A systematic study of the whole genome sequence of *Amycolatopsis methanolica* strain 239T provides an insight into its physiological and taxonomic properties which correlate with its position in the genus*.

Biao Tang a, b, Feng Xie c, Wei Zhao b, Jian Wang c, Shengwang Dai c, Huajun Zheng d, Xiaoming Ding a, Xufeng Cen b, Haican Liu e, Yucong Yu d, Haokus Zhou f, Yan Zhou a, d, Lixin Zhang c, *** Michael Goodfellow g, ** Guo-Ping Zhao a, b, d, f, *

a State Key Laboratory of Genetic Engineering, Department of Microbiology, School of Life Sciences and Institute of Biomedical Sciences, Fudan University, Shanghai, 200438, China
b CAS-Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200093, China
c CAS-Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, No. 1 Beichen West Road, Chaoyang District, Beijing, 100101, China
d Shanghai-MOST Key Laboratory of Disease and Health Genomics, Chinese National Human Genome Center at Shanghai, Shanghai, 201203, China
e State Key Laboratory for Infectious Diseases Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China
f Department of Microbiology and Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China
g School of Biology, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK

**ARTICLE INFO**

Article history:
Received 29 October 2015
Received in revised form 1 April 2016
Accepted 18 May 2016

Keywords:
* Amycolatopsis methanolica
* Complete genome sequence
* One carbon metabolism
* Sub-generic phyletic clades
* AOS
* ATS
* AMS

**ABSTRACT**

The complete genome of methanol-utilizing *Amycolatopsis methanolica* strain 239T was generated, revealing a single 7,237,391 nucleotide circular chromosome with 7074 annotated protein-coding sequences (CDSs). Comparative analyses against the complete genome sequences of *Amycolatopsis japonica* strain MG417-CF17T, *Amycolatopsis mediterranei* strain U32 and *Amycolatopsis orientalis* strain HCCB10007 revealed a broad spectrum of genomic structures, including various genome sizes, core/quasi-core/non-core configurations and different kinds of episomes. Although polyketide synthase gene clusters were absent from the A. mediterranei genome, 12 gene clusters related to the biosynthesis of other specialized (secondary) metabolites were identified. Complete pathways attributable to the facultative methylotrophic physiology of *A. methanolica* strain 239T, including both the *mdo/mscR* encoded methanol oxidation and the *hps/lpi* encoded formaldehyde assimilation via the ribulose monophosphate cycle, were identified together with evidence that the latter might be the result of horizontal gene transfer. Phylogenetic analyses based on 16S rDNA or orthologues of *AMETH_3452*, a novel actinobacterial class-specific conserved gene against 62 or 18 *Amycolatopsis* type strains, respectively, revealed three major phyletic lineages, namely the mesophilic or moderately thermophilic *A. orientalis* subclade (AOS), the mesophilic *Amycolatopsis taiwanensis* subclade (ATS) and the thermophilic *A. methanolica* subclade (AMS). The distinct growth temperatures of members of the subclades correlated with corresponding genetic variations in their encoded compatible solutes. This study shows the value of integrating conventional taxonomic with whole genome sequence data. © 2016 Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* 1. Introduction

The genus *Amycolatopsis* [1] belongs to the family *Pseudonocardiae* [2,3] which is a member of the order *Pseudonocardiales* [4] in the class *Actinobacteria* [5]. This genus currently encompasses

http://dx.doi.org/10.1016/j.synbio.2016.05.001

**2015-2016 Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).**
65 validly published species (http://www.bacterio.net/amycolatopsis.html) including the type and sole representative of Amycolatopsis methanolica, a facultative methylotrophic actinobacterium with a tortuous taxonomic pedigree. The organism was initially classified as Streptomyces sp. strain 239 [6,7], but was moved first to the genus Nocardia [8] and then to the family Pseudonocardiaceae [9] prior to its assignment to the genus Amycolatopsis based on a combination of genotypic and phenotypic criteria [10].

Like most Amycolatopsis strains, Amycolatopsis methanolica strain 239 is Gram-positive, non-acid-fast, and forms aerial and substrate hyphae that fragment into squarish elements. It contains meso-diaminopimelic acid (meso-A₂pm), arabinose and galactose in the wall peptidoglycan, major amounts of di- and tetrahydrogenated menaquinones, phosphatidylethanolamine as the diagnostic phospholipid, ISO-C₁₆:₀ as the predominant fatty acid, but lacks mycolic acids [11]. The genus Amycolatopsis can be distinguished from other genera classified in the family Pseudonocardiaceae using genus-specific oligonucleotide primers [12] and a broad range of genotypic and phenotypic markers [2,3].

The genus Amycolatopsis contains alkaliphilic, endophytic, halophilic, mesophilic, pathogenic and thermophilic species [11,13–15]. Over the years, a gradually increasing number of Amycolatopsis species were assigned to several multi- and single-membered subclades based on gyrB, recN and 16S rRNA gene sequence analyses [11,16,17] though most species belong to two major subclades represented by the earliest described species, namely A. methanolica (AMS) and Amycolatopsis orientalis (AOS). Several AOS species have been shown to synthesize antibiotics, notably A. orientalis, the type species of the genus, which produces vancomycin [1,18] and Amycolatopsis mediterranei which produces rifamycin [1,19]. In contrast, A. methanolica seems to have a markedly different metabolism as it is a facultative methylotroph that synthesizes few, if any, antibiotics. It is difficult to distinguish between AMS and AOS species using phenotypic properties though strains in the former group grow well at 45 °C [10,11,20–25] hence species classified in the AMS can be considered as thermophilic actinobacteria [26,27]. AMS strains are of potential value in biotechnology as vehicles for fermentative overproduction of aromatic amino acids [10,28] and as agents of bioremediation [25,29].

Data derived from whole-genome sequences are being used increasingly to clarify relationships between prokaryotes, including filamentous actinobacteria which are difficult to resolve using established procedures [30,31]. Complete genome sequences are available for Amycolatopsis japonica strain MG417-CF17 which produces (S,S)-N,N’-ethylenediamineuccinacid, a potential phospholipase inhibitor [32]; for the rifamycin SV-producer, A. mediterranei strain U32 [33] and for the vancomycin-producing A. orientalis strain HCCB10007 [34]. Here, we present the complete genome sequence of A. methanolica strain 239. We show that it contains genes encoding one-carbon metabolic pathways and 12 potential specialized (secondary) metabolic biosynthetic gene clusters thereby providing a further insight into the taxonomy of the genus Amycolatopsis. Comparative genomic analyses undertaken between the A. methanolica genome and those of the AOS strains, namely A. japonica, A. mediterranei and A. orientalis, revealed a broad spectrum of genomic structures. In addition, when the phylogenetic structure of the genus Amycolatopsis was revisited based on in silico 16S rDNA and orthologues of the actinobacterial class-specific conserved gene (AMETH_3452), a third phyletic branch, the Amycolatopsis taiwanensis subclade (ATS) and some yet-to-be clearly delineated minor groups were highlighted.

2. Results and discussion

2.1. The complete genome of A. methanolica strain 239 provides a high quality alternative genetic blueprint for the genus Amycolatopsis

The A. methanolica chromosome is circular (Fig. 1A) and can thereby be distinguished from the linear chromosomes of Streptomyces species, but not from those of representatives of the family Pseudonocardiaceae (Actinosynnema mirum, A. japonica, A. mediterranei, A. orientalis, Pseudonocardia dioxanivorans Saccharomonospora viridis and Saccharopolyspora erythraea) as shown in Table 1. The genome of A. methanolica strain 239 (CP009110) consists of 7,237,391 bp and is much smaller than those of the A. japonica (8.9 Mbp), A. mediterranei (10.2 Mb) and A. orientalis (8.9 Mbp) strains; its G − C content is 71.53 mol% which is comparable to experimentally determined values (Fig. 1A) and is in the same range as those of other Amycolatopsis strains [32–34] (Table 1).

The A. methanolica genome contains three rRNA operons as opposed to four found in the A. japonica, A. mediterranei and A. orientalis genomes. The positions of all of the copies of the 16S rRNA gene in the genomes of these strains, together with those in the genomes of S. viridis DSM 43017 and S. erythraea NRRL 2338, representative species of the two most closely related genera, are listed in Table S1 (Supplementary Table 1). All three rRNA operons found in the A. methanolica genome share common genomic positions with those of the S. viridis strain with respect to the consistency of their neighbouring genes, though only two of them are in the same positions as those in the A. japonica, A. mediterranei, A. orientalis and S. erythraea genomes. The genome of A. methanolica strain 239 has 51 tRNA genes including a tRNA^Sec encoding selenocysteine. This tRNA^Sec gene is located immediately downstream of selBA and upstream of selD. It is transcribed in the same DNA strand of selBA while in the opposite strand of selD. We also found a gene in the A. methanolica genome, namely AMETH_3396, that encodes a formate dehydrogenase α subunit equipped with a Sec-encoding UGA codon, a Sec insertion sequence (SEcis) element and a stem-loop structure required for the incorporation of Sec into proteins [33]. This gene is similar to that found in the genome of A. mediterranei (Table S2).

Similar to the two p MEA100-like integrated plasmids found in the A. mediterranei genome, one integrated p MEA300-like plasmid with 13,285 bp was found at the 10 Kb position of the A. methanolica genome (Fig. 1A). As previously reported [35,36], this p MEA300-like plasmid might be a conjugative element present mostly in an integrated state at a single site in the chromosome, though it can also replicate autonomously. The attB site for the integration of the plasmid was located in the f 10-r tRNA region. This type of integrated plasmid is not found in the A. orientalis genome though the latter does have a free plasmid, pXL100, which has not been detected in any of the other sequenced Amycolatopsis strains. Interestingly, a predicted prophage with 37,822 bp encoding 52 ORFs (AMETH_6251 to AMETH_6302), most of which were annotated as hypothetical proteins, is present in the A. methanolica 239 genome as a tandem repeat positioned at the genome coordinate around 6.5 Mbp (Fig. 1A). Prophages have not been found in the genomes of A. mediterranei, A. orientalis or A. japonica.

By employing the three criteria defined in the study of the A. mediterranei genome [33], genome-wide comparisons were made against the complete genome sequences of several Pseudonocardiaceae strains, namely A. mirum DSM 43827, A. orientalis HCCB10007, A. mediterranei U32, P. dioxanivorans CB1190, S. viridis DSM 43017 and S. erythraea NRRL 2338 [37]. These results revealed a highly conserved core region ranging from 0 to 2.1 Mb
and 4.8 e 7.2 Mb (right arm) in the A. methanolica chromosome (total of ~4.5 Mb, Figs. 1A, 2A, and 3). Although ortholog ordering of quite a few consecutive genomic segments in the non-core region of A. methanolica are apparently co-linear in the quasi-core or core regions of the A. mediterranei and A. orientalis genomes (Fig. 3A and B), there is no obvious “quasi-core” region identified within the non-core region ranging from 2.1 to 4.8 Mb (total of ~2.7 Mb) of its genome. This situation is quite different from that found in the genomes of A. mediterranei, A. orientalis and S. viridis (Fig. 2). On the other hand, this co-linearity analysis shows that A. methanolica strain 239T and S. viridis DSM 43017T share the highest consistency throughout the whole genome, even more so than against the A. mediterranei and A. orientalis genomes. This phenomenon may simply reflect the smaller genome size of S. viridis (~4.2 Mb) compared to the genomes of the two other Amycolatopsis strains, especially in the non-core region (~1 Mb in S. viridis) (Fig. 3C). In contrast, extremely low co-linearity was found between the genomes of the A. methanolica and S. erythraea strains (Fig. 3D). Consequently, by increasing the number of genomes of various sizes from different phyletic lineages within the genus Amycolatopsis, the genome configuration aspect of “core” may be considered as the large consecutive genomic segment extending from oriC in both directions, a segment that is relatively stable and hence has fewer inserted sequences compared with the “non-core” region. The concept of “quasi-core” may be seen as part of the core near the replication terminus in ancestral species, but more accessible to multiple horizontal gene transfer (HGT) and chromosomal rearrangement events during evolution. Further, the boundaries of core and quasi-core regions against the non-core region are “relative”, i.e., they are more or less artificial and apparently variable depending on the pair of genomes being compared.
The replication origin \((oriC)\) of the *A. methanolica* chromosome was established based on GC skew [38], and its adjacent dnaA gene was chosen as the starting point for numbering the 7074 predicted protein-coding sequences (CDs) (Fig. 1A). Among all of the CDSS, 68.28% of them were assigned to a functional category in the Cluster of Orthologous Groups (COGs, Table S3). The orthologues between the *A. methanolica* genome and those of *A. japonica*, *A. mediterranei* and *A. orientalis* are 4107 (49.5% of all proteins of *A. japonica*), 4158 (51.4% of all proteins of *A. mediterranei*) and 4168 (45.1% of all proteins of *A. orientalis*), respectively. The ratios of genes encoded for amino acid transport and metabolism, as well as for energy production and conversion in the *A. methanolica* genome are higher than those found in the closely related genomes, a conserved gene (Fig. 2), especially in the families Pseudonocardiaceae, *A. japonica*, *A. mediterranei* and *A. orientalis* are 4107 (49.5% of all proteins of *A. japonica*), 4158 (51.4% of all proteins of *A. mediterranei*) and 4168 (45.1% of all proteins of *A. orientalis*), respectively. The ratios of genes encoded for amino acid transport and metabolism, as well as for energy production and conversion in the *A. methanolica* genome are higher than those found in the closely related genomes, a conserved gene (Fig. 2), especially in the families Pseudonocardiaceae, *A. japonica*, *A. mediterranei* and *A. orientalis*.

In addition to a comparison of chromosomal configurations of the closely related genomes, a conserved gene (AMETH_3452) spanning nucleotides 3,578,958 to 3,579,767 in the *A. methanolica* genome was found to encode a hypothetical protein of 269 amino acid residues (Fig. 2A, Fig. S1). It is significant that this gene is orthologous among all of the genomes belonging to the family Pseudonocardiaceae (Fig. S1, e.g., AMED_4666, AORI_3898, AJAP_19695, SVIR_19880, SACE_3783, AMIR_3419, and PSED_3357). When all of the sequenced genomes of bacteria were analyzed, genes orthologous to AMETH_3452 were identified in almost all of the sequences from species classified in the class Actinobacteria but not in those of other classes of prokaryotes (ref to Fig. S1). Furthermore, it is particularly interesting to note that most of these sequences are located in the middle of the circular chromosomes opposite to oriC (Fig. 2), especially in the families Corynebacteriaceae, Micrococccaceae, Mycobacteriaceae and Pseudonocardiaceae, regardless of significant variations in the genome sizes of their constituent species (Fig. 2B, Table S4). This hypothetical protein is predicted to be soluble and cytoplasmic based on comparison of sequence similarity and secondary structural content (Fig. S1). Although the function of this CDS is unknown, it turned out to be a precise molecular marker for intra-generic species delineation of members of the family Pseudonocardiae as shown in the next section.

A draft *A. methanolica* genome (AQLU00000000.1) consisting of two contigs, 6,804,661 bp and 392,199 bp each, was released in 2013. We have determined the differences between the sequences of the two versions (AQLU00000000.1 and CP009110) (Fig. S2) and found that there are 47 single nucleotide polymorphisms (SNPs), 30 indels and a large insertion of 37,822 bp located in the genomic coordinate around 6.5 Mb in genome CP009110, which is actually the prophage mentioned above but shown as a single copy rather than a tandem repeat. In addition, contig 1 in AQLU00000000.1 shows inversions of two fragments compared to that of genome CP009110.

### 2.2. Amycolatopsis species may be classified into three major subclades and some minor groups based on phylogenetic analyses of 16S rDNA sequences, growth physiology and genomic characteristics

*Amycolatopsis* species can be assigned to several multi-membered and single-membered subclades/groups based on 16S rDNA sequences, as well as on gyrB and recN sequences, two highly conserved eubacterial orthologous genes [11,16,17]. Most of the species fall into the *A. methanolica* (AMS) and the *A. orientalis* (AOS) subclades, the members of which share similar taxonomic properties though they have been reported to have different temperature profiles [10,23,24].

We generated an *Amycolatopsis* phylogenetic tree based on 16S rDNA genes (data from EzTaxon [39]) of the type strains of 62 *Amycolatopsis* species using *S. viridis* DSM 43017\(^T\) as the outgroup strain (Fig. 4A). The initial AOS subclade has grown since its inception and now comprises 36 mesophilic species and 13 moderately thermophilic species which can grow at 45 °C. This subclade is generally stable and apparently accommodates groups A-G [16,17] and groups H-I and J from this study (Fig. 4A). The remaining taxon, the F group, can be divided readily into two well delineated taxa, phygetic group F1 composed of *Amycolatopsis helvola* TT-99-32\(^T\) [40], *Amycolatopsis pigmentata* TTO-43\(^T\) [40], and *A. taiwanensis* O345M-7\(^T\) [41] and phygetic group F2, composed of *A. methanolica* and related species (Fig. 4A). It can be seen from Fig. 4B that in contrast to group F1, only members of the group F2 grow at 45 °C and in the presence of 5%, w/v NaCl, these properties still need to be acquired for *Amycolatopsis thermophilus* GY088\(^T\) (see Fig. 4B). Consequently, another independent sub-generic lineage, the *A. taiwanensis* subclade (ATS/F1), is proposed to cover these three species thereby emphasizing their phylogenetnic and physiological differences from members of the AMS/F2 subclade, which is represented by *A. methanolica* [10] and comprises 10 thermophilic species capable of growing at temperatures above 50 °C except for *Amycolatopsis endophytica* KLBMP 122\(^T\) [42], which grows at 45 °C.

It is well known that compatible solutes can stabilize biological membranes and protect cells and cell components from freezing.
Consequently, genes related to the synthesis of several compatible solutes, such as ectoine/hydroxyectoine, glycine, betaine and trehalose were sought in the AMS genomes of A. methanolica, Amycolatopsis sp. ATCC 39116 (AFWY00000000) and Amycolatopsis thermodiluviana N1165 (NZ_AXBH00000000) versus the ATS genome.
A. taiwanensis DSM 45107T NZ_JAFB00000000 (Table S5). It is significant that the OtsA-OtsB, TreY-TreZ, TreS and TreP pathways for trehalose biosynthesis were found in the three AMS genomes while the ATS genome has only the OtsA-OtsB and TreP pathways suggesting that the absence of TreY-TreZ and TreS pathways may contribute to the reduction of trehalose accumulation in ATS strains. In addition, several transporters encoded by \textit{ssuB} (AMETH_2747), \textit{proP} (AMETH_5526, AMETH_6721) and \textit{betP} (AMETH_3300, AMETH_4479) and genes encoding five betaine aldehyde dehydrogenases (AMETH_0085, AMETH_2645, AMETH_5372, AMETH_5920, AMETH_5579) were missing in the ATS strains thereby accounting for their salt and heat intolerance.

The whole genome sequences of \textit{A. japonica} strain MG417-CF17, \textit{A. mediterranei} strain U32, \textit{A. methanolica} strain 239T and \textit{A. orientalis} strain HCCB10007 have multiple copies of 16S rRNA genes with various levels of identity between selected pairs of rDNAs. Specifically, the sequences of the three copies of 16S rDNAs in the \textit{A. methanolica} genome are identical but distinct from those in the genomes of the other strains. The sequences of the 16S rRNA genes of the \textit{A. japonica}, \textit{A. mediterranei} and \textit{A. orientalis} strains share identities ranging from 98.61 to 100% compared to other 16S rDNAs in the same genome (Fig. 5A). In particular, the sequence identity for some inter-species pairs of 16S rDNA between the \textit{A. orientalis} and \textit{A. japonica} strains are higher than those of the corresponding intra-species pairs. Specifically, HCCB10007_R022 is much closer to MG417-CF17_AJAP r14255 (99.52%) and MG417-
CF17_AJAP r30385 (99.52%) than to HCCB10007_R016 (98.91%); while MG417-CF17_AJAP r33495 is much closer to HCCB10007_R045 (99.66%) and HCCB10007_R041 (99.66%) than to MG417-CF17_AJAP r10595 (99.2%) (Fig. 5A). This variation needs to be addressed as it may influence the structure of Amycolatopsis 16S rRNA gene trees. We conducted a phylogenetic analysis using all 15 copies of the 16S
rDNAs encoded in the genomes of the *A. japonica*, *A. methanolica*, *A. mediterranei* and *A. orientalis* strains (Fig. 5B) and found that certain copies of the 16S rDNAs from the genomes of the *A. japonica* and *A. orientalis* strains were cross-clustered while those of the *A. mediterranei* and *A. methanolica* strains clearly clustered independently. This result is not surprising and can be attributed to the different levels of sequence similarities between pairs of the 16S rDNAs from the different species versus sequence diversities between pairs of 16S rDNAs within a species (Fig. 5A).

We also analyzed the phylogeny of the genus *Amycolatopsis* using the fifteen 16S rDNA sequences of the four *Amycolatopsis* species and 16S rRNA gene sequences from another 55 *Amycolatopsis* species as shown in Fig. S3. It can be seen that while different copies of the 16S rDNAs selected from the *A. japonica* and *A. orientalis* genomes affect their positions in the phylogenetic tree they had no effect on the assignment of strains to the three *Amycolatopsis* subclades. This result provides further evidence that phylogenetic relationships between strains at the sub-generic level can be influenced by the presence of multiple copies of diverse 16S rRNA genes present in the constituent species [53–55].

The latest September 2015 version of the 16S rDNA based All-Species Living Tree accommodated the 62 *Amycolatopsis* species.
mentioned above and divided them into two lineages separated by some species of *Prauserella* and *Saccharomonospora* (http://www.arb-silva.de/fileadmin/silva_databases/living_tree/LTP_release_123/LTPsi23_SSU_tree.pdf). The smaller lineage encompassed 13 species, including *A. methanolica* and *A. taiwanensis* i.e., the representatives of the AMS and ATS subclades, respectively. The larger phylectic line contained 52 species, including *A. orientalis*, the representative of the AOS subclade. This tree is consistent with our division of the genus into three major subclades (Fig. 4A) though the intrusion of the *Prauserella* and *Saccharomonospora* species needs an explanation.

The All-Species Living Tree [56] based on the Silva database [57] was designed to include all of the sequenced type strains of hitherto published species of *Archaea* and *Bacteria* based on their 16S rRNA gene sequences. To encompass such a broad range of bacterial species into a single tree to capture the higher taxonomic groups and clusters, different sequence filters were used to exclude some highly variable regions of the aligned 16S rRNA gene sequences [56]. However, exclusion of too many variable regions may reduce the phylogenetic resolution at lower taxonomic ranks, such as at generic and species levels. In order to construct a more reliable phylogenetic tree of the targeted bacterial group(s) instead of considering all of the taxa in the All-Species Living Tree, we performed a phylogenetic analysis focused on the relevant groups (the three genera in this case) using complete 16S rRNA gene sequences without using the sequence filters from the Silva database. This approach enabled us to retain all of the informative aligned sequences needed to construct a better phylogentic tree of the targeted bacterial taxa (*Amycolatopsis*, *Prauserella* and *Saccharomonospora* in this case) as shown in Fig. S4.

Currently, complete and draft genomes of 18 *Amycolatopsis* strains have been deposited in the NCBI database. We generated phylogenetic trees based on these strains using either gyrB or recN genes alone or in combination with 16S rDNA genes and used corresponding data from species representing four other genera classified in the family *Pseudonocardia*ae as the outgroup, an approach that has rarely been applied [16,17]. The results were completely unexpected, *Amycolatopsis halophila* YIM 93222T did not cluster with the other *Amycolatopsis* species but with *Saccharomonospora* and *Saccharopolyspora* species (Fig. S5). The same problem was encountered when the phylogenetic tree was constructed based on the tail-head linked 95 single-copy orthologous genes encoding conserved functional proteins annotated from a total of 123 actinobacterial genomes (Table S6, Fig. S6). However, this approach worked perfectly well when restricted to the three representative *Amycolatopsis* species and representative species of five genera belonging to the family *Pseudonocardiaeae* (Fig. S7). We speculate that this phenomenon is due to the gyrB and recN sequences not being sufficiently variable to delineate between species of closely related genera, notably genera belonging to the same family. Consequently, phylogenetic trees based on the newly identified, conserved class-specific AMETH_3452 orthologues (Fig. 2, Fig. S1, Table S4), with or without the corresponding 16S rDNA sequences from the complete or draft genomes of the 18 *Amycolatopsis* species were constructed and compared to the corresponding 16S rDNA gene tree (Fig. 6). In all three trees, the species representing the five genera classified in the family *Pseudonocardiaeae* clustered as expected. It was particularly interesting that all but one of the 18 *Amycolatopsis* species were segregated into AOS (13 species, covering groups A, B, C, D and E), AMS (3 species of group F2) and ATS (1 species of group F1) lineages, the exception was *A. halophila* (Fig. 4A).

It is particularly interesting that the type strain of *A. halophila* formed a distinct phylectic line that was more closely related to the *A. methanolica* and *A. taiwanensis* strains than to representatives of the *A. orientalis* subclade (Fig. 6). However, this result is not so surprising as the *A. halophila* and *Amycolatopsis salitolerans* strains formed a distinct group (designated as J), supported by a 100% bootstrap value, towards the periphery of the *A. orientalis* 16S rRNA subclade and hence is quite close to the group F strains (Fig. 4). The *A. halophila* and *A. salitolerans* strains also share distinctive physiological features as they are thermophilic and grow at high salt concentrations [58,59]. The *A. halophila* strain has a particularly small genome (5.55 Mb) with very few gene clusters annotated that are likely to encode for specialized metabolites. These genomic characteristics show a greater similarity to the genomes of F group species than to AOS species, the former have smaller genomes (8.78–7.24 Mbps against 10.86–8.53 Mbps) and fewer gene clusters (16–12 against 55–17) than the latter (Table S7 and Fig. 6). These data clearly show that the *A. halophila* and *A. salitolerans* strains belong to a subclade that lies closer to the F group than to the AOS subclade (Fig. S5 and Fig. S6).

The systematic analyses illustrated above underpin the long-standing view that 16S rDNA-based phylogenetic analysis is reliable at and above the generic level(s) but this is not always so at the species level where relationships may not be in sync with corresponding phenotypic data. By the same token, gyrB and recN sequence-based phylogenies [16,17] failed to distinguish between the species from clusters related taxa among the genera of the family *Pseudonocardiaeae* considered here (Fig. S5). On the other hand, the AMETH_3452-based tree in this study provides an alternative perspective by employing a class-specific orthologues gene that encodes for a hypothetical CDS, in this instance good correlation was found between the phylogenetic and corresponding physiologic data (Fig. 6).

2.3. Whole genome sequence of *A. methanolica* strain 2396 reveals the genetic basis for its special primary and specialized metabolic characteristics

Although *A. methanolica* strain 2396 has been reported to produce few, if any, specialized metabolites [60], a total of 12 gene clusters putatively encoding for such metabolites were identified in its genome (Fig. 1B) using antiSMASH3 [61] and BAGEL3 [62] software. It is interesting that all 12 gene clusters are highly conserved among the three AMS species with known complete or draft genome sequences (*A. methanolica* strain 2396, *Amycolatopsis* sp. ATCC 39116 and *Amycolatopsis thermoflava* strain N1165T); while, except for *oth* and *ectABC* (see later), neither the gene clusters *per se* nor their chromosomal loci are conserved in the three completely sequenced and well annotated representative genomes of AOS species [32–34]. These findings can be attributed to the close phylectic relationships found between the AMS species, while further analysis shows that species in the AOS generally have a greater potential for the synthesis of specialized metabolites than corresponding data drawn from the ATS and AMS strains (Table S7). In particular, few polyketyde synthases (PKSs) and/or non-ribosomal peptide synthetases (NRPSs) were identified in the three AMS genomes (1 NRPS in *A. methanolica* strain 2396, 1 PKS and 1 NRPS in *Amycolatopsis* sp. ATCC 39116, 3 PKSs and 1 NRPS in *A. thermoflava* strain N1165T), which is significantly different from corresponding data in the AOS genomes.

Among the 12 potential specialized metabolites, two classes of siderophore biosynthetic gene clusters were identified. The NIS (NRPS Independent Siderophore) gene cluster encodes the key enzyme NIS synthase, which is derived from esterified or amidated derivatives of carboxylic acid and belongs to group type C [63] based on phylogenetic analysis. The only NRPS cluster in the genome of *A. methanolica* strain 2396 (AMys. AMETH_0591 - AMETH_0609) with 39,563 bp shows 96.5% and 97.0% identity with a similar cluster identified in *Amycolatopsis* sp. ATCC 39116 and
Fig. 6. Neighbour-joining trees based on AMETH_3452 orthologues with (A) or without (B) the corresponding 16S rDNA sequences, and the 16S rDNA sequences alone (C) from the complete or draft genomes of the 18 Amycolatopsis type strains. The AOS/ATS/AMS classification and the A-F and J groupings are shown. Half square brackets in red denote the AMS branches, the green color the ATS branch and the blue color the AOS branches. Numbers at the nodes are bootstrap values based on 1000 replicates. The scale bar indicates 0.02, 0.05 or 0.005 nucleotide substitutions per site.
A. thermofoflava strain N1165\textsuperscript{7}, respectively. The genes encoded in this cluster are complete and the NRPS products have recently been identified and characterized as amymcheinlin [60].

Three ribosomally synthesized and post-translational modified peptides (RiPPs) containing Lasso, Lan1 and Lan2 have low homology to known sequences indicating that their products are likely to be novel. Terp2 and Terp3 were recognized as gene clusters responsible for isorenieratene, heneicosane biosynthesis, respectively based on antiSMASH data; Terp2 probably accounts for the yellow colonies of A. methanolica strain 239\textsuperscript{3}, while Terp1 and Terp4 may be new types of terpenes. The bacteriocin biosynthetic gene cluster in A. methanolica strain 239\textsuperscript{3} seems to be conserved in Amycolatopsis sp. ATCC 39116 [64], Amycolatopsis sp. AA4 (NZ-A-CVE00000000), Amycolatopsis sp. MJM2582 (JPLW00000000) [65] and in the type strains of Amycolatopsis azurea DSM 43854\textsuperscript{2} (NZ_ANNMG00000000), Amycolatopsis decaplanina DSM 44594\textsuperscript{2} (NZ_AOH00000000) [66], Amycolatopsis lurida NRRL 2430\textsuperscript{T} (CP007219) [67], Amycolatopsis japonica MG417-CF17\textsuperscript{T} and Amycolatopsis nigriscens CSC17a-90\textsuperscript{N} (NZ_ARVW00000000).

The ectABC gene cluster involved in the synthesis of ectoine has been isolated and characterized for a number of strains belonging to the Actinobacteria, Firmicutes and Proteobacteria. The ectoine dehydrogenase gene (ectD) in A. methanolica strain XMU15T (CM001439.1) is responsible for the expression of ectoine to hydroxyectoine; both ectoine and hydroxyectoine are involved in thermoprotection [68,69]. The ectoine gene cluster has been detected in the genomes of many Amycolatopsis species, including A. taiwanensis DSM 45107\textsuperscript{T} (ATS), A. japonica (AJAP_02100_409 - AJAP_02085_406), A. mediterranei (AMED_8594 - AMED_8597) and A. orientalis (AORI_7380 - AORI_7383) (AOS). For the AOS species, the protein EctABC of A. methanolica strain 239\textsuperscript{3} has 99% similarity to that of the other two AOS species, i.e., A. thermofoflava strain N1165\textsuperscript{7} and Amycolatopsis sp. ATCC 39116; associated CDSs are also highly conserved in these strains (Fig. 8F).

One carbon metabolism was recognized as a key characteristic of A. methanolica strain 239\textsuperscript{T}. However, KEGG mapping of the whole genome sequence of this strain showed that it lacked the gene encoding for the well-known methanol dehydrogenase (e.g., cytochrome c-dependent or NAD-dependent) used to oxidize methanol [70,71]. Instead, an mdo gene (AMETH_5577) was identified in A. methanolica strain 239\textsuperscript{3} as previously described [72,73] as with the well-known methanol dehydrogenase (FDFAIDH/MscR, EC: 1.2.1.46) [74]. Though a glutathione-independent formaldehyde dehydrogenase (GDFAIDH, EC: 1.2.1.46) is encoded by AMETH_1319 in the genome of A. methanolica strain 239\textsuperscript{3}, its activity in this strain is doubtful [74,76] as glutathione has not been detected in actinobacteria [77,78]. In this study, we confirmed the proposition that glutathione is unlikely to be synthesized in A. methanolica strain 239\textsuperscript{3} due to the lack of the gene gshB which encodes one of the two key enzymes, i.e., glutathione synthetase [79].

Here we propose that A. methanolica strain 239\textsuperscript{T} assimilates formaldehyde via the ribulose monophosphate (RuMP) cycle rather than by the serine pathway [76], as indicated by KEGG mapping of its whole genome sequence. In this pathway, formaldehyde is fixed with ribulose 5-phosphate (RuSP) to form D-ribulose-5-phosphate (Hulp) catalyzed by 3-hexulose-6-phosphate synthase (HPS), and is then isomerized to fructose 6-phosphate (F6P) catalyzed by 3-hexulose-6-phosphate isomerase (HPI) (Fig. 7A). The A. methanolica genome has one gene (AMETH_4517) encoding HPI and two identical genes (AMETH_4519, AMETH_4538) encoding HPS (Fig. 7B), indicating an efficient RuMP pathway for F6P supply.

The genome of A. mediterranei strain U32 does not encode any of the genes for the above mentioned key enzymes for C1 utilization (Mdo, FDFAIDH/MscR, HPS or HPI), while the genomes of A. japonica strain MG417-CF17\textsuperscript{T} and A. orientalis strain HCCB10007 only encode genes for FDFAIDH/MscR. The genomes of other AMS strains, namely A. thermofoflava N1165\textsuperscript{7} and Amycolatopsis sp. ATCC 39116 [64] contain two key genes (mdo and mscR) for methanol oxidation, but neither hps nor hpi genes have been found in their partial genome sequences (Fig. 7B). Therefore, it seems unlikely that other members of the thermophilic AMS will be methylophilic, mainly due to the absence of the RuMP cycle for assimilation of C1 intermediates. In contrast with AT5 where none of the four genes have been detected in the genome of its representative species, namely A. taiwanensis, the mscR gene has been detected in the genomes of a few species of AOS, but not for A. mediterranei as mentioned above (Fig. 7B). It is particularly interesting that all four genes are found in the partially sequenced genome of Amycolatopsis benzoaotylitica AK 16/65\textsuperscript{T} (ARPK0000000001.0), a member of the AOS subclade (Fig. 7B), a result which suggests it may be a facultative methylophilic.

As a facultative methylotrophic, genomic annotation confirmed that A. methanolica strain 239\textsuperscript{3} is able to generate F6P from carbon sources other than C1 substrates via the RuMP cycle, such as glucose via glycolysis and the KDPG (Etnner-Doudoroff pathway) and TA (pentose phosphate) pathways (Fig. 7A). For the downstream assimilation reactions that occur in the TCA cycle, it is significant that the A. methanolica strain lacks 2-ketogulutarate dehydrogenase, but as in other Amycolatopsis species has 2-oxoglutarate synthase encoded by korA (AMETH_0502) and korB (AMETH_0501) (Fig. 7A) instead.

The hps and hpi genes of A. methanolica strain 239\textsuperscript{T} are located in a cluster that extends from AMETH_4515 to AMETH_4540 and which encodes a few genes for the TA and glycolysis pathways. An analysis of the phylogeny of the protein sequences of HPS and HPI orthologs when present in the same strain shows that while these genes have been evolving more or less concomitantly (Fig. 8A and B), similar clusters are absent in most of the genomes, as exemplified by actinomycetes such as Arthrobacter aurescens strain TC1 (CP004474.1) [80] and Rhodococcus jostii strain RHA1 (CP000431.1) [81,82] and by the firmicute Bacillus subtilis strain BSN5 (CP002468.1) [83] (Fig. 8C). However, in the incompletely sequenced genomes of two representatives of the family Pseudonocardiaeae, namely A. benzoaotylitica strain AK 16/65\textsuperscript{S} (ARPK000000001.1) and Saccharamonaspora marina strain XMU15\textsuperscript{T} (CM001439.1), we identified the conserved clusters albeit with evidence of considerable rearrangements (Fig. 8C). In addition, although neither the genes of hps nor hpi nor this hps-hpi related cluster are found in the three complete genomes of the AOS strains, orthologous genes, gene operons and gene clusters encoding RuMP-related carbon metabolic pathways for F6P generation are well conserved (Table 2). However, as previously noted [84], sequence divergence (35.4% identity) and conditional expression differences between the A. methanolica ATP-dependent 6-phosphofructokinase (ATP-PFK, AMETH_4515) versus its isotopic PPI-dependent 6-phosphofructokinase (PPI-PFK, AMETH_4897) infers an HGT event for the acquisition of the ATP-PFK. Meanwhile, genes of this RuMP cluster encoding fructose 1,6-bisphosphatase II (glpX, AMETH_4516 and AMETH_4530) and the fructose-bisphosphate aldolase (ALDO, AMETH_4531) are also specific for A. methanolica strain 239\textsuperscript{T}, i.e., <40% identity compared to the corresponding orthologs of all of the four completely sequenced Amycolatopsis strains (data not shown), thereby supporting the case for HGT.

On the other hand, in the same RuMP cluster, certain genes, such
as those encoding transketolase (tkt AMETH_4518), ribulose-phosphate 3-epimerase (rpe AMETH_4520) and ribose 5-phosphate isomerase (rpi AMETH_4529), have corresponding iso-enzymes with similar functions in other parts of the genomes of both A. methanolica strain 239 and the three AOS strains with high similarities (>60% identity, Table S9). These genes are highly similar to corresponding orthologs in other strains belonging to the class Actinobacteria (Fig. 8C) and hence are likely to be specific to this taxon and integrated into the hps-hpi RuMP cluster in ancestral strain(s) of the family Pseudonocardiaceae.

Utilization of methanol, formaldehyde and formic acid as sole one-carbon substrates by A. methanolica strain 239 was retested after whole genome annotation and KEGG mapping (see Materials and Methods). After incubation for 14 days, the bacterium grew well in liquid synthetic media (HMM) supplemented with 0.1%, 0.2% and 0.4% methanol as sole carbon sources. In contrast, the strain failed to grow in the same basal media supplemented with the same concentrations of either formaldehyde or formic acid as sole carbon sources (Fig. S9, S10). Consequently, we believe that A. methanolica strain 239 might not be able to tolerate the toxicity caused by environmental formaldehyde or formic acid even though they are intermediates of methanol metabolism. To test this hypothesis, the non-growing cultures were re-spread on GT plates with starch as the carbon source followed by incubation at 28 °C for 3 days. Two non-growing cultures, i.e., A. methanolica in HMM media without carbon source additives or with formic acid as the sole carbon source, grew well on the GT plate; while the formaldehyde culture failed to grow (Table S8). These findings indicate that A. methanolica stain 239 is unlikely to tolerate the toxic effect of formaldehyde while its ability to metabolize formic acid may be
too weak to support healthy growth.

3. Conclusions and perspectives

*Amycolatopsis methanolica* strain 239T (CP009110) is the first member of the AMS species to be completely sequenced thereby providing a high quality alternative genetic blueprint for the taxonomically diverse genus *Amycolatopsis*. Comparative and evolutionary genomic analyses based on this study allowed several major conclusions to be drawn:

1. The genomes of *Amycolatopsis* species are circular ranging in size from 7 Mbp to 10 Mbp (ref to Table S7). In addition to a highly conserved and consistent core region with few inserted sequences extending from oriC in both directions of the *A. methanolica* genome, certain consecutive segments of the non-core region are generally co-linear with those from other species of the genus. The AMETH_3452 orthologous genes identified in all of the sequenced genomes of actinobacteria were mainly located in the middle of the circular chromosomes opposite to the oriC (Fig. 2); while the so-called “quasi-core” regions are proposed to be part of the core in ancestral actinobacterial species near the replication termini (refer to Figs. 2 and 3). These regions are likely to be more accessible to multiple HGT and chromosomal rearrangement events and to adaptation under different environmental conditions thereby ensuring that a few segments of the core genome are relatively stable.

2. In light of the systematic search for molecular markers capable of delineating between the 62 *Amycolatopsis* species and representatives of 5 other genera classified in the family *Pseudonocardaceae*, three major *Amycolatopsis* subclades (AMS, AOS and ATS) and some minor taxa, such as the J group, were
Table 2
Comparison of genes involved in the pentose phosphate pathway and glycolysis/gluconeogenesis in Amycolatopsis genomes.

| Pathway/Genome | A. methanolica (in RuMP cluster) | A. methanolica (not in RuMP cluster) | A. Mediterranei | A. orientalis | A. Japonica |
|----------------|---------------------------------|--------------------------------------|----------------|--------------|-------------|
| Pentose phosphate pathway | pfk AMETH_4515 (ATP-pfk) | AMETH_4897 (PPi-pfk) | AMED_2460 | AORI_2423 | AJAP_26710 |
| Glycolysis/gluconeogenesis | AMETH_4516 | AMETH_4530 | AMED_6811 | AORI_1928 | AJAP_06835 |
| | AMETH_2243, AMETH_2246 | AMETH_2245, AMETH_2246 | AMETH_4529 | AMETH_4531 | |
| | AMETH_5084 | AMETH_50B5 | AMER_2263 | AORI_5659 | AJAP_11180 |
| | AMETH_2262 | AMETH_2263 | AMER_2809 | AORI_2796 | AJAP_24810 |
| | AMETH_4518 | AMETH_4092 | AMER_2771 | AORI_2761 | AJAP_25039 |
| | AMETH_4520 | AMETH_4130 | AMER_6811 | AORI_6417 | AJAP_25985 |
| | AMETH_4259 | AMETH_1699 | AMER_7455 | AORI_7455 | AJAP_38990 |
| | AMETH_11180 | AMETH_11180 | AMER_24820 | AORI_2793 | AJAP_24825 |
| | AMETH_11185 | AMETH_11185 | AMER_24835 | AORI_2791 | AJAP_24835 |
| | AMETH_0006 | AMETH_0003 | AMER_0003 | AORI_0003 | AJAP_00020 |
| | AMETH_4093 | AMETH_2808 | AMER_2795 | AORI_2795 | AJAP_24815 |
| | AMETH_4094 | AMETH_2807 | AMER_2794 | AORI_2794 | AJAP_24820 |
| | AMETH_4095 | AMETH_2806 | AMER_2793 | AORI_2793 | AJAP_24825 |
| | AMETH_4097 | AMETH_2804 | AMER_2791 | AORI_2791 | AJAP_24835 |
| | AMETH_5852 | AMETH_6059 | AMER_6881 | AORI_6881 | AJAP_04695 |
| | AMETH_5853 | AMETH_6058 | AMER_6881 | AORI_6881 | AJAP_04695 |
| | AMETH_5854 | AMETH_6121 | AMER_6121 | AORI_5185 | AJAP_13540 |
| | AMETH_6122 | AMER_6122 | AORI_5185 | AORI_5185 | AJAP_13540 |

Each line is the corresponding ortholog.

proposed based on phylogenetic analyses of 16S rDNA and AMETH_3452 orthologues (against the 18 Amycolatopsis species with available genomic sequences) as well as on their genome characteristics and growth physiology (Figs. 4 and 6). The constituent species of the subclades share distinctive temperature profiles and some genomic features. Thus, the 10 AMS species are thermophilic with smaller genome sizes, as are the 2 species of group J; the 3 ATS species are mesophilic, while the AOS species encompass 36 mesophilic and 11 moderately thermophilic strains, many of which have a large number of gene clusters that may produce various kinds of specialized metabolites (Fig. 6, Table S7). The separation of the F group into gene clusters that may produce various kinds of specialized metabolites (Fig. 6, Table S7). The separation of the F group into gene clusters that may produce various kinds of specialized metabolites (Fig. 6, Table S7). The separation of the F group into gene clusters that may produce various kinds of specialized metabolites (Fig. 6, Table S7). The separation of the F group into gene clusters that may produce various kinds of specialized metabolites (Fig. 6, Table S7). The separation of the F group into gene clusters that may produce various kinds of specialized metabolites (Fig. 6, Table S7).

3. This genomic study revealed the complete pathways for the utilization of methanol as a sole carbon source in A. methanolica strain 239T including both mdo and mscr encoded methanol oxidation and hps and hpi encoded formaldehyde assimilation via the RuMP cycle (Fig. 7). Although phenotypic verification is still required, it seems likely that A. benzoatilytica AK 16/65T, a member of the AOS subclade, may prove to be a facultative methylothroph as all four of the essential genes required for methanol utilization are encoded in its partially sequenced genome (ARP000000001) (Fig. 7B). However, methylothrophy is unlikely to be a common trait within the genus Amycolatopsis as none of the other available Amycolatopsis genomes have the complete set of the four genes required for methanol utilization. Indeed, the RuMP cluster including hps and hpi defined in the genomes of the A. methanolica, A. benzoatilytica and S. marina strains seem to have been acquired by HGT, an event that probably occurred in ancestral strain(s) of the family Pseudonocardiaceae (Figs. 7 and 8).

4. It was also shown that A. methanolica strain 239T is potentially able to produce 12 different kinds of specialized metabolites though not polyketides (Fig. 1B). These gene clusters are highly conserved among the three AMS genomes; while, except for oth and ectABC, none of them are found in the three completely sequenced and well annotated representative genomes of the AOS species.

In summary, the complete sequencing, de novo assembly and comparative annotation of the A. methanolica genome greatly improves our systematic understanding of the genus Amycolatopsis with respect to the spectrum of genomic structures and configurations, genus-focused phylogeny and sub-generic classification, as well as to its primary and specialized metabolism. It is, once again, evident that systems biology guided genomic analyses based on the complete genetic information of strains representing diverse prokaryotic species is the key to revitalizing and even to revolutionizing the taxonomy of microorganisms in the genomic era [85].

4. Materials and methods

Genome sequencing and assembly: Amycolatopsis methanolica strain 239T was sequenced using a whole genome shotgun strategy with the Roche 454 GS FLX Titanium System [86]. A total of 64 contigs (length>500 bp) with a total size of 7.23 Mb were assembled from 561,423 reads (average length 430 bp) by the Newbler V2.3 Program of the 454 suite package, providing a 28-fold coverage of the whole genome. The relationships between the contigs were determined by using ContigScape plugin [87] and by reference to the A. Mediterranei U32 genome. The gaps were filled by sequencing PCR products. The final sequence assembly was performed using the phred/phrap/consed package [88–90]. Sanger-based sequencing was employed to facilitate gap closure and to amend low-quality regions (score < 40). Finally, a consensus sequence containing 7,237,391 bp with an estimated error rate of 0.8 per 100,000 bases provided 28-fold coverage. The complete
sequence of the chromosome was deposited in the GenBank database under accession number CP009110.

**Genome annotation and analysis:** Putative protein-coding sequences were predicted based on results from Glimmer 3.02 [91], Genemark [92] and Z-Curve [93]. The CDS annotation was based on the BLASTP results obtained with the CDD and the databases of KEGG [94] and NR; manual corrections were also implemented. The tRNA genes were predicted directly with tRNAscan-SE v1.123 [95]. The genome-wide colinearity analysis was performed using the MUMmer 3.22 Project [96]. The gene clusters for specialized metabolism were predicted using antiSMASH 3 [61]. The critical genes that determined the existence, conformation, and chain lengths of compounds were further elucidated through literature consultation, sequence alignment, domain comparison, and/or phylogenetic analysis. All of the BLASTP analyses used a threshold E value of 1e-5, and protein length diversity was not less than 0.5. Potential specialized metabolite gene clusters of all 18 *Amycolatopsis* genomes that have been predicted using antiSMASH 3. RfPPs analysis were predicted by BAGEL3 [62].

**Circular map, comparative analysis and assignment of common orthologs:** Genome circular maps were generated using GenomeViz [97] and DNAPlotter [98]. Comparative analyses of genomes from strains classified in the family Pseudonocardiales were achieved using Mauve 2.3.1 genome alignment software [99]; the genoPlotR package [100] was used in post-processing. Orthologous analysis data were submitted to the MBGD platform, a database for comparative analysis of complete microbial genomes [101] with default parameters (http://mbgd.genome.ad.jp). Further, all of the annotated proteins of the 123 genomes belonging to the phylum Actinobacteria including the genera Acidimicrobiurn, Conexibacter and Rubrobacter which were assigned to independent classes in the phylum Actinobacteria in the most recent edition of Bergey’s Manual of Systematic Bacteriology (Fig. 57), were downloaded from ftp.ncbi.nlm.nih.gov/genomes/Bacteria and clustered using the BLASTCLUST [102] program under the conditions of a minimum of 30% identity and 70% length coverage. After a further strict screen, 95 conserved and single-copy orthologous genes were selected, but neither *gyrB* nor *recN* genes once used in phylogenetic analysis of *Amycolatopsis* species [16,17] were included in the list as shown in Table S6. These genes were not considered because there are two *gyrB* genes in *Gardinerella vaginalis* ATCC 14019 (NC_014644, HM Prep0421_20130, HM Prep0421_20607) while the *recN* gene of *Mycobacterium smegmatis* MC2 155 (MSMEG_3749) contains a frame shift mutation within its coding sequence which is not the result of sequencing error (CP000480).

### 4.1. Phylogenetic analyses

#### 4.1.1. Phylogenetic analysis based on 16S rDNA of 62 Amycolatopsis type strains

The 16S rRNA gene sequences of the 62 *Amycolatopsis* type strains were obtained from the EzTaxon server. The MUSCLE algorithm [103] from MEGAS [104] was used for sequence alignments. The neighbour-joining (NJ) [105] method in the MEGA5 software was used to construct the phylogenetic trees with *M. smegmatis* as an outgroup. The topology of the phylogenetic trees were achieved using the bootstrap resampling method with 1000 repeats from the MEGAS package.

#### 4.1.2. Phylogenetic tree of 13 AOS, 1 ATS and 3 AMS and group J based on head-tail linked 95 core protein sequences

Phylogenetic trees were based on the 95 orthologous protein sequences (Table S6). Each orthologous protein was aligned using the RAxML package [108], and the reliability of each branch was tested by 1000 bootstrap replications. Finally, the tree file was processed by using the MEGA5 package [104]. The accession numbers of the genomes used in this tree are JNY00000000, JMQI00000000, ARBHH00000000, CP002000, ARPK00000000, JNY00000000, ARF00000000, ANMG00000000, CP007219, AOH00000000, CP008953, CP003410, JAFB00000000, AFVW00000000, CP009110, AXHH00000000, CP001683, AM420293, CP001630 and CP002593.

**Experimental analysis for one-carbon source utilization by *A. methanolica* strain 239**

The *A. methanolica* strain was grown on an agar plate (20 g/L starch, 0.5 g/L L-asparagine, 1 g/L KNO3, 0.5 g/L KH2PO4, 3 g/L NaCl, 0.5 g/L MgSO4 7H2O, 1 g/L CaCO3, and 2 g/L agar, pH 7.5) at 28 °C for 3 days. A single colony from the plate was inoculated into 40 mL of ISP2 broth (0.4% yeast extract, 1% malt extract, and 0.4% glucose) and incubated at 28 °C for 4 days. To test for the survival of the strain, we put 2 mL of the seed culture in endopod tube, washed twice (centrifuged before resuspension) in HMM medium (prepared by mixing (1) 100 mL phosphate salts solution (25.3 g/L KH2PO4 and 22.5 g/L Na2HPO4), (2) 100 mL sulfate salts solution (5 g/L (NH4)2SO4 and 2 g/L MgSO4 7H2O), (3) 799 mL of deionized water, and (4) 1 mL of trace metal solution (0.177 g/L ZnSO4 7H2O, 1.46 g/L CaCO3, 2H2O, 0.107 g/L MnCl2 4H2O, 2.496 g/L FeSO4 7H2O, 0.177 g/L (NH4)2MoO4 2H2O, 0.374 g/L CuSO4 5H2O, 0.238 g/L CoCl2 6H2O and 0.1 g/L Na2WO4 2H2O) followed by resuspension with 1 mL HMM. Next, 40 µL of the resultant suspension was inoculated into 15 mL of HMM broth supplemented with methanol, formaldehyde or formic acid (pH adjusted to 7.5 before use) as the sole carbon sources, these experiments were carried out in triplicate concentrations of the carbon sources, namely 0.1%, 0.2% or 0.4%; and HMM without any additives was used as the control. For the volatile methanol, continuous feeding was performed every two days. Optical density at 600 nm (OD600) was measured every second day. After 14 days of incubation, 10 µL of each broth was spread over the GT plates to evaluate the survival of *A. methanolica* strain 239 in HMM broth supplemented with methanol, formaldehyde or formic acid as sole carbon sources, and in HMM broth without any additives serving as the controls.

**Statistical analysis**

The boxplot chart and scatter diagram were prepared using R-i386 – 3.2.1.

**Accession numbers**

The genome sequence has been deposited at the NCBI under the GenBank accession number [CP009110].

**Competing interests**

The authors declare that they have no competing interests.
References

[1] Lechevalier M, Prauser H, Labeda D, Ruan J. Two new genera of nocardioform actinomycetes: Amycolatopsis gen. nov. and Amycolatopsis gen. nov. Int J Syst Evol Microbiol 1986;36:29–37.

[2] Labeda DP, Goodfellow M, Chun J, Zhi XY, Li WJ. Reassessment of the systematics of the suborder Pseudonocardineae: transfer of the genera within the family Actinomycetaceae Labeda and Kroppenstedt 2000 emend. Zhi et al. 2009 into an emended family Pseudonocardineae Embley et al. 1989 emend. Zhi et al. 2009. Int J Syst Evol Microbiol 2011;61:1259–64.

[3] Labeda D, Goodfellow M, Family I. Pseudonocardineae embl. Smida and Stackebrandt 1989, 2059V, emend. Labeda, Goodfellow, Chun, Zhi and Li. 2010a. In: Goodfellow M, Kämpfer P, Busse H-J, Trujillo ME, Suzuki K-I, Ludwig W, et al., editors. Bergey’s manual of systematic bacteriology, second ed., vol. 5, Part B. New York: Springer; 2012. p. 1302–5.

[4] Labeda D, Goodfellow M. Order XIII. Pseudonocardiaceae ord. nov. In: Goodfellow M, Kämpfer P, Busse H-J, Trujillo ME, Suzuki K-I, Ludwig W, et al., editors. Bergey’s manual of systematic bacteriology, second ed., vol. 5, Part B. New York: Springer; 2012. p. 1301.

[5] Stackebrandt E, Rainey FA, Ward-Rainey NL. Proposal for a new hierarchic classification system. Actinobacteria classis nov. Int J Syst Evol Microbiol 1997;47:479–91.

[6] Kato N, Tsuji K, Tani Y, Ogata K. Methanol-utilizing actinomycete. J Ferment Technol 1974;52:917–20.

[7] Kato N, Tsuji K, Ohashi H, Tani Y, Ogata K. Two assimilation pathways of C2-compound in Streptomyces sp. no. 239 during growth on methanol. Agric Biol Chem 1977;41:29–34.

[8] Hazeu W, Debruyn JC, Vandijken JP. A new actinomycete. Int J Syst Bacteriol 1988;38:291–6.

[9] Embley TM, Smida J, Stackebrandt E. Reverse transcriptase sequencing of 16S rRNA from Faenia rectivirgula, Pseudonocardia thermophila and Sarcina polyaspora hirns, three type IV actinomycetes which lack mycolic acids. J Gen Microbiol 1988;134:261–6.

[10] de Boer L, Dijkhuizen L, Vandenbroucke-Grauls CMEM. The actinomycete. Int J Syst Evol Microbiol 2008;58:2389–426.

Abbreviations

NRPS Non-ribosomal peptide synthetase
PPP Pentose phosphate pathway
RuMP pathway Ribulose monophosphate pathway
AMS Amycolatopsis methanolica subclade
AOS Amycolatopsis orientalis subclade
ATS Amycolatopsis taizensis subclade
Ru5P ribulose-5-phosphate
G6P glucose-6-phosphate
H6P hexulose-6-phosphate
F6P fructose-6-phosphate
FBP fructose-1,6-bisphosphate
GAP glyceraldehyde-3-phosphate
PEP phosphoenolpyruvate
Pyr pyruvate
1,3 PG 1,3-diphosphoglycerate
Rib P ribose-5-phosphate
X5P xylulose-5-phosphate
S7P sedoheptulose-7-phosphate
E4P erythrose-4-phosphate
AcCoA acetyl-CoA
Cis-Aco Cis-aconitate
iso-Cit isocitrate
Oxa-suc Oxalosuccinate
Suc-CoA succinate-CoA
Oxa oxaloacetate
HPS hexulose phosphate synthase
HPI hexulose phosphate isomerase
KDPG 2-Dehydro-3-deoxy-6-phospho-D-gluconate
Glu-1,5-Lac-6P D-Glucono-1,5-lactone 6-phosphate
Glu-6P 6-Phospho-D-gluconate.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.syntbio.2016.05.001.

Acknowledgements

This work was supported in part by grants from the National Basic Research Program of China (2012CB721102, 2013CB734000), the Natural Science Foundation for Youth (31300034) and the National Natural Science Foundation of China (31270056, 31430004 and 31421061). LX.Z. is an Awardee of the National Distinguished Young Scholar Program in China (31250002).
Tang SK, Wang Y, Guan TW, Lee JC, Kim CJ, Li WJ, Chen JZ, Miao XY, Xu M, He JL, Xie Y, Wu XW, et al. Intra-genomic heterogeneity in 16S rRNA genes in strictly anaerobic clinical isolates from peri-tonsillar abscesses. Int J Syst Evol Microbiol 2014;64:1036–41.

Stegmann E, Albersmeier A, Spohn M, Gert H, Weber T, Wohlflein W, et al. Complete genome sequence of the actinobacterium Amycolatopsis japonica MAC417-CP17 (−DMS 44213) producing (5,5′-N-ethyl-pimelimidoyl) acid. J Biotechnol 2014;188:46–7.

Zhao W, Zhong Y, Yuan H, Wang J, Zheng H, Wang Y, et al. Complete genome sequence of the rifamycin SV-producing Amycolatopsis mediterranei U32 revealed its genetic characteristics in phylogeny and metabolism. Cell Res 2010;20:609–11.

Xu L, Huang H, Wei W, Zhong Y, Tang B, Yuan H, et al. Complete genome sequence and comparative genomic analyses of the vancomycin-producing Amycolatopsis orientalis. BMC Genomics 2014;15:363.

Vrijhoef J, Madon J, Diikhuizen L. A plasmid from the methylophilic actinomycete Amycolatopsis methanolica capable of site-specific integration. J Bacteriol 1994;176:7087–90.

Vrijhoef J, Jelinková M, Hessels GJ, Diikhuizen L. Identification of the methanol-utilizing plasmid pME3A00 of the methylophilic actinomycete Amycolatopsis methanolica. Mol Microbiol 1995;18:21–31.

Olynyk M, Samborskyi M, Lester JB, Mironenko T, Scott N, Dickens S, et al. Complete genome sequence of the thermocyanococcus bacterium Saccharophagus degradans E1994. Nucleic Acids Res 2012;40:1059–70.

Grigoriev A. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res 1998;26:2286–90.

Kim CS, Cho YJ, Lee DK, Yang SH, Kim M, Na H, et al. Introducing ESTaxon-e: a prokaryotic 16S rRNA gene database with phylogenies that repre- sent uncultured species. Int J Syst Evol Microbiol 2012;62:716–21.

Tamura T, Isida Y, Otaguro M, Suzuki K. Amycolatopsis helveolus sp. nov. and Amycolatopsis pigmentum sp. nov., isolated from soil. Int J Syst Evol Microbiol 2010;60:2629–33.

Tseng M, Yang SF, Li WJ, Jiang CL. Amycolatopsis taiwanensis sp. nov., from soil. Int J Syst Evol Microbiol 2006;56:1811–5.

Miao Q, Qin S, Biao GX, Yuan B, Xing K, Zhang YJ, et al. Amycolatopsis endophytica sp. nov., a novel endophytic actinomycete isolated from oil-seed plant Jatropha curcas L. Ant Van Leeuwenhoek 2011;100:333–9.

Tripler HG, Galinski EA. Biosynthesis and fate of compatible solutes in extremely halophilic phototrophic eubacteria. FEMS Microbiol Lett 1990;75:253–5.

Müller F, Trüper HG, Galinski EA. Biosynthesis and fate of compatible solutes in extremely halophilic phototrophic eubacteria. FEMS Microbiol Lett 1990;75:253–5.

Vrkalj C, Bremer E. Uptake and synthesis of compatible solutes as microbial signaling. Annu Rev Plant Biol 2008;59:417–39.

Casp R, Altman T, Billingston R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 2014;42:D539–47.

Van Oppen PW, Van Beunen J, Duine JA. Nicotineprotein [NAD(P)]- containing alcohol/aldohexose dehydrogenases. Purification and characterization of a novel type from Amycolatopsis methanolica. Eur J Biochem 1993;212:819–26.

Bystrykh LV, Vonck J, van Bruggen EG, van Beunen J, samyn B, Govorukhina NL, et al. Electro magnetic analysis and structural character- ization of novel NAD(P)H-containing metalloenzyme: NN-dimethyl-4-nitro- soaniline oxidoreductases from the gram-positive methylotrophic bacteria Amycolatopsis methanolica and Mycobacterium gastri MB19. J Bacteriol 1993;175:1814–22.

Norin A, Van Oppen PW, Perssma SR, Persson B, Duine JA, Jornvall H. Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium- chain dehydrogenase/reductase, phylogenetically links different eukaroytic alcohol dehydrogenases–primary structure, conformational modelling and functional correlations. Eur J Biochem 1997;248:280–8.

Van Oppen PW, Jansen KM, Duine JA. Characterization of novel NAD(P)H- dependent formaldehyde dehydrogenases for formate and formate esters in Amycolatopsis methanolica. Characteriza- tion of a molybdenum enzyme active with formate esters and aldehydes. Eur J Biochem 1992;206:519–25.

Vorholta J. Colator-dependent pathways of formaldehyde oxidation in methylococci. Arch Microbiol 2002;178:39–49.

Newton GL, Arnold K, Price MS, Sherrill C, Delcardayre SB, Aharonowitz Y, et al. Distinct families of microorganisms: mycothiole is a major thiol in most actinomycetes. J Bacteriol 1996;178:1990–5.

Newton GL, Buchmeier N, Fahey RC. Biosynthesis and functions of mycothiol, the unique protective thiol of actinomycetes. Microbiol Mol Biol Rev 2001;65:99–174.

Jez JM, Cahoone RE. Kinetic mechanism of glutathione synthetase from Bacillus subtilis. J Biol Chem 2004;279:4276–31.

Morganin DF, Shapar N, Daughtrey SC, DeBoer RT, Emerson JB, Shvartsbeyn A, et al. Secrets of soil survival revealed by the genome sequence of Arthro- bacter acetones'. PLoS Genet 2006;2:e214.

McLeod MP, Warren RL, Hiayo WW, Araki N, Myhre M, Fernandes C, et al. The complete genome of Rhodococcus sp. RH1 provides insights into a catalytic powerhouse. Proc Natl Acad Sci U S A 2009;106:15382–7.

Jones AL, Davies J, Fukuda M, Brown R, Lim J, Goodfellow M. Ant Van Leeuwenhoek 2014;106:365.

Xue F, Dai SW, Sheu YJ, Lin Y, et al. A new salticly synthase AmS is identified for siderophores biosynthesis in Amycolatopsis medicina 239. Appl Microbiol Biotechnol 2015;99:5895–905.

Weber T, Blin K, Duddela S, Krug D, Kim BU, Brculcoci R, et al. AntiSMASH 3.0: Comprehensive analysis of non-ribosomal peptide synthetases. Nucleic Acids Res 2013;41:W448–53.

Challis GL. A widely distributed bacterial pathway for siderophore biosyn- thesis independent of nonribosomal peptide synthetases. ChemBioChem 2008;9:601–11.

Kwon MJ, Cheng J, Yang SH, Lee DK, Suh JW, Hong HJ. Draft genome sequence of ristocetin-producing strain Amycolatopsis sp. strain MJM2582 isolated in South Korea. Genome Announc 2014;2.

Kaur N, Kumar S, Bala M, Raghava GP, Mayilraj S. Draft genome sequence of Amycolatopsis decaplanina strain DSM 45405. Genome Announc 2013;1:e00138.13.

Kwon MJ, Hong HJ. Draft genome sequence of Amycolatopsis lurida NRRL 2430, producer of the glycopeptide family bacterial antibiotic. Genome Announc 2015;3:e00675.14.

Garcia-Estepa R, Argandona M, Reina-Bueno M, Capote N, Iglesias-Fuera F, Nieto JJ, et al. The ecE gene, which is involved in the synthesis of the compatible solute hydroxyectoine, is essential for thermoprotection of the halophilic bacterium Chromohalobacter salexigens. J Bacteriol 2006;188:3774–85.

Vargas C, Argandona M, Reina-Bueno M, Rodriguez-Moya J, Fernandez- Aunsson C, Nieto JJ. Unraveling the adaptation responses to osmotic and temperature stress in Chromohalobacter salexigens, a bacterium with broad salinity tolerance. Saline Syst 2008;4:14.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M, REGEO. Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 1999;27:29–34.

Caspi R, Altman T, Billingston R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 2014;42:D539–47.

Van Oppen PW, Van Beunen J, Duine JA. Nicotineprotein [NAD(P)]- containing alcohol/aldohexose dehydrogenases. Purification and characterization of a novel type from Amycolatopsis methanolica. Eur J Biochem 1993;212:819–26.
of ATP- and PP(i)-dependent phosphofructokinase isoenzymes in the methylotrophic actinomycete Amycolatopsis methanolica. J Bacteriol 2001;183:7231–40.

[85] Zhi XY, Zhao W, Li WJ, Zhao GP. Prokaryotic systematics in the genomics era. Ant Van Leeuwenhoek 2012;101:21–34.

[86] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005;437:376–80.

[87] Tang B, Wang Q, Yang M, Xie F, Zhu Y, Zhuo Y, et al. ContigScape: a Cytoscape plugin facilitating microbial genome gap closing. BMC Genomics 2013;14:285.

[88] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998;8:175–85.

[89] Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Res 1998;8:195–202.

[90] Gordon D. Viewing and editing assembled sequences using Consed. Curr Protoc Bioinforma 2003 [Chapter 11]:Unit 11.2.

[91] Aggarwal G, Ramaswamy R. Ab initio gene identification: prokaryote genome annotation with GeneScan and Glimmer. J Biosci 2002;27:7–14.

[92] Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 2000;28:W451–4.

[93] Song K. Recognition of prokaryotic promoters based on a novel variable-window Z-curve method. Nucleic Acids Res 2005;33:W451–4.

[94] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snORFs. Nucleic Acids Res 2005;33:W686–9.

[95] Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinforma 2003 [Chapter 10]: Unit 10.3.

[96] Chai R, Hain T, Chakraborty T. GenomeViz: visualizing microbial genomes. BMC Bioinforma 2004;5:198.

[97] Carver T, Thomson N, Bealby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.

[98] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:394–403.

[99] Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 2010;26:2334–5.

[100] Uchiyama I, Mihara M, Nishide H, Chiba H. MBGD update 2013: the microbial genome database for exploring the diversity of microbial world. Nucleic Acids Res 2013;41:D631–5.

[101] Elizabeth Cha I, Rouxhka EC. Comparison of current BLAST software on nucleotide sequences. IPDPS 2005;19:8.

[102] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.

[103] Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 2007;56:564–77.

[104] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. Bioinformatics 2007;23:2947–8.

[105] Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–90.

[106] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–25.

[107] Goodfellow M, Kämpfer P, Busse H-J, Trujillo ME, Suzuki K-i, Ludwig W, et al., editors. Bergey's manual of systematic Bacteriology. second ed., vol. 5, Part B. New York: Springer; 2012. p. 1967.