Genotype–phenotype correlations within the Geodermatophilaceae

Maria del Carmen Montero-Calasanz1,2*, Adnan Yaramis2, Manfred Rohde3, Peter Schumann4, Hans-Peter Klenk2 and Jan P. Meier-Kolthoff5

1IFAPA Las Torres-Andalusian Institute of Agricultural and Fisheries Research and Training, Junta de Andalucía, Seville, Spain, 2School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, 3Central Facility for Microscopy, HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany, 4Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, 5Department Bioinformatics and Databases, Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

The integration of genomic information into microbial systematics along with physiological and chemotaxonomic parameters provides for a reliable classification of prokaryotes. *In silico* analysis of chemotaxonomic traits is now being introduced to replace characteristics traditionally determined in the laboratory with the dual goal of both increasing the speed of the description of taxa and the accuracy and consistency of taxonomic reports. Genomics has already successfully been applied in the taxonomic rearrangement of *Geodermatophilaceae* (Actinomycetota) but in the light of new genomic data the taxonomy of the family needs to be revisited. In conjunction with the taxonomic characterisation of four strains phylogenetically located within the family, we conducted a phylogenetic analysis of the whole proteomes of the sequenced type strains and established genotype–phenotype correlations for traits related to chemotaxonomy, cell morphology and metabolism. Results indicated that the four isolates under study represent four novel species within the genus *Blastococcus*. Additionally, the genera *Blastococcus*, *Geodermatophilus* and *Modestobacter* were shown to be paraphyletic. Consequently, the new genera *Trujillonella*, *Pleomorpha* and *Goekera* were proposed within the *Geodermatophilaceae* and *Blastococcus endophyticus* was reclassified as *Trujillonella endophytica* comb. nov., *Geodermatophilus daqingensis* as *Pleomorpha daqingensis* comb. nov. and *Modestobacter deserti* as *Goekera deserti* comb. nov. Accordingly, we also proposed emended descriptions of *Blastococcus aggregatus*, *Blastococcus jejuensis*, *Blastococcus saxobsidens* and *Blastococcus xanthilyniticus*. *In silico* chemotaxonomic results were overall consistent with wet-lab results. Even though *in silico* discriminatory levels varied depending on the respective chemotaxonomic trait, this approach is promising for effectively replacing and/or complementing chemotaxonomic analyses at taxonomic ranks above the species level. Finally, interesting but previously overlooked insights regarding morphology and ecology were revealed by the presence of a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly in all representatives of...
the family. A rich carbon metabolism including four different CO₂ fixation pathways and a battery of enzymes able to degrade complex carbohydrates were also identified in Blastococcus genomes.

KEYWORDS
Trujillonella, Pleomorpha, Goekera, phylogenetic systematics, in silico chemotaxonomy.

Introduction

The family Geodermatophilaceae, belonging to order Geodermatophilales (Sen et al., 2014), was initially proposed by Normand et al. (1996), confirmed by Stackebrandt et al. (1997), formally described by Normand (2006) and later emended by Zhi et al. (2009) and Montero-Calasanz et al. (2017). The family accommodates the genera Blastococcus (Ahrens and Moll, 1970; Urzi et al., 2004; Lee, 2006; Hezbri et al., 2016), Geodermatophilus, the type genus (Luedemann, 1968; Montero-Calasanz et al., 2017; Montero-Calasanz, 2020a), Klenkia (Montero-Calasanz et al., 2017), and Modestobacter (Mevs et al., 2000; Reddy et al., 2007; Xiao et al., 2011; Qin et al., 2013; Montero-Calasanz et al., 2019).

The genus Blastococcus consists of Gram-reaction-positive aerobic bacteria that can be distinguished by the presence of individual motile rod-shaped cells and/or non-motile cocci (Lee, 2006) with a tendency to form aggregates (Urzi et al., 2004; Hezbri et al., 2016). At the time of writing, the genus comprises 10 validly named species including B. aggregatus (type species), isolated from sediment from Baltic Sea (Ahrens and Moll, 1970), B. atacamensis, isolated from an extreme hyper-arid Atacama Desert soil (Castro et al., 2018), B. capsensis, isolated from an ancient Roman pool in Tunisia (Hezbri et al., 2016), B. colisei, isolated from the ruins of a Roman amphitheatre in Tunisia (Hezbri et al., 2017), B. deserti, isolated from Gurbantunggut Desert in China (Yang et al., 2019), B. endophyticus, isolated from Chinese medicinal plant leaves (Zhu et al., 2013; Hezbri et al., 2016), B. jejuensis, isolated from soil in Korea (Lee, 2006; Hezbri et al., 2016), B. litoris, isolated from sea-tidal flat sediment (Lee et al., 2018), B. saxobsidens, isolated from an archaeological site in Greece (Urzi et al., 2004), and B. xanthinilyticus, isolated from a marble sample collected from the Bulla Regia monument in Tunisia (Hezbri et al., 2018).

The integration of genomic information into microbial systematics (Klenk and Göker, 2010) along with physiological and chemotaxonomic parameters enables a reliable classification of prokaryotes (Chun et al., 2018). Laboratory-based DNA–DNA hybridisation methods have already been routinely replaced by overall genome relatedness indices (OGRi; Chun et al., 2018) of which digital DNA:DNA hybridization was shown to be the best-performing method (Meier-Kolthoff et al., 2013a). The use of in silico analysis for chemotaxonomic traits is now being introduced to replace characteristics traditionally determined in the laboratory (Amaral et al., 2014; Fotedar et al., 2020; Lawson et al., 2020) with the dual goal of both increasing the speed of the description of taxa and the accuracy and consistency of taxonomic reports (Chun and Rainey, 2014; Whitman, 2015).

Genomics has already successfully been applied in the taxonomic rearrangement of Geodermatophilaceae (Montero-Calasanz et al., 2017) although other research studies based on genomic data are still scarce in the family. In particular, only two papers were published so far about the genus Blastococcus including genomic and proteomic mining of B. saxobsidens DD2 and genome mining of B. atacamensis which revealed numerous genes involved in stress response and adaptation to harsh habitats (Chouaia et al., 2012; Sghaier et al., 2016) apart from an exceptional potential to produce novel natural compounds (Castro et al., 2018).

Based on genome-scale data and phenotypic evidence according to the principles of phylogenetic systematics (Hennig, 1965; Wiley and Lieberman, 2011), this study characterises four novel species in the genus Blastococcus, introduces the new genera Trujillonella, Pleomorpha, and Goekera in Geodermatophilaceae and reclassifies Blastococcus endophyticus as Trujillonella endophytica comb. nov., Geodermatophilus daqingensis as Pleomorpha daqingensis comb. nov. and Modestobacter deserti as Goekera deserti comb. nov., all representing type species of the new genera. Accordingly, we also propose emended descriptions within Blastococcus. Additionally, we emphasise how genomics could effectively replace and/or complement some of the phenotypic analyses routinely carried out in microbial systematics.

Materials and methods

Isolation

Strains AT 7-1⁴, AT 7-8, AT 7-14⁴, and AT 7(-2)-11⁴ were isolated during a screening of microorganisms from soil samples from the Atacama Desert (Chile) in the 90s and deposited by Prof. Fred A. Rainey (University of Alaska Anchorage, United States) at Leibniz Institute DSMZ. On the other hand, strain GIS⁴ was isolated by Eppard et al. (1996) from leachate of a landfill in Vancouver (Canada). All strains were subsequently accessed in the
DSMZ open collection as DSM 44268^T, DSM 44269, DSM 44270^T, DSM 44272^T, and DSM 44205^T, respectively.

**Genotypic analysis**

Genomic DNA extraction, PCR-mediated amplification of the 16S rRNA gene, and purification of the PCR product were carried out as previously described by Rainey et al. (1996). For genome sequencing, strains were cultivated in GYM Streptomyces broth (DSMZ medium 65, pH 6.8 ± 0.2; https://www.dsmz.de/collection/catalogue/microorganisms/culture-technology/list-of-media-for-microorganisms) at 28 °C and genomic DNA was extracted by using JetFlex™ Genomic DNA Purification Kit (Thermo Fisher Scientific) following the manufacturer’s instructions. The project information is available through the Genomes Online Database (Mukherjee et al., 2017). The draught genomes were generated at the DOE Joint Genome Institute (JGI) as part of the Genomic Encyclopedia of Archaeal and Bacterial Type Strains, Phase II (KMG-II): “From individual species to whole genera” (Kyripides et al., 2014) and ACTINO1000: “Exploiting the genomes of the Actinobacteria: plant growth promoters and producers of natural products and energy relevant enzymes united in a taxonomically unresolved phylum,” following the same protocol as in Nouioui et al. (2017). All genomes were annotated using the DOE-JGI annotation pipeline (Huntemann et al., 2015; Chen et al., 2016) and released through the Integrated Microbial Genomes system (Chen et al., 2017). In addition, genomes were uploaded to RAST (Aziz et al., 2008; Brettin et al., 2015) and analysed through the SEED viewer (Overbeek et al., 2014). CRISPRFinder (Grissa et al., 2007) was used to identify CRISPR elements. Further details about sequencing projects are summarised in the Supplementary Table S1.

Phylogenetic analysis of the 16S rRNA gene sequences from the type strains of all species with validly named genera in Geodermatophilaceae, as well as those from Antricoccus suffuscus DSM 100065^T, Cryptosporangium arvum DSM 44712^T, Cryptosporangium aurantiacum DSM 46144^T, Epidermidibacterium keratini JCM 31644^T, Longivirga aurantiaca NBRC 112237^T and Sporichthya polymorpha DSM 43042^T for use as outgroup, was conducted as previously described (Göker et al., 2011; Montero-Calasanz et al., 2014). Pairwise 16S rRNA gene sequence similarities were calculated as recommended by Meier-Kolthof et al. (2013b) to determine strains with ≥99.0% similarity, between which (digital) DNA-DNA hybridization experiments were conducted for clarifying species affiliation. Genome-scale phylogenies were inferred from the available Geodermatophilaceae (and outgroup) whole proteome sequences using the high-throughput version (Meier-Kolthof et al., 2014a) of the Genome BLAST Distance Phylogeny (GBDP) approach (Auch et al., 2010) in conjunction with FastME (Lefort et al., 2015). The tree was rooted according to Nouioui et al. (2018). The entire approach is implemented in the Type (Strain) Genome Server (TYGS; Meier-Kolthof and Göker, 2019; Meier-Kolthof et al., 2022). The GBDP tree restricted to the well-supported branches (≥95% pseudo-bootstrap support) was used as a backbone constraint in a subsequent 16S rRNA gene sequence analysis to integrate information from genome-scale data (Hahnke et al., 2016). Digital DNA-DNA hybridisations were conducted using the recommended settings of the Genome-To-Genome Distance Calculator (GGDC) version 3.0 (Meier-Kolthof et al., 2013a) as implemented in the TYGS (Meier-Kolthof and Göker, 2019; Meier-Kolthof et al., 2022). The G+C content was calculated from the genome sequences as described by Meier-Kolthof et al. (2014b). The genomic homogeneity of strains DSM 44268^T and DSM 44269 was examined by ribotyping and matrix-assisted laser-desorption/ionisation time-of-flight (MALDI-TOF) mass spectra (MS). Automated ribotyping of PuII-digested samples was carried out using the RiboPrinter microbial characterisation system (Qualicon, DuPont) as described in Bruce (1996) and Stackebrandt et al. (2002). MALDI-TOF analysis was performed according to Tóth et al. (2008).

**Phenotypic analysis**

**Morphological and physiological tests**

Morphological features of colonies were examined by stereomicroscope on GYM Streptomyces agar at 28 °C and observed periodically during 7 days. Bacterial cell morphology and motility was determined by phase contrast microscopy with a magnification of 100x and with a field-emission scanning electron microscope (FE-SEM Merlin, Zeiss, Germany). Catalase production was tested on slides with the addition of some drops of 3% H₂O₂ solution into a loop of bacterial biomass. It was considered positive when the production of bubbles was observed after mixing. Oxidase activity was determined by obtaining blue-purple colour after addition of some drops of 1% (w/v) solution of N,N,N′,N′-tetramethyl-p-phenylenediamine (Sigma-Aldrich) to a loop of bacterial biomass spread on filter paper. Gram reaction was performed according to the procedure developed by Gregersen (1978). API ZYM system (bioMérieux) was used to determine enzymatic activities according to the manufacturer’s protocol at 28 °C after 24h incubation. Homogeneous bacterial suspensions of strains DSM 44205^T, DSM 44268^T, DSM 44269, DSM 44270^T, and DSM 44272^T along with the reference type strains in the genus Blastococcus were prepared in viscous inoculating fluid (IF C) at 86% T (Transmittance) for DSM 46835^T, 83% T for B. capsensis DSM 46835^T, and 90% T for DSM 44270^T, B. colisei DSM 46837^T, and B. xanthinilyticus DSM 46842^T, 90% T for strains DSM 44268^T, DSM 44269, and DSM 44205^T, 91% T for DSM 44272^T, and 95% T for all the others reference strains. Then those were inoculated in GEN III inoculating fluid (IF C) at 86% T (Transmittance) for DSM 44272^T, 90% T for all the others reference strains. Then those were inoculated in GEN III Microplates per duplicate and incubated for 10 days at 28 °C in an Omnilog device (BIOLOG Inc., Hayward, CA, United States). Data were analysed using the opm package R v.1.3.72 (Vaas et al., 2012, 2013). Reactions that reflected different behaviour between replicates were regarded as ambiguous. Temperature ranges were evaluated on GYM Streptomyces medium for 15 days at 4, 10, 15,
20, 25, 30, 37, 40 and 45°C. pH ranges were determined using modified ISP2 medium (Montero-Calasanz et al., 2015) from 4.5 to 12.5 in increments of 0.5 pH unit adjusting with NaOH or HCl since the use of a buffer system inhibited the growth of strains (Montero-Calasanz et al., 2014). pH of solidified plates was measured before inoculation and when recording results (15 days post-inoculation). Hydrolysis of specific substrates such as tyrosine (Gordon and Smith, 1955), casein, starch, xanthine as well as hypoxanthine was also tested as outlined by Montero-Calasanz et al. (2015).

**Chemotaxonomic tests**

For chemotaxonomic analysis, strains were grown in GYM *Streptomyces* broth with shaking (120 rpm) at 28°C for 7 days. Bacterial biomass was collected and subsequently freeze dried. Extraction of whole-cell amino acids and sugars was carried out as developed by Lechevalier and Lechevalier (1970) followed by thin layer chromatography (TLC) analysis and identification (Stanek and Roberts, 1974). Analysis of peptidoglycan hydrolysates (6N HCl, 100°C for 16h) was performed on cellulose TLC plates as described by Schleifer and Kandler (1972). Menaquinones (MK) analyses (Kroppenstedt, 1982) were accomplished as indicated by Collins et al. (1977) following the identification of MK by high-performance liquid chromatography (HPLC; Kroppenstedt, 1982). Polar lipids were extracted and identified by 2D TLC as outlined by Minnikin et al. (1984) with modifications proposed by Kroppenstedt and Goodfellow (2006). Dragendorff’s reagent (Merck Millipore, 102035) was additionally sprayed to identify choline-containing lipids (Tindall, 1990). For fatty acid analysis, cell biomass grown on PYGV (DSMZ 621) agar plates for 16 days at 20°C was harvested and fatty acid methyl esters were extracted according to Sassar (2001). Microbial Identification System (MIDI) Sherlock Version 6.1 (method TSBA40, ACTIN6 database) was used for performing data analysis. All the chemotaxonomic analyses for strains were carried out under standardised conditions.

**Results and discussion**

**Basic genome statistics**

Standard draft genomes were obtained for strains DSM 44205\(^1\), DSM 44268\(^\), DSM 44270\(^\), and DSM 44272\(^\) and type strains of species *B. aggregatus*, *B. endophyticus* and *B. xanthinilyticus*. An improved high-quality genome draft according to GOLD sequencing quality standards (Mukherjee et al., 2017) was also obtained for both *B. colisei* DSM 46837\(^2\) and *B. saxobisids* DSM 44509\(^\) (Supplementary Table S1). The genome sizes ranged from 4.0 Mbp (strain DSM 44205\(^1\)) to 5.1 Mbp (strains DSM 44272\(^\) and DSM 46837\(^\)), with an average genome size for all sequenced *Blastococcus* strains of 4.6 ± 0.4 Mbp (Supplementary Table S2). The genomic G + C content of the sole chromosome in the type strains varied between 72.5–74.6%.

Deviations from the published G + C contents of reference strains were all below the 1% threshold determined within species (Meier-Kolthoff et al., 2014b). In accordance with the genome sizes, the number of protein-coding genes ranged from 3,939 for strain DSM 44205\(^1\) to 5,007 for strain DSM 44272\(^\), not observing a *priori* any association between genome size and habitat of isolation. Both tRNA and rRNA genes represented between 1.3% (47 tRNA and five rRNA genes) in *B. xanthinilyticus* DSM 46842\(^2\) and 2.2% (70 tRNA and nine rRNA genes) in DSM 44205\(^1\) of the whole genome sequences. A single copy of the 16S rRNA gene was identified for strains DSM 44270\(^\), DSM 44272\(^\), *B. aggregatus* DSM 4725\(^2\), and *B. xanthinilyticus* DSM 46842\(^2\). Three almost identical 16S rRNA genes copies were annotated in the genome sequences of strain DSM 44205\(^1\), *B. colisei* DSM 46837\(^2\), *B. endophyticus* DSM 45413\(^2\), and *B. saxobisids* DSM 44509\(^\), and four almost identical ones in the genome sequence of strain DSM 44268\(^\) (intragenomic heterogeneity of 16S rRNA genes is below the 1% level for species delimitation; Meier-Kolthoff et al., 2013a). Previous studies already reported low copy numbers (2–4) for slow-growing actinobacteria (Sun et al., 2013; Vetoryský and Baldrían, 2013) isolated from oligotrophic environments (Klappenbach et al., 2000). The number of rRNA operon copies is suggested to be related to the life strategy of bacteria and how quickly they respond to favourable changes in environmental conditions and nutrient availability (Klappenbach et al., 2000). Hence, tRNA copy number appears to be correlated with growth speed. The number of identified tRNAs ranged from 47 to 49 in all the *Blastococcus* representatives except in the genome sequence of strain DSM 44205\(^1\) that contained 70 tRNAs. The 20 standard tRNA genes were present in all genomes but in a different distribution. For example, four tRNA-Arg gene copies (six gene copies for strain DSM 44205\(^1\)) versus one tRNA-Tyr gene copy. Such uneven distribution supports the co-adaptation between tRNA abundance and specific codon usage of each organism as a higher ratio of optimal codons may facilitate the transcription efficiency (Du et al., 2017). Out of the 70 (DSM 44205\(^1\)) and 48 (B. endophyticus) annotated tRNAs, seven (DSM 44205\(^1\)) and one (B. endophyticus), respectively, were identified as pseudogene(s). The latter are known to be dysfunctional in translation but to be potentially involved in modulating other processes such as antibiotic and cell wall biosynthesis and gene expression (Rogers et al., 2012). One copy of nonsense suppressor tRNA (tRNA amber gene) was also identified in the repertoires of DSM 44270\(^\), DSM 44272\(^\), and *B. colisei*, which could prevent the premature termination of translation caused by nonsense mutations (Ko et al., 2013). The number of pseudogenes was within the expected values for free-living prokaryotes (0%–5%; Liu et al., 2004) and varied from zero (for example: DSM 44205\(^1\), DSM 44272\(^\), B. endophyticus, B. xanthinilyticus or *B. atacamensis*, amongst others) to 123 in *B. colisei* DSM 46837\(^\) (2.5%). Each genome contained a largely unique set of pseudogenes, with over half of those annotated as “hypothetical” proteins, suggesting that pseudogenes in *Blastococcus* are formed and eliminated relatively rapidly as previously reported by Kuo and Ochman (2010) for
Salmonella genomes. In addition, the number of paralogous genes varied from 437 (11%) in strain DSM 44205 to 1,601 (32%) in B. colisei DSM 46837; the signal peptide percentage was 3.0%–5.8% whilst the percentage of transmembrane proteins ranged from 19.7% to 24.7%. According to CRISPRFinder, the number of “questionable” Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) ranged from zero to seven (Supplementary Table S2). Additional analyses to test the genetic environment often associated with CRISPR structures should nevertheless be carried out to confirm the presence of a naturally occurring genome editing system in Blastococcus representatives.

The average percentage of genes with a predicted function in the Clusters of Orthologous Groups (COGs) database was 63.3 ± 2.3% (ranged from 2,558 to 3,117 genes; Supplementary Table S2). As expected, functional COG profiles of Blastococcus strains are similar (Supplementary Figure S1; Supplementary Table S3).

**Phylogenetic analysis**

The phylogenetic tree inferred by whole proteomes of the sequenced type strains and strains under study revealed the affiliation of DSM 44268, DSM 44270, DSM 44272 and DSM 44205 to the genus Blastococcus (Figure 1). The subtree containing the genera Geodermatophilus, Modestobacter and Blastococcus was
not monophyletic which was caused by three misclassified species. First, B. endophyticus DSM 45413\textsuperscript{3} appeared with maximum support as a sister group of Geodermatophilus and was not phylogenetically placed within the maximally supported subtree harbouring all the other Blastococcus type strains. Second, G. daqingensis DSM 104001\textsuperscript{2} represented a maximally supported sister group to the large subtree comprising all other Geodermatophilus and Blastococcus type strains. Third, M. deserti CPCC 205119\textsuperscript{11} formed a deeply-branching and maximally supported sister group of Kleina.

The constrained 16S rRNA phylogenetic tree (Figure 2) based on gene sequences of all validly published species in the family and strains under study, including strain DSM 44269, confirmed the observations from the proteome-based GBDP analysis. Moreover, strains DSM 44268\textsuperscript{1} and DSM 44269 were placed with high support in a clade together with strains DSM 44205\textsuperscript{2} and B. aggregatus DSM 4725\textsuperscript{5}.

The 16S rRNA gene sequence similarity values between (i) DSM 44268\textsuperscript{1} and DSM 44269, (ii) between DSM 44205\textsuperscript{2}, DSM 44270\textsuperscript{7} and G. aggregatus DSM 4725\textsuperscript{5}, and, (iii) between DSM 44270\textsuperscript{7} and B. litoris GP-S2-8\textsuperscript{10}, ranged above the Actinobacteria-specific 16S rRNA threshold of 99.0% (accepting an error probability of 1%; Meier-Kolthoff et al., 2013b).

It was thus necessary to investigate the species status via digital DDH (dDDH). The dDDH values between strain DSM 44268\textsuperscript{1} and strains DSM 44205\textsuperscript{2} (30.8%), DSM 44270\textsuperscript{7} (23.8%), and B. aggregatus DSM 4725\textsuperscript{5} (30.3%) and between B. litoris GP-S2-8\textsuperscript{10} and strains DSM 44270\textsuperscript{7} (26.9%) resulted in values below the 70% threshold throughout, thus indicating distinct species (Wayne et al., 1987).

In addition, the MALDI-TOF dendrogram (Supplementary Figure S2) confirmed the affiliation of strains DSM 44268\textsuperscript{1} and DSM 44269 to the same species showing a 100% similarity regarding mass spectra patterns of proteins and separated from those under study at a distance level ~900 arbitrary units. The riboprobing patterns of the DSM 44268\textsuperscript{1} and DSM 44269 isolates were also highly similar (98.3% similarity; Supplementary Figure S3) but differentiated them as clonal variants of each other.

**Phenotypic-genotypic correlations**

**Morphology and physiology**

All target strains were aerobic, Gram-reaction positive and showed coccolid cells with a tendency to form aggregates as previously described by Urzi et al. (2004) for the genus Blastococcus (Supplementary Figure S4). In addition, under studied culture conditions, reproduction by budding was predominant although bacterial binary fission was also observed (Supplementary Figure S4). It was already indicated for strain DSM 44205\textsuperscript{2} by Eppard et al. (1996). Effectively, the presence of the cell division protein FtsZ and related genes such as the actin-like protein MreB were annotated in all genomes but no known mechanisms involved in division via budding were identified (Supplementary Table S4). Whilst FtsZ is an essential structural homologue of the eukaryotic cytoskeletal element tubulin (Lowe and Amos, 1998) and involved in forming a cytokinetic ring at the division site (Wang and Lutkenhaus, 1993), MreB is mostly essential for cell elongation and maintenance of the cell (Rivas-Marin et al., 2020). Mechanisms of bacterial division by budding are still poorly understood in evolutionary cell biology (Rivas-Marin et al., 2020). Geodermatophilaceae representatives are described as aerobic bacteria (Montero-Calasanz, 2020b). Nevertheless, some microaerophilic representatives isolated from the oxic/anoxic interface of marine sediments were also reported (Ahrens and Moll, 1970). To date no anaerobic isolate was described, although the presence of reductases using anaerobic terminal electron acceptors such as arsenate, ferredoxin or flavodoxins as well as a repertoire of genes involved in the fermentation of propanoate, butanoate and pyruvate, amongst others, in the Blastococcus genomes could indicate that anaerobic respiration may be possible. It was already suggested by Sghaier et al. (2016).

*In vivo* motility was not observed in any of the studied strains. The presence of homologues involved in flagellar biosynthesis and function and chemotaxis in all the genome sequences, nevertheless, suggests that Blastococcus strains may be flagellated. Chemotaxis and motility may have been influenced by culture conditions. In fact, Ahrens and Moll (1970) suggested that the motile rod stage in B. aggregatus could be favoured under microaerophilic conditions. Later studies carried out by Holt and Chaubal (1997) for Salmonella pullorum strains, confirmed that flagellar synthesis is affected by agar concentration, carbohydrate concentration and type, and incubation temperature. The presence of flagellar gene clusters is consistent with observed motility by type strains of B. jejuensis and B. saxobsidens (Urzi et al., 2004; Lee 2006) and other strains in the family (Montero-Calasanz et al., 2017). In fact, flagellin synthesis was identified as the most highly expressed protein in the proteogenome of Blastococcus sp. DD2 (Sghaier et al., 2016). The core set of flagellar structural genes for Geodermatophilaceae representatives seems to be largely consistent with that described by Liu and Ochman (2007) for the bacterial domain. Twenty-eight different genes comprise specific proteins that form the filament (\textit{flic}, two copies for strain DSM 44205\textsuperscript{2}, DSM 44270\textsuperscript{7}, and B. xanthinilyticus), the hook–flagament junction (\textit{flgE} and \textit{flgK}), the hook (\textit{flgE}), the rod (\textit{fliB} and \textit{fliC}), the MS ring (\textit{fliF}; absent in DSM 44272\textsuperscript{5}), the C ring (\textit{fliG}, \textit{fliM} and \textit{fliN}), the motor (\textit{motA} and \textit{motB}), and the export apparatus (\textit{flhA}, \textit{fliB}, \textit{fliP}, \textit{fliQ}, and \textit{fliR}). In addition, \textit{fliG} encoding for the hook-capping protein, which is required for flagellar assembly (Liu and Ochman, 2007), and other broadly distributed flagellar structural genes such as \textit{fliD} (filament cap), \textit{fliE}, \textit{fliK} and \textit{fliH} (hook) as well as, \textit{fliZ}, \textit{fliS} and \textit{fliO} are also part of the core set in Geodermatophilaceae (Supplementary Table S4). The L and P ring proteins \textit{FliH} and \textit{FliI} are absent in Gram-positive bacteria. Chemotaxis, on the other hand, may be controlled by the universal two-component system. Methyl-accepting chemotaxis proteins (MCP) would control autophosphorylation of a cytoplasmic histidine kinase (CheA), via a coupling protein (CheW), which would activate the response regulator CheY and, consequently, interact with the switch...
mechanism in the flagellar motor (Szurmant and Ordal, 2004).

Surprisingly, a dedicated ribose-binding protein (ribose transport system substrate-binding protein, RbsB) was annotated in the genome sequences of DSM 44205T, DSM 44268T, and DSM 44270T suggesting a key role of ribose as chemoeffector in those strains at low concentrations (Galloway and Furlong, 1977).
In addition, *tad* (tight adherence) genes were unambiguously annotated in all the studied genomes (Supplementary Table S4). The *tad* genes encode the machinery required for the assembly of adhesive Flp (fimbrial low-molecular-weight protein) pili (Tomich et al., 2007). Those are essential for biofilm formation, colonisation, DNA transfer and pathogenesis and have implications in functions such as phage binding, rough colony morphology, and twitching motility (Proft and Baker, 2009). The presence of *Tad* complexes in *Geodermatophilaceae* genomes could be important for successful colonisation of altered stones and patina formation (Eppard et al., 1996; Urzi et al., 2001). Howbeit, the presence and abundance of *tad* genes varied amongst strains suggesting that strains may present different adherence-related phenotypes as previously demonstrated by Kachlany et al. (2000) and Perez et al. (2006).

Production of spores was not observed in any of the isolates under the given culture conditions. Nevertheless, the unambiguous presence of single copies of *SsgA*-like proteins (SALPs) in all studied genomes suggested that sporulation may be a typical feature previously overlooked in *Blastococcus* spp. and that, similar to *Geodermatophilaceae* spp., the presence of SALPs may be affected by growth conditions (Ishiguro and Wolfe, 1970). SALPs occur exclusively in morphologically complex actinomycetes where they play an important role in morphogenesis and control of cell division (Jakimowics and van Wezel, 2012) and are essential for sporulation (van Wezel et al., 2000). Furthermore, the presence of single SALPs (presumably and invariably *SsgB*) in *Blastococcus saxobidens* DD2 and other *Geodermatophilaceae* representatives was already described by Girard et al. (2013) and is suggested as a link between the number of SALPs and the complexity of the developmental process (Traag and van Wezel, 2008), suggesting that these actinomycetes would produce single spores.

Colonies were opaque with a moist surface and regular margin and varied from coral (strain DSM 44268° and DSM 44205°) to brownish-orange in colour (Strains DSM 44270° and DSM 44272° and DSM 44269). Accordingly, phytoene synthase (*crtB*; EC 2.5.1.32) and phytoene desaturase (*crtI*; 4 step enzyme; EC 1.3.99.28, EC 1.3.99.29, EC 11.3.99.31), the key enzymes involved in the biosynthesis of the colourful carotenoids ζ-carotene, neurosporene, and lycopene (Paniagua-Michel et al., 2012) were identified in all the genomes. Yet, the putative operon mentioned by Sghaier et al. (2016) for *B*. sp. DD2 consisting of *hopC* [squalene-associated FAD-dependent desaturase (fragment)], *isPA* (generallypterol prophophosphate synthase), *shc* (squalene-hopene cyclise), *hpnH* (hopsanid biosynthetic associated radical SAM protein HpnH), was not identified in these genomes, except for *crtB* and *ilvC* (keto-acid reductoisomerase; Supplementary Table S4).

Strains grew well on GYM *Streptomyces* and PYGV agar media. They tolerated temperature ranging from 15°C to 37°C, except strain DSM 44269, which only tolerated a more restricted range from 20°C to 37°C. All strains showed optimal growth at 25°C–30°C. In addition, growth was observed at pH 6.0–12.0 for strains DSM 44268° and DSM 44269, at pH 5.0–11.5 for strain DSM 44205°, at 5.0–11.0 for strain DSM 44270°, and at 6.5–11.0 for strain DSM 44272°. All strains had an optimal pH range from 6.5 to 8.0. Regarding tolerance to salt, strains DSM 44269 and DSM 44272° were the least tolerant (0%–1% NaCl, although at 4% NaCl the Biolog system determined an ambiguous result for DSM 44272°), followed by both DSM 44268° and DSM 44205° tolerating 0%–4% NaCl, and strain DSM 44270°, which tolerated up to 8% NaCl. A summary of selected differential phenotypic characteristics is presented in Table 1 (for an overview of phenotypic profiles in *Blastococcus* see Supplementary Figure S5; Supplementary Table S5). Details about core metabolism will be provided later on in this manuscript.

### Chemotaxonomy

Whole-cell hydrolysates of all strains showed meso-diaminopimelic (DAP) acid (Cell wall type III; Lechevalier and Lechevalier, 1970) as the diagnostic diaminoc acid being consistent with the family *Geodermatophilaceae* (Montero-Calasanz, 2020b) and genomic data (presence of *MurE* encoding for UDP-N-acetylmuramoylalanyl-D-glutamate–2,6-diaminopimelate ligase [6.3.2.13] and absence of UDP-N-acetylmuramoyl-L-allyl-D-glutamate—L-lysine ligase [6.3.2.7]; Supplementary Table S4). In addition, D-allyl-D-alanine carboxypeptidase [3.4.16.4] was automatically annotated in all *Blastococcus* representative genomes except in *B. atacamensis* P6° and in most *Geodermatophilaceae* representatives. The presence of diaminopimelate epimerase (*dapE*), responsible for the interconversion of the LL- and meso-isomers of DAP (Anita et al., 1957), was also annotated in all genomes including in the genome of the LL- and meso-isomers of DAP (Montero-Calasanz, 2020b).

As suggested by Hezbr et al. (2016) for the emendation of the genus *Blastococcus*, whole-cell sugar analysis displayed a basic pattern consisting of ribose, glucose, and mannose. Additionally, galactose was identified in all strains under study except in the profile of strain DSM 44205°. Similarly, arabinose was found as part of the sugar cell wall composition in all studied strains excluding strain DSM 44270°. Xylose was detected in strain DSM 44205°.

The predominant menaquinone (MK) for strains DSM 44205°, DSM 44268°, DSM 44269, DSM 44270°, and DSM 44272° was MK-9(H2) 75.9%, 92.6%, 56.7%, 78%, and 62.9%, respectively) in agreement with what was described for *Geodermatophilaceae*.
| Characteristic          | DSM 44205<sup>T</sup> | DSM 44268<sup>T</sup> | DSM 44269 | DSM 44270<sup>T</sup> | DSM 44272<sup>T</sup> |
|------------------------|------------------------|------------------------|------------|------------------------|------------------------|
| Motility               | −                      | +                      | +          | −                      | +                      |
| Oxidation of:          | +/−                    | +                      | +/−        | +/−                    | −/−                    |
| Stachyose              | +                      | +                      | +/−        | +                      | +/−                    |
| α-D-Lactose            | −                      | +                      | +/−        | +                      | +/−                    |
| D-Melibiose            | +/−                    | +/−                    | +/−        | +                      | +/−                    |
| N-Acetyl-D-            | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| Galactosamine          | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| D-Mannose              | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| L-Rhamnose             | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| L-Aspartic acid        | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| Quinic acid            | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| Quinic acid            | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| Citric acid            | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| α-Keto-Glutaric acid   | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| D-Malic acid           | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| L-Malic acid           | +/−                    | +/−                    | +/−        | +/−                    | −/−                    |
| Sugar                  | DPG, PC, PE, PI        | DPG, PC, PE, PI        | DPG, PC, PE, PI | DPG, PC, PE, PI        | DPG, PC, PE, PI        |
| Xylose                 | DPG, PC, PE, PI        | DPG, PC, PE, PI        | DPG, PC, PE, PI | DPG, PC, PE, PI        | DPG, PC, PE, PI        |
| Arabinose              | DPG, PC, PE, PI        | DPG, PC, PE, PI        | DPG, PC, PE, PI | DPG, PC, PE, PI        | DPG, PC, PE, PI        |
| Mannose                | DPG, PC, PE, PI        | DPG, PC, PE, PI        | DPG, PC, PE, PI | DPG, PC, PE, PI        | DPG, PC, PE, PI        |
| Glucose                | DPG, PC, PE, PI        | DPG, PC, PE, PI        | DPG, PC, PE, PI | DPG, PC, PE, PI        | DPG, PC, PE, PI        |
| Fatty acid<sup>a</sup> | iso-C<sub>10:0</sub>   | iso-C<sub>10:0</sub>   | iso-C<sub>10:0</sub> | iso-C<sub>10:0</sub>   | iso-C<sub>10:0</sub>   |
| Fatty acid<sup>a</sup> | anteiso-C<sub>10:0</sub> | anteiso-C<sub>10:0</sub> | anteiso-C<sub>10:0</sub> | anteiso-C<sub>10:0</sub> | anteiso-C<sub>10:0</sub> |
| Fatty acid<sup>a</sup> | 9-methyl-C<sub>16:0</sub> | 9-methyl-C<sub>16:0</sub> | 9-methyl-C<sub>16:0</sub> | 9-methyl-C<sub>16:0</sub> | 9-methyl-C<sub>16:0</sub> |
| Fatty acid<sup>a</sup> | 11, 12, 13, 15, 17 : 0  | 11, 12, 13, 15, 17 : 0  | 11, 12, 13, 15, 17 : 0  | 11, 12, 13, 15, 17 : 0  | 11, 12, 13, 15, 17 : 0  |

<sup>a</sup> Strains: 1, strain DSM 44205<sup>T</sup>; 2, strain DSM 44268<sup>T</sup>; 3, strain DSM 44269; 4, strain DSM 44270<sup>T</sup>; 5, strain DSM 44272<sup>T</sup>; 6, B. aggregatrus DSM 4725<sup>T</sup>; 7, B. atacamensis NRRL B-65468<sup>T</sup>; 8, B. capsulatus DSM 46835<sup>T</sup>; 9, B. collos DSM 46837<sup>T</sup>; 10, B. endophyticus DSM 45413<sup>T</sup>; 11, B. psudoporins DSM 19597<sup>T</sup>; 12, B. stoeckelii DSM 45409<sup>T</sup>; 13, B. xanthinilyticus DSM 46842<sup>T</sup>. +, positive reaction; −, negative reaction; +/−, ambiguous; DPG, diphostatidylglycerol; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PC, phosphatidylcholine; MK, menaquinones.
MK-9(H₉) was however also found in minor levels in the strains isolated from Atacama soils (DSM 44268, 0.7%; DSM 44269, 17.4%; DSM 44270, 7.4%; DSM 44272, 20.2%). Minor amounts of MK-9(H₉) were also observed in the MK profiles of strains DSM 44269, DSM 44270, and DSM 44272 (1.6, 1.3, and 15.0%, respectively). MK-8(H₈) was also identified in strains DSM 44205 (5.5%) and DSM 44270 (2.9%). But MK-10(H₁₀) was only found in strains DSM 44268 (4.7%) and DSM 44269 (0.6%) and MK-9(H₉) in strain DSM 44270 (1.1%). The presence of quinones other than MK-9(H₉) is commonly observed in other representatives in Geodermatophilaceae (Montero-Calasanz et al., 2017; Montero-Calasanz, 2020b, 2021), although it is worth mentioning that this is the first report of MK-10(H₁₀) in members of the genus Blastococcus. The biosynthesis of menaquinoins, via chiorosmate derived from Shikimate pathway (genes: MenA-G), was confirmed by genomic data in the order Geodermatophilales. Curiously, the fotalosine pathway, the other major type of menaquinone biosynthetic pathway (Seto et al., 2008) was also annotated in genomes of Cryptosporangium arvum and Crytosporangium aurantiaca type strains (Supplementary Table S4). MenJ encoding for menaquinone-9-β-reductase [EC:1.3.99.38] and implicated in menaquinone side chain saturation (Upadhyay et al., 2015) was misidentified by databases as a geranylgeranyl reductase family protein but was verified in all genomes after BLAST using UniProtKB:P9WNY8 (MENJ_MYCTO). Moreover, the presence of 1,4-dihydroxy-2-naphthoate polyprenyltransferase [EC:2.5.1.74] and demethylmenaquinone methyltransferase [EC:2.1.1.163] in all genomes confirmed the potential of Geodermatophilales representatives to synthesise MK 8, 9, and 10. Regarding the elongation of isoprenoid side chains, genes uppS and hepST, encoding for ditarans,polyis-undecaprenyl-diphosphate synthase (2E,6E)-farnesyl-diphosphate specific) and polyproprenyl diphosphate synthase (annotated here as geranylgeranyl pyrophosphate synthase), respectively, were annotated in all genomes. But the length and the degree of saturation of the C-3 isoprenyl side chains were impossible to predict. MKs consist of repeated isoprenop s subunits and the exact number is determined by the synthase encoded by the particular microbe and a microbe-dependent molecular ruler mechanism that involves bulky amino acid residues blocking the enzyme active sites to stop chain elongation (Han et al., 2015). The inability to predict the isoprenyl chain length of menaquinones was already reported by Baek et al. (2018).

Regarding fatty acid profiles, the five strains qualitatively showed similar patterns to those found in other representatives in Blastococcus (Table 1), but the predominant fatty acids significantly varied amongst them. Dominant fatty acids for strains DSM 44268 and DSM 44269 (>5%) were C₁₄:0 (36.2% and 37.2%, respectively), iso-C₁₅:0 (19.3% and 11.3%, respectively), iso-C₁₆:0 (6.2% and 8.6%, respectively), and C₁₆:1ω7c (7.8% and 7.2%, respectively). Additionally, DSM 44268 showed iso-C₁₆:1ω7c (5.4%) and DSM 44269 displayed some minor amounts of C₁₈:0 (5.6%) and C₁₈:1ω9c (6.8%). Those latter fatty acids were also present in DSM 44268 and iso-C₁₆:1ω7c H in DSM 44269 but at levels below 5%. On the other hand, strain DSM 44270 had the major fatty acids (>5%) iso-C₁₅:0 (32.0%), C₁₇:0 (27.0%), iso-C₁₆:1ω7c H (8.7%), iso-C₁₆:0 (7.3%), and C₁₄:0 (6.0%). A similar composition was observed for strain DSM 44271 although those quantitatively varied as follows: iso-C₁₆:0 (24.0%), C₁₅:0 (23.0%), C₁₇:0 (13.7%), iso-C₁₅:0 (10.8%), and iso-C₁₆:1ω7c H (8.1%). Finally, the major fatty acids of strain DSM 44205 (>5%) consisted of iso-C₁₆:0 (23.4%), anteiso-C₁₇:0 (19.8%), and anteiso-C₁₇:1 (11.9%), iso-C₁₆:1ω7c H (11.1%), and C₁₄:0 (6.7%). The complete gene set (Fab cluster) related to fatty acids metabolism (type II fatty acid synthase (FAS II) system) was present in all the studied genomes. Similar to the biosynthesis of MKs, fatty acid synthesis occurs via recurring reactions. It is thus not possible to predict just by using genomic data modifications that result in desaturations and/or elongations (López-Lara and Geiger, 2019).

The major polar lipids were diphosphatidylglycerol (DPG), phosphatidylethanolamine (PE), phosphatidylcholine (PC), and phosphatidylinositol (PI; Figure 3). It was consistent with those already described by Hezbri et al. (2016). Significant amounts of phosphatidylglycerol (PG) were also identified in all the isolates except in strain DSM 44205. The inconsistent presence of PG was already noted for other species in the family (Montero-Calasanz et al., 2014, 2015, 2017; Hezbri et al., 2015a,b, 2016). Similarly, the characteristic glycophosphoinositol (GPI; annotated as GPL in Hezbri et al., 2016) as well as the reproductive presence of two unidentified polar lipids (PL2 and PL3) was detected in strains DSM 44268, DSM 44269, DSM 44270, DSM 44272 but not in DSM 44205. However, a minor amount of an unidentified glycolipid (GL4) was found in the polar lipids pattern of strain DSM 44205. Three different ones (GL1, GL2, and GL3) were also identified in the profile of strain DSM 44268 but not in strain DSM 44269. In contrast, an unidentified aminolipid (AL1) was displayed by strain DSM 44269 but not by DSM 44268. One having a similar chromatographic mobility was present in strain DSM 44270, too. Genomic data were consistent with those obtained experimentally with some minor discrepancies. In particular, the enzymes phosphatidylcholine synthase and cardiolipin synthase [EC:2.7.8.40] required in the synthesis of PC and cardiolipin (DGP), respectively, were present in all the genomes. As expected, phosphatidylinosyl synthase was missing in Modestobacter and Kleinkia type species. CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase [EC:2.7.8.50] involved in the synthesis of phosphatidyl-glycerophosphate, and also in fatty acids metabolism, was present in all species although phosphatidylglycerophosphate synthase [EC:3.1.3.27] required to the final conversion to PG was just annotated in DSM 44272 and in B. litoris. Phosphatidylerine desacbreoxylase [EC 4.1.1.65] required to synthesise phosphatidylethanolamine was not annotated in M. multiseptatus. CDP-diacylglycerol—inositol 3-phosphatidyltransferase [EC:2.7.8.11] needed for the synthesis of PI was annotated in all genomes. Finally, some other enzymes involved in the biosynthesis of other polar lipids were also annotated in the studied genomes suggesting that those may
correspond to those minor unidentified lipids shown on the TLC plate.

**Carbon metabolism**

As expected, and in agreement with results determined by GEN III Biolog System (Supplementary Table S5), Blastococcus spp. show aerobic microbial respiration potentially driven by atmospheric CO₂ concentration and as production platforms for a wide range of bioproducts from CO₂ (Katsyv and Müller, 2020).

A complete tetrahydrofolate (THF) methyl-branch along with the presence of carbon monoxide dehydrogenases, required for the carbonyl-branch, were annotated in all genomes. These microorganisms could also be able to produce acetate as end product (acetogenesis). But no acetyl-CoA synthase [2.3.1.169] homologue was identified in this study. And neither an autotroph nor an acetogen was described in the family so far. Overall, the presence of a number of mechanisms related to chemolithoautotrophy denotes that Blastococcus representatives could play a key role in the global carbon cycle. Microorganisms able to efficiently capture and store CO₂ are prime candidates to be used as biological approaches in the reduction of the atmospheric CO₂ concentration and as production platforms for a wide range of bioproducts from CO₂.

Apart from D-glucose as carbon source, genomics data identified other potential carbon sources such as fructose (fructokinase mediated: EC 2.7.1.4); α-galactose (galk, galactokinase); D-galactose (galM, Aldose 1-epimerase; EC 5.1.3.3); which is absent in DSM 44205; L-gulose (gul, RGN gluconolactonase; EC 3.1.1.17); 1-butanol (alcohol deshydrogenase; EC 1.1.1.-), only present in B. aggregatus; D-mannose-6P/1P (manB phosphomannomutase; EC 5.4.2.8/manC mannose-1-phosphate guanylyltransferase; EC 2.7.7.13), absent in B. aggregates; glycine via GDC glycine dehydrogenase and glycine hydroxymethyltransferase (glyA, EC 2.1.2.1); L-glutamate (gdHA, glutamate dehydrogenase; EC 1.4.1.3); L-serine (glycine hydroxymethyltransferase; glyA, EC 2.1.2.1), isocitrate (isocitrate dehydrogenase), absent in DSM 4272²; formamide (formamidase; EC 3.5.1.49) present in DSM 4245², DSM 4427², and B. aggregatus; trehalose (treS, maltose α-D-glucosyltransferase; EC 4.99.16 and D-glucose α,α-trehalose phosphorylase; EC 2.4.1.65); L-lactate (lldD, L-lactate dehydrogenase; EC 1.1.2.3), just present in B. endophyticus; sucrose (malZ, α-glucosidase; EC 3.2.1.20); maltose (maltokinase; EC 2.7.1.175), absent in B. xanthinilyticus; isomaltose (oligo-1,6-glucosidase), amongst others, that may also be utilised by Blastococcus strains. These results supported the presence of a rich carbohydrate metabolism in the genus and were largely in correlation with those showed by GEN III Biolog and API System (See Supplementary Figure S5; Supplementary Tables S4, S5).

Additionally, Blastococcus genomes revealed extracellular enzymes involved in the degradation of complex carbohydrates, which are closely linked to the decay of organic residues, transformation of native soil organic matter, mineralisation of plant nutrients, and soil aggregation (Balezentiene, 2012). For example,
α-glucosidases [EC 3.2.1.20], involved in the cellulose hydrolysis, were observed in all studied genomes (three copies in *B. xanthinilyticus*); β-glucosidases [EC 3.2.1.21] were also annotated in all of them except in *B. colisei* and *B. xanthinilyticus*, having three gene copies in DSM 44205\(^{T}\), DSM 44268\(^{T}\), and *B. saxobsidens*. Moreover, all studied genomes contained α-amylases [EC 3.2.1.1]. Extracellular enzymes handling the almost complete starch saccharification (Tu and Miles, 1976), but enzymatic activity was not recorded in any of them according to the APIM ZYM system; the presence of glycosidase (glucose hydrolyase; annotated as α-amylase by Pfam database) was, nevertheless, only annotated in *B. endophyticus*. Finally, chitinases [EC 3.2.1.14] were annotated in *B. endophyticus* and *B. saxobsidens*. Cellulases (β-1,4-endoglucan hydrolyase; EC 3.2.1.4) and 1,4-β-celllobiohydrolase [EC 3.2.1.91] were absent. On the other hand, lipolytic enzymes are an important group of biotechnologically relevant enzymes presenting a key role in regulating the levels of hydrocarbons in soil (Gupta et al., 2004). Extracellular lipases such as the well-known triacylglycerol acylhydrolase (here as tryacylglycerol lipase, EC 3.1.1.3) was only annotated in DSM 44268\(^{T}\). In addition, lysophospholipases [EC 3.1.1.5] and monoglyceride lipases (EC 3.1.1.23) were present in all the genomes analysed except DSM 44272\(^{T}\) and *B. endophyticus*. Polyhydroxyalkanoic acid (PHA) syntheses (EC 2.3.1.-), the key enzyme in the biosynthesis of PHAs, a class of aliphatic polyesters that are generally regarded as a carbon and energy reserve material in bacteria and archaea (McCool and Cannon, 2001), were also identified in all the genomes in a number of copies that ranged from one to six with the exception of DSM 44272\(^{T}\). Again, large correlations but also discrepancies were observed between genomic and experimental data. For example, the hydrolysis of xanthine by *B. xanthinilyticus*, previously described by Hezbri et al. (2018), was confirmed by the xanthine deshydrogenase, but experimentally it could not be proven for strains DSM 44205\(^{T}\) and DSM 44268\(^{T}\) and *B. endophyticus* although such enzyme was also annotated in their genomes. Similar inconsistencies were also identified for other unambiguously annotated enzymes such as α-glucosidase which was not expressed under standardised experimental conditions in strains DSM 44205\(^{T}\), DSM 44268\(^{T}\), DSM 44270\(^{T}\), DSM 44272\(^{T}\), and *B. aggregatus* or β-glucosidase which was only expressed by *B. endophyticus* and *B. saxobsidens* (for more examples see Supplementary Figure S5; Supplementary Tables S4, S5).

**Final remarks and taxonomic consequences**

The integration of genomic information into microbial systematics have proven to be a reproducible, reliable and highly informative approach to reveal phylogenetic relationships amongst prokaryotes (Chun and Rainey, 2014; Chun et al., 2018). Montero-Calasanz et al. (2017) already applied phylogenomics in combination with the identification of diagnostic features in the taxonomic rearrangement of *Geodermatophilaceae*. In light of new genomic information now available, the taxonomic status of the family was revisited in this study. The whole proteome-based phylogenomic tree of the sequenced type strains and strains under study along with phenotypic tests showed with strong support that DSM 44268\(^{T}\), DSM 44270\(^{T}\), DSM 44272\(^{T}\), and DSM 44205\(^{T}\) represent novel species within the genus *Blastococcus*. Additionally, the non-monophyly of *Blastococcus*, *Geodermatophilus* and *Modestobacter* was revealed, i.e., *B. endophyticus*, *G. daqingensis* and *M. deserti* formed separate lineages within *Geodermatophilaceae*. Based on principles of phylogenetic systematics and applying taxonomic conservatism, we propose the following reclassifications: *Blastococcus endophyticus* as the type species of the new genus *Trijillonella* gen. nov., which the name *Trijillonella endophytica* comb. nov. is given; *Geodermatophilus daqingensis* as the type species of the new genus *Pleomorpha* gen. nov., which the name *Pleomorpha daqingensis* comb. nov. is given; *Modestobacter deserti* gen. nov., as the type species of the new genus *Goekera* which the name *Goekera deserti* comb. nov. is given. Finally, based on the additional taxonomic data now available we propose the emendation of *B. aggregatus*, *B. jejuensis*, *B. saxobsidens* and *B. Xanthinilyticus* to include accession numbers for the whole genome sequences, genome size and genomic G + C content and insights. The use of genomic data for the characterisation of novel species in *Geodermatophilaceae* also revealed interesting insights regarding their metabolism and ecology. In particular, the number of 16S rRNA and tRNA gene copies was suggested to be related to their slow growth speed and their life strategy supporting their prevalence in oligotrophic environments such as desert soils and decayed monuments (Neilson et al., 2012; Giongo et al., 2013). In addition, the presence of a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly in all representatives of the family highlighted potential yet unobserved features that would enable us to better understand their metabolism, morphogenesis and ecology. Combining high-throughput phenotype analyses with genomics also supported the uncovering of a rich carbon metabolism in *Blastococcus* spp. Results obtained from GEN III microplates and API System were largely correlated with what was observed in the genomes. In this way, *Blastococcus* spp. were described as aerobic bacteria with the potential to present anaerobiosis and fixate CO₂ by using four potential pathways and to degrade a range of complex carbohydrates. Such metabolic versatility would support their success in colonising a variety of biotopes and a potential key role in the global carbon cycle and as biological tools to reduce the atmospheric CO₂ concentration and produce bioproducts from CO₂.

Even though chemotaxonomic procedures marked a turning point in prokaryotic systematics (Goodfellow et al., 2012), results can be influenced by cultivation conditions (i.e., fatty acid analysis) or by inexperienced staff (i.e., regarding the interpretation of polar lipids patterns on 2-D TLC plates; Sutcliffe et al., 2013). In favour of a revitalization of bacterial taxonomy, *in silico* chemotaxonomic analyses were already used in some studies to complement and, in some cases, replace characteristics traditionally determined in the laboratory (Amaral et al., 2014; Fotedar et al., 2020;
Lawson et al., 2020). Here, we attempted to determine in silico cell-wall peptidoglycan, respiratory quinone, fatty acids and polar lipid patterns of Geodermatophilaceae. Overall, genomic data were consistent with the results obtained from the laboratory. Nevertheless, the amount of information that can be predicted based on genomic data varied depending on the chemotaxonomic trait. For instance, the in silico polar lipid analyses were largely in line with those obtained experimentally and allowed us to distinguish at the genus level by the presence/absence of the gene encoding for phosphatidylcholine synthase. On the other hand, the presence of the DAP as the diagnostic diamino acid over L-lysine in the whole-cell hydrolysates was confirmed but the in silico procedure failed in differentiating microorganisms incorporating LL-DAP rather than meso-DAP (discrimination at order level). Regarding both MK and fatty acids analysis, final products could not be predicted due to the nature of biosynthetic pathways. Nevertheless, sufficient information would, for example, be gathered from Shikimate pathway along with the absence of fucitolalose pathway to establish certain taxonomic discrimination at order level. The use of genomic data for species delineation and the impractical value of chemotaxonomic data when genomic data are available are widely known (Nouioui et al., 2018). Nevertheless, as argued by Vandamme and Sutcliffe (2021), knowledge of cellular components can be valuable to establish discriminative thresholds when applied at taxonomic ranks above the species level. However, improvements in analytical methods leading to the use of automated and higher-resolution technologies such as mass spectrometry lipidomic methods (Rashid et al., 2017) should first be introduced and deployed. Our results here proved the use of in silico chemotaxonomic analysis as a simple and fast approach to supplement phylogenetic results and to provide sufficient resolution to support the affiliation of novel isolates into appropriate taxonomic groups. Its routine application would undoubtedly increase the speed of taxonomic descriptions and the accuracy and consistency of taxonomic reports.

In summary, the integration of genomics into the systematics of Geodermatophilaceae not only allowed for more stable and reliable taxonomic arrangements but also showed the feasibility for a potential replacement of wet-lab chemotaxonomy. Second, it provided deeper insights into the molecular mechanisms behind phenotypic features and revealed potentially overlooked ones that could be key to an understanding of their evolution, ecology and biotechnological potential.

Description of Trujillella gen. nov.

Trujill.e.la. N.L. fem. n. Trujillella, named in honour of Martha E. Trujillo in recognition of her contributions to microbial systematics, mainly on Actinobacteria, on Bergey’s Manual trust, and as the Editor-in-chief of the International Journal of Systematic and Evolutionary Microbiology.

Cells are aerobic, non-motile, non-spore-forming, Gram-stain positive, catalase-positive and oxidase-negative. Cells occur singly, in pairs or in tetrads, often tending to form aggregates. The peptidoglycan in the cell-wall contains meso-diaminopimelic acid. The predominant menaquinone is MK-9(H₈), with MK-8 and MK-9(H₉) as minor components. The basic polar lipid profile contains diphasphatidylglycerol, phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol. The major fatty acids are iso-C₁₅:₀, iso-C₁₇:₀ and C₁₆:₀:9c. The basic whole-cell sugar pattern includes arabinose and galactose. The genomic G+C content is 71%–72%.

The type species of Trujillella is Trujillella endophytica, sp. nov.

Description of Trujillella endophytica comb. nov.

T. en.do.phy’ti.ca (Gr. pref. Endo-, within; Gr. neut. n. phyton, plant; L. fem. Adj. suff. –ica, adjectival suffix used with the sense of belonging to; N.L. fem. Adj. endophytica, within plant, endophytic, pertaining to the isolation from plant tissues).

Basonym: Blastococcus endophyticus Zhu et al. (2013) emend. Hezbri et al. (2016).

The properties are as given in the species description by Zhu et al. (2013) and emendation by Hezbri et al. (2016) with the following modification. The genomic G+C content is 74.6%. The genome size is 4.9 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C₄-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain DSM 45413² is FOE00000000.

The type strain YIM 68236 = CCTCC AA 209045 = DSM 45413² = KCTC 19998 was isolated from healthy leaves of Camptotheca acuminata collected in Yunnan Province, southwest China.

Description of Pleomorpha gen. nov.

Ple.o.mor’pha. Gr. adv. Pleon more; Gr. fem. n. morphe, shape or form; N.L. fem. n. Pleomorpha, organism showing multiple forms.

Pleiomorphic, motile, spore-forming, aerobic, Gram-stain positive cells. Those occurs singly or associated in aggregates. The peptidoglycan in the cellwall contains meso-diaminopimelic acid. The predominant menaquinone is MK-9(H₈). The basic polar lipid profile contains diphasphatidylglycerol, phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol. The major fatty acids are iso-C₁₅:₀ and iso-C₁₇:₀. The basic whole-cell sugar pattern includes galactose, glucose and xylose. The genomic G+C content is 73%–74%.

The type species of Pleomorpha is Pleomorpha daqingensis, sp. nov.
Description of Pleomorpha daqingensis comb. nov.

P. da.qing.en’sis (N.L. masc./fem. Adj, daqingensis, pertaining to Daqing city, China, where the type strain was isolated).

Basonym: Geodermatophilus daqingensis Wang et al. (2017).

The properties are as given in the species description by Wang et al. (2017) with the following modification. The genomic G + C content is 73.6%. The genome size is 5.4 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carboxylic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain CPCC 205119T is JAA6KW00000000.

The type strain CPCC 205119T = DSM 104001T was isolated from pretroleum-contaminated soil in Daqing city, China.

Description of Goekera gen. nov.

Goekera. N.L. fem. n. Goekera, named in honour of Markus Göker in recognition of his contributions to microbial systematics, including work on Actinobacteria, on the List of Prokaryotic Names with Standing in Nomenclature (LPSN), and as a member of the Judicial Commission.

Cells are motile, non-spore-forming, aerobic, Gram-stain positive, catalase and oxidase positive cocci and/short rods. Bud-like structure was observed for some cells. The peptidoglycan in the cell-wall contains meso-diaminopimelic acid. The predominant menaquinone is MK-9(H8), with MK-8(H6) as a minor component. The basic polar lipid profile contains diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylglycositol, phosphatidylinositol mannoside. The major fatty acids are C16:1ω7c, iso-C16:0, C16:0, iso-C15:0 and C16:0ω7c. The basic whole-cell sugar pattern includes arabinose, glucose and ribose. The genomic G + C content is 74%–75%. The type species of Goekera is Goekera deserti, sp. nov.

Description of Goekera deserti comb. nov.

G. deser’ti (L. gen. Neut. n. deserti, of a desert, where the organisms were acquired).

Basonym: Modestobacter deserti Jiang et al. (2021).

The properties are as given in the species description by Jiang et al. (2021) with the following modification. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carboxylic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain CPCC 205119T is JAA6KW00000000.

The type strain CPCC 205119T = DSM 104001T was isolated from pretroleum-contaminated soil in Daqing city, China.

Emended description of Blastococcus aggregatus

The properties are as given in the species description by Ahrens and Moll (1970) and emendations by Urzi et al. (2004) and Hezbri et al. (2016) with the following emendation. The genomic G + C content is 73.3%. The genome size is 4.6 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle, and carboxylic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of the type strain DSM 4725T is OBQ00000000.

The type strain B15T = ATCC 25902T = DSM 4725T = JCM 12602T = NCIMB 1849T was isolated from Baltic Sea.

Emended description of Blastococcus jejuensis

The properties are as given in the species description by Lee (2006) and emendation by Hezbri et al. (2016) with the following emendation. The presence of phosphatidylethanolamine listed by Lee (2006) and confirmed by Hezbri et al. (2016) was incorrectly annotated by the latter and it should be recognised as hydroxypatidylethanolamine accordingly to its chromatographic mobility.

The type strain KST3-10T = DSM 19597T = JCM 15614T = KCCM 42251T = NRRL B-24440T was isolated from sand sediment of a beach in Jeju, Korea.

Emended description of Blastococcus saxobsidens

The properties are as given in the species description by Urzi et al. (2004) and emendation by Hezbri et al. (2016) with the following emendation. The genomic G + C content is
Emended description of *Blastococcus xanthilyniticus*

The properties are as given in the species description by Hezbi et al. (2018) with the following emendation. The genomic G+C content is 74.4%. The genome size is 4.6 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellar synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of the type strain DSM 44509 is SHKV01000001. The type strain BC444 = DSM 44509 = JCM 13239 = NRRL B-24246 was isolated from the surface of marble and calcareous stones in Italy.

Description of *Blastococcus aurantiacus* sp. nov.

*Blastococcus aurantiacus* (au.ran.ti.a' cus. N.L. masc. Adj. *aurantiacus* orange-coloured, referring to the orange colour of the colonies).

Colonies are bright red orange-coloured, opaque with a dry surface and regular margin. Cells are Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 µm in diameter) with a tendency to form aggregates. It grows in aerobiciosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spor-forming but a repertoire of genes related to flagellar synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative. Temperature and pH ranges are 15–37°C (optimal range) and 6.0–12.0 (optimum 6.5–8.0), respectively. NaCl is not needed for growth. It can grow between 0% and 4% NaCl (w/v; optimal range) but not at 8%. It grows well on GYM starch, xanthine, and hypoxanthine are negative.

The accession number for the whole genome sequence of the type strain AT 7-1 (=DSM 44268 = JCM 18931) isolated from in the Atacama Desert (Chile) is MH479060. The IMG accession number for the whole genome sequence of strain DSM 44268 is 2599183538.
Description of *Blastococcus fimeti* sp. nov.

*Blastococcus fimeti* (fi.me’ti. L. neut. Gen. n. *fimeti*, of or from a dunghill).

Colonies are pink red-coloured, opague with a greasy surface and irregular margin. Cells are Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 μm in diameter) with a tendency to form aggregates. It grows in aerobiciosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative and aesculin positive. Temperature and pH ranges are 15°C–37°C (optimal range) and 6.0–12.0 (optimum 7.0–9.0), respectively. NaCl is not needed for growth. It can grow between 0–4% NaCl (w/v; optimal range) but not at 8%. It grows well on GYM Streptomyces, R2A (DSMZ Medium 830), tryticate soy agar (TSA), Luedemann (DSMZ medium 877), GPHF (DSMZ medium 553), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, α-fucosidase. According to the Biolog System, it oxidises: dextrin, D-maltose, D-trehalose, D-cellobiose, sucrose, turanose, β-methyl-D-glucoside, D-salicin, N-acetyl-D-glucosamine, N-acetyl-β-D-mannosamine, D-glucose, 3-O-methyl-D-glucose, L-fucose, glycerol, D-glucose-6-phosphate, rifampicin SV, L-glutamic acid, methyl pyruvate, D-lactic acid methyl ester, citric acid, α-keto-glutaric acid, nalidixic acid, lithium chloride, potassium tellurite, tween 40, γ-amino-β-pyruvic acid, β-hydroxy-butyric acid, α-keto-butyric acid, acetotrophic acid, propionic acid, acetate acid, aztreonam, butyric acid, and sodium bromate but not D-raffinose, α-D-lactose, N-acetyl-D-galactosamine, N-acetyl-neuraminic acid, D-mannose, L-rhamnose, fusidic acid, D-serine (inhibitory concentration), D-arabitol, myo-inositol, D-fructose-6-phosphate, D-aspartic acid, D-serine, treleandomycin, minocycline, glycine-proline, L-alanine, L-arginine, L-histidine, L-proglutamatic acid, L-serine, lincomycin, guanidine, hydrochloride, niaproof, L-galactonic acid-γ-lactone, muccic acid, quinic acid, D-saccaric acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, L-lactic acid, L-malic acid, bromo-succinic acid and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-C₁₇:₀ anteiso-C₁₇:₀ anteiso-C₁₇:₁ C, iso-C₁₇:₀ H and C₁₈:₀ 9c. The cell wall peptidoglycan contains meso-diaminopimelic acid. Ribose, xylose, arabinose, mannose, and glucose are the whole-cell sugars. Polar lipid profile consists of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine and phosphatidylinositol. The high-quality draft genome of strain DSM 44205 was resolved to 30 scaffolds consisting of 4,044,261 bp, with a G+C content of 73.4%, 3,853 candidate protein-coding genes, 70 tRNA genes, and nine rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain G1S (=DSM 44205 = CECT 8406) isolated from seepeage water of a dumping ground in Vancouver (Canada) is MH479059. The IMG accession number for the whole genome sequences of strain DSM 44205 is 2599185193.

Description of *Blastococcus haudaquaticus* sp. nov.

*Blastococcus haudaquaticus* (hau.da.qua’ti.cus. L. adv. Haud, not at all, by no means; L. masc. Adj. *aquaticus*, living, growing, or found in or by the water, aquatic; N.L. masc. Adj. *haudaquaticus*, growing far away from or without any water).

Colonies are reddish-brown coloured, opague with a mucoid surface and regular margin. Cells are aerobic, Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 μm in diameter) with a tendency to form aggregates. It grows in aerobiciosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative. Temperature and pH ranges are 15–37°C (optimal range) and pH 5.0–11.0 (optimum 6.5–8.0), respectively. NaCl is not needed for growth. It can grow between 0% and 8% NaCl (w/v; optimal range). It grows well on GYM Streptomyces, R2A (DSMZ Medium 830), tryticate soy agar (TSA), Luedemann (DSMZ medium 877), GPHF (DSMZ medium 553), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, trypsin, and cystine acrylamidase. According to the Biolog System, it oxidises: dextrin, D-maltose, D-trehalose, D-cellobiose, sucrose, turanose, β-methyl-D-glucoside, D-salicin, N-acetyl-D-glucosamine, N-acetyl-β-D-mannosamine, D-glucose, 3-O-methyl-D-glucose, L-fucose, glycerol, D-glucose-6-phosphate, rifampicin SV, L-glutamic acid, methyl pyruvate, D-lactic acid methyl ester, citric acid, α-keto-glutaric acid, nalidixic acid, lithium chloride, potassium tellurite, tween 40, γ-amino-β-pyruvic acid, β-hydroxy-butyric acid, α-keto-butyric acid, acetotrophic acid, propionic acid, acetate acid, aztreonam, butyric acid, and sodium bromate but not D-raffinose, α-D-lactose, N-acetyl-D-galactosamine, N-acetyl-neuraminic acid, D-mannose, L-rhamnose, fusidic acid, D-serine (inhibitory concentration), D-arabitol, myo-inositol, D-fructose-6-phosphate, D-aspartic acid, D-serine, treleandomycin, minocycline, glycine-proline, L-alanine, L-arginine, L-histidine, L-proglutamatic acid, L-serine, lincomycin, guanidine, hydrochloride, niaproof, L-galactonic acid-γ-lactone, muccic acid, quinic acid, D-saccaric acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, L-lactic acid, L-malic acid, bromo-succinic acid and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-C₁₆:₀ anteiso-C₁₇:₀ anteiso-C₁₇:₁ C, iso-C₁₇:₀ H and C₁₈:₀ 9c. The cell wall peptidoglycan contains meso-diaminopimelic acid. Ribose, xylose, arabinose, mannose, and glucose are the whole-cell sugars. Polar lipid profile consists of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine and phosphatidylinositol. The high-quality draft genome of strain DSM 44205 was resolved to 30 scaffolds consisting of 4,044,261 bp, with a G+C content of 73.4%, 3,853 candidate protein-coding genes, 70 tRNA genes, and nine rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain G1S (=DSM 44205 = CECT 8406) isolated from seepeage water of a dumping ground in Vancouver (Canada) is MH479059. The IMG accession number for the whole genome sequences of strain DSM 44205 is 2599185193.
L-arginine, L-pyroglutamic acid, lincomycin, guanidine hydrochloride, niaprof, L-galactonic acid-γ-lactone, D-glucuronic acid, glucuronanomide, quinic acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, 16:0-acetic acid, 16:0-propionic acid, 16:0-valeric acid, 17:0-valeric acid, 18:0-stearic acid, D-glucuronic acid, glucuronamide, α-keto-glutaric acid, D-malic acid, 3-hydroxy-butyric acid, sodium D-lactate, D-sorbitol, D-fructose, D-galactose, sodium lactate, D-serine, α-keto-gluconic acid, D-mannitol, D-arabitol, myo-inositol, glycerol, L-pyroglutamic acid, pectin, D-glucuronic acid, quinic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, N-alanine, sodium D-lactate, myo-inositol, L-fucose, L-rhamnose, inosine, fusidic acid, D-serine (inhibitory concentration), D-aspartic acid, D-serine, troleandomycin, rifamycin SV, minocycline, gelatin, glycine-proline, L-histidine, L-serine, lincomycin, guanidine hydrochloride, niaprof, L-galactonic acid, D-arabitol, D-mannitol, D-arabitol, myo-inositol, glycerol, L-pyroglutamic acid, citric acid, D-malic acid, L-malic acid, bromo-succinic acid, γ-amino-β-butyric acid, and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-C₁₅:₀, iso-C₁₆:₁, H, iso-C₁₆:₀, 17:0 3–OH, C₁₈:₃ 9c, 18:1 9c, and C₁₆:₁ 11c. The cell wall peptidoglycan contains meso-diaminopimelic acid. Ribose, mannose, galactose, and glucose are the whole-cell sugars. Polar lipids consist of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylcholine, and phosphatidylinositol. The high-quality draft genome of strain DSM 44270 was resolved to 11 scaffolds consisting of 4,512,672 bp, with a G + C content of 73.29%, 4,410 candidate protein-coding genes, 48 tRNA genes, and six rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain AT 7-14² (=DSM 44270²=JCM 18932²) isolated from soil in the Atacama Desert (Chile) is MH479062. The IMG accession number for the whole genome sequences of strain DSM 44270² is 2728369258.

Description of Blastococcus mobilis sp. nov.

Blastococcus mobilis (mo’bi.lis. L. masc. Adj. mobilis movable, motile).

Colonies are reddish-brown coloured, opaque with a mucoid surface and regular margin. Cells are aerobic, Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 μm in diameter) with a tendency to form aggregates. It grows in aerobiosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative. Temperature and pH ranges are 15–37°C (optimal range) and pH 6.5–11.0 (optimum 6.0–10.0), respectively. NaCl is not needed for growth. It can grow between 0% and 1% (w/v; optimal range) but not at 4%. It grows well on GYM Streptomyces, R2A (DSMZ Medium 830), Luedemann (DMSZ medium 877), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, and valine arylamidase. According to the Biolog System, it oxidises: dextrin, D-maltose, sucrose, turanose, β-methyl-D-glucoside, D-glucose, D-mannose, D-fructose, D-galactose, sodium lactate, D-sorbitol, D-fructose, D-galactose, sodium lactate, D-sorbitol, D mannitol, D-arabitol, myo-inositol, glycerol, L-pyroglutamic acid, pectin, D-glucuronic acid, quinic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, N-alaldixic acid, potassium tellurite, tween 40, α-hydroxy-butyric acid, β-hydroxy-butyric acid, β-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, aztreonam, butyric acid, and sodium bromated but not D-raffinose, α-D-lactose, D-melibiose, N-acetyl-D-galactosamine, N-acetyl-neuraminic acid, L-fucose, L-rhamnose, inosine, fusidic acid, D-serine (inhibitory concentration), D-aspartic acid, D-serine, troleandomycin, rifamycin SV, minocycline, gelatin, glycine-proline, L-histidine, L-serine, lincomycin, guanidine hydrochloride, niaprof, L-galactonic acid-γ-lactone, music acid, D-saccharic acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, citric acid, D-malic acid, L-malic acid, bromo-succinic acid, γ-amino-β-butyric acid, and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-C₁₅:₀, iso-C₁₆:₁, H, iso-C₁₆:₀, 17:0 3–OH, C₁₈:₃ 9c, 18:1 9c, and C₁₆:₁ 11c. The cell wall peptidoglycan contains meso-diaminopimelic acid. Rhamnose, ribose, mannose, arabinose, galactose, and glucose are the whole-cell sugars. Polar lipid profile consists of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylcholine, and phosphatidylinositol. The high-quality draft genome of strain DSM 44272 was resolved to 88 scaffolds consisting of 5,094,633 bp, with a G + C content of 72.5%, 4,938 candidate protein-coding genes, 49 tRNA genes, and six rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain AT 7-2-111² (=DSM 44272²=JCM 18933²) isolated from soil in the Atacama Desert (Chile) is MH479063. The IMG accession number for the whole genome sequences of strain DSM 44272² is 2724679778.

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

MdCM-C designed the study. MdCM-C, AY, MR, and PS performed experiments. MdCM-C, AY, and JM-K performed bioinformatics analyses and wrote the manuscript. H-PK funded...
Conflict of interest

The 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.

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.2022.975365/full#supplementary-material.
Katsyv, A., and Müller, V. (2020). Overcoming energetic barriers in acetogenic C1 conversion. Front. Bioeng. Biotechnol. 8:621166. doi: 10.3389/fbioe.2020.621166

Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. S. (2000). rRNA s copu number reflects ecological strategies of bacteria. App. Environ. Microbiol. 66, 1328–1333. doi:10.1128/AEM.66.4.1328-1333.2000

Klenk, H.-P., and Göker, M. (2010). En route to a genome-based classification of Archaea and bacteria? Syst. Appl. Microbiol. 33, 175–182. doi:10.1016/j.syapm.2010.03.003

Ko, J.-H., Montero Llopis, P., Heinritz, J., Jacobs-Wagner, C., and Sil, D. (2013). Suppression of amber codons in Caulobacter crescentus by the orthogonal Escherichia coli Histidyl-tRNA synthetase/tRNA1000 pairing. PLoS One 8: e83630. doi:10.1371/journal.pone.0083630

Kroppenstedt, R. M. (1982). Separation of bacterial maconategories by HPLC using reverse phase (RP18) and a silver loaded ion exchanger as stationary phases. J. Liquid Chromatography 5, 2359–2367. doi:10.1080/014820678076540

Kuo, C.-H., and Ochan, H. (2010). The extinction dynamics of bacterial pseudogenes. PLoS Genet. 6:e1001050. doi:10.1371/journal.pgen.1001050

Kyrpides, N. C., Hugenholtz, P., Eisen, J. A., Woyke, T., Göker, M., Parker, C. T., et al. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol. 12:e1001920. doi:10.1371/journal.pbio.1001920

Lawson, P. A., Patel, N. B., Mohammed, A., Moore, E. R. B., Lo, A. S., Sardi, A., et al. (2020). Parapseudonocardia muciniphila gen. nov., sp. nov., a member of the family Chitimphagaceae isolated from a human peritoneal tumour and reclassification of Pseudobacterium ginsenosidum as Pseudonocardia ginsenosidum comb. nov. Int. J. Syst. Evol. Microbiol. 70, 3639–3646. doi:10.1099/ijsem.0.004204

Lechevalier, M. P. L. H., and Lechevalier, H. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. Int. J. Syst. Evol. Microbiol. 20, 433–445. doi:10.1099/00207713-20-4-433

Lee, S. D. (2006). Blastococcus jejuensis sp. nov., an actinomycete from beach sediment, and emended description of the genus Blastococcus. Microbiol. Immunol. 50, 819–826. doi:10.1001/0028-282X.50.8.819

Lee, D. W., Lee, H., Kwon, B. O., Kim, J. S., Kim, B. S., et al. (2018). Blastococcus litoris sp. nov., isolated from sea-tidal flat sediment. Int. J. Syst. Evol. Microbiol. 68, 3435–3440. doi:10.1099/ijsem.0.003004

Liu, Y., Harrison, P. M., Kunin, V., and Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol. 5:1715

Lefort, V., Desper, R., and Gascuel, O. (2019). “Chapter 3: formation of fatty acids, ” in The Prokaryotes, eds. M. Dworkin and S. Falkow (New York, NY: Springer). pp. 579–592.

Lee, D. W., Lee, H., Kwon, B. O., Kim, J. S., Kim, B. S., et al. (2018). Blastococcus litoris sp. nov., isolated from sea-tidal flat sediment. Int. J. Syst. Evol. Microbiol. 68, 3435–3440. doi:10.1099/ijsem.0.003004

Liu, Y., Harrison, P. M., Kunin, V., and Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol. 5:1715

Lefort, V., Desper, R., and Gascuel, O. (2019). “Chapter 3: formation of fatty acids, ” in The Prokaryotes, eds. M. Dworkin and S. Falkow (New York, NY: Springer). pp. 579–592.

Liu, Y., and Ochan, H. (2007). Stepwise formation of the bacterial flagellar system. Proc. Natl. Acad. Sci. U. S. A. 104, 7116–7121. doi:10.1073/pnas.0702646104

López-Lara, I. M., and Geiger, O. (2019). “Chapter 3: formation of fatty acids,” in Biogenesis of Fatty Acids, Lipids and Membranes, ed. O. Geiger (Springer). p. 475.

Lowe, J., and Amos, L. A. (1998). Crystal structure of the bacterial cell division protein FtsZ. Nature 391, 203–206. doi:10.1038/34472

Luedemann, G. M. (1968). Geodermatophilus sp. nov., a new genus the Dermatophilaceae (Actinomyceteta). J. Bacteriol. 96, 1848–1858. doi:10.1128/JB.96.11.1848-1858.1968

McCoo, G. J., and Cannon, M. C. (2001). PhaC and PhaR are required for polyhydroxyalkanoic acid synthase activity in Bacillus megaterium. J. Bacteriol. 183, 4233–4243. doi:10.1128/JB.183.11.4233-4243.2001

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013a). Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Biol. 14:640. doi:10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Göker, M., Spröer, C., and Klenk, H.-P. (2013b). When should a DDH experiment be mandatory in microbial taxonomy? Arch. Microbiol. 190, 413–418. doi:10.1007/s00203-013-0888-4

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2014a). Highly parallelized inference of large genome-based phylogenies. Genome Biol. Evol. 6, 413–418. doi:10.1093/gbe/evu043

Meier-Kolthoff, J. P., Klenk, H.-P., and Göker, M. (2014b). 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/ijsem.0.056994-0

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
gamma-proteobacterium isolated from Wohlfahrtia magnifica (Diptera: Sarcophagidae). Int. J. Syst. Evol. Microbiol. 58, 976–981. doi: 10.1099/ijs.0.65324-0

Traga, B. A., and van Wezel, G. P. (2008). The SigA-like proteins in actinomycetes: small proteins up to a big task. Antonie van Leeuwenhoek 94, 85–97. doi: 10.1007/s10482-008-9225-3

Trujillo, M. E., and Normand, P. (2019). Sportichthya in Bergy’s Manual of Systematics of Archaea and Bacteria. John Wiley and Sons, Ltd. Chichester.

Upadhyay, A., Fontes, F. L., Gonzalez-Jarero, M., McNeil, M. R., Crans, D. C., Kacson, M., et al. (2015). Partial saturation of menaquinone in mycobacterium tuberculosis: function and essentiality of a novel reductase. Menf. ACS Cent. Sci. 1, 292–302. doi: 10.1021/acscentsci.5b00212

Urzi, C., Brunetti, L., Salamone, P., Sorlini, C., Stackebrandt, E., and Daffonchio, D. (2001). Biodiversity of geodermatophilaceae isolated from altered stones and monuments in the mediterranean basin. Environ. Microbiol. 3, 471–479. doi: 10.1046/j.1462-2920.2001.00217.x

Urzi, C., Salamone, P., Schumann, P., Rohde, M., and Stackebrandt, E. (2004). Blastococcus saxobidens sp. nov., and emended descriptions of the genus Blastococcus Ahrens and Möll 1970 and Blastococcus aggregatus Ahrens and Möll 1970. Int. J. Syst. Evol. Microbiol. 54, 253–259. doi: 10.1099/ijs.0.02745-0

Vaas, L. A. I., Sikorski, J., Hofner, B., Fiebig, A., Buddruhs, N., Klenk, H. P., et al. (2013). Opm: an R package for analysing OmniLog® phenotype microarray data. Bioinformatics 29, 1823–1824. doi: 10.1093/bioinformatics/btt291

Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M., and Klenk, H. P. (2012). Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. PLoS One 7:e34846. doi: 10.1371/journal.pone.0034846

Van Wezel, G. P., van der Meulen, J., Swamoto, S., Luiten, R. G., Koerten, H. K., and Kraal, B. (2000). SsgA is essential for sporulation of Streptomyces coelicolor A3(2) and affects hyphal development by stimulating septum formation. J. Bacteriol. 182, 5653–5662. doi: 10.1128/JB.182.20.5653-5662.2000

Vandamme, P., and Sutcliffe, I. (2021). Out with the old and in with the new: time to rethink twentieth century chemotaxonomic practices in bacterial taxonomy. Int. J. Syst. Evol. Microbiol. 71.005127. doi: 10.1099/ijsem.0.005127

Vetrovský, T., and Baldwin, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS One 8:e57923. doi: 10.1371/journal.pone.0057923

Wang, X., and Lilkenhaus, J. (1993). The FtsZ protein of Bacillus subtilis is localized at the division site and has GTPase activity that is dependent upon FtsZ concentration. Mol. Microbiol. 9, 435–442. doi: 10.1111/j.1365-2958.1993.tb01705.x

Wang, Y., Zhang, L., Zhang, X., Huang, J., Zhao, Y., Zhao, Y., et al. (2017). Geodermatophilus daqingensis sp. nov., isolated from petroleum-contaminated soil. Antonie van Leeuwenhoek 110, 803–809. doi: 10.1007/s10482-017-0853-3

Wayne, L. G., Brenner, D. J., Colewell, R. R., Grignon, P. A. D., Kandler, O., Krichevsky, M. I., et al. (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464. doi: 10.1099/00207713-37-4-463

Whitman, W. B. (2015). Genome sequences as the type material for taxonomic descriptions of prokaryotes. Syst. Appl. Microbiol. 38, 217–222. doi: 10.1016/j.syapm.2015.02.003

Wiley, E. O., and Lieberman, B. B. S. (2011). Phylogenetics: Theory and Practice of Phylogenetic Systematics, 2nd. Hoboken, NJ: John Wiley and Sons.

Xiao, J., Luo, Y., Xu, J., Xie, S., and Xu, J. (2011). Modestobacter marinus sp. nov., a psychrotolerant actinobacterium from deep-sea sediment, and emended description of the genus Modestobacter. Int. J. Syst. Evol. Microbiol. 61, 1710–1714. doi: 10.1099/ijs.0.023085-0

Yang, Z. W., Asem, M. D., Li, X., Li, L. Y., Salam, N., Alkalifah, D. H. M., et al. (2019). Blastococcus deserti sp. nov., isolated from a desert sample. Arch. Microbiol. 201, 193–198. doi: 10.1007/s00203-018-1604-1

Zhu, W. Y., Zhai, Y., Qin, Y. L., Xiong, Z. J., Zhang, D. F., Klenk, H. P., et al. (2013). Blastococcus endophyticus sp. nov., an actinobacterium isolated from Camptotheca acuminata. Int. J. Syst. Evol. Microbiol. 63, 3269–3273. doi: 10.1099/ijs.0.049239-0

Zhi, X. Y., Li, W. J., and Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. Int. J. Syst. Evol. Microbiol. 59, 589–608. doi: 10.1099/ijs.0.65780-0