the best model and parameters of amino acid substitution via ModelFinder (Kalyaanamoorthy et al., 2017). The maximum-likelihood phylogenomic tree was generated using IQ-TREE v.2.2.0 (Minh et al., 2020), and the tree branch support was inferred using the ultrafast bootstrap approximation (Hoang et al., 2018).

For phylogenetic analyses based on photosynthetic genes, protein-coding sequences were predicted using Prodigal v.2.6.3 (Hyatt et al., 2010) and annotated with eggNOG-mapper v.2.1.9 (Cantalapiedra et al., 2021). Translated sequences of *pufLM*, *puhA*, and *bchXYZ* photosynthetic genes were aligned with Muscle v.3.8.31 (Edgar, 2021) and then concatenated with SeqKit v.2.3.0 (Shen et al., 2016). The best-fit model of amino acid sequence evolution was calculated using ProTest v.3.4.2 (Abascal et al., 2005). The maximum-likelihood tree (using the WAG model) with 1,000 pseudoreplicates was inferred by RAxML v.8.2.12 (Stamatakis, 2014).

2.5. Phenotypic and physiological characterization

Colony morphology and pigmentation were observed in solid SWOM after 6 days of incubation, and cell features, such as shape and motility, were examined by phase contrast microscopy (Olympus BX41) from a 72-h liquid culture incubated at 25°C with agitation. The optimal growth conditions were determined in the following ranges: temperature, 4–55°C, intervals of 10°C; NaCl, 0–20% (w/v), intervals of 0.5%; and pH, 5.0–10.0, intervals of 1.0. To test pH and salinity ranges, a basal liquid medium supplemented with 0.1% (w/v) of yeast extract was used, while temperature was tested in SWOM. For pH tests, the medium was adjusted to 3.5% (w/v) of NaCl (sea salts, Sigma-Aldrich, MO, USA) and buffered with MES (pH 5.0–6.0), MOPS (pH 6.5–7.0), Tris (pH 7.5–8.5), or CHES (pH 9.0–10.0) at a final concentration of 50 mM. Anaerobic growth was determined by incubation on MA and SWOM agar plates at 25°C for 72 h in the AnaeroGen™ system (Oxoid). Catalase activity was assessed by adding a 3% (w/v) H₂O₂ solution to colonies on a solid medium. Oxidase activity was examined using sticks containing tetramethyl-*p*-phenylenediamine (PanReac AppliChem, Darmstadt, Germany). Other biochemical and physiological tests were evaluated following the characterization methods described by Barrow and Feltham (1993). Hydrolysis of Tween 80, gelatin, starch, DNA, casein, and aesculin, indole production from tryptophan, methyl red, and Voges–Proskauer tests, Simmons’s citrate utilization, production of H₂S, urease, nitrate and nitrite reduction, arginine dihydrolase and anaerobic growth with 5% DMSO were tested by supplementing the appropriated medium with 3.5% (w/v) of NaCl. The metabolic profile was evaluated using the microbial ID assay GEN III MicroPlate™ (Biolog), which determines the phenotypic pattern by assimilating a panel of sugars, alcohols, organic acids, and amino acids as a sole source of carbon and energy. As a reference,

we used *Roseibaca ekhonensis* CECT 7235^T. All phenotypic and physiological tests were determined in triplicate.

2.6. Chemotaxonomic analysis

Polar lipid analysis was carried out by Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (Braunschweig, Germany) services using the deposited strain V10^T (=DSM 112951^T). In brief, polar lipids were extracted using the modified Bligh and Dyer method (Bligh and Dyer, 1959) and separated by two-dimensional silica gel thin-layer chromatography. Total lipids were revealed by spraying molybdate-phosphoric acid with specific reagents to detect defined functional groups.

The cellular fatty acid methyl ester profile of strain V10^T was determined at the CECT, Spanish Type Culture Collection (Valencia, Spain), following the protocol recommended by the MIDI Microbial Identification System (Sasser, 1990). The biomass of strain V10^T for the fatty acid determination was obtained from a culture on MA after incubation for 72 h at 30°C. The cellular fatty acid content was analyzed by gas chromatography with an Agilent 6850 gas chromatograph and identified according to the TSBA6 method using the Microbial Identification Sherlock software package (MIDI Inc., 2012).

2.7. Identification of BGCs, diversity, and evolution

The main objective of this study was to investigate the secondary/specialized metabolites of BGCs to infer the biosynthetic potential of *Roseibaca* sp. V10^T and the related members of the family *Paracoccaceae* from an evolutionary perspective. The biosynthetic profiling study involved genome functional annotation using Prokka v.1.14.6 (Seemann, 2014); identification of secondary metabolite gene clusters with antiSMASH v.7.0 (RRID: SCR_022060) (Blin et al., 2023); gene cluster similarity analysis by means of BiG-SCAPE (Biosynthetic Gene Similarity Clustering and Prospecting Engine) (RRID: SCR_022561) v.1.1.2, and CORASON (CORE Analysis of Syntenic Orthologs to prioritize Natural Product Biosynthetic Gene Clusters) (Navarro-Muñoz et al., 2020) tools. The core genes within the BGCs detected by antiSMASH were used to compute cluster sequence similarities by comparison with experimentally characterized genes encoding the biosynthesis of known chemical molecules annotated in the Minimum Information about a Biosynthetic Gene Cluster (MIBiG 3.0) repository (Terlouw et al., 2023). Furthermore, based on sequence similarity networks that encode the biosynthesis of highly similar or identical molecules, BGC sequences from each genome were linked to enzyme phylogenies to create gene cluster families (GCFs). The application of this pipeline allowed us to obtain a global biosynthetic profile to predict the potential of all species to produce novel compounds.

To determine the BGC homologies with known patented proteins, the core sequence within each BGC was translated into amino acids and further searched against non-redundant protein sequence (nr) and patented protein sequence (pataa) databases

using BLASTp+ (RRID: SCR_001010). Assigning proteins to families was performed using the Pfam (RRID: SCR_004726) database (Mistry et al., 2021).

3. Results

3.1. The strain V10^T is a novel photosynthetic bacterial species

The initial molecular identification of the strain V10^T based on the 16S rRNA prokaryotic marker sequence and subsequent genome-based phylogeny (calculated after concatenation of the translated 595 single-copy core genes) and OGRIs confirmed its taxonomic affiliation. The species characterization was completed with the polyphasic study, which included phenotypic and chemotaxonomic approaches.

According to the almost complete 16S rRNA gene sequence (1432 bp) (NCBI accession number: MW785571) comparison, strain V10^T belongs to the genus *Roseibaca*, showing a top sequence identity of 97.8% with *Roseibaca ekhonensis*, the single validly described bacterial name of this genus. This percentage is below the current threshold of 98.65%, which is widely accepted for species delineation (Chun et al., 2018), suggesting that this bacterium might represent a novel taxon within the genus *Roseibaca*.

The phylogenetic tree based on the 16S rRNA gene sequences (Supplementary Figure S1) showed that strain V10^T is clustered with species of the genus *Roseibaca*, although independently and distantly branched. The other two closest genera were *Rhodobaca* and *Roseinatronobacter*. A rather similar topology was retrieved after the core-genome phylogenomic reconstruction, but with strain V10^T clustering together to “*Roseibaca calidilacus*” and *Roseibaca ekhonensis*, while *Roseibaca* sp. Y0-43 falls into a separate branch (Figure 1). This phylogenomic tree is coupled to their BGC profile and discussed in Section 3.2. The clade made up of species of the genus *Roseibaca* includes bacteria derived from marine/saline/hypersaline sources, while the genera *Roseinatronobacter* and *Rhodobaca* harbor alkaliphilic species derived from soda lakes.

Since the species of the genera *Roseibaca*, *Rhodobaca*, and *Roseinatronobacter* comprise Aerobic Anoxygenic Phototrophic (AAP) bacteria, a phylogenomic reconstruction based on the photosynthetic genes encoding proteins for bacteriochlorophyll production (*pufL*, *pufM*, *puhA*, *bchX*, *bchY*, and *bchZ*) was inferred. In the tree obtained from the concatenated alignments of these protein-translated photosynthetic genes (Supplementary Figure S2), it was observed that there is a relationship among AAP species, where strain V10^T has a similar classification to the previous phylogenetic analyses. Clustered with the species *Roseibaca* and close to the most related species, the strain V10^T also constitutes a photosynthetic microorganism.

The OGRi values between *Roseibaca* sp. V10^T and the most closely related type strain, *Roseibaca ekhonensis* CECT 7235^T, were 24.1% for dDDH and 81.8% for ANiB. Both OGRi are below the current thresholds for species circumscription, that is, 70% dDDH (Auch et al., 2010) and 95–96% ANI (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Chun and Rainey, 2014). Similarly, the OGRi values of *Roseibaca* sp.

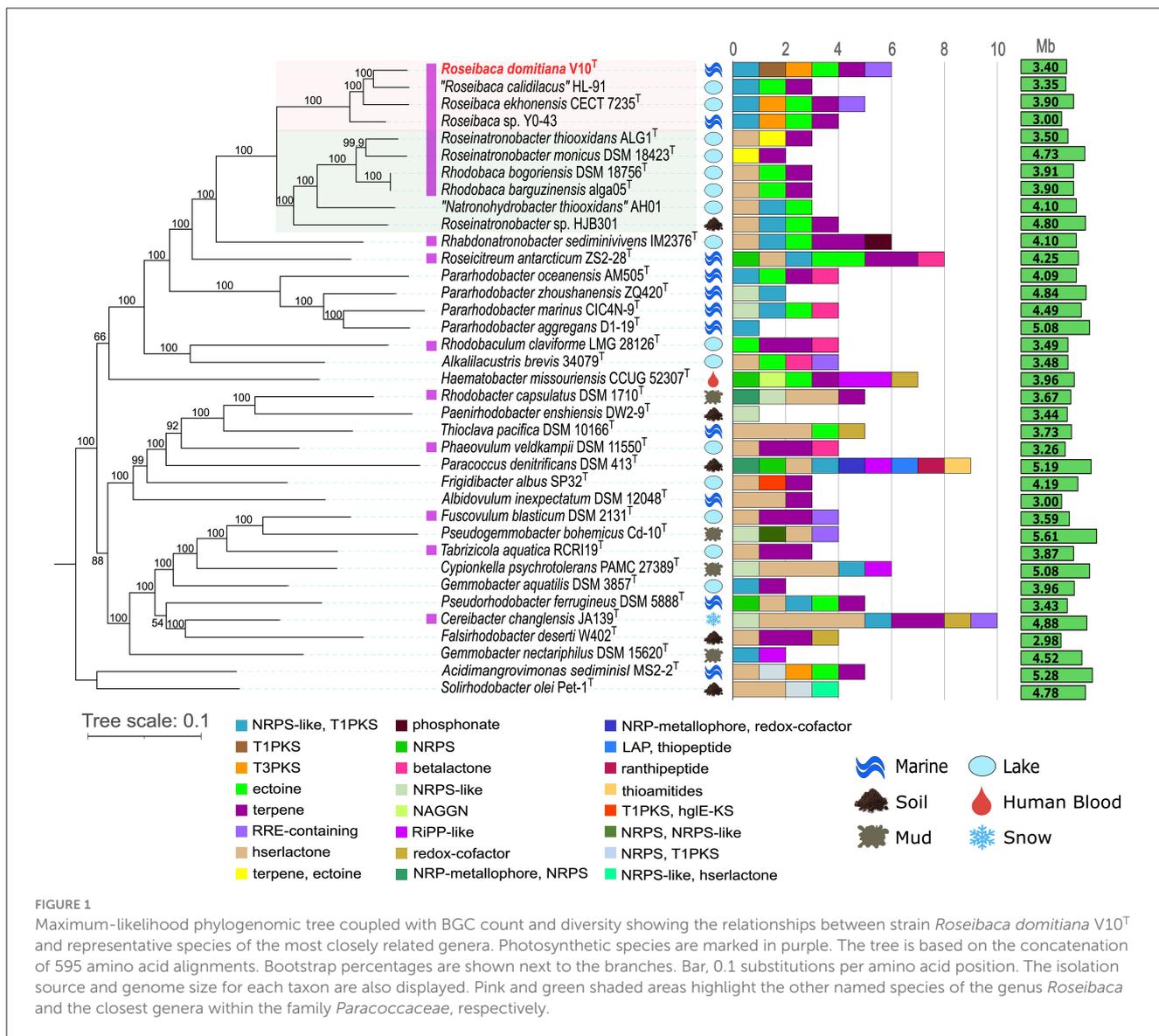
V10^T with respect to the non-type strains of species “*Roseibaca calidilacus*” HL-91 and *Roseibaca* sp. Y0-43 were under these thresholds, as shown in the heatmap (Figure 2).

Overall, single-gene, bacteriochlorophyll multi-locus, and core-genome phylogenies, as well as OGRi results, clearly support that strain *Roseibaca* sp. V10^T constitutes a novel species of the genus *Roseibaca*. Additionally, dDDH and ANiB data displayed in Figure 2 point to taxonomic inconsistencies within the family *Paracoccocaceae*, in particular the two species of the genus *Rhodobaca*, *R. barguzinensis* and *R. bogorensis*, both exceeding the species cutoff boundaries with a 100% identity. According to these indices and the very close relationship between them observed in the phylogenomic study, these strains are members of the same species as previously reported (Liang et al., 2021); thus, we confirm that these species names are synonyms.

Phenotypically, *Roseibaca* sp. V10^T forms pale to pink-pigmented colonies after 4 days of aerobic and light-dark cycle incubation on natural SWOM at ~25°C. Pigments are not produced when the strain grows in the absence of vitamin solution (B complex) and limiting light exposure; thus, this explains why this bacterium can display a fading pigmentation from pink to pale. Physiologically, strain V10^T is halophilic and alkali-tolerant, not being able to grow with <3% (w/v) NaCl and outside the pH range of 6.0 to 9.0. Cells are Gram-stain-negative rods with a size ranging from 1.5 to 3.0 μm in length and 0.5 μm in width, presenting gliding motility with a tendency to occur in long irregular chains, showing a budding process and exhibiting globular shapes (Supplementary Figure S3).

The metabolic profile evaluated by means of the Biolog phenotypic system (Supplementary Figure S4) showed that *Roseibaca* sp. V10^T has low nutritional requirements, using few carbon and energy sources. The chemotaxonomic study revealed that the polar lipid pattern (Supplementary Figure S5) consists of phosphatidylglycerol as a major polar lipid, diphosphatidylglycerol, an aminolipid, a phospholipid, and an unidentified lipid. In contrast to *R. ekhonensis*, phosphatidylethanolamine, and phosphatidylcholine were absent in *Roseibaca* sp. V10^T. The cellular fatty acid composition comprises C_{18:1} ω7c/C_{18:1} ω6c as the major (>79.7%) phospholipid fatty acid (PLFA), followed by minor fatty acids (< 8%), namely C_{16:1} ω7c/C_{16:1} ω6c and C_{19:1} ω6c/C_{19:0} cyclo. Further physiological and chemotaxonomic data are detailed in the species description and the differential table with respect to the most closely related type species, *Roseibaca ekhonensis* (Table 1).

Based on the phylogenetic/phylogenomic reconstructions and the genomic comparisons, together with the physiological features, we can conclude that the strain V10^T constitutes a novel species within the genus *Roseibaca*, for which the name *Roseibaca domitiana* sp. nov. is proposed. The type strain is V10^T, and it is publicly available in the following culture collections: Spanish Type Culture Collection (CECT) as CECT 30319^T; German Collection of Microorganisms and Cell Cultures GmbH (DSMZ) as DSM 112951^T; and Laboratory of Microbiology of Ghent University/Belgian Coordinated Collections of Microorganisms (LMG/BCCM) as LMG 32429^T.



3.1.1. Description of *Roseibaca domitiana* sp. nov.

Roseibaca domitiana (do.mi.ti.à'na. L. fem. adj. *domitiana*, belonging to the Domitian littoral).

Cells are Gram-stain-negative rods, measuring 1.5–3.0 × 0.5 μm in size. Colonies are pale to pink-pigmented, 2 mm in diameter, with a cream consistency and smooth aspect on seawater oligotrophic medium (SWOM) after 4 days of incubation at 25°C. Similar features are observed in marine agar. The pink pigment is reduced when growth occurs in the absence of vitamins and light. The bacterium shows gliding motility and forms long irregular chains exhibiting a budding process with globular shapes in 4-day-old cultures: aerobic anoxygenic phototrophic (AAP), bacteriochlorophyll *a*-producing, and facultative photoheterotrophic bacterium.

Halophilic and alkali-tolerant, it is able to grow in a range from 3 to 10% (w/v) NaCl at 10 to 30°C and pH 6.0 to 9.0, with optimal growth at 3.5% (w/v) NaCl at 25°C and pH 7.5. Catalase

and oxidase are negative. Nitrate is reduced, but nitrite is not. Tween 80, starch, DNA, aesculin, urea, casein, and gelatin are not hydrolyzed. Indole and H₂S are not produced. Voges-Proskauer and methyl red tests yield negative results. The organism utilizes various compounds including D-fructose 6-PO₄, D-glucose 6-PO₄, L-glutamic acid, L-histidine, glucuronamide, L-lactic acid, α-keto-glutaric acid, L-malic acid, α-hydroxy-butyric acid, β-hydroxy-D, L-butyric acid, α-keto-butyric acid, and acetoacetic acid as its main sources of carbon and energy. It demonstrates the ability to grow in the presence of sodium butyrate, 5% DMSO, and the following antibiotics: troleandomycin, rifamycin, minocycline, lincomycin, vancomycin, nalidixic acid, and aztreonam. The polar lipid profile consists of phosphatidylglycerol as a major lipid, diphosphatidylglycerol, an aminolipid, a phospholipid, and an unidentified lipid. The cellular fatty acid composition comprises C_{18:1} ω7c/C_{18:1} ω6c as the major fatty acid (>79.7%). The DNA G+C content is 60.5 mol% (genome).

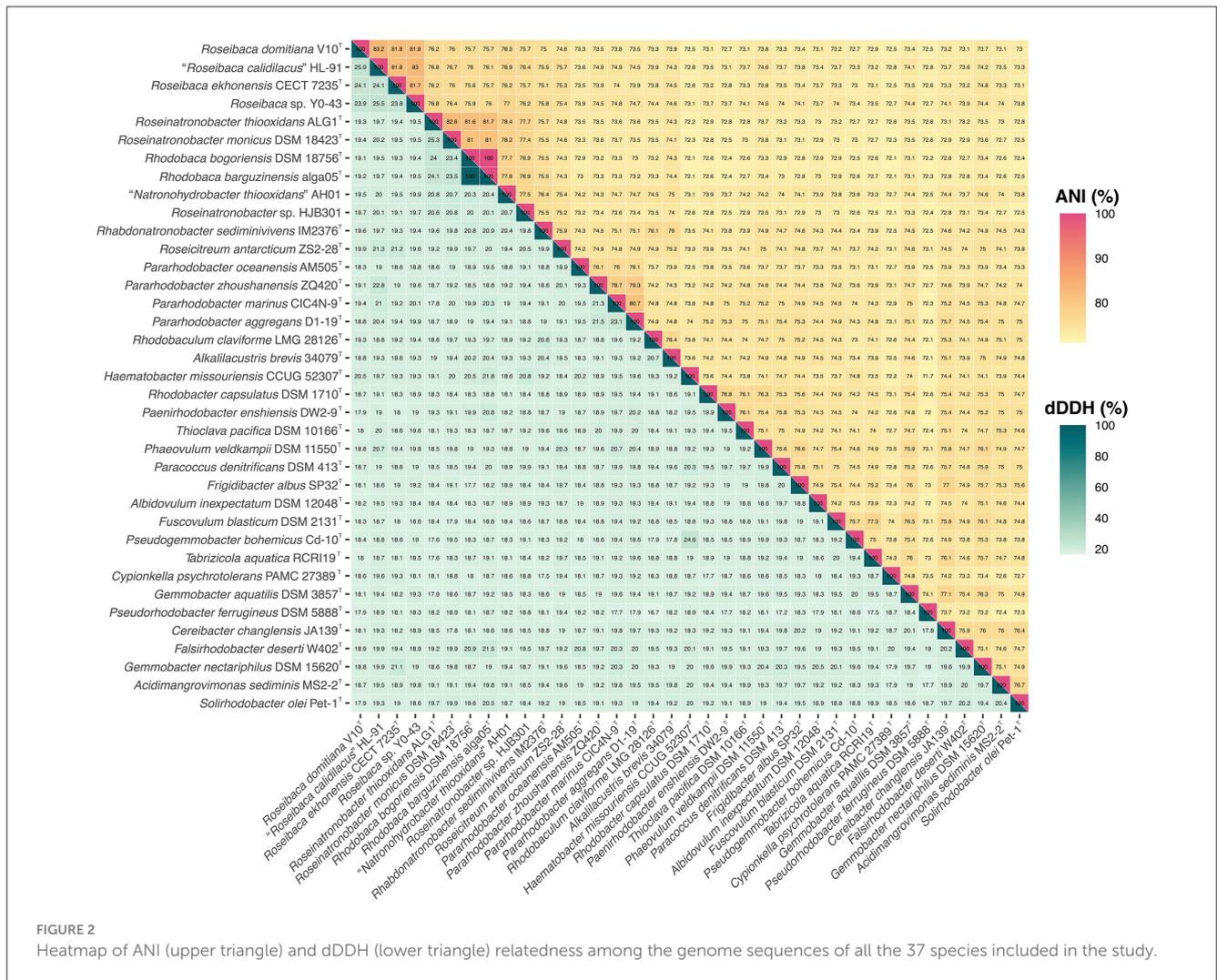


FIGURE 2 Heatmap of ANI (upper triangle) and dDDH (lower triangle) relatedness among the genome sequences of all the 37 species included in the study.

The type strain is V10^T (= CECT 30319^T = DSM 112951^T = LMG 32429^T), isolated from seawater in the Domitian littoral, Italy. The GenBank/EMBL/DBJ accession numbers for the 16S rRNA gene and the whole-genome sequences of the type strain are MW785571 and JALZWP000000000, respectively.

3.2. Biosynthetic gene profiling and genomic potential

3.2.1. Phylogenetic distance is uncorrelated with the BGC profile

The theoretical biosynthetic profile of *Roseibaca domitiana* and related strains was determined by assessing the secondary/specialized metabolism through BGC predictions. Then, the identified BGCs were coupled to the whole-genome-based phylogeny, which allowed us to establish the relationships between phylogenetic distance and BGC diversity.

The BGC predictions obtained by the antiSMASH tool revealed that the species of the family *Paracoccaceae* investigated in this study (those closely related to *Roseibaca domitiana*) harbored a low BGC count, ranging from one to 10 BGC regions per

genome and spanning a total of 24 different types of clusters (Figure 1). The studied strains belonging to different genera within the *Paracoccaceae* came from different environmental sources, especially from marine and aquatic environments, such as alkaline and hypersaline lakes. The phylogenomic tree, coupled with the BGC count, showed that each species possesses a different BGC profile, except for the genus *Rhodobaca*, which shared the same profile. In all cases, the observed patterns were independent of both the isolation source and the genome size. The global profile also displayed a predominance of some BGC types, such as homoserine lactone (hserlactone) and terpene; this latter was often identified in photosynthetic species that produce bacteriochlorophyll (marked in purple), and it is linked to the ability to synthesize carotenoid pigments, one of the main features of the purple bacteria here studied. Less abundant types of BGCs were the hybrid NRPS/TIPKS and ectoine, both distributed across species of different genera, though the latter was identified in some species of aquatic, soil, or clinical origin, such as *Hematobacter missouriensis*. The presence of NRPS/TIPKS is linked to the ability to synthesize bioactive molecules, while ectoine is related to the osmoprotection to survive in saline environments, as in the case of most of the species in the study.

TABLE 1 Differential features between *Roseibaca* sp. V10^T and *Roseibaca ekhonensis* CECT 7235^T.

Feature	<i>Roseibaca</i> sp. V10 ^T	<i>Roseibaca ekhonensis</i> CECT 7235 ^T
Cell size [width × length; μm]	0.5 × 1.5–3.0	1.0 × 2.5
NaCl (% w/v) growth range	3–10%	0–4%
Temperature range	10 – 30 °C	10 – 30 °C
Motility (gliding)	+	–
Colony pigmentation	Pink pale/pink	Pink
pH range	6.0 – 9.0	7.0 – 9.5
Utilization of:		
D-maltose	–	+
D-fructose 6-PO ₄	+	–
Gentiobiose	+	–
D-melibiose	+	–
Tween 40	–	+
D-galacturonic acid	–	+
D-glucuronic acid	–	+
β-hydroxy-D, L-butyric acid	+	–
L-lactic acid	+	–
Citric acid	–	+
N-acetyl D-galactosamine	–	+
N-acetyl-D-glucosamine	–	+
Chemotaxonomy		
Phosphatidylethanolamine (PE)	–	+ *
Phosphatidylcholine (PC)	–	+ *
Major fatty acid	C _{18:1} ω7c/C _{18:1} ω6c	C _{18:1} ω7c *
DNA G+C [mol%, genome]	60.5	61.6

All data from this study unless otherwise indicated. +, positive reaction; –, negative reaction.

*Data taken from the author's description (Labrenz et al., 2009).

It is important to note that most of the BGCs belonging to the categories PKS and NRPS are present as hybrid clusters. This means that the cluster is composed of more than one type of BGC located or overlapped in the same region of the genome, and, therefore, hybrids were classified into different categories according to the combination of BGC types they were made of.

Focusing on the genus *Roseibaca*, all its four cultivated members possessed a common biosynthetic profile consisting of three shared BGCs: NRPS-like/T1PKS hybrid, ectoine, and terpene. With the exception of “*R. calidilacus*,” T3PKS was also present in species of this genus. Additionally, only *R. domitiana* and *R. ekhonensis* shared the RRE-containing cluster. Remarkably, the new species proposed here, *R. domitiana*, and the only validly named one, *R. ekhonensis*, had a very similar biosynthetic pattern despite having been isolated from different environments, the former from a marine habitat and the latter from a hypersaline lake. On the other hand, the two validly

named species of *Rhodobaca* shared the same profile. However, this was not unexpected considering that both species are synonyms in previously reported studies and in our comparative genomic results. Regarding the relationship between BGC content and genome size, no direct association was observed since one of the largest genomes analyzed, i.e., *Pararhodobacter aggregans* D1-19^T (5.08 Mb), exhibited only one BGC, while some of the smallest (< 4.0 Mb), i.e., *Roseibaca domitiana* V10^T harbored six BGCs.

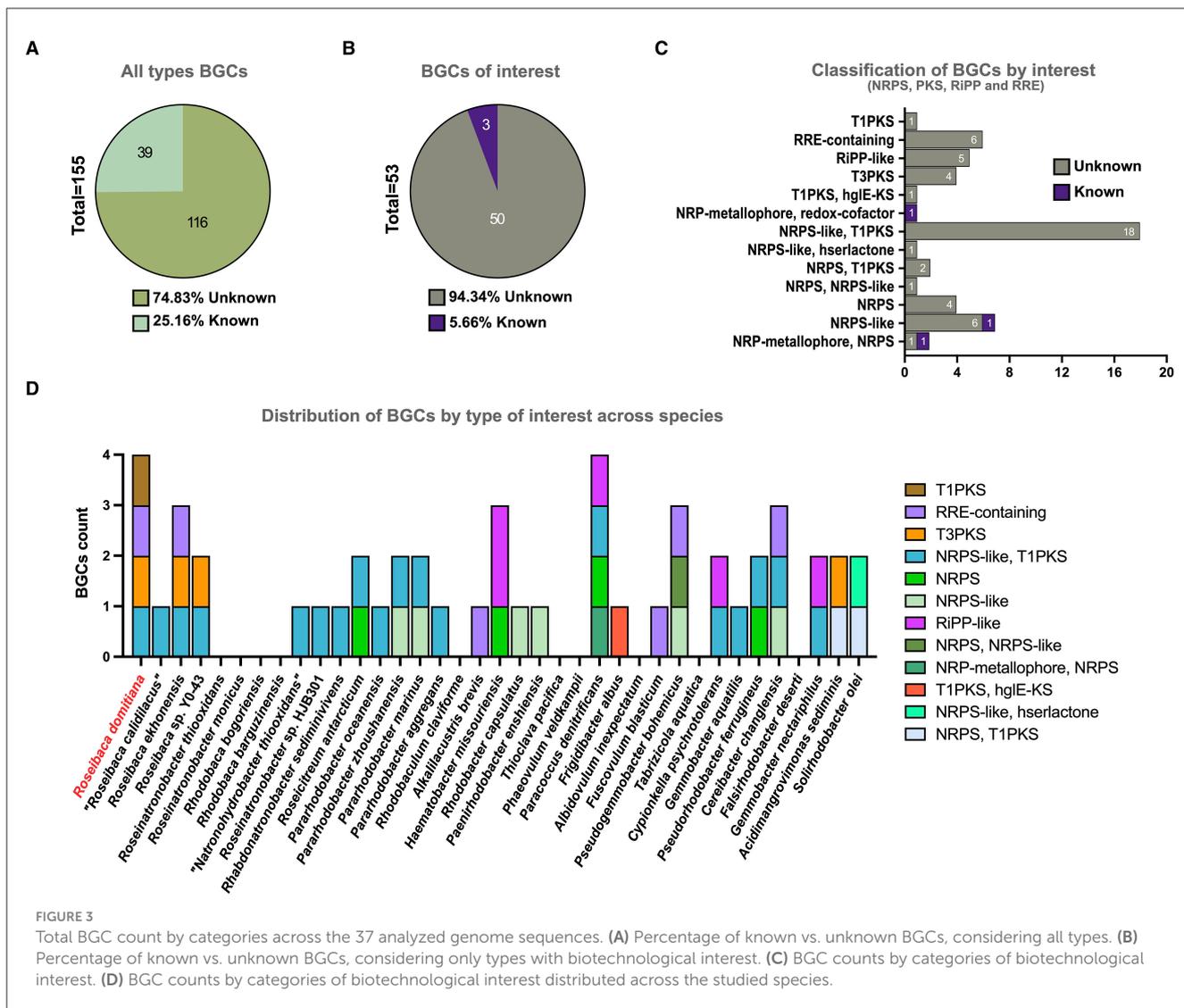
Overall, most of the studied species displayed a differential BGC profile that was not related to the phylogenetic distance among the taxa. In addition, it was found that the source of isolation and the size of the genome are not linked to the biosynthetic profile of the species within the family *Paracoccocoeae*. Only a resolute BGC pattern that allowed us to distinguish between the photosynthetic genera *Roseibaca*, *Roseinatronobacter*, and *Rhodobaca* could be established. This observation will be strengthened when more genomes of species in these genera become available.

3.2.2. The vast majority of BGCs are involved in the synthesis of unknown molecules

The biosynthetic potential of *Roseibaca domitiana* and its closely related species was determined based on the novelty of the predicted BGCs in each genome, based on the homology with characterized genes encoding the synthesis of known chemical molecules present in the MIBiG repository and verified *in vitro*.

A total of 155 BGCs assigned to 24 different categories were predicted in the 37 analyzed genomes, as aforementioned. A total of 74.83% of the detected BGCs are unknown (Figure 3A) since no similarities were found with any known cluster involved in the synthesis of chemical entities. Focusing on the BGC types of biotechnological interest (i.e., NRPS, PKS, RiPP, and RRE), 53 BGCs were identified across the studied genomes, 94.34% of which were completely unknown (Figures 3B, C). These promising novel BGCs were distributed among 27 out of the 37 analyzed species, which may represent potential producers of new molecules (Figure 3D). Concerning the type of these new BGCs, the most predominant was the hybrid NRPS-like/T1PKS since it was detected in 18 species. Furthermore, this hybrid BGC was found in all the species of *Roseibaca*, so far representing a common feature of the members of this genus.

Regarding *Roseibaca domitiana*, a total of six BGCs were predicted (terpene, T3PKS, NRPS-like/T1PKS, ectoine, RRE-containing, and T1PKS), of which only terpene and ectoine showed high levels of sequence similarity (100 and 80%, respectively) with known clusters that encode the synthesis of carotenoids (squalene/phytoene) and ectoine, respectively. On the contrary, T3PKS, RRE-containing, and T1PKS did not match with any known cluster or had a low similarity (>8%) with known BGCs present in the MIBiG database, as in the case of the hybrid NRPS-like/T1PKS (Supplementary Table 2). It is important to consider the relevance of T1PKS since it is the unique BGC that



distinguishes *Roseibaca domitiana* with respect to other species of the genus.

3.2.3. Novel BGCs of *Roseibaca domitiana* are conserved only at the genus level

As has been previously indicated, *Roseibaca domitiana* V10^T possesses six BGC types, of which four are considered new: NRPS-like/T1PKS, T1PKS, T3PKS, and RRE-containing. Here, we determine the similarity/diversity of these BGCs and the evolutionary relationships among the others predicted across the studied species to evaluate their biosynthetic potential.

Their diversity was evaluated by classifying them into GCFs based on sequence similarity networks and further linking them to enzyme phylogenies. Likewise, the evolutionary relationships among these BGCs were inferred through their core sequence and additional multi-locus phylogenies that allowed us to identify the BGC conservancy in *Roseibaca domitiana* V10^T with respect to those of other species, therefore establishing their biosynthetic potential.

The similarity network analysis carried out with BiGSCAPE showed that the BGCs from *Roseibaca* form five GCFs of a different class (Figure 4A): FAM_00146 and FAM_00119, made up of all species of this genus and belonging to terpene and ectoine, respectively; FAM_00135, made of three species excluding *Roseibaca* sp. Y0-43 belonging to the NRPS-T1PKS hybrid class; and finally, FAM_00117 and FAM_00120, both made of *R. domitiana* V10^T and *R. ekhonensis* CECT 7235^T belonging to T3PKS and RRE categories, respectively.

Since the networking is based on the similarity of BGCs, the families here obtained are not made of the species that contain the same type of cluster, as could be expected in the case of the hybrid NRPS-T1PKS predicted in all strains of *Roseibaca*, and T3PKS predicted also in *Roseibaca* sp. Y0-43 but not classified within the same family. It is important to note that no similarities were found for T1PKS of *R. domitiana* V10^T among the studied species; therefore, it remains unclassified without forming a GCF. As displayed in the GCF absence/presence matrix (Figure 4A), *R. domitiana* V10^T only shared core enzyme-coding genes with species of its own genus, and, in the case of the unclassified BGC T1PKS, this remains exclusive of *R. domitiana* V10^T.

A BIGSCAPE: Similarity Networks

Networks by Class	Family	Gene Cluster Families absence/presence			
		Strains			
		<i>Roseibaca domitiana</i> V10 ^T	<i>Roseibaca ekhonensis</i> CECT 7235 ^T	<i>*Roseibaca calidifilacus*</i> HL-91	<i>Roseibaca</i> sp. Y0-43
Terpene	FAM_00146	1	1	1	1
Ectoine	FAM_00119	1	1	1	1
*NRPS-T1PKS	FAM_00135	1	1	1	0
*T3PKS	FAM_00117	1	1	0	0
*RRE	FAM_00120	1	1	0	0

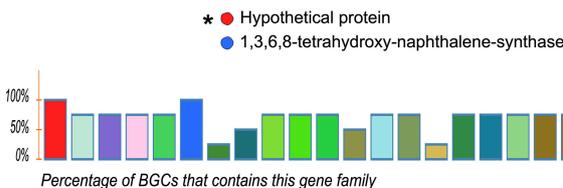
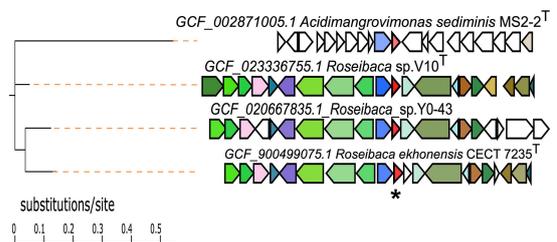
* New BGCs

Note: No similarities were found for T1PKS present only in *R. domitiana* V10^T

B CORASON: Evolutionary analysis of new BGCs present in *Roseibaca domitiana* V10^T

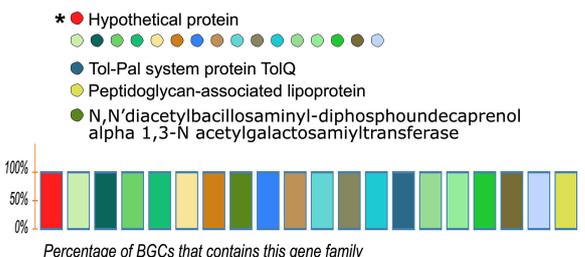
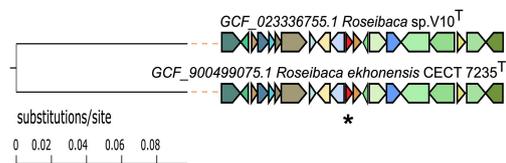
T3PKS

* Query locus: BMJDCPAI_00537

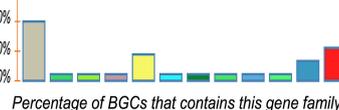
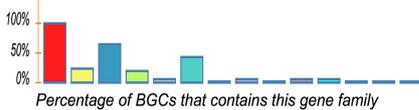
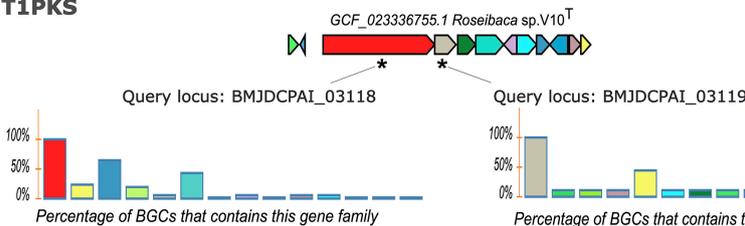


RRE

* Query locus: BMJDCPAI_02120



T1PKS



- * L-threonine 3-dehydrogenase
- Putative oxidoreductase
- Hypothetical protein
- * 8-amino-7-oxononanoatesynthase
- Polysialic acid transport ATP-binding protein KpsT

FIGURE 4

Similarity networking and evolutionary analysis of BGCs. (A) Absence/presence matrix of GCFs across species of the genus *Roseibaca*. (B) Multi-locus phylogeny based on the core sequence of the three new BGCs across studied species, (the query is specified with an asterisk). The charts show the level of synteny between species when an evolutionary relationship is found. In colored circles, the domains that cover 100% (highly conserved). In the case of T1PKS, those less conserved domains are also shown, highlighting the uniqueness of this BGC. T1PKS is displayed as a singleton (evolutionarily unrelated to other species).

Once homologous BGCs based on protein domain content and sequence identity were identified, we addressed the biosynthetic potential from an evolutionary perspective by using CORASON software. Through this tool, a phylogenetic reconstruction of BGCs was obtained based on syntenic orthologs in the BGC core shared between species (Figure 4B). In this way, the phylogenies obtained from the four novel BGCs (NRPS-T1PKS, T3PKS, RRE, and T1PKS) predicted in *R. domitiana* V10^T showed that the hybrid NRPS-T1PKS was present and conserved in all species of *Roseibaca*, but non-confined to this genus, as it is partially shared with some species from other genera (Supplementary Figure S6). As observed in the phylogenetic reconstruction of T3PKS, this

cluster is present and highly conserved in *R. domitiana* V10^T, *R. ekhonensis*, and *Roseibaca* sp. Y0-43, which exhibits a high synteny and shares more than 60% of this BGC. Similar is the case with RRE, which is present only in the genomes of *R. domitiana* V10^T and *R. ekhonensis* and shares 100% of the genes that conform to this cluster, which means that this BGC is fully conserved in both species. Regarding T1PKS (predicted only in *R. domitiana* V10^T), the phylogenetic reconstruction of this BGC was based on two genes (core and additional ones) as queries to investigate its presence and conservation in the genomes of the species under study. Both genes confirmed that T1PKS is unique to *Roseibaca domitiana* V10^T, since this cluster is present

in other species with low or no homology, which indicates that this BGC may be present in multiple genomes, but the genes involved in the cluster are not conserved in these genomes (Supplementary Figure S7).

Finally, to reveal if these BGCs from *R. domitiana* codify enzymes matching already available commercial products or patented proteins, their core amino acid sequence was BLASTp+ searched against the database of patented proteins. The results indicated that the novel BGC-derived proteins only matched <58% similarity with active patents (Table 2). In detail, T3PKS and T1PKS are remotely related to patents with plant growth promotion applications, while NRPS-T1PKS and RRE are related to the antitumoral argyran synthetic pathway enzymes, indicating that *R. domitiana* has the potential to synthesize new molecules, possibly with similar functions.

4. Discussion

Marine microbiota is considered among the most promising natural resources for new drug discovery in biomedical and biotechnological fields. However, a large proportion of the microbial diversity remains hidden in terms of cultivation. Hence, the main challenge is indeed uncovering and isolating this “microbial dark matter”. Once this goal is achieved, what is sought is not only to understand the natural environment to improve and simplify access to microbial resources but also to perform a deep and effective study of the physiology, ecological role, and possible applications of these novel isolated microorganisms. The biotechnological potential of new strains requires a comprehensive evaluation of their biosynthetic abilities before undertaking efforts that may entail the rediscovery of already known natural products, a frequent issue in drug discovery projects.

The results obtained here are derived from seasonal and diel samplings conducted in the Tyrrhenian Sea in Italy. This strategy was employed to isolate new microbial species that belong to the as-yet uncultured marine microbiota. In this way, we achieved the isolation of strain V10^T, initially identified as a member of *Roseibaca*, a little-known genus represented by a single species, *Roseibaca ekhonensis*, recovered from a thermal hypersaline lake in Antarctica (Labrenz et al., 2009). Our phylogenomic and polyphasic studies have revealed that strain V10^T represents a new species within this genus; we propose naming it *Roseibaca domitiana* sp. nov. Given that this strain was isolated from winter nocturnal samples of the coastal surface, while *R. ekhonensis* was cultured from the depths of a hypersaline lake, the isolation of strain V10^T supports the use of night-time and seasonal sampling as a strategy for the cultivation of new microorganisms, an approach that we successfully used in a previous study and allowed the isolation of *Leeuwenhoekiella parthenopeia*, a new rare biosphere bacterium that inhibits the proliferation of tumor cells (Gattoni et al., 2023).

In addition to the only recognized species of *Roseibaca* (*R. ekhonensis*), our study also included uncharacterized cultivated strains reported as “*Roseibaca calidilacus*” HL-91 isolated from a microbial mat of a hypersaline thermal lake (Romine et al., 2017) and *Roseibaca* sp. Y0-43 of marine origin (unpublished).

Considering that these cultivated species come from extreme aquatic environments, *Roseibaca domitiana* represents the only marine species characterized so far. Thus, the genus *Roseibaca* groups marine and polyextremophile (halophilic, thermophilic, and alkali-tolerant) microorganisms and, as such, inhabits marine, saline, hypersaline, and thermal aquatic systems.

The metabolic versatility of *R. domitiana* V10^T is striking since it is an aerobic anoxygenic phototrophic bacterium that could also grow in the absence of light and vitamin B complex, but not without NaCl. One of its most remarkably phenotypic features is the potential to synthesize the carotenoid pigment bacteriochlorophyll due to the presence of the *pufL*, *pufM*, and *pufA* genes, encoding, respectively, the L, M, and H subunits of the photosynthetic reaction center proteins, which enable aerobic anoxygenic photoheterotrophy (Pradella et al., 2004; Zheng et al., 2011). The production of the carotenoid pigment is conditioned by light exposure, with pink (with light) and pale (without light) colonial phenotypes. We also found the *pufLM* and *pufA* genes in the genome sequences of the phylogenetically closest species, such as those of the genera *Roseibaca*, *Roseinatronobacter*, and *Rhodocaca*, whose light-related physiology might be similar to that of strain V10^T.

R. domitiana is a prokaryote that can grow in a wide range of saline and alkaline conditions, although it utilizes a restricted number of substrates. It belongs to the purple nonsulfur bacteria (PNSB), and its light behavior is compatible with this group of microorganisms, known as nature’s preeminent photoheterotrophs, capable of photoautotrophy and with diverse capacities for dark metabolism and growth (Madigan and Jung, 2009). We may hypothesize that its main role in the marine environment would be contributing to the recycling of organic matter and the carbon cycle (Koblížek, 2015; Piwosz et al., 2022).

The biosynthetic gene profiling of the specialized metabolism of *Roseibaca domitiana* and its closest relatives revealed that there is a typical BGC pattern shared among the members of the genera *Roseibaca*, *Roseinatronobacter*, and *Rhodobaca*, consisting of two BGC types, ectoin, and terpene. The first is related to the production of ectoin, a compatible solute that acts as an osmoprotectant to counteract salt stress (Roessler and Muller, 2001), and the second is involved in the synthesis of carotenoids produced by photosynthetic bacteria (Higuchi-Takeuchi and Numata, 2019). The presence of both of these types of BGCs concurs with their physiology to thrive in the environment where they inhabit.

Since BGC occurrence and distribution are homogenous in these three genera, *Roseibaca*, *Roseinatronobacter*, and *Rhodobaca*, only in these cases may the BGC allow differentiation at the genus level. A similar finding has also been reported for the species of the actinobacterial genus *Rhodococcus* (Undabarrena et al., 2021). Conversely, the other analyzed genera within the family *Paracoccaceae* showed heterogeneous BGC profiles unrelated to the phylogenetic distance between them, meaning that each species possesses its own diversity of BGC and could synthesize very different molecules. This non-conserved pattern has been previously found for the species of the genus *Streptomyces*, the most prolific source of natural products (Belknap et al., 2020; Alam et al., 2023), and other actinobacteria, such as *Saccharomonospora* (Ramírez-Durán et al., 2021).

TABLE 2 Amino acid sequence identity of novel BGCs from *Roseibaca domitiana* V10^T searched against a patent protein database.

Region/Type	Query locus tag/encoded protein		Location (size)	Patent hit/similarity	Application (status)/reference
2.1 T3PKS	BMJDPAI_00536 1,3,6,8-tetrahydroxynaphthalene synthase		124,396 - 125,436 (1,041 nt)	US 7314974 57.40%	Expression of microbial proteins in plants for the production of plants with improved properties (active) (Cao et al., 2008)
2.2 NRPS-T1PKS	NRPS	BMJDPAI_00765 D-alanine-D-alanyl carrier protein ligase	350,790 - 355,316 (4,527 nt)	US 6833447 41.03%	EP 2297181 B1 Synthetic pathway enzymes for the production of argyrins (active) (Müller et al., 2013)
	T1PKS	BMJDPAI_00767 hypothetical protein	356,309 - 362,641 (6,333 nt)	US 6833447 44.87%	
9.1 RRE-containing	BMJDPAI_02120 hypothetical protein		106,516 - 106,791 (276 nt)	US 6833447 29.23%	
26.1 T1PKS	BMJDPAI_03118 L-threonine 3-dehydrogenase		2,994 - 10,304 (7,311 nt)	US 8343764 36.74%	Genes encoding glutamine synthetase and use for plant improvement (active) (Abad et al., 2013)

Regarding the novelty of the BGCs of biotechnological interest (NRPS, PKS, RRE, and RiPP) of all the studied species, the vast majority possess an uncovered biosynthetic potential since 94.34% of the identified BGCs encode genes involved in the synthetic pathways of chemically unknown compounds. Besides, the most predominant BGC categories found were hybrid NRPS-like and T1PKS, which means that most species could be potential synthesizers of antibiotic and antitumor peptides (Frattaruolo et al., 2017; Nesic et al., 2022).

With respect to *R. domitiana*, it harbored six BGCs unrelated or fairly distant to those of the MIBiG database that could, therefore, be considered novel BGCs. The evolutionary analysis of these new BGCs allowed us to identify two of them (RRE-containing and T3PKS) as highly conserved in *R. domitiana* and *R. ekhonensis*, while another one, i.e., T1PKS, was exclusive to *R. domitiana* and is not related to any known BGC. The latter can be considered “new among the new ones.” The amino acid sequence derived from these three BGCs showed an identity of less than 58% with active patented proteins. Specifically, the hybrid NRPS-T1PKS and RRE-containing types displayed very low similarities with enzymes from a European patent involved in the synthetic pathways for the production of argyrins, a potent antitumor under study (Sasse et al., 2002; Müller et al., 2013). Likewise, the T3PKS type was remotely related to a patent involved in plant production improvement in agricultural crops (Cao et al., 2008). The exclusive BGC, T1PKS of *R. domitiana*, was related to genes encoding a glutamine synthetase patented to improve plant production (Abad et al., 2013).

In summary, our study permitted us to conclude that the seasonal and diel sampling approaches increase the probability of isolating new species with not-yet-described biochemical and physiological features. Besides, the exploration of secondary/specialized metabolism combined with phylogenomic and evolutionary studies is shown to be a valuable strategy to assess the biosynthetic potential of new microbial groups, either at the species level or in an overall context. Finally, this and other recent studies (Ramírez-Durán et al., 2021; Gattoni et al., 2023) show that rare and minority bacteria harbor untapped metabolic potential for synthesizing novel, biotechnologically significant

compounds. A good example is the new marine bacterium *R. domitiana* V10^T, which is theoretically able to produce plant-promoting compounds and antitumor molecules, which merits prioritizing future *in vitro* studies. The research conducted here encourages the exploitation of microbial minorities as promising, realistic sources for new compounds in agricultural and pharmaceutical applications.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

GG and PC: investigation, formal analysis, and writing of the original draft manuscript. GG, FDC, RRH, ABF, SG-F, NS-M, and PC: methodology, software, and visualization. GG and FDC: data curation. RRH, NS-M, AV, and PC: validation. NS-M, AV, and PC: funding acquisition, project administration, and resources. PC: conceptualization and supervision. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

This study was conducted under the frame of the project BluePharma grant 2018-PDR-00533 funded by Fondazione CON IL SUD (Italy) to PC and supported by the following grants: PAPIIT IA103822 funded by UNAM to NS-M, PID2020-118136GB-I00 funded by the Spanish Ministry of Science and Innovation MCIN/AEI/10.13039/501100011033, and from the Junta de Andalucía (P20_01066 and BIO-213), both with FEDER funds to AV.

Acknowledgments

We thank Aharon Oren (the Hebrew University of Jerusalem) for his help on the nomenclature of the new species.

Conflict of interest

ABF was employed by the Bioinsectis SL.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Abad, M. S., Coffin, M. A., and Goldman, B. S. (2013). *Genes Encoding Glutamine Synthetase and Uses for Plant Improvement*. Washington, DC: U.S. Patent and Trademark Office.
- Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105. doi: 10.1093/bioinformatics/bti263
- Alam, K., Islam, M. M., Islam, S., Hao, J., Abbasi, M. N., Hayat, M., et al. (2023). Comparative genomics with evolutionary lineage in *Streptomyces* bacteria reveals high biosynthetic potentials. *World J. Microbiol. Biotechnol.* 39, 64. doi: 10.1007/s11274-022-03433-y
- Auch, A. F., von Jan, M., Klenk, H.-P., and 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
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barrow, G. I., and Feltham, R. K. A. (1993). "Characterization tests," in *Cowan and Steel's Manual for the Identification of Medical Bacteria*, eds. G. I. Barrow and R. K. A. Feltham (Cambridge University Press), 219–238.
- Bauman, K. D., Shende, V. V., Chen, P. Y.-T., Trivella, D. B. B., Gulder, T. A. M., Vellalath, S., et al. (2022). Enzymatic assembly of the salinosporamide γ -lactam- β -lactone anticancer warhead. *Nat. Chem. Biol.* 18, 538–546. doi: 10.1038/s41589-022-00993-w
- Belknap, K. C., Park, C. J., Barth, B. M., and Andam, C. P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* 10, 2003. doi: 10.1038/s41598-020-58904-9
- Berdy, B., Spoering, A. L., Ling, L. L., and Epstein, S. S. (2017). In situ cultivation of previously uncultivable microorganisms using the ichip. *Nat. Protoc.* 12, 2232–2242. doi: 10.1038/nprot.2017.074
- Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37, 911–917. doi: 10.1139/y59-099
- Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* 51, W46–W50. doi: 10.1093/nar/gkad344
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 4, e1660. doi: 10.7717/peerj.1660
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Cao, Y., Hinkle, G., Slater, S. C., Chen, X., and Goldman, B. S. (2008). *Expression of Microbial Proteins in Plants for Production of Plants With Improved Properties*. Washington, DC: U.S. Patent and Trademark Office.
- Chen, R., Wong, H. L., Kindler, G. S., MacLeod, F. I., Benaud, N., Ferrari, B. C., et al. (2020). Discovery of an abundance of biosynthetic gene clusters in Shark Bay microbial mats. *Front. Microbiol.* 11, 1950. doi: 10.3389/fmicb.2020.01950
- Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., 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
- Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijse.0.054171-0
- Corral, P., de la Haba, R. R., Infante-Domínguez, C., Sánchez-Porro, C., Amoozegar, M. A., Papke, R. T., et al. (2018). *Halorubrum chaoviator* Mancinelli et al. 2009 is a later, heterotypic synonym of *Halorubrum ezzemoulense* Kharroub et al. 2006. Emended description of *Halorubrum ezzemoulense* Kharroub et al. 2006. *Int. J. Syst. Evol. Microbiol.* 68, 3657–3665. doi: 10.1099/ijsem.0.003005
- de la Haba, R. R., Corral, P., Sánchez-Porro, C., Infante-Domínguez, C., Makkay, A. M., Amoozegar, M. A., et al. (2018). Genotypic and lipid analyses of strains from the archaeal genus *Halorubrum* reveal insights into their taxonomy, divergence, and population structure. *Front. Microbiol.* 9, 512. doi: 10.3389/fmicb.2018.00512
- Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K.-S., Haines, R. R., Tchalukov, K. A., et al. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* 10, 963–968. doi: 10.1038/nchembio.1659
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R. C. (2021). High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Bioinformatics* 5, 1–9. doi: 10.1101/2021.06.20.449169
- Eren, A. M., Kiehl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., et al. (2020). Community-led, integrated, reproducible multi-omics with anvio. *Nat. Microbiol.* 6, 3–6. doi: 10.1038/s41564-020-00834-3
- Eustáquio, A. S., Nam, S.-J., Penn, K., Lechner, A., Wilson, M. C., Fenical, W., et al. (2011). The discovery of salinosporamide k from the marine bacterium "*Salinispora pacifica*" by genome mining gives insight into pathway evolution. *Chem. Bio.* 12, 61–64. doi: 10.1002/cbic.201000564
- Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2003). Salinosporamide a: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinispora*. *Angew. Chem. Int. Ed.* 42, 355–357. doi: 10.1002/anie.200390115
- Frattaruolo, L., Lacroix, R., Cappello, A. R., and Truman, A. W. (2017). A genomics-based approach identifies a thioviridamide-like compound with selective anticancer activity. *ACS Chem. Biol.* 12, 2815–2822. doi: 10.1021/acscchembio.7b00677
- Gattoni, G., de la Haba, R. R., Martín, J., Reyes, F., Sánchez-Porro, C., Feola, A., et al. (2023). Genomic study and lipidomic bioassay of *Leeuwenhoekella parthenopeia*: a novel rare biosphere marine bacterium that inhibits tumor cell viability. *Front. Microbiol.* 13, 1090197. doi: 10.3389/fmicb.2022.1090197

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1238779/full#supplementary-material>

- Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Müller, R., Medema, M. H., et al. (2022). *Compendium of Secondary Metabolite Biosynthetic Diversity Encoded in Bacterial Genomes*. Available online at: <https://www.biorxiv.org/content/10.1101/2021.08.11.455920v3> (accessed March 29, 2022).
- Göker, M. (2022). Filling the gaps: missing taxon names at the ranks of class, order and family. *Int. J. Syst. Evol. Microbiol.* 72, 638. doi: 10.1099/ijsem.0.005638
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijse.0.64483-0
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Haro-Moreno, J. M., López-Pérez, M., de la Torre, J. R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018). Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 6, 128. doi: 10.1186/s40168-018-0513-5
- Higuchi-Takeuchi, M., and Numata, K. (2019). Marine purple photosynthetic bacteria as sustainable microbial production hosts. *Front. Bioeng. Biotechnol.* 7, 258. doi: 10.3389/fbioe.2019.00258
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi: 10.1186/1471-2105-11-119
- Infante-Domínguez, C., de la Haba, R. R., Corral, P., Sanchez-Porro, C., Arahál, D. R., and Ventosa, A. (2020). Genome-based analyses reveal a synonymy among *Halorubrum distributum* Zvyagintseva and Tarasov 1989; Oren and Ventosa 1996, *Halorubrum terrestre* Ventosa et al. 2004, *Halorubrum arcis* Xu et al. 2007 and *Halorubrum litoreum* Cui et al. 2007. Emended description of *Halorubrum distributum* Zvyagintseva and Tarasov 1989; Oren and Ventosa 1996. *Int. J. Syst. Evol. Microbiol.* 70, 1698–1705. doi: 10.1099/ijsem.0.003956
- Jensen, P. R., Moore, B. S., and Fenical, W. (2015). The marine actinomycete genus *Salinispora*: a model organism for secondary metabolite discovery. *Nat. Prod. Rep.* 32, 738–751. doi: 10.1039/C4NP00167B
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kloosterman, A. M., Shelton, K. E., Van Wezel, G. P., Medema, M. H., and Mitchell, D. A. (2020). Rre-finder: a genome-mining tool for class-independent ripp discovery. *mSystems* 5, e00267–e00220. doi: 10.1128/mSystems.00267-20
- Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* 39, 854–870. doi: 10.1093/femsre/fuv032
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Nat. Acad. Sci.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Labrenz, M., Lawson, P. A., Tindall, B. J., and Hirsch, P. (2009). *Roseibaca ekhonensis* gen. nov., sp. nov., an alkali-tolerant and aerobic bacteriochlorophyll *a*-producing alphaproteobacterium from hypersaline Ekho Lake. *Int. J. Syst. Evol. Microbiol.* 59, 1935–1940. doi: 10.1099/ijse.0.016717-0
- Lamy, D., Jeanthon, C., Cottrell, M. T., Kirchman, D. L., Van Wambeke, F., Ras, J., et al. (2011). Ecology of aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 8, 973–985. doi: 10.5194/bg-8-973-2011
- Liang, K. Y. H., Orata, F. D., Boucher, Y. F., and Case, R. J. (2021). Roseobacters in a Sea of poly- and paraphyly: whole genome-based taxonomy of the family *Rhodobacteraceae* and the proposal for the split of the “Roseobacter clade” into a novel family, *Roseobacteraceae* fam. nov. *Front. Microbiol.* 12, 683109. doi: 10.3389/fmicb.2021.683109
- Lodhi, A. F., Zhang, Y., Adil, M., and Deng, Y. (2018). Antibiotic discovery: combining isolation chip (Ichip) technology and co-culture technique. *Appl. Microbiol. Biotechnol.* 102, 7333–7341. doi: 10.1007/s00253-018-9193-0
- 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. *GigaSci* 1, 18. doi: 10.1186/2047-217X-1-18
- Madigan, M. T., and Jung, D. O. (2009). “An overview of purple bacteria: systematics, physiology, and habitats,” in *The Purple Phototrophic Bacteria*, eds. C. N. Hunter, F. Daldal, M. C. Thurnauer, and J. T. Beatty (Dordrecht: Springer Netherlands), 1–15.
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11, 625–631. doi: 10.1038/nchembio.1890
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinf.* 14, 60. doi: 10.1186/1471-2105-14-60
- Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and 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
- Meier-Kolthoff, J. P., and Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 10, 2182. doi: 10.1038/s41467-019-10210-3
- MIDI Inc. (2012). *Sherlock Microbial Identification System Operating Manual*. Available online at: http://midi-inc.com/pdf/Sherlock_MIS_Operating_Manual.pdf
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913
- Müller, R., Wenzel, S., and Garcia, R. (2013). *Synthetic Pathway Enzymes For The Production Of Argyrins*. Washington, DC: U.S. Patent and Trademark Office.
- Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* 16, 60–68. doi: 10.1038/s41589-019-0400-9
- Nesic, M., Ryyffel, D. B., Maturano, J., Shevlin, M., Pollack, S. R., Gauthier, D. R., et al. (2022). Total synthesis of darobactin *a*. *J. Am. Chem. Soc.* 144, 14026–14030. doi: 10.1021/jacs.2c05891
- Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71, 5056. doi: 10.1099/ijsem.0.005056
- Pan, G., Xu, Z., Guo, Z., Ma, M., Yang, D., Zhou, H., et al. (2017). Discovery of the leinamycin family of natural products by mining actinobacterial genomes. *Proc. Natl. Acad. Sci. U.S.A.* 114. doi: 10.1073/pnas.1716245115
- Paoli, L., Ruscheweyh, H.-J., Forneris, C. C., Hubrich, F., Kautsar, S., Bhushan, A., et al. (2022). Biosynthetic potential of the global ocean microbiome. *Nature* 607, 111–118. doi: 10.1038/s41586-022-04862-3
- Parkinson, E. I., Tryon, J. H., Goering, A. W., Ju, K.-S., McClure, R. A., Kembell, J. D., et al. (2018). Discovery of the tyrobutaine natural products and their biosynthetic gene cluster *via* metabologenomics. *ACS Chem. Biol.* 13, 1029–1037. doi: 10.1021/acscchembio.7b01089
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Piwosz, K., Villena-Alemany, C., and Mujakić, I. (2022). Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME J* 16, 1046–1054. doi: 10.1038/s41396-021-01142-2
- Pradella, S., Allgaier, M., Hoch, C., Päufer, O., Stackebrandt, E., and Wagner-Döbler, I. (2004). Genome organization and localization of the *pufIm* genes of the photosynthesis reaction center in phylogenetically diverse marine *Alphaproteobacteria*. *Appl. Environ. Microbiol.* 70, 3360–3369. doi: 10.1128/AEM.70.6.3360-3369.2004
- Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H
- Ramírez-Durán, N., de la Haba, R. R., Vera-Gargallo, B., Sánchez-Porro, C., Alonso-Carmona, S., Sandoval-Trujillo, H., et al. (2021). Taxogenomic and comparative genomic analysis of the genus *Saccharomonospora* focused on the identification of biosynthetic clusters PKS and NRPS. *Front. Microbiol.* 12, 603791. doi: 10.3389/fmicb.2021.603791
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *PNAS* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Rössler, M., and Müller, V. (2001). Osmoadaptation in bacteria and archaea: common principles and differences. *Environ. Microbiol.* 3, 743–754. doi: 10.1046/j.1462-2920.2001.00252.x
- Romine, M. F., Rodionov, D. A., Maezato, Y., Osterman, A. L., and Nelson, W. C. (2017). Underlying mechanisms for syntrophic metabolism of essential enzyme cofactors in microbial communities. *ISME J* 11, 1434–1446. doi: 10.1038/ismej.2017.2
- Sánchez-Andrea, I., Florentino, A. P., Semerel, J., Strepis, N., Sousa, D. Z., and Stams, A. J. M. (2018). Co-culture of a novel fermentative bacterium, *Lucifera butyrica* gen. nov. sp. nov., with the sulfur reducer *Desulfurella amilsii* for enhanced sulfidogenesis. *Front. Microbiol.* 9, 3108. doi: 10.3389/fmicb.2018.03108
- Sasse, F., Steinmetz, H., Schupp, T., Petersen, F., Memmert, K., Hofmann, H., et al. (2002). Argyrins, immunosuppressive cyclic peptides from myxobacteria. I. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* 55, 543–551. doi: 10.7164/antibiotics.55.543

- Sasser, M. (1990). *Bacterial Identification by Gas Chromatographic Analysis of Fatty Acid Methyl Esters (GC-FAME)*. Newark, NY: Microbial ID.
- Sato-Takabe, Y., Nakao, H., Kataoka, T., Yokokawa, T., Hamasaki, K., Ohta, K., et al. (2016). Abundance of common aerobic anoxygenic phototrophic bacteria in a coastal aquaculture area. *Front. Microbiol.* 7, 1996. doi: 10.3389/fmicb.2016.01996
- Scherlach, K., and Hertweck, C. (2021). Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* 12, 3864. doi: 10.1038/s41467-021-24133-5
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* 11, e0163962. doi: 10.1371/journal.pone.0163962
- Simon, M., Scheuner, C., Meier-Kolthoff, J. P., Brinkhoff, T., Wagner-Döbler, I., Ulbrich, M., et al. (2017). Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* 11, 1483–1499. doi: 10.1038/ismej.2016.198
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stewart, E. J. (2012). Growing unculturable bacteria. *J. Bacteriol.* 194, 4151–4160. doi: 10.1128/JB.00345-12
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569
- Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., et al. (2023). MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.* 51, D603–D610. doi: 10.1093/nar/gkac1049
- Undabarrena, A., Valencia, R., Cumsille, A., Zamora-Leiva, L., Castro-Nallar, E., Barona-Gomez, F., et al. (2021). *Rhodococcus* comparative genomics reveals a phylogenomic-dependent non-ribosomal peptide synthetase distribution: insights into biosynthetic gene cluster connection to an orphan metabolite. *Microb. Genomics* 7, 621. doi: 10.1099/mgen.0.000621
- Ward, A., and Allenby, N. (2018). Genome mining for the search and discovery of bioactive compounds: the Streptomyces paradigm. *FEMS Microbiol. Lett.* 12, 240. doi: 10.1093/femsle/fny240
- Yang, Q., Cahn, J. K. B., Piel, J., Song, Y.-F., Zhang, W., and Lin, H.-W. (2022). Marine sponge endosymbionts: structural and functional specificity of the microbiome within *Euryspongia arenaria* cells. *Microbiol. Spectr.* 10, e02296–e02221. doi: 10.1128/spectrum.02296-21
- Yoon, S.-H., Ha, S.-M., 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
- Zheng, Q., Zhang, R., Koblížek, M., Boldareva, E. N., Yurkov, V., Yan, S., et al. (2011). Diverse arrangement of photosynthetic gene clusters in aerobic anoxygenic phototrophic bacteria. *PLoS ONE* 6, e25050. doi: 10.1371/journal.pone.0025050