Proposal of Carbonactinosporaceae fam. nov. within the class Actinomycetia. Reclassification of Streptomyces thermoautotrophicus as Carbonactinospora thermoautotrophica gen. nov., comb. nov. Systematic and Applied Microbiology, 44 (4). p. 126223. ISSN 0723-2020

Published by: Elsevier

URL: https://doi.org/10.1016/j.syapm.2021.126223 <https://doi.org/10.1016/j.syapm.2021.126223>

This version was downloaded from Northumbria Research Link: http://nrl.northumbria.ac.uk/id/eprint/46505/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher’s website (a subscription may be required.)
Proposal of Carbonactinosporaceae fam. nov. within the class Actinomycetia. Reclassification of Streptomyces thermoautotrophicus as Carbonactinospora thermoautotrophica gen. nov., comb. nov

Camila Gazolla Volpiano a,1, Fernando Hayashi Sant’Anna b,1, Fábio Faria da Mota c, Vartul Sangal d, Iain Sutcliffe d, Madhaiyan Munusamy e, Venkatakrishnan Sivaraj Saravanan f, Wah-Seng See-Toog g, Luciane Maria Pereira Passaglia a, Alexandre Soares Rosado b,1,6

a Departamento de Genética and Programa de Pós-graduação em Genética e Biologia Molecular, Instituto de Biociências, 9500, Bento Gonçalves Ave, Porto Alegre, RS, Brazil
b PROADI-SUS, Hospital Moinhos de Vento, 630, Ramiro Barcelos Porto Alegre, RS, Brazil
c Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, 4365, Brasil Ave, Rio de Janeiro, RJ, Brazil
d Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom
e Temasek Life Sciences Laboratory, I Research Link, National University of Singapore, Singapore 117604, Singapore
f Department of Microbiology, Indira Gandhi College of Arts and Science, Kathirkamam, Pondicherry, India
g Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia
h LEMM, Laboratory of Molecular Microbial Ecology, Institute of Microbiology Paulo de Góes, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
i BESE, Biological and Environmental Sciences and Engineering Division, KAUST, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

Abbreviations: AAI, Average Amino Acid Identity; AF, Alignment Fractions; ANI, Average Nucleotide Identity; ANIb, ANI using the BLASTn; BBHs, bidirectional best hits; CDSs, protein-coding sequences; dDDH, digital DDH; gANL, whole-genome ANL; GBDP, Genome BLAST Distance Phylogeny; HGT, horizontal gene transfer; KOs, KEGG (Kyoto Encyclopedia of Genes and Genomes) orthologous (KOs); LPSN, List of Prokaryotic names with Standing in Nomenclature; LTP, Living Tree Project; MiSI, Microbial Species Identification; ML, Maximum Likelihood; NISEs, Non-Homologous Isofunctional Enzymes; POCP, Percentage of Conserved Proteins; TYGS, the Type (Strain) Genome Server.

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Introduction

The genus *Streptomyces* Waksman and Henrici 1943, belonging to order *Streptomyces*, family *Streptomycetaceae* within the class *Actinomycetia* (former *Actinobacteria*) [1], is one of largest bacterial genera as currently defined, with more than 600 validly named species according to the List of Prokaryotic names with Standing in Nomenclature (https://lpsn.dsmz.de/genus/streptomyces). Members of the genus are highly significant because of their complex lifecycles, including sporation, which have made them important as model organisms for studies of bacterial genetics and ecology [2-4]; and because members of the genus are highly proficient producers of secondary metabolites of biomedical and biotechnological interest, notably antibiotics and anticancer compounds [5-7]. Given the size of the genus *Streptomyces*, attempts have been made to determine internal structure of the genus through phylogenetic characterization of species groups, and its relationships to other taxa within the family *Streptomycetaceae* [8–11], although more comprehensive phylogenomic studies are required to resolve interspecies and suprageneric structure [6,11].

Genomic metrics have become the gold standard for defining taxonomic ranks, especially species designations among prokaryotes as they provide a reproducible, reliable, and a highly informative means to infer relatedness directly between genomes sequences [12,13]. In particular, average nucleotide identity (ANI) using the BLASTn algorithm (AN Ib) and the Genome BLAST Distance Phylogeny (GBDP)-based digital DDH (dDDH) methods have been widely used to determine species boundaries and confirm identification [12,14,15]. The use of phylogenetic analyses in addition to Average Amino acid Identity (AAI), Percentage of Conserved Proteins (POCP), and whole-genome ANI (gANI) coupled with Alignment Fractions (AF) metrics have been also proposed to demarcate genus and higher taxa [13,16-18].

*Streptomyces thermoautotrophicus* Gadkari et al. 1991 has been suggested to merit generic status since it does not cluster within the sensu stricto *Streptomyces* clade in a tree inferred with the GBDP using formula $dS$ [11] and was located apart from the clade containing members of the family *Streptomycetaceae* in a phylogeny from 14 well-conserved proteins [19]. This species was described by Gadkari et al. [20] based on characteristics of a single strain, UBT1T, isolated from soil covering a charcoal burning pile. Strain UBT1T is of interest as a sporulating aerobic thermophile, exhibiting growth at 40–68 °C, likely reflecting its isolation source. It was claimed to be a CO- and H$_2$-oxidizing obligate chemolithoautotrophic bacterium [20] and to produce a biochemically distinct, oxygen insensitive nitrogenase [21,22]. Later, MacKellar et al. [19] isolated a second ‘*S. thermoautotrophicus*’ strain, H1, from another burning charcoal pile near an active coal seam fire. Multiple CO dehydrogenase gene clusters have been identified in the genomes of strains UBT1T and H1; however, genes encoding nitrogenase enzymes seem to be absent. In addition, strains H1 and UBT1T were unable to grow on Noble agar or to incorporate $^{15}$N$_2$ into biomass, besides growing heterotrophically on pyruvate. As a result, MacKellar et al. [19] proposed the reclassification of ‘*S. thermoautotrophicus*’ as non-diazotrophic, facultative chemolithoautotrophic bacteria. Nevertheless, the chemolithoautotrophic metabolism of ‘*S. thermoautotrophicus*’ distinguishes it from members of the genus *Streptomyces* [11]. In addition, the presence of eleven (UBT1T) or nine (H1) biosynthetic gene clusters for secondary metabolites in the ‘*S. thermoautotrophicus*’ genomes [19] is relatively low compared to other members of *Streptomyces* [5], reflecting the comparatively small genome sizes (~5 Mb) of these two strains. Finally, the circularity of the H1 genome distinguishes it from *Streptomyces sensu stricto*, where most genomes are linear [23].

Although earlier studies have presented convincing evidence that ‘*S. thermoautotrophicus*’ should be reclassified into a novel genus [11,19], formal taxonomic proposals to achieve this have not been made due to concerns about the type strain availability and ambiguity concerning its suprageneric relationships. Here, we revisit the taxonomy of ‘*S. thermoautotrophicus*’ and propose its reclassification as *Carbonactinospora thermoautotrophica* gen. nov., comb. nov., within the *Carbonactinosporaceae* fam. nov.

Material and methods

16S rRNA sequence identity analysis

The 16S rRNA gene sequence data for the type strains within the class *Actinomycetia* were retrieved from the SILVA SSU r138.1 database [24] (link to the full license: https://creativecommons.org/licenses/by/4.0/legalcode). The search criteria in the Living Tree Project (LTP) dataset were set as follows: “*Actinobacteria*” in the taxonomy field, sequence length $>$1400 nucleotides, sequence quality $>$90, and type strains (search term “[TI]” in the strain field). Sequences were downloaded as an alignment in FASTA format containing gaps. As in the SILVA database, *Acidothermus cellulolyticus* is classified within the order *Frankiales* [25], it was manually corrected to *Acidothermales*.

The 16S rRNA genes sequences of strains ‘*S. thermoautotrophicus*’ UBT1T [20] and H1 [19] were extracted from the RefSeq genome assemblies GCF_001543895 and GCF_001543925, respectively, and subsequently aligned using SINA 1.2.11 [26]. A consensus alignment between the genomic ‘*S. thermoautotrophicus*’ and the SILVA alignments was obtained. Positions containing gaps were removed from the alignment and an identity matrix for the resulting 858 positions was computed using Bioedit v. 7.0.5.3. The Python library Seaborn v. 0.11.1 was utilized for building box-plots of the 16S rRNA identities for each order within the class *Actinomycetia*. Additionally, sequence identity was assessed by comparing the 16S rRNA sequences of the UBT1T and H1 strains with the sequences from EzBioCloud [27], a quality-controlled 16S rRNA server database.

Phylogenetic analyses

Genera within the class *Actinomycetia* were identified in the lineage file available in https://github.com/zyxue/ncbitax2lin (v. 2019-02-20), which was generated from the NCBI taxonomy dump. Subsequently, the type species of each genus were retrieved according to the names provided on LPSN using the script “get_t_type_genus.py” (available at https://github.com/fhsantanna/bioinfo_scripts). All the available proteomes for *Actinomycetia* type species in the NCBI Assembly RefSeq database were downloaded for further analyses. The proteomes of the *Embleya* and *Trebonia* type species were later included manually since these genera were not available in the lineage file. Lastly, the proteome of UBT1T and H1 were included in the final sequence dataset.

Two different approaches were carried out for the phylogenetic reconstruction based on the concatenated alignment of orthologous proteins. The first approach utilized the AMPHORA2 [28] pipeline for the identification of universal taxonomic markers in the *Actinomycetia* proteomes. For this purpose, the “phylogenomics-tools” scripts were utilized [29]. The markers dnaG, inFC, nusA, pgk, pyrG, rplK, rpoB, rpsC, and smpB were excluded from the analyses as they were present in either multiple copies or at low representation among the type species. A total of 253 taxa
remained in the dataset after the exclusion of proteomes that did not present the final 22 markers (Supplementary Material). After, each marker protein was aligned using MUSCLE [30] v. 3.8.31 and concatenated. Positions containing gaps were excluded and the final 3184 amino acids alignment was utilized as input for the phylogenetic reconstruction based on the Maximum Likelihood (ML) method in the PhyML 3.0 server [31]. The substitution model was selected based on the Akaike Information Criterion to select LG + G + I as the best model, with an estimated gamma shape parameter of 0.829 and an estimated proportion of invariable sites of 0.165. Branch support was assessed using aLRT SH-like [32].

The second approach was a protein-based core genome phylogeny using a de novo identification of phylogenetic markers. Core ortholog groups of the previously selected strains were identified using bidirectional best hits (BBHs) algorithm implemented in GET_HOMOLOGUES [33] pipeline build 31072020, excluding inparalogs and using minimal blast searches. Once the core proteins were identified, GET_PHYLOMARKERS [34] v. 2.2.8.1 was used with default parameters (-R 1 -t PROT options) for finding optimal ortholog clusters for phylogenomic reconstruction. This approach is based on three main filters: exclusion of alignments containing recombinant sequences, removal of reconstructions that deviate from expectations of the multispecies coalescent, and elimination of poorly resolved gene trees. Top-scoring gene alignments were concatenated into a supermatrix, which was utilized to estimate the species-tree with the ML method. The phylogenetic trees were processed with Newick utilities [35], whose functionalities include taxa renaming and tree pruning (i.e. removing clades and only keeping those of interest).

In order to obtain a genome tree using GBDP, the genome sequence data were uploaded to TYGS, the Type (Strain) Genome Server [36]. In brief, the determination of closest type strain genomes was done in two complementary ways: first, the UBT1T and H1 genomes were compared against all type strain genomes available in the TYGS database via the MASH algorithm, a fast approximation of intergenomic relatedness [37], and, then the ten type strains with the smallest MASH distances were chosen for each S. thermoaerotrophicus genome. Second, an additional set of ten closely related type strains was determined via the 16S rDNA gene sequences. These were extracted from UBT1T and H1 genomes using RNAmmer [38] and each sequence was subsequently BLAST searched [39] against the 16S rDNA gene sequence of each of the currently 13,011 type strains available in the TYGS database. This was used as a proxy to find the best 50 matching type strains (according to the bitscore) for each S. thermoaerotrophicus genome and to subsequently calculate precise distances using the GBDP approach under the algorithm ‘coverage’ and distance formula $d_5$ [40]. For the calculation, local-alignment programs are used to align a genome X against a genome Y, and vice versa, producing a set of high-scoring segment pairs (HSPs). These matches are then transformed to a single distance value $d(X, Y)$ by applying the formula $d_5$, which is calculated as two times the sum of identical base pairs over all HSPs ($2 \times C_i$) divided by the total length of all HSPs found in both genomes ($H_N + H_S$) [41,42], rescaled for phylogenetic inference and with branch support values based on resampling [43]. These distances were finally used to determine the 10 closest type strain genomes for each of the S. thermoaerotrophicus genomes. For the GBDP tree reconstruction, all pairwise comparisons among the set of genomes were conducted using GBDP under the algorithm ‘trimming’ and distance formula $d_5$. The resulting distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.4 including SPR postprocessing [44]. Branch support was inferred from 100 pseudo-bootstrap replicates each. The trees were rooted at the midpoint [45] and visualized with PhyD3 [46].

Genome and proteomic metrics

Proteomic and genomic relatedness metrics were computed comparing S. thermoaerotrophicus to type species from the order Streptosporangiales, and the genera Acidothermus, Catenulispora, Frankia, Micromonaspora, Pseudonocardia, Sporichthya and Streptomyces.

gANI and AF values were obtained by the Microbial Species Identifier (MSI) method using ANICalculator 2014-127 v. 1.0 (https://ani.jgi.doe.gov/html/home.php?page=introduction). gANI is calculated for a pair of genomes by averaging the nucleotide identity of orthologous genes identified as BBHs, which are the genes that show >70% sequence identity and >70% alignment of the shorter gene. AF is calculated as a fraction of the sum of the lengths of BBH genes divided by the sum of the lengths of all genes in a genome [47].

POCP values were obtained with the script “POCP.sh” (available at https://figshare.com/articles/POCP_calculation_for_two_genomes/4577953/1), which was written based on Quin et al. [17]. For POCP calculation, the conserved proteins between a pair of genomes are determined by aligning all the protein sequences of a genome X against a protein’s sequences from a genome Y, using the BLASTP aligner. Conserved proteins are defined as presenting a match with an <1e^-5 E value, > 40% of sequence identity, and >50% of an alignable region of the query protein sequence. The POCP (X, Y) % is calculated as $\frac{C_x + C_y}{(T_x + T_y)} \times 100$, where C represents the conserved number of proteins and T represents the total number of proteins on the respective genome.

AAI analyses were performed using the script “aai.rb” implemented in the Enveomics Collection [48]. For AAI calculation, the conserved genes between a pair of genomes are determined by aligning all the protein-coding sequences (CDs) of a genome X against a translated database of genome Y, using the TBLASTN aligner. Conserved CDs are defined as presenting >30% of sequence identity at the amino acid level and >70% of an alignable region of the query CDS sequence. The matching segment from the genomic sequence is extracted and the reverse search with BLASTX is used to determine the presumably orthologous fraction of conserved genes between the two genomes (two-way BLAST). The two-way AAI (X, Y) % is measured by the average amino acid identity of all two-way BLAST conserved genes between the genomes, as computed by the BLAST algorithm [49]. For evaluating the AAI diversity between S. thermoaerotrophicus, Streptomyces, and Streptomyces, and Streptomyces type strains, all available proteomes from these taxa were utilized for AAI computation as described above. As a control, Streptomyces albus, the type species of the genus Streptomyces, was compared to the same taxa.

Scatter plots showing the relationship between AAI and POCP and between AAI and AF were generated using the Python library Seaborn v. 0.11.1.

Taxonomic profiling of proteomes

AAI-profiler, which is a webserver dedicated to taxonomic identification [50], was employed to perform proteome-wide sequence searches using S. thermoaerotrophicus UBT1T and H1 genomes. AAI-profiler computes AAI between a query proteome and all target species in the UniProt database [50]. Each protein is binned considering the taxonomic attribution of the closest counterpart in the database. A taxonomic profile of the proteome of interest is built considering the counts of the target taxa, and these frequencies are weighted by the percent identity of the match to the query. Given that S. thermoaerotrophicus is already included in the AAI-profiler database, the taxonomic profile excluded hits of the top-ranked taxon, which are from S. thermaaerotrophicus itself.
**KEGG orthologous analysis**

To find metabolic divergences between ‘S. thermoautotrophicus’ and other *Streptomyces* spp., the predicted amino acid sequences from the genomes of strains UBT1T and H1 were compared to the KEGG (Kyoto Encyclopedia of Genes and Genomes) orthologous (KOs) belonging to the 71 *Streptomyces* spp. obtained from KEGG database [51] (Supplementary Material).

**Results and discussion**

**Diversity of the 16S rRNA genes from ‘S. thermoautotrophicus’**

To evaluate the taxonomic position of ‘S. thermoautotrophicus’ within Actinomycetia, we first conducted a 16S RNA gene identity sequence analysis of strains UBT1T and H1. As previously reported [19], the genome of UBT1T contains three 16S rRNA genes, two of which are identical to each other (locus tags TH66_RS04095 and TH66_RS03010) whilst the other is divergent (TH66_RS22860), presenting 94% identity to the other two. The genome assembly of H1 contains two 16S rRNA genes, one of which (LI90_RS08655) is identical to the TH66_RS04095/TH66_RS03010 pair, and the other (LI90_RS18525) is identical to the divergent copy TH66_RS22860.

The presence of multiple 16S RNA gene copies within a single bacterial genome has been observed before. Indeed, bacteria can harbour more than 20 copies of this marker gene [52]. The presence of intragenomic heterogeneity of 16S rRNA ≥ 6% was also reported in some thermophiles, such as the *Firmicutes* members *Desulfotomaculum kuznetsovii* DSM 6115T and *Thermoanaerobacter tengcongensis* MB4T, and the *Actinobacteria* member *Thermobispora bispora* DSM 43833T [53,54]. This may constitute an ecological strategy [55–57] to adapt the bacterial cellular machinery to perform under different temperatures [58], with different copies being...
functional under different environmental conditions [59]. In addition to the biases introduced from PCR [60,61], the presence of multiple different 16S rRNA gene copies is another strong argument against relying only on 16S rRNA gene phylogeny in species delineation in traditional polyphasic approach.

To identify 16S rRNA gene relatedness at the genus level, each copy from UBT1T and H1 was compared to 2,792 16S rRNA sequences from type strains of Actinomycetia species available in the SILVA database. According to this analysis, TH66_RS04095/T TH66_RS03010/LI90_RS08655 exhibit identities above the 94.5%.

Fig. 2. Multiprotein phylogenies of Actinomycetia type-species and ‘S. thermoautotrophicus’ UBT1T and H1. (A) Phylogenetic reconstruction based on the core-proteome proteins of Actinomycetia type species using ML (IQ-TREE). Clade support values are shown next to the nodes. The first value corresponds to approximate Bayes branch support values and the second one to the UFBoot (Ultra Fast bootstrap) values. (B) Phylogenetic reconstruction of AMPHORA2 proteins using PhyML. aLRT values greater than 70% are shown next to the nodes. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. Rubrobacter radiotolerans DSM 5868T is the outgroup. Only representative taxa are shown in these trees, which are the pared-down version of the complete trees (253 genomes) shown in Figs. S1–S3.

Fig. 3. Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from ‘S. thermoautotrophicus’ UBT1T and H1 genome sequences. The branch lengths are scaled in terms of GBDP distance formula $d_5$. The numbers above branches are GBDP pseudo-bootstrap support values >60% from 100 replications, with an average branch support of 66%. The tree was rooted at the midpoint.
Acidothermus and ifida, Thermobispora, and H1, and the latter strains share a last common ancestor with Streptomyces from 87 non-genus circumscription threshold [62] with 16S rRNA sequences. The more divergent TH66_RS22860/LJ90_RS18525 copies did not belong to any recognized phyotypes at the genus level when compared with sequences from the Actinomyceta dataset (Fig. 1B), and even with the 65,797 entries in the EzBioCloud 16S rRNA database (Table S1). In both analyses, A. cellulolyticus stood out in presenting 93.7% identity to these divergent copies.

The current understanding of the evolutionary forces shaping the genomes of Actinomyceta is limited [63]; however, McDonald and Currie [64] analyzed 122 Streptomyces genomes and found that the acquisition and retention of genes through horizontal gene transfer (HGT) are surprisingly rare in this genus. Considering these findings, one of these 16S rRNA sequences can be assumed to be ancestral to strains UBT1 and H1 while the other in each genome seems to be the product of a more recent duplication, rather than an HGT event.

Fig. 4. Distribution of proteomic/genomic metrics between ‘S. thermoautotrophicus’ and Streptosporangiales/Streptomyces type strains. (A) AAI vs POCP. (B) gANI vs AF. Symbols representing each comparison are depicted in the legend box. Dashed lines represent genus circumscription thresholds. The gray box represents the region where comparisons present proteomic/genomic values that fall inside the genus limits.

Phylogenetic placement of ‘S. thermoautotrophicus’ within Actinomyceta

A multigene-based phylogenetic approach should be the choice for defining genera or higher taxa according to the minimal standards for the use of genome data for the taxonomy of prokaryotes [13]. Thus, to identify the current closest relatives to ‘S. thermoautotrophicus’ UBT1 and H1, two different approaches were employed to reconstruct the evolutionary history of UBT1, H1 and an additional set of 251 type species of Actinomyceta with genomes/proteomes available. We first reconstructed a ML phylogenetic tree with the concatenated protein sequences from 22 conserved single-copy genes identified in the assemblies with the AMPHORA2 pipeline. In addition, we performed a de novo approach for the identification of nine ortholog genes/ubiquitous proteins in the Actinomyceta type species genomes which were appropriate for phylogenomic analysis, with only three of them, encoding proteins of the 50S ribosomal subunit, also present in the AMPHORA2 dataset. Both phylogenetic reconstructions (Fig. 2 and Figs. S1–S3) infer that the genus Streptomyces does not form a clade with UBT1 and H1, and the latter strains share a last common ancestor with A. cellulolyticus and members of the Streptosporangiales clade, thus belonging to a deeply branching lineage.

In the previous phylogenomic analysis that included ‘S. thermoautotrophicus’, MacKellar et al. [19] highlighted the unusual position of the UBT1 and H1 genomes as being closely related to Acidothermus and Streptosporangiales (Streptosporangium, Thermobifida, Thermobispora, and Thermomonospora), and distinct from the clade containing the families Streptomycetaceae and Catenulisporaceae. Therefore, the authors proposed that UBT1 and H1 do not belong to the genus Streptomyces and instead are nearer to families including Acidothermales and Streptosporangiacae. The proposal of a generic status for ‘S. thermoautotrophicus’ was also supported by Nouioui et al. [11], in a tree inferred with GBDP formula d5 [11], where UBT1 branched away from core Streptomyces before Kitasatospora and Streptacidiphilus, forming a sister group to the core Streptomyces-Kitasatospora-Streptacidiphilus clade. The position inferred by Nouioui et al. [11], however, conflicts with MacKellar et al. [19] and our phylogenetic reconstructions based on ML estimations (Fig. 2), where UBT1 forms a sister group with Acidothermus and members of Streptosporangiales.
phylogenetic approaches measure distances based on variation in the nucleotide or amino acid sequences at each site, or the presence or absence of indels, upon an implicit or explicit mathematical model describing the evolution, namely, Bayesian and ML approaches [66]. As exemplified here, the occurrence of incongruence among different tree reconstruction methods are well-known [67,68]. However, we note that the Genome Taxonomy Database (GTDB, release 06-15202) tree places ‘S. thermautotrophicus’ in the order Streptomyces, thus being congruent with the GBDP tree (Fig. S4). The GTDB approach is based on genome trees inferred with FastTree from an aligned concatenated set of up to 120 single copy marker proteins tree [69,70].

According to the phylogenies demonstrated here, strains UBT1T and H1 have a distinct phylogenetic position within the class Actinomycetia, clearly belonging to a novel family. However, further studies are needed to resolve the ambiguity over the placement of the family, which may represent a novel order.

Genus delineation for UBT1T and H1 using genomic and proteomic metrics

Despite the advancements in resolving species delineation and the use of genome data to reconstruct the phylogenetic relationship of microorganisms, there is no consensus on the incorporation of genomic metrics and cutoffs to demarcate genera and higher taxa. Nevertheless, different metrics that measure proteomic and genomic relatedness to demarcate genera have been proposed on the basis of AAI [16] and POCP [17]. Recently, Barco et al. [18] utilized the MiSI method [47] for genus delineation, and they verified that the gANI and AF mean values for genus infection points in Bacteria are 73.1% and 0.333, respectively. Thus, we have applied these approaches to evaluate ‘S. thermautotrophicus’ UBT1T and H1, Streptomyces, Acidothermus, and Streptosporangiales genomes in detail within the taxonomic context of genus.

In the comparison of the closely related Actinomycetia to UBT1T and H1, different Streptosporangiales genomes presented the highest POCP values while some Streptomyces genomes present the highest AAI values (Fig. 4A). According to the AAI measure, Streptomyces megarosporus NRRL B-16372T is a closely related strain to H1 with 59.0% AAI and 45.1% POCP values, while Streptomyces vitaminophilus ATCC 31673T is closely related to UBT1T, presenting 58.9% AAI and 44.4% POCP. According to the POCP metric, Thermomonospora catenispore 3–22–3T (Streptosporangiales) is closely related to both H1 and UBT1T presenting 46.2 and 48.0% POCP, and 58.3 and 58.5% AAI, respectively. The comparisons of ‘S. thermautotrophicus’ with Acidothermus presented even lower values of ~57.3% AAI and 39.3% POCP. Nevertheless, none of the obtained values surpassed the recommended 65 to 72% [16] and 50% [17] thresholds for the delineation of genera using AAI and POCP metrics, respectively. As expected, S. albus was unambiguously grouped with Streptomyces sensu stricto, while the comparisons of UBT1T and H1 strains to Streptosporangiales and Streptomyces appeared to be distinct from S. albus vs Streptosporangiales.

Given the proteomic similarity of some Streptomyces and Streptosporangiales genomes to ‘S. thermautotrophicus’, we further explored the proteomic similarity between UBT1T and H1 to 122 Streptosporangiales and 223 Streptomyces genomes. Comparing Streptomyces species to UBT1T and H1, respectively, we found an AAI of 57.20 ± 0.49 (% mean ± SD) and 57.15 ± 0.5, and the number of common proteins to be 2471 ± 116 and 2315 ± 104. For Streptosporangiales, we found an AAI value of 55.8 ± 1.5 and 55.6 ± 1.5, and the number of common proteins to be 2430 ± 204 and 2255 ± 184. In this analysis, we also did not find any AAI values ≥ 65% to the ‘S. thermautotrophicus’ strains.

In the gANI(AF) analysis (Fig. 4B), similarly to the POCP vs AAI correlation plot, S. albus was grouped with Streptomyces as expected, while the comparisons of UBT1T and H1 to Streptosporangiales were intermixed, and the two strains are clearly distinct from Streptomyces. Although some type species from Streptosporangium, Catenulispora, Frankia, Micromonospora, Pseudonocardia, Sporichthya and Streptomyces present gANI values that surpass 73.1% in relation to UBT1T and H1, these comparisons do not surpass the minimum AF requirement for genus definition i.e. gANI and AF are inconsistent. While gANI represents the identity of orthologous genes identified as BBHs using similarity searches, the AF is a complementary measure of the minimum amount that genomes must overlap [47]. If the homologous regions are short with respect to the total length of the genomes, as might be seen following a HGT event, then ANI values may be high even though the bacteria are distantly related. The comparison of UBT1T and H1 with A. cellulolyticus presented 73.0% (±0.15) gANI (AF) (Tables S2 and S3).

The genomic and proteomic metrics results together demonstrated the substantial difference between ‘S. thermautotrophicus’ and other Actinomycetia members. Sequences from strains UBT1T and H1 are clearly below the established cut-off values (gANI-AF: 73.1%–0.333; AAI: 65–72%; POCP: 50%) for defining bacterial genera, strongly suggesting they represent a novel taxon within Actinomycetia.
To evaluate the taxonomic composition of the 'S. thermoautotrophicus' proteomes, we used strain UBT1\textsuperscript{T} and H1 protein sequences as queries at AAI-profiler for homology searches in the UniProt database. As demonstrated in Fig. 5, Streptomyces proteomes were the top hit for only ~36% of the query proteins from strains UBT1\textsuperscript{T} and H1, while ~19% of them matched to Streptosporangiaceae order proteomes. The other query proteins are distributed among different orders of the Actinomycetia.

The apparent mosaic nature of the UBT1\textsuperscript{T} and H1 genomes reflects the underrepresentation of closely related strains in the public sequence databases rather than HGT. Despite the rapid expansion in number of sequenced bacterial and archaeal genomes in the past decade [27,71,72] along with the number of species names validly published [12], understudied groups are often represented by a single family [73–77], along with a few or no genomes present in nucleotide databases. This bias is evident to [Table 1]. UBT1\textsuperscript{T} can also be distinguished from these families based on discontinuous distribution of chemotaxonomic markers, notably cell wall amino acids, menaquinones, and diagnostic sugars in whole cell hydrolysates, in addition to the presence of spores and colony morphology.

**Phenotypic distinctness of 'S. thermoautotrophicus'**

The metabolic distinctiveness of 'S. thermoautotrophicus' UBT1\textsuperscript{T} and H1 was predicted based on genome comparisons with 71 Streptomyces spp. KO profiles available in the KEGG database. Additional discriminative phenotypic properties were retrieved from the literature for closely related Actinomycetia species.

When compared to UBT1\textsuperscript{T} and H1, 101 KOs were exclusively present among the Streptomyces spp. profiles (Table S4). On the other hand, 136 KOs were exclusively present in the UBT1\textsuperscript{T} and H1 profiles (Table S5), including a nitrate/nitrite sensor two-component system (narNKP) and multiple genes related to carbon metabolism, such as ribulose-bisphosphate carboxylase (rbcLS), glucose/mannose-6-phosphate isomerase, phosphoenolpyruvate carboxykinase, PKF 6-phosphofructokinase 1, fructose 1,6-bisphosphate aldolase/phosphatase, fructose-bisphosphate aldolases, classes I and II. Many exclusive KOs and some Non-Homologous Isofunctional Enzymes (NISEs) cases observed between UBT1\textsuperscript{T} and H1 and other Streptomyces spp. suggest evolutionary divergences in their metabolisms and distant common ancestors. NISEs are evolutionarily unrelated enzymes that catalyze the same biochemical reactions [78]. For example, exclusive KOs for UBT1\textsuperscript{T} and H1 (K01754) and for other Streptomyces spp. (K01752) are related to the same L-serine = pyruvate + NH\textsubscript{4} enzyme reaction (RO0220) but were exclusively found in each group. While UBT1\textsuperscript{T} and H1 have some exclusive enzymes, including RubiscO, related to a carbon autotrophic lifestyle, the other KEGG from Streptomyces spp. showed some exclusive KOs related to a heterotrophic lifestyle, including gluABC.

The major characteristic that differentiates UBT1\textsuperscript{T} from Acidothermaceae, Nocardiopsaceae, Streptomycetaceae, Streptosporangiaceae, Thermomonosporaceae, and Trebionibaccae is its unique ability to grow chemolithotrophically on CO or CO\textsubscript{2} and H\textsubscript{2} (Table 1). UBT1\textsuperscript{T} can also be distinguished from these families based on the discontinuous distribution of chemotaxonomic markers, notably cell wall amino acids, menaquinones, and diagnostic sugars in whole cell hydrolysates, in addition to the presence of spores and colony morphology.
Conclusions

Based on the genetic and phenotypic distinctness presented above, we conclude that the chemolithotrophic strains ‘S. thermoautotrophicus’ UBT1T and H1 represent a novel genus, consistent with previous observations [11,19], and for which we propose the name Carbonactinospora thermoautotrophica gen. nov., comb. nov. (Table 2). Our additional phylogenomic analysis indicate that the genus Carbonactinospora should be placed in a novel family, Carbonactinosporaceae fam. nov. In accordance with the current GTDB taxonomy (Fig. S4), the family Carbonactinosporaceae is placed within the order Streptomyctetales, but we note that there are ambiguities in phylogenomic analyses (Figs. 2 and 3) that warrant further studies.

Description of Carbonactinosporaceae fam. nov. (Car.bon.ac.ti.no.sp'o.ra.ce’ae. N.L. fem. n. Carbonactinospora, type genus of the family: -aceae, ending to denote a family: N.L. fem. pl. n. Carbonactinosporaceae, the Carbonactinospora family).

Gram-stain positive, mycelium-forming sporulating bacteria. Carbonactinosporaceae represents a distinct Actinomycetae phylogenetic lineage based on multigene-based phylogenetic analyses. The type genus is Carbonactinospora.

Funding

This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil), the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and KAUST, Saudi Arabia (BAS/1/1096-01-01).

Acknowledgement

We thank Dr Imen Nouioui (Leibniz Institute DSMZ, Germany) for arranging deposit of strain UBT1T (=DSM 100163T) in the Korean Collection of Type Cultures. We thank Aharon Oren (Hebrew University of Jerusalem, Israel) for his advice on the formation of the Latin names proposed.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.syapm.2021.126223.
References

[1] Salam, N., Jao, J.Y., Zhang, X.T., Li, W.J. (2020) Update on the classification of higher ranks in the phylum Actinobacteria. Int. J. Syst. Evol. Microbiol. 70, 1339–1355.
[2] Chater, K.F. (2016) Recent advances in understanding Streptomyces. Front. Microbiol. 7, 2533–2544.
[3] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[4] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[5] Konstantinidis, K.T., Tiedje, J.M. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[6] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[7] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[8] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[9] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[10] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[11] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[12] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[13] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[14] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[15] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[16] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[17] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[18] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[19] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[20] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[21] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[22] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[23] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[24] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[25] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[26] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[27] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[28] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[29] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.

[30] Nodwell, J.R. (2019) Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes. Microbiologia 165, 953–955.
[31] Konstantinidis, K.T. (2007) Prokaryotic taxonomy and phylogeny:扶贫腼rivey revisited: advances in the family Streptomycesaceae. J. Bacteriol. 190, 1113–1124.
