Synthetic and Systems Biotechnology 1 (2016) 169–186

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: http://www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology/

A systematic study of the whole genome sequence of *Amycolatopsis methanolica* strain 239^T 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 [a], Haokui 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, 200031, 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 239^T 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-CF17^T, *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. methanolica* 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 239^T, including both the *mdo*/*mscR* encoded methanol oxidation and the *hps*/*hpi* 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.

1. Introduction

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

☆ Peer review under responsibility of KeAi Communications Co., Ltd.

* Corresponding author. State Key Laboratory of Genetic Engineering, Department of Microbiology, School of Life Sciences and Institute of Biomedical Sciences, Fudan University, Shanghai, 200438, China.
** Corresponding author.
*** Corresponding author.
E-mail addresses: lzhang03@gmail.com (L. Zhang), m.goodfellow@ncl.ac.uk (M. Goodfellow), gpzhao@sibs.ac.cn (G.-P. Zhao).

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

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^T is Gram-positive, non-acid-fast, and forms aerial and substrate hyphae that fragment into squarish elements. It contains *meso*-diaminopimelic acid (*meso*-A_2pm), arabinose and galactose in the wall peptidoglycan, major amounts of di- and tetrahydrogenated menaquinones, phosphatidylethanolamine as the diagnostic phospholipid, *iso*-$C_{16:0}$ 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 strains [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^T which produces (*S,S*)-*N,N'*-ethylenediaminedisuccinic acid, 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^T. 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^T 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^T (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 genes in the genomes of these strains together with those in the genomes of *S. viridis* DSM 43017^T and *S. erythraea* NRRL 2338^T, 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^T 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 pMEA100-like integrated plasmids found in the *A. mediterranei* genome, one integrated pMEA300-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 pMEA300-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 Ile-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^T 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^T, *A. orientalis* HCCB10007, *A. mediterranei* U32, *P. dioxanivorans* CB1190^T, *S. viridis* DSM 43017^T and *S. erythraea* NRRL 2338^T [37]. These results revealed a highly conserved core region ranging from 0 to 2.1 Mb

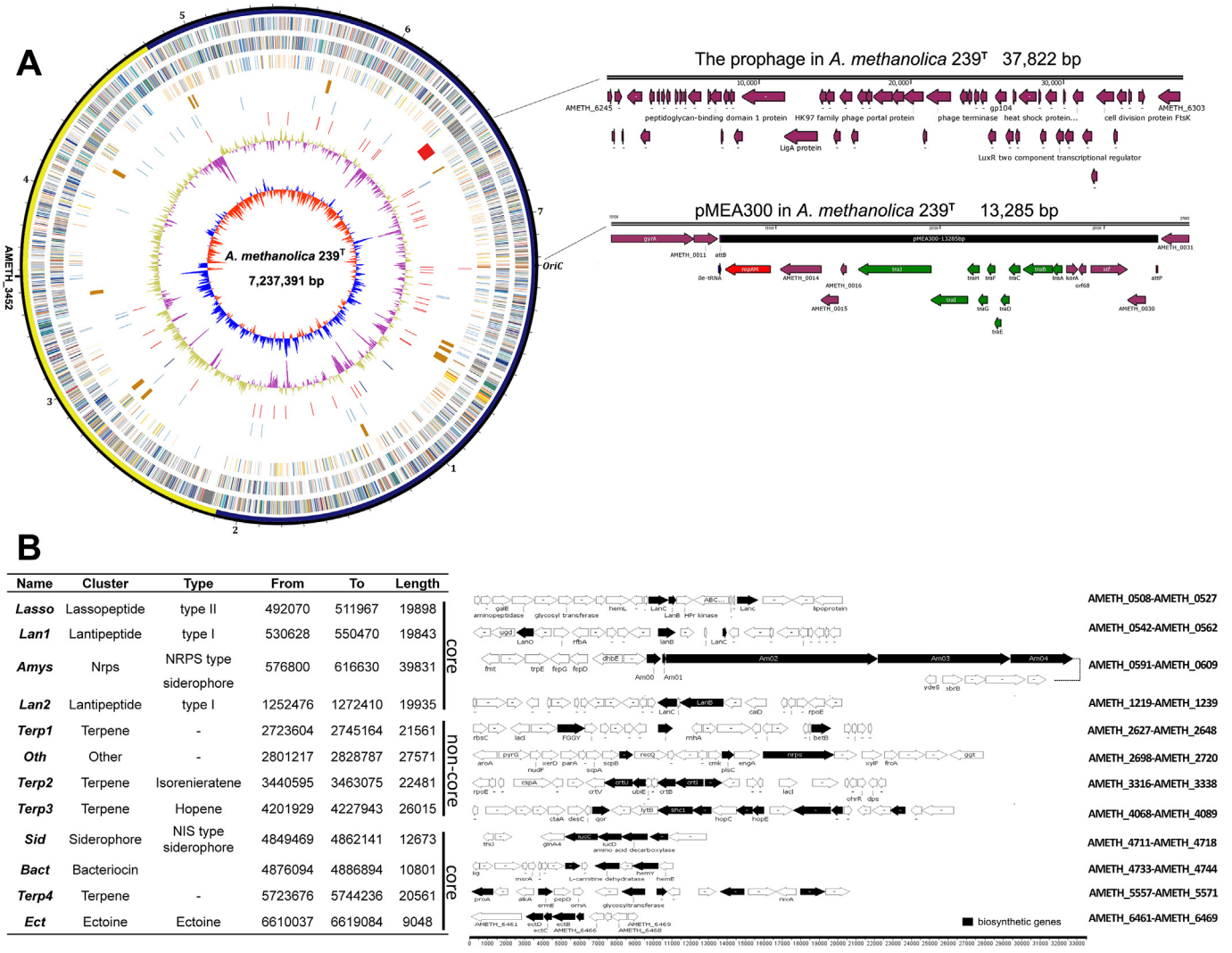


Fig. 1. Genome atlas of *A. methanolica* strain 239^T with two characteristic episomes enlarged (A) and the gene clusters encoding biosynthetic functions for specialized metabolites (B). (A) The large circle represents the chromosome: the outer scale is numbered in megabases and indicates the core (blue) and non-core (yellow) regions. The circles are nominated to start from the outside in. The genes in circles 1 and 2 (forward and reverse strands, respectively) are color-coded according to COG functional categories. Circle 3 shows selected essential genes (cell division, replication, transcription, translation, and amino-acid metabolism; the paralogs of essential genes in the non-core regions are not included), circle 4 the phage (red), specialized metabolite clusters (orange) and integrated plasmid pMEA300 (pink), circle 5 the mobile genetic elements (transposases), circle 6 the RNAs, circle 7 the GC content with a calculated ratio of 71.53 mol % and circle 8 the GC bias (blue, values > 0; red, values < 0). **The right panel of (A) demonstrates the two characteristic episomes of the *A. methanolica* genome of strain 239^T, *i.e.*, the prophage with its 52 ORFs (upper) and the pMEA300-like integrated plasmid (lower). (B) Gene clusters in the genome of *A. methanolica* strain 239^T encoding biosynthetic functions for specialized metabolites**. The corresponding biosynthesis related genes are dark colored.

(left arm) and 4.8–7.2 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 239^T and *S. viridis* DSM 43017^T 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.

Table 1
General features of *Pseudonocardiaceae* genomes.

Species/Strains (accession number)	*A. orientalis* HCCB10007 (CP003410)	*A. mediterranei* U32 (CP002000)	*A. japonica* MG417-CF17^T (CP008953)	*A. methanolica* 239^T (CP009110)	*S. viridis* DSM 43017^T (CP001683)	*S. erythraea* NRRL 2338^T (NC_009142)	*A. mirum* DSM 43827^T (CP001630)	*P. dioxanivorans* CB1190^T (CP002593)
Length (bp)	8,948,591	10,236,715	8,961,318	7,237,391	4,308,349	8,212,805	8,248,144	7,096,571
GC content	69.01%	71.30%	68.89%	71.53%	67.32%	71.15%	73.71%	73.31%
Total proteins	8168	9228	8298	7074	3828	7197	6916	6495
Proteins with function	5512	6441	4227	4830	2719	4979	4423	4642
Proteins with function %	67.50%	69.85%	50.94%	68.28%	71.02%	69.18%	63.95%	71.47%
Average ORF size (bp)	992.3	990.1	981.8	897.3	978.1	969.2	1039.7	967.6
Average intergenic region size (bp)	103.3	119.2	98.38	115.1	148.4	172.9	153.8	126
Coding density	90.57%	89.30%	90.89%	88.71	86.9	84.9	87.2	88.6
rRNA operons	4	4	4	3	3	4	5	3
tRNA genes	50	52	55	51	49	50	60	50

T: type strain.

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 (CDSs) (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 *A. mediterranei* and *A. orientalis* genomes.

In addition to a comparison of chromosomal configurations of the closely related genomes, a conserved gene (*AMETH_3452*) spanning nucleotides 3,578,958 to 3,579767 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_19800*, *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*, *Micrococcaceae*, *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 *Pseudonocardiaceae* as shown in the next section.

A draft *A. methanolica* genome (AQUL00000000.1) 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 (AQUL00000000.1 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 AQUL00000000.1 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, AMETH_3452 orthologous 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 rRNA 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, phyletic group F1 composed of *Amycolatopsis helveola* TT-99-32^T [40], *Amycolatopsis pigmentata* TT00-43^T [40], and *A. taiwanensis* 0345M-7^T [41] and phyletic 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 thermophila* 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 phylogenetic 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 1221^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,

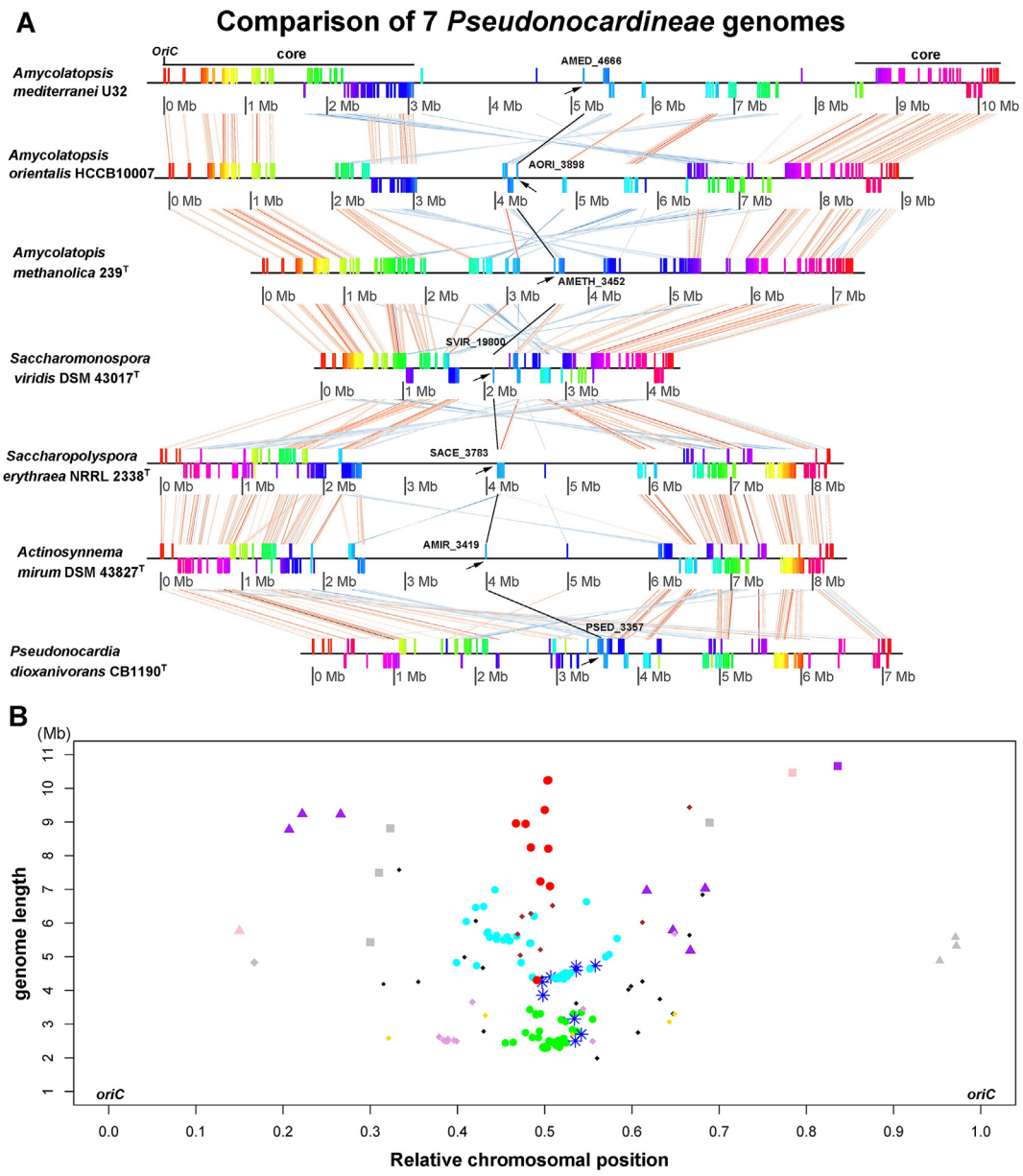


Fig. 2. Comparative analyses of 7 *Pseudonocardiaceae* genomes (A) and the chromosomal loci distribution of the highly conserved *AMETH_3452* gene among actinobacterial genomes (B). In **panel A**, parallel straight lines represent the genomes of *A. orientalis* HCCB10007 (CP003410.1), *A. mediterranei* U32 (CP002000.1), *A. methanolica* 239^T (CP009110), *S. viridis* DSM 43017^T (CP001683.1), *S. erythraea* NRRL 2338^T (AM420293.1), *A. mirum* DSM 43827^T (CP001630.1) and *P. dioxanivorans* CB1190^T (CP002593.1), and are drawn to scale with *oriC* located at the very end of the lines. Vertical short bars representing different conserved genes are marked with distinct colors to demonstrate their genomic loci; the latter bars are connected by corresponding colored thin lines. The segregation of core, non-core and quasi-core regions are shown as clusters of conserved orthologous genes. Black arrows highlight the genomic loci of the highly conserved *AMETH_3452* analogs located in the middle of the 7 genomes and are connected by thick black lines. In **panel B**, the distribution of chromosomal loci of the highly conserved genes orthologous to *AMETH_3452* among species of the class *Actinobacteria* with circular genomes is plotted against their corresponding genome sizes. The horizontal coordinate represents the relative chromosomal positions of the genomes normalized to 0.0–1.0 with *oriC* located at both ends; the genomic loci of the conserved genes orthologous to *AMETH_3452* are shown as scattered points. This conserved gene is located in the middle of the circular chromosomes (*i.e.*, close to the 0.5 locus) in species that mainly belong to four taxa, namely the families *Pseudonocardiaceae* (12 genomes, dot, red), *Corynebacteriaceae* (50 genomes, dot, green), *Micrococcaceae* (9 genomes, asterisk, blue) and *Mycobacteriaceae* (59 genomes, dot, cyan). In contrast, the gene is located close to *oriC* in the species belonging to the families *Acidimicrobiaceae* (rhombus, gray). *Catenulisporaceae* (square, pink), *Geodermatophilaceae* (triangle, gray), *Micromonosporaceae* (triangle, purple), *Nocardiopsaceae* (triangle, pink) and *Streptomycetaceae* (only *Streptomyces violaceusniger* Tu 4113) (square, purple). This gene is located between these two positions in species belonging to the families *Frankiaceae* (square, gray), *Microbacteriaceae* (rhombus, gold), *Nocardiaceae* (rhombus, brown), *Propionibacteriaceae* (rhombus, plum), *Actinomycetaceae, Beutenbergiaceae, Cellulomonadaceae, Dermabacteraceae, Dermacoccaceae, Glycomycetaceae, Intrasporangiaceae, Jonesiaceae, Nakamurellaceae, Nocardioidaceae, Promicromonosporaceae, Sanguibacteraceae, Segniliparaceae, Thermomonosporaceae*, *Tsukamurellaceae* and in *Thermobispora bispora*, a member of the genus *Thermobispora* that was once classified in the family *Pseudonocardiaceae* but is now known to form a deep lineage in the 16S rRNA actinobacterial gene tree hence its reclassification as an order *incertae sedis* in the current edition of Bergey's Manual of Systematic Bacteriology [109] (small dot, black). Detailed information on all of these genomes is given in **Table S4**.

desiccation, high temperature and oxygen radicals [43–47]. 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 thermoflava* N1165^T (NZ_AXBH00000000) versus the ATS genome

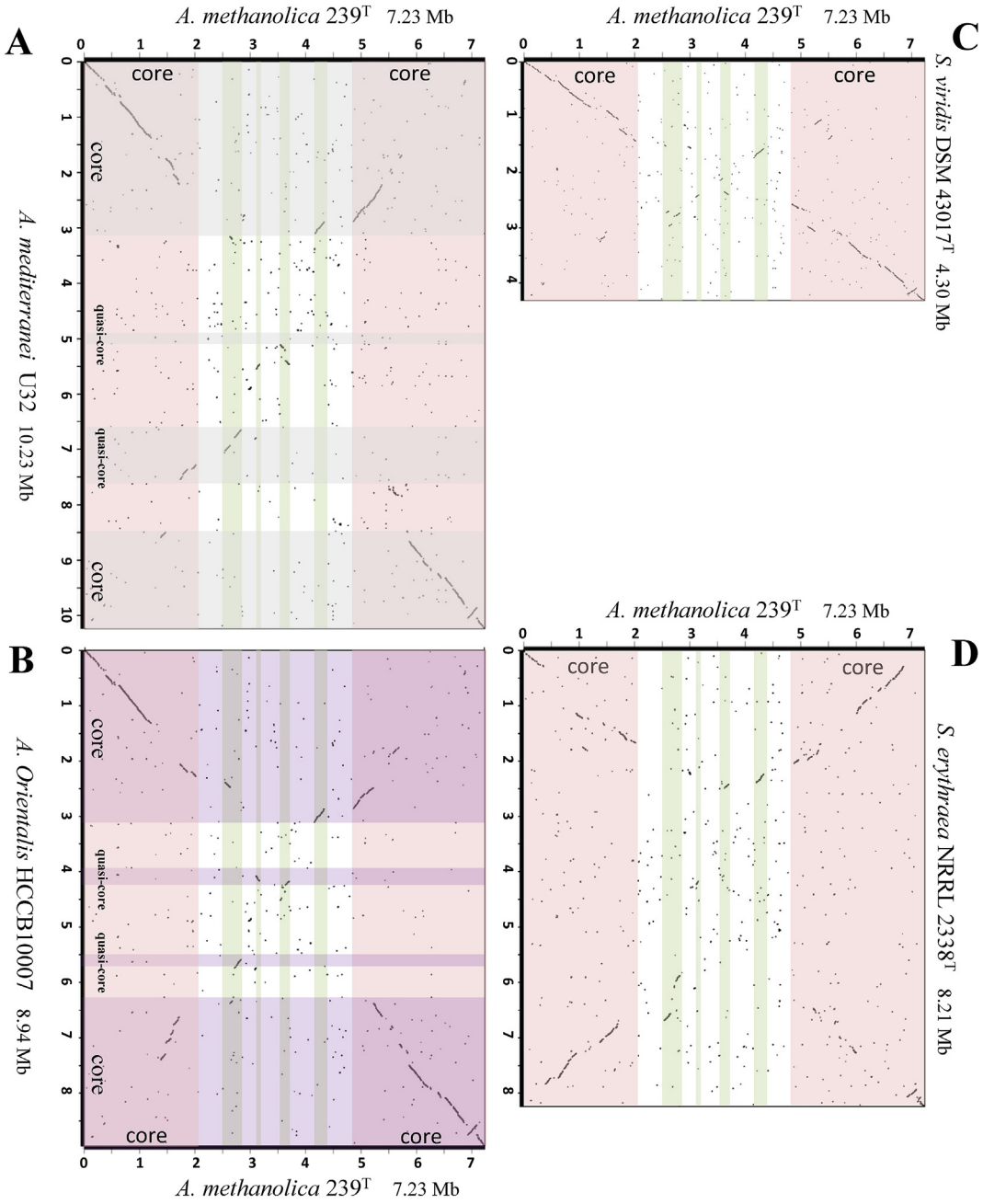
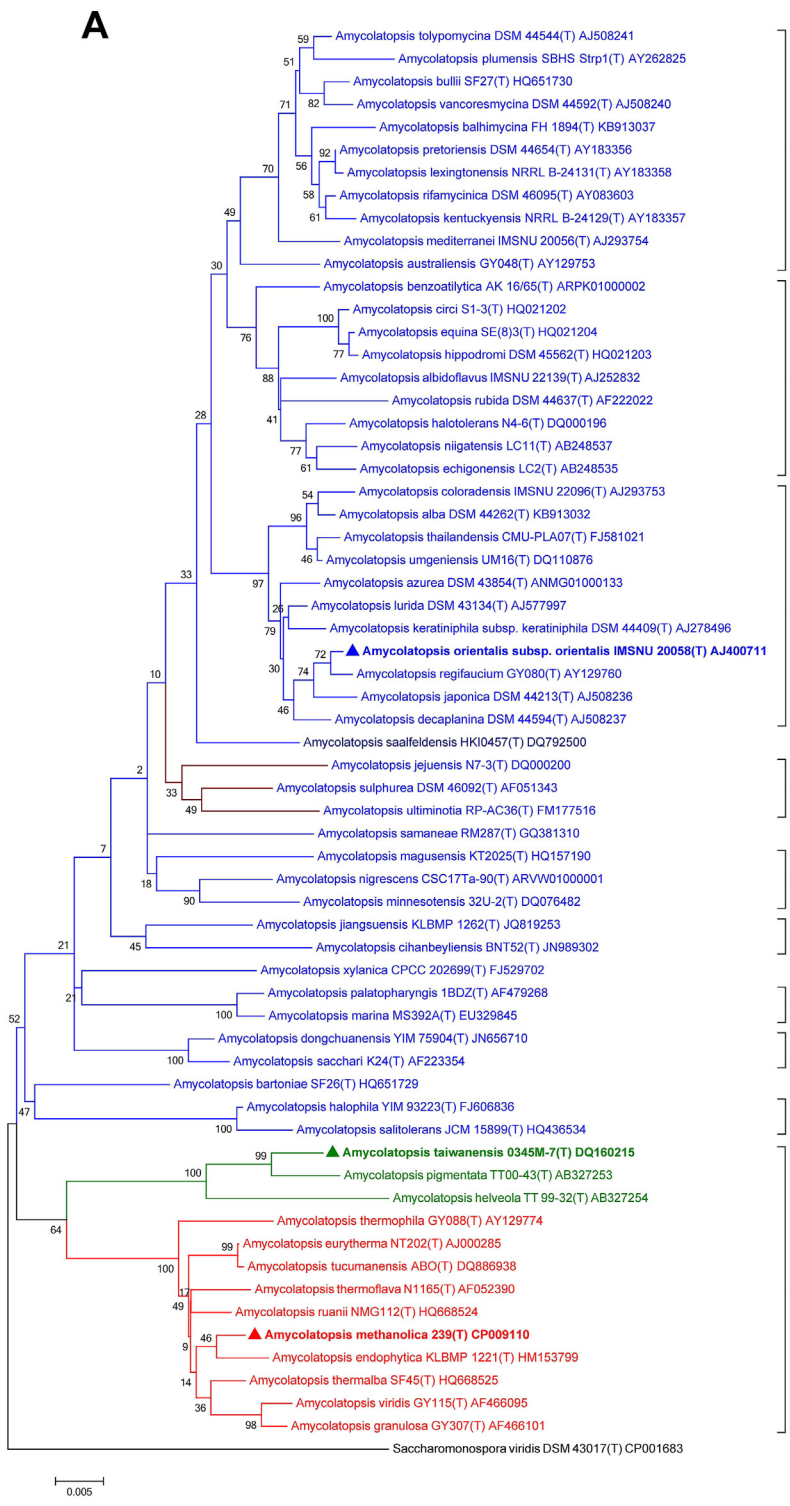


Fig. 3. Collinearity analyses of the genome of *A. methanolica* strain 239^T against the genomes of *A. mediterranei* strain U32 (A), *A. orientalis* strain HCCB10007 (B), *S. viridis* DSM 43017^T (C), and *S. erythraea* NRRL 2338^T (D), respectively. The ortholog ordering in the defined core region of the *A. methanolica* genome is highlighted in light red while the corresponding consecutive genomic segments in the non-core region, which are matched with the quasi-core regions of the corresponding genomes, are shown in light green.

of *A. taiwanensis* DSM 45107^T NZ_JAFB00000000 (Table S5). It is significant that the OtsA-OtsB [48], TreY-TreZ [49], TreS [48] and TreP [50–52] 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 *ssuB* (*AMETH_2747*), *proP* (*AMETH_5526*, *AMETH_6721*) and *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 *A. japonica* strain MG417-CF17^T, *A. mediterranei* strain U32, *A. methanolica* strain 239^T and *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 *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 *A. japonica*, *A. mediterranei* and *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 *A. orientalis* and *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-*



B

Growth in 10°C	Growth in 45°C	Growth in 5% NaCl	GC%
ND	+	+	ND
-	-	-	ND
-	-	-	ND
-	-	+	72.0
+	-	ND	70.8
-	-	+	ND
-	+	+	ND
+	-	W	71.8
-	+	+	ND
+	-	W	71.3
ND	+	-	ND
+	-	+	70.0
-	-	+	ND
-	-	+	ND
-	-	+	ND
-	-	+	68.5
+	-	+	67.4
+	-	+	72.5
+	+	+	72.4
+	+	+	72.7
+	-	+	66
-	-	-	68.7
+	-	+	67
-	+	+	ND
+	-	+	68.9
+	-	+	67
+	-	+	ND
+	-	W	69
+	-	+	ND
+	-	W	68.9
+	-	+	ND
W	-	-	ND
+	-	-	69.1
-	-	-	ND
+	-	+	67.5
-	-	+	71.7
-	-	+	70.8
-	-	+	70.0
+	-	+	69.5
-	-	+	70.4
-	-	+	68.8
+	+	+	65.9
+	-	+	65.8
+	+	+	70.1
+	+	+	68.5
+	+	+	ND
-	-	-	ND
-	+	+	67.8
-	+	+	66.4
-	-	-	68.9
-	-	-	67.2
-	-	-	68.9
+	+	ND	ND
-	+	+	73.1
-	+	+	ND
-	+	+	71.6
-	+	+	ND
+	+	+	71.5
-	+	+	73.3
-	+	+	ND
+	+	+	ND
-	+	-	ND

Fig. 4. Neighbour-joining phylogenetic tree based on 16S rDNA of 62 *Amycolatopsis* type strains employing *S. viridis* DSM 43017^T as the outgroup (A) together with corresponding growth and temperature tolerance properties (B) In panel **A**, numbers at the nodes are percentage bootstrap values based on 1000 resampled datasets. The scale bar indicates 5 nucleotide substitutions per 1000 nucleotides. The subclades are designated AOS, ATS and AMS, respectively. Subclade groups (A–G) defined by Everest and Meyers [16] and Everest et al. [17] are shown, the F group is divided into F1 and F2. The subclades H-J were defined in this study; the J group lies at the periphery of AOS subclade hence the corresponding vertical bar is dashed. In **panel B**, the phenotypic characteristics of the 62 *Amycolatopsis* type strains are listed according to their phylogenetic locations. All of the growth condition data were taken from the initial published descriptions of the species. Abbreviations: +, positive growth, w, weak growth, -, no growth and ND, no data available.

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

A

The identities among all 16s sequences of *A. mediterranei* U32, *A. orientalis* HCCB10007, *A. methanolica* 239 and *Amycolatopsis japonica* MG417-CF17^T

Number	AMED_R16	AMED_R25	AMED_R39	AMED_R49	AORI_R016	AORI_R022	AORI_R041	AORI_R045	AMETH_R14	AMETH_R17	AMETH_R33	AJAP_r10595	AJAP_r14255	AJAP_r30385	AJAP_r33495
Transcriptional direction	+	+	-	-	+	+	-	-	+	+	+	+	+	-	-
location	core	core	quasi-core	quasi-core	core	core	quasi-core	core	core	core	non-core	ND	ND	ND	ND
AMED_R16	100														
AMED_R25	99.93	100													
AMED_R39	99.93	99.86	100												
AMED_R49	99.93	99.86	100	100											
AORI_R016	97.2	97.27	97.13	97.13	100										
AORI_R022	97.2	97.2	97.13	97.13	98.91	100									
AORI_R041	97.2	97.2	97.13	97.13	99.04	98.63	100								
AORI_R045	97.2	97.2	97.13	97.13	99.04	98.63	100	100							
AMETH_R14	94.47	94.41	94.47	94.47	94.46	94.32	94.19	94.19	100						
AMETH_R17	94.47	94.41	94.47	94.47	94.46	94.32	94.19	94.19	100	100					
AMETH_R33	94.47	94.41	94.47	94.47	94.46	94.32	94.19	94.19	100	100	100				
AJAP_r10595	97.06	97.06	96.99	96.99	98.77	98.5	98.84	98.84	94.59	94.59	94.59	100			
AJAP_r14255	97.19	97.19	97.13	97.13	98.43	99.52	98.08	98.08	94.52	94.52	94.52	98.74	100		
AJAP_r30385	97.19	97.19	97.13	97.13	98.43	99.52	98.08	98.08	94.52	94.52	94.52	98.74	100	100	
AJAP_r33495	97.4	97.4	97.33	97.33	98.7	98.29	99.66	99.66	94.45	94.45	94.45	99.2	98.61	98.61	100

B

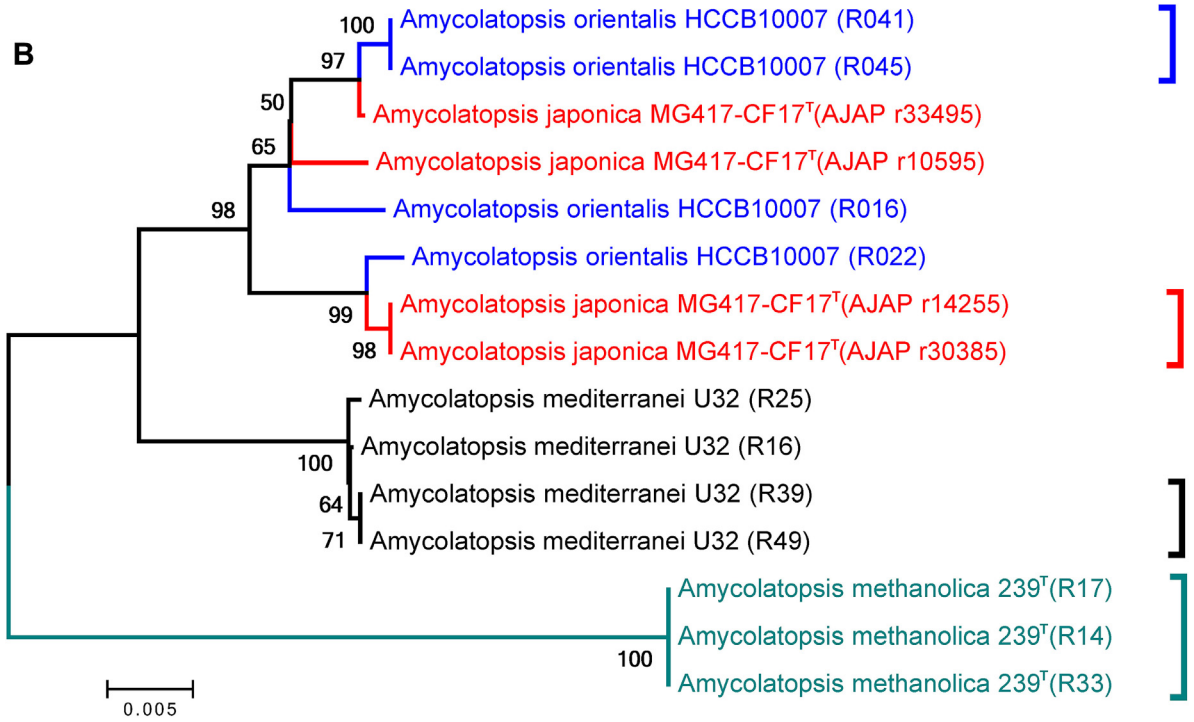


Fig. 5. Pairwise comparison of 16S rDNA sequences from *A. mediterranei* strain U32, *A. orientalis* strain HCCB10007, *A. japonica* strain MG417-CF17^T *and A. methanolica* strain 239^T (A) and their overall phylogenetic relationships (B). In panel **A**, the percentage of sequence identity is shown for each pair of the different copies of 16S rDNAs. The brackets on the diagonal line of the table represent comparisons between the same 16S rDNAs. The pairs circled by triangles represent intra-species comparisons while those circled by squares represent inter-species comparisons comparable to those of the corresponding intra-species comparisons shown as square circled data inside the triangle. In **panel B**, the small differences between the intra- and inter-species 16S rDNA comparisons accounted for the abnormal clustering of the corresponding species in the phylogenetic tree are shown.

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/LTPs123_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 phyletic 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 phylogenetic 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 *Pseudonocardiaceae* as the outgroup, an approach that has rarely been applied [16,17]. The results were completely unexpected, *Amycolatopsis halophila* YIM 93223^T 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 head-tail 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 *Pseudonocardiaceae* (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 *Pseudonocardiaceae* 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).

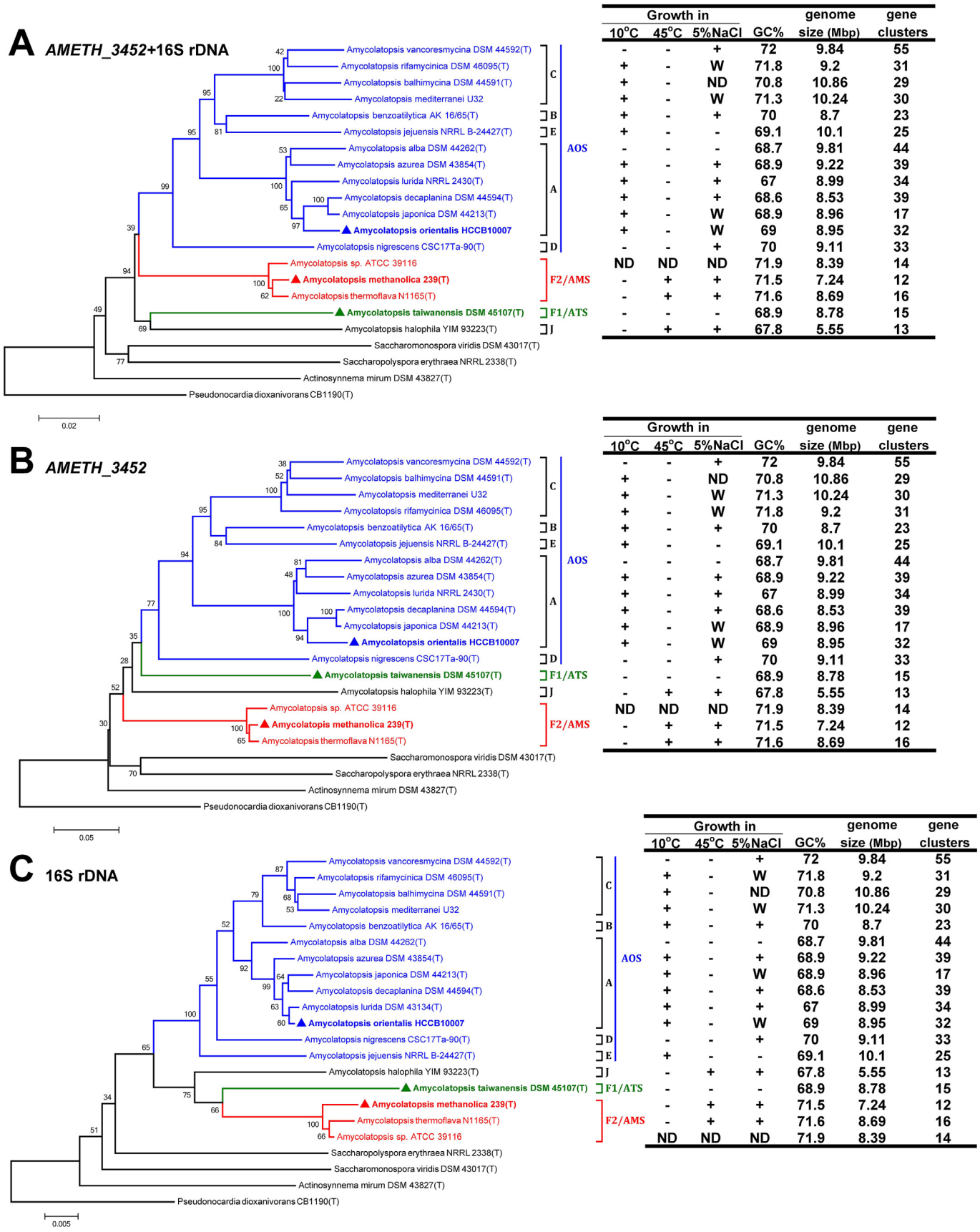
It is particularly interesting that the type strain of *A. halophila* formed a distinct phyletic line that was more closely related to the *A. methanolica* and *A. taiwanensis* strains than to representatives of the *A. orientalis* subclade (Fig. 6). However, this result is not so surprising as the *A. 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 cluster related taxa among the genera of the family *Pseudonocardiaceae* 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 physiological data (Fig. 6).

2.3. Whole genome sequence of *A. methanolica* strain 239^T reveals the genetic basis for its special primary and specialized metabolic characteristics

Although *A. methanolica* strain 239^T 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 239^T, *Amycolatopsis* sp. ATCC 39116 and *Amycolatopsis thermoflava* strain N1165^T); 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 phyletic 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 polyketide synthases (PKSs) and/or non-ribosomal peptide synthetases (NRPSs) were identified in the three AMS genomes (1 NRPS in *A. methanolica* strain 239^T, 1 PKS and 1 NRPS in *Amycolatopsis* sp. ATCC 39116, 3 PKSs and 1 NRPS in *A. thermoflava* strain N1165^T), 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 (**N**RPS **I**ndependent **S**iderophore) 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 239^T (*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



	Growth in				genome	gene
	10°C	45°C	5%NaCl	GC%	size (Mbp)	clusters
Amycolatopsis vancoresmycina DSM 44592(T)	-	-	+	72	9.84	55
Amycolatopsis rifamycinica DSM 46095(T)	+	-	W	71.8	9.2	31
Amycolatopsis balhimycina DSM 44591(T)	+	-	ND	70.8	10.86	29
Amycolatopsis mediterranei U32	+	-	W	71.3	10.24	30
Amycolatopsis benzoatilytica AK 16/65(T)	+	-	+	70	8.7	23
Amycolatopsis jejuensis NRRL B-24427(T)	+	-	-	69.1	10.1	25
Amycolatopsis alba DSM 44262(T)	-	-	-	68.7	9.81	44
Amycolatopsis azurea DSM 43854(T)	+	-	+	68.9	9.22	39
Amycolatopsis lurida NRRL 2430(T)	+	-	+	67	8.99	34
Amycolatopsis decaplanina DSM 44594(T)	+	-	+	68.6	8.53	39
Amycolatopsis japonica DSM 44213(T)	+	-	W	68.9	8.96	17
Amycolatopsis orientalis HCCB10007	+	-	W	69	8.95	32
Amycolatopsis nigrescens CSC17Ta-90(T)	-	-	+	70	9.11	33
Amycolatopsis sp. ATCC 39116	ND	ND	ND	71.9	8.39	14
Amycolatopsis methanolica 239(T)	-	+	+	71.5	7.24	12
Amycolatopsis thermoflava N1165(T)	-	+	+	71.6	8.69	16
Amycolatopsis taiwanensis DSM 45107(T)	-	-	-	68.9	8.78	15
Amycolatopsis halophila YIM 93223(T)	-	+	+	67.8	5.55	13

	Growth in				genome	gene
	10°C	45°C	5%NaCl	GC%	size (Mbp)	clusters
Amycolatopsis vancoresmycina DSM 44592(T)	-	-	+	72	9.84	55
Amycolatopsis balhimycina DSM 44591(T)	+	-	ND	70.8	10.86	29
Amycolatopsis mediterranei U32	+	-	W	71.3	10.24	30
Amycolatopsis rifamycinica DSM 46095(T)	+	-	W	71.8	9.2	31
Amycolatopsis benzoatilytica AK 16/65(T)	+	-	+	70	8.7	23
Amycolatopsis jejuensis NRRL B-24427(T)	+	-	-	69.1	10.1	25
Amycolatopsis alba DSM 44262(T)	-	-	-	68.7	9.81	44
Amycolatopsis azurea DSM 43854(T)	+	-	+	68.9	9.22	39
Amycolatopsis lurida NRRL 2430(T)	+	-	+	67	8.99	34
Amycolatopsis decaplanina DSM 44594(T)	+	-	+	68.6	8.53	39
Amycolatopsis japonica DSM 44213(T)	+	-	W	68.9	8.96	17
Amycolatopsis orientalis HCCB10007	+	-	W	69	8.95	32
Amycolatopsis nigrescens CSC17Ta-90(T)	-	-	+	70	9.11	33
Amycolatopsis taiwanensis DSM 45107(T)	-	-	-	68.9	8.78	15
Amycolatopsis halophila YIM 93223(T)	-	+	+	67.8	5.55	13
Amycolatopsis sp. ATCC 39116	ND	ND	ND	71.9	8.39	14
Amycolatopsis methanolica 239(T)	-	+	+	71.5	7.24	12
Amycolatopsis thermoflava N1165(T)	-	+	+	71.6	8.69	16

	Growth in				genome	gene
	10°C	45°C	5%NaCl	GC%	size (Mbp)	clusters
Amycolatopsis vancoresmycina DSM 44592(T)	-	-	+	72	9.84	55
Amycolatopsis rifamycinica DSM 46095(T)	+	-	W	71.8	9.2	31
Amycolatopsis balhimycina DSM 44591(T)	+	-	ND	70.8	10.86	29
Amycolatopsis mediterranei U32	+	-	W	71.3	10.24	30
Amycolatopsis benzoatilytica AK 16/65(T)	+	-	+	70	8.7	23
Amycolatopsis alba DSM 44262(T)	-	-	-	68.7	9.81	44
Amycolatopsis azurea DSM 43854(T)	+	-	+	68.9	9.22	39
Amycolatopsis japonica DSM 44213(T)	+	-	W	68.9	8.96	17
Amycolatopsis decaplanina DSM 44594(T)	+	-	+	68.6	8.53	39
Amycolatopsis lurida DSM 43134(T)	+	-	+	67	8.99	34
Amycolatopsis orientalis HCCB10007	+	-	W	69	8.95	32
Amycolatopsis nigrescens CSC17Ta-90(T)	-	-	+	70	9.11	33
Amycolatopsis jejuensis NRRL B-24427(T)	+	-	-	69.1	10.1	25
Amycolatopsis halophila YIM 93223(T)	-	+	+	67.8	5.55	13
Amycolatopsis taiwanensis DSM 45107(T)	-	-	-	68.9	8.78	15
Amycolatopsis methanolica 239(T)	-	+	+	71.5	7.24	12
Amycolatopsis thermoflava N1165(T)	-	+	+	71.6	8.69	16
Amycolatopsis sp. ATCC 39116	ND	ND	ND	71.9	8.39	14

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. thermoflava strain N1165^T, respectively. The genes encoded in this cluster are complete and the NRPS products have recently been identified and characterized as amychelins [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 and hopene biosynthesis, respectively based on antiSMASH data; *Terp2* probably accounts for the yellow colonies of *A. methanolica* strain 239^T, while, *Terp1* and *Terp4* may be new types of terpenes. The bacteriocin biosynthetic gene cluster in *A. methanolica* strain 239^T seems to be conserved in *Amycolatopsis* sp. ATCC 39116 [64], *Amycolatopsis* sp. AA4 (NZ_ACEV00000000), *Amycolatopsis* sp. MJM2582 (JPLW00000000) [65] and in the type strains of *Amycolatopsis azurea* DSM 43854^T (NZ_ANMG00000000), *Amycolatopsis decaplanina* DSM 44594^T (NZ_AOHO00000000) [66], *Amycolatopsis lurida* NRRL 2430^T (CP007219) [67], *Amycolatopsis japonica* MG417-CF17^{T32} and *Amycolatopsis nigrescens* CSC17Ta-90^T (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 hydroxylase gene (*ectD*) is responsible for the conversion 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^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 AMS species, the protein EctABCD of *A. methanolica* strain 239^T has 99% similarity to that of the other two AMS species, *i.e.*, *A. thermoflava* strain N1165^T and *Amycolatopsis* sp. ATCC 39116; associated CDSs are also highly conserved in these strains (Fig. S8).

One carbon metabolism was recognized as a key characteristic of *A. methanolica* strain 239^{T10}. 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^T as previously described [72,73] _ENREF_72. This gene encodes methanol:*N,N*-dimethyl-4-nitrosoaniline oxidoreductase (EC: 1.1.99.37) which may act on the CH–OH group of donors. The subsequent oxidization of formaldehyde [74–76] and formate is likely to be catalyzed by NAD/mycothiol-dependent formaldehyde dehydrogenase (FD-FA1DH/MscR, EC: 1.1.1.306) encoded by *AMETH_3767* and formate dehydrogenase (EC: 1.2.1.2) encoded by *AMETH_3397* (*fdh*) and *AMETH_3398* (*fdoH*), respectively (Fig. 7A). Though a glutathione-independent formaldehyde dehydrogenase (GD-FA1DH, EC: 1.2.1.46) is encoded by *AMETH_1319* in the genome of *A. methanolica* strain 239^T, 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^T due to the lack of the gene *gshB* which encodes one of the two key enzymes, *i.e.*, glutathione synthetase [79].

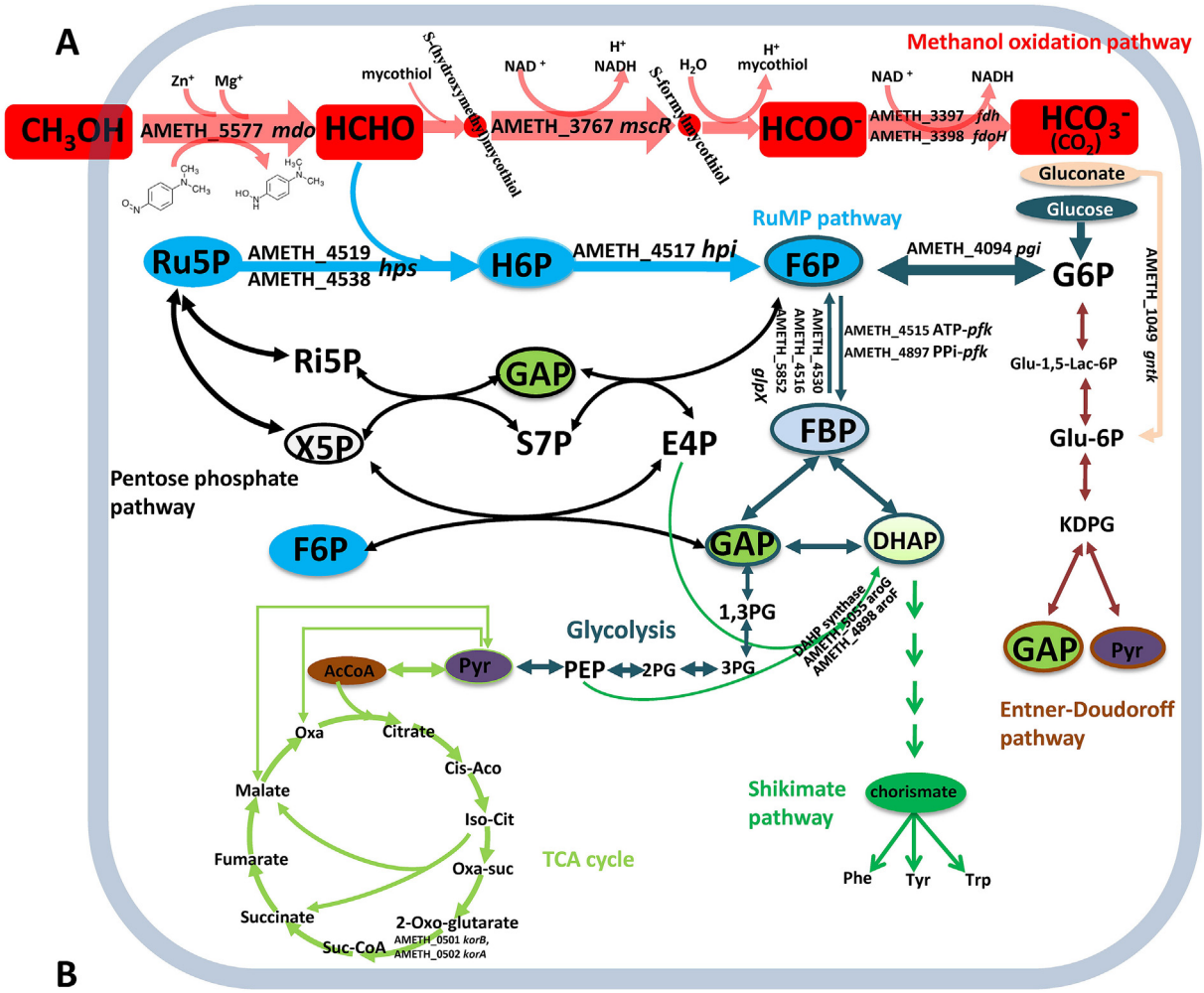
Here we propose that *A. methanolica* strain 239^T assimilates formaldehyde *via* the ribulose monophosphate (RuMP) cycle rather than by the serine pathway [76], as indicated by KEGG mapping of its whole genome sequence. In this pathway, formaldehyde is fixed with ribulose 5-phosphate (Ru5P) to form D-*arabino*-3-hexulose-6-phosphate (Hu6P) 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, FD-FA1DH/MscR, HPS or HPI), while the genomes of *A. japonica* strain MG417-CF17^T and *A. orientalis* strain HCCB10007 only encode genes for FD-FA1DH/MscR. The genomes of other AMS strains, namely *A. thermoflava* N1165^T 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 methylotrophic, mainly due to the absence of the RuMP cycle for assimilation of C1 intermediates. In contrast with ATS 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 benzoatilytica* AK 16/65^T (ARPK00000000.1), a member of the AOS subclade (Fig. 7B), a result which suggests it may be a facultative methylotroph.

As a facultative methylotroph, genomic annotation confirmed that *A. methanolica* strain 239^T is able to generate F6P from carbon sources other than C1 substrates *via* the RuMP cycle, such as glucose *via* glycolysis and the KDPGA (Entner-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-ketoglutarate dehydrogenase but, as in other *Amycolatopsis* strains 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^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 (CP000474.1) [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 *Pseudonocardiaceae*, namely *A. benzoatilytica* strain AK 16/56^T (ARPK00000000.1) and *Saccharomonospora marina* strain XMU15^T (CM001439.1), we identified the conserved clusters albeit with evidence of considerable rearrangements (Fig. 8C). In addition, although neither the genes of *hps* or *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 isogenic 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^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



	AMS			ATS	AOS			
	A. methanolica	*A.* sp. ATCC 39116 *	*A. thermoflava**	*A. taiwanensis**	*A. mediterranei*	*A. orientalis*	*A. japonica*	*A. benzoatilytica**
mdo	AMETH_5577	Y	Y	not found	N	N	N	Y
mscR	AMETH_3767	Y	Y	not found	N	AORI_4709	AJAP_16090_3150	Y
hpi	AMETH_4517	not found	not found	not found	N	N	N	Y
hps	AMETH_4419, AMETH_4538	not found	not found	not found	N	N	N	Y

Fig. 7. Schematic representation of the pathways for methanol, glucose and gluconate metabolism and for the biosynthesis of aromatic amino acids in *A. methanolica* strain 239^T (A) and comparison of the corresponding genes against those of other *Amycolatopsis* strains (B). *AMETH_5577* represents methanol:*N*,*N*-dimethyl-4-nitrosoaniline oxidoreductase, K17067, EC:1.2.99.4/EC:1.1.99.37; *AMETH_3767* NAD/mycothiol-dependent formaldehyde dehydrogenase, K00153, EC:1.1.1.306; *AMETH_3397* formate dehydrogenase, K00123, EC:1.2.1.2, and *AMETH_3398* formate dehydrogenase iron-sulfur subunit, K00124, EC:1.2.1.2. In **panel B**, **Y** denotes the identified genes, **N** denotes the unidentified genes in complete genomes, and **ND** denotes the genes that are not detected in the incomplete genomes.

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^T 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^T 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^T 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^T is unlikely to tolerate the toxic effect of formaldehyde while its ability to metabolize formic acid may be

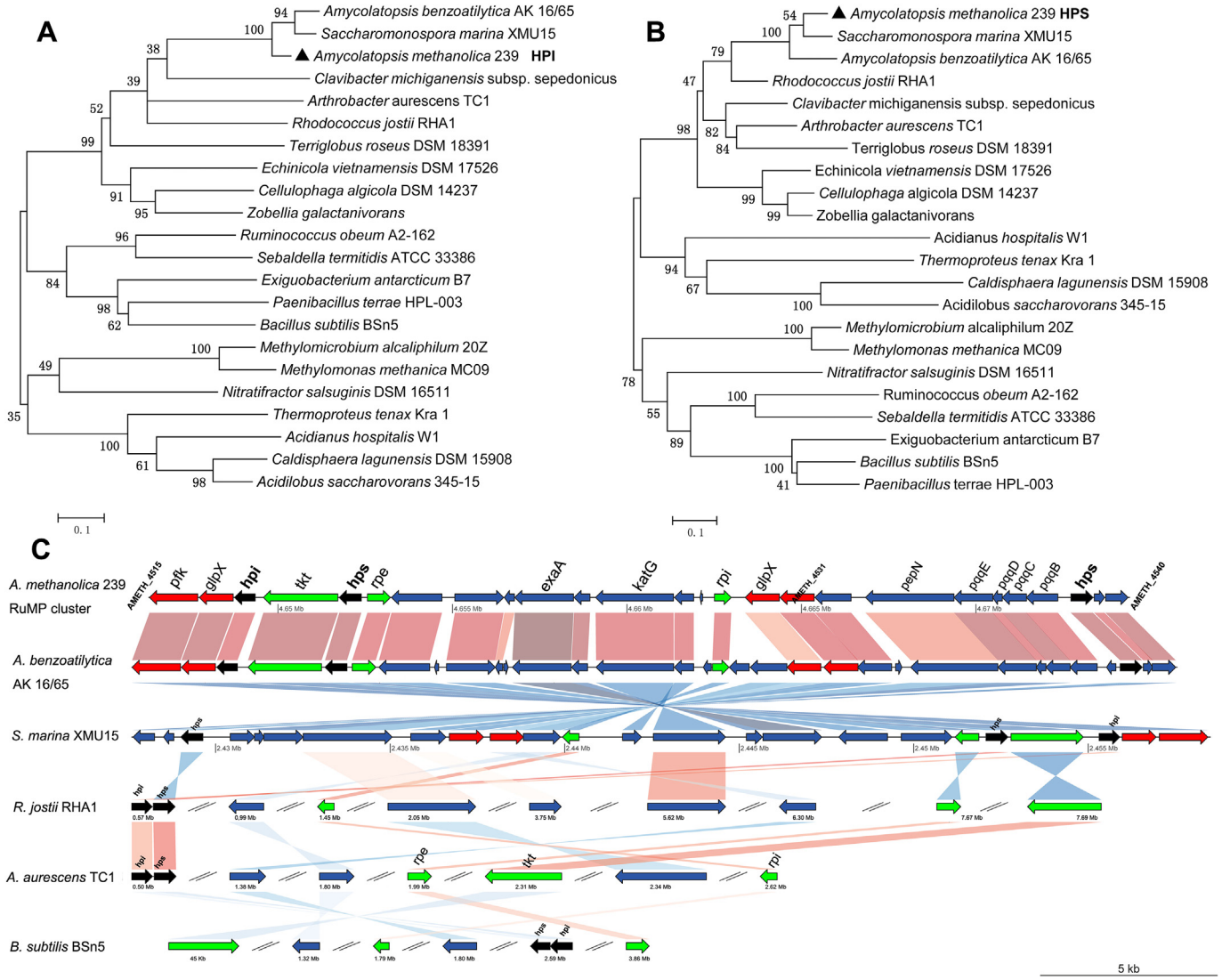


Fig. 8. Phylogenetic tree of HPI (A) and HPS (B) protein sequences and comparison of their encoding cluster in *A. methanolica* strain 239^T against those in the two *Pseudonocardiaceae* strains with the same two proteins (C). Maximum-likelihood tree **(A and B)** showing 6-phospho-3-hexuloisomerase (HPI) and 3-hexulose-6-phosphate synthase (HPS) in relation to the orthologous proteins selected from organisms which have available whole genome sequences and the CDSs for both HPS and HPI. The protein sequences of HPS and HPI from *A. methanolica* strain 239^T are marked by solid triangles. Numbers at the nodes are percentage bootstrap values based on a maximum-likelihood analysis of 1000 resampled datasets. Bar indicates 0.1 substitutions per site. **(C)** Comparison of chromosomal loci of genes involved in RuMP and related carbon metabolism from *A. methanolica* strain 239^T, *A. benzoatilytica* AK 16/56^T, *S. marina* XMU15, *R. jostii* RHA1, *Arthrobacter aurescens* TC1 and *B. subtilis* BSn5. The clusters from *A. methanolica*, *A. benzoatilytica* and *S. marina* strains are conserved. Genes marked in red are specific for the corresponding strains, those in green have corresponding orthologs in most of the actinobacteria, the *hps* and *hpi* genes are marked in black.

too weak to support healthy growth.

3. Conclusions and perspectives

A. methanolica strain 239^T (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 *Pseudonocardiaceae*, 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.

		A. methanolica (in RuMP cluster)	*A. methanolica* (not in RuMP cluster)	*A. mediterranei*	*A. orientalis*	*A. japonica*
Pentose phosphate pathway (complete)	*pfk*	*AMETH_4515* (ATP-*pfk*)	*AMETH_4897* (PPi-*pfk*)	*AMED_2460*	*AORI_2423*	*AJAP_26710*
	tkt		*AMETH_2245*, *AMETH_2246*			
				AMED_0651, *AMED_0652*		
			AMETH_5084, *AMETH_5085*	*AMED_2262*, *AMED_2263*	*AORI_5659*, *AORI_5660*	*AJAP_11180*, *AJAP_11185*
		AMETH_4518	*AMETH_4092*	*AMED_2809*	*AORI_2796*	*AJAP_24810*
	rpe	*AMETH_4520*	*AMETH_4130*	*AMED_2771*	*AORI_2761*	*AJAP_25030*
	rpi	*AMETH_4529*	*AMETH_1699*	*AMED_6811*	*AORI_1928*	*AJAP_29895*
					AORI_6417	*AJAP_06835*
			AMETH_2243, *AMETH_6524*		*AORI_7455*	*AJAP_38990*
	gnd		*AMETH_0006*	*AMED_0003*	*AORI_0003*	*AJAP_00020*
	talA		*AMETH_4093*	*AMED_2808*	*AORI_2795*	*AJAP_24815*
	pgi		*AMETH_4094*	*AMED_2807*	*AORI_2794*	*AJAP_24820*
	zwf		*AMETH_4095*	*AMED_2806*	*AORI_2793*	*AJAP_24825*
	pgl		*AMETH_4097*	*AMED_2804*	*AORI_2791*	*AJAP_24835*
Glycolysis/gluconeogenesis	*glpX*		*AMETH_5852*	*AMED_8059*	*AORI_6881*	*AJAP_04695*
		AMETH_4516 *AMETH_4530*				
	fbaA		*AMETH_6838* *AMETH_4655*	*AMED_9110*	*AORI_7850*	*AJAP_40945*
			AMETH_6121	*AMED_3923*	*AORI_5186*	*AJAP_13540*
		AMETH_4531				

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 ATS/F1 and AMS/F2 subclades is supported by differences in either the numbers or types of their transporters and biosynthetic genes involved in compatible solutes that are likely to be an expression of their adaption to hot saline environments.

3. This genomic study revealed the complete pathways for the utilization of methanol as a sole carbon source in *A. methanolica* strain 239^T 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 16/65^T, a member of the AOS subclade, may prove to be a facultative methylotroph as all four of the essential genes required for methanol utilization are encoded in its partially sequenced genome (ARPK00000000.1) (Fig. 7B). However, methylotrophy 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 239^T 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 239^T 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 v1.23 [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. RiPPs analysis were predicted by BAGEL3 [62].

Circular map, comparative analysis and assignment of common orthologues: Genome circular maps were generated using GenomeViz [97] and DNAplotter [98]. Comparative analyses of genomes from strains classified in the family *Pseudonocardiaceae* 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 *Acidimicrobium*, *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. S7), 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 *Gardnerella vaginalis* ATCC 14019 (NC_014644, HMPREF0421_20130, HMPREF0421_20607) while the *recN* gene of *Mycobacterium smegmatis* MC^2 155 (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 stains were obtained from the EzTaxon server. The MUSCLE algorithm [103] from MEGA5 [104] was used for sequence alignments. The neighbour-joining (NJ) [105] method in the MEGA5 software was used to construct the phylogenetic trees with *S. viridis* DSM 43017^T as an outgroup. The topology of the phylogenetic trees were evaluated using the bootstrap resampling method with 1000 repeats from the MEGA5 package.

4.1.2. Phylogenetic tree of 13 AOS, 1 ATS and 3 AMS and group J based on gyrB, 16S rDNA-gyrB, recN, 16S rDNA-recN gene sequences

Concatenated 16S rDNA-*gyrB and* 16S rDNA-*recN* gene sequences were generated by joining individual *gyrB* or *recN* gene sequences to the end of corresponding 16S rDNA sequences. The *gyrB* and *recN* gene sequences were full-length and were extracted from the genomes. The phylogenetic tree was constructed as described above.

4.1.3. Phylogenetic trees 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 MUSCLE3.5 [103]. All of the alignment files were joined in series and the results generated by Gblocks 0.9b [106]. The alignment format was transformed by Clustal X2.1 [107]. The phylogenetic trees were constructed using the maximum-likelihood (ML) method from 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 JNYY00000000, JMQI00000000, ARBH00000000, CP002000, ARPK00000000, JNYZ00000000, ARAF00000000, ANMG00000000, CP007219, AOHO00000000, CP008953, CP003410, JAFB00000000, AFWY00000000, CP009110, AXBH00000000, CP001683, AM420293, CP001630 and CP002593.

Experimental analysis for one-carbon source utilization by *A. methanolica* strain 239^T: The *A. methanolica* strain was grown on an GT plate (20 g/L starch, 0.5 g/L L-asparagine, 1 g/L KNO_3, 0.5 g/L $K_2HPO_4 \cdot 3H_2O$, 0.5 g/L NaCl, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 1 g/L $CaCO_3$, 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 eppendorf tubes, washed twice (centrifuged before resuspension) in HMM medium (prepared by mixing (1) 100 mL phosphate salts solution (25.3 g/L K_2HPO_4 and 22.5 g/L Na_2HPO_4), (2) 100 mL sulfate salts solution (5 g/L $(NH_4)_2SO_4$ and 2 g/L $MgSO_4 \cdot 7H_2O$), (3) 799 mL of deionized water, and (4) 1 mL of trace metal solution (0.177 g/L $ZnSO_4 \cdot 7H_2O$, 1.466 g/L $CaCl_2 \cdot 2H_2O$, 0.107 g/L $MnCl_2 \cdot 4H_2O$, 2.496 g/L $FeSO_4 \cdot 7H_2O$, 0.177 g/L $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, 0.374 g/L $CuSO_4 \cdot 5H_2O$, 0.238 g/L $CoCl_2 \cdot 6H_2O$ and 0.1 g/L $Na_2WO_4 \cdot 2H_2O$) 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 (OD_{600}) 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^T 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.

Authors' contributions

MG proposed the research project and GPZ conceived and designed the research strategy. BT and JW were responsible for sequencing, finishing and the annotations. BT performed and contributed to the annotation data processing and analysis. BT, FX, WZ, SWD, HCL, XMD, YCY, XFC, YZ and HKZ carried out the experiments and data analyses. BT, FX, MG and WZ were involved in drafting the manuscript. GPZ and LXZ supervised the experimental work and participated in preparing the manuscript. MG revised the article. All authors read and approved the final manuscript.

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 (31125002).

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 taiwanensis* 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
Ri5P	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.synbio.2016.05.001.

References

[1] Lechevalier M, Prauser H, Labeda D, Ruan J. Two new genera of nocardioform actinomycetes: *Amycolata* 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 *Actinosynnemataceae* Labeda and Kroppenstedt 2000 emend. Zhi et al. 2009 into an emended family *Pseudonocardiaceae* 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. Pseudonocardiaceae embley, Smida and Stackebrandt 1989, 205^{VP}, 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. *Pseudonocardiales* 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 C_1-compound in *Streptomyces* sp. no. 239 during growth on methanol. Agric Biol Chem 1977;41:29–34.
[8] Hazeu W, Debruyn JC, Vandijken JP. *Nocardia* sp. 239, a facultative methanol utilizer with the ribulose monophosphate pathway of formaldehyde fixation. Arch Microbiol 1983;135:205–10.
[9] Embley TM, Smida J, Stackebrandt E. Reverse transcriptase sequencing of 16S ribosomal RNA from *Faenia rectivirgula, Pseudonocardia thermophila* and *Saccharopolyspora hirsuta*, three wall type IV actinomycetes which lack mycolic acids. J Gen Microbiol 1988;134:961–6.
[10] de Boer L, Dijkhuizen L, Grobben G, Goodfellow M, Stackebrandt E, Parlett JH, et al. *Amycolatopsis methanolica* sp. nov., a facultatively methylotrophic actinomycete. Int J Syst Bacteriol 1990;40:194–204.
[11] Tan GYA, Goodfellow M, Genus V. Amycolatopsis Lechevalier, Prauser, Labeda and Ruan 1986, 34^{VP} emend. Lee 2009, 1403. 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. 1334–58.
[12] Tan GY, Ward AC, Goodfellow M. Exploration of *Amycolatopsis* diversity in soil using genus-specific primers and novel selective media. Syst Appl Microbiol 2006;29:557–69.
[13] Camas M, Sahin N, Sazak A, Sproer C, Klenk HP. *Amycolatopsis magusensis* sp. nov., isolated from soil. Int J Syst Evol Microbiol 2013;63:1254–60.
[14] Everest GJ, le Roes-Hill M, Omorogie C, Cheung SK, Cook AE, Goodwin CM, et al. *Amycolatopsis umgeniensis* sp. nov., isolated from soil from the banks of the Umgeni River in South Africa. Ant Van Leeuwenhoek 2013;103:673–81.
[15] Xing K, Liu W, Zhang YJ, Bian GK, Zhang WD, Tamura T, et al. *Amycolatopsis jiangsuensis* sp. nov., a novel endophytic actinomycete isolated from a coastal plant in Jiangsu, China. Ant Van Leeuwenhoek 2013;103:433–9.
[16] Everest GJ, Meyers PR. The use of *gyrB* sequence analysis in the phylogeny of the genus *Amycolatopsis*. Ant Van Leeuwenhoek 2009;95:1–11.
[17] Everest GJ, Cook AE, Kirby BM, Meyers PR. Evaluation of the use of *recN* sequence analysis in the phylogeny of the genus *Amycolatopsis*. Ant Van Leeuwenhoek 2011;100:483–96.
[18] Brigham RB, Pittenger RC. *Streptomyces orientalis*, n. sp., the source of vancomycin. Antibiot Chemother N 1956;6:642–7.
[19] Margalith P, Beretta G. Rifomycin. XI. Taxonomic study on *Streptomyces mediterranei* nov. sp. Mycopathol Mycol Appl 1960;13:321–30.
[20] Zucchi TD, Tan GY, Goodfellow M. *Amycolatopsis thermophila* sp. nov. and *Amycolatopsis viridis* sp. nov., thermophilic actinomycetes isolated from arid soil. Int J Syst Evol Microbiol 2012;62:168–72.
[21] Zucchi TD, Tan GY, Bonda AN, Frank S, Kshetrimayum JD, Goodfellow M. *Amycolatopsis granulosa* sp. nov., *Amycolatopsis ruanii* sp. nov. and *Amycolatopsis thermalba* sp. nov., thermophilic actinomycetes isolated from arid soils. Int J Syst Evol Microbiol 2012;62:1245–51.
[22] Zucchi TD, Bonda AN, Frank S, Kim BY, Kshetrimayum JD, Goodfellow M. *Amycolatopsis bartoniae* sp. nov. and *Amycolatopsis bullii* sp. nov., mesophilic actinomycetes isolated from arid Australian soils. Ant Van Leeuwenhoek 2012;102:91–8.
[23] Chun J, Kim SB, Oh YK, Seong CN, Lee DH, Bae KS, et al. *Amycolatopsis thermoflava* sp. nov., a novel soil actinomycete from Hainan Island, China. Int J Syst Bacteriol 1999;49(Pt 4):1369–73.
[24] Kim B, Sahin N, Tan GY, Zakrzewska-Czerwinska J, Goodfellow M. *Amycolatopsis eurytherma* sp. nov., a thermophilic actinomycete isolated from soil. Int J Syst Evol Microbiol 2002;52:889–94.
[25] Albarracin VH, Alonso-Vega P, Trujillo ME, Amoroso MJ, Abate CM. *Amycolatopsis tucumanensis* sp. nov., a copper-resistant actinobacterium isolated from polluted sediments. Int J Syst Evol Microbiol 2010;60:397–401.
[26] Cross T. Thermophilic actinomycetes. J Appl Bacteriol 1968;31:36–53.
[27] Brock T. Introduction: an overview of the thermophiles. In: Brock T, editor. Thermophiles: general, molecular and applied microbiology. New York: Wiley; 1986. p. 1–17.
[28] Abou-Zeid A, Euverink G, Hessels GI, Jensen RA, Dijkhuizen L. Biosynthesis of l-phenylalanine and l-tyrosine in the actinomycete *Amycolatopsis methanolica*. Appl Environ Microbiol 1995;61:1298–302.
[29] Albarracin VH, Winik B, Kothe E, Amoroso MJ, Abate CM. Copper bioaccumulation by the actinobacterium *Amycolatopsis* sp. AB0. J Basic Microbiol 2008;48:323–30.

[30] Girard G, Traag BA, Sangal V, Mascini N, Hoskisson PA, Goodfellow M, et al. A novel taxonomic marker that discriminates between morphologically complex actinomycetes. Open Biol 2013;3:130073.

[31] Girard G, Willemse J, Zhu H, Claessen D, Bukarasam K, Goodfellow M, et al. Analysis of novel kitasatosporae reveals significant evolutionary changes in conserved developmental genes between *Kitasatospora* and *Streptomyces*. Ant Van Leeuwenhoek 2014;106:365–80.

[32] Stegmann E, Albersmeier A, Spohn M, Gert H, Weber T, Wohlleben W, et al. Complete genome sequence of the actinobacterium *Amycolatopsis japonica* MG417-CF17^T (=DSM 44213^T) producing (S,S)-N, N '-ethylenediaminedisuccinic acid. J Biotechnol 2014;189:46–7.

[33] 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;20:1096–108.

[34] 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.

[35] Vrijbloed JW, Madon J, Dijkhuizen L. A plasmid from the methylotrophic actinomycete *Amycolatopsis methanolica* capable of site-specific integration. J Bacteriol 1994;176:7087–90.

[36] Vrijbloed JW, Jelinkova M, Hessels GI, Dijkhuizen L. Identification of the minimal replicon of plasmid pMEA300 of the methylotrophic actinomycete *Amycolatopsis methanolica*. Mol Microbiol 1995;18:21–31.

[37] Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, et al. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL 23338. Nat Biotechnol 2007;25:447–53.

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

[39] Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol 2012;62:716–21.

[40] Tamura T, Ishida Y, Otoguro M, Suzuki K. *Amycolatopsis helveola* sp. nov. and *Amycolatopsis pigmentata* sp. nov., isolated from soil. Int J Syst Evol Microbiol 2010;60:2629–33.

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

[42] Miao Q, Qin S, Bian GK, 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.

[43] Trúper HG, Galinski EA. Biosynthesis and fate of compatible solutes in extremely halophilic phototrophic eubacteria. FEMS Microbiol Lett 1990;75:247–54.

[44] da Costa MS, Santos H, Galinski EA. An overview of the role and diversity of compatible solutes in *Bacteria* and *Archaea*. Adv Biochem Eng Biotechnol 1998;61:117–53.

[45] Kempf B, Bremer E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch Microbiol 1998;170:319–30.

[46] Welsh DT. Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. FEMS Microbiol Rev 2000;24:263–90.

[47] Roessler M, Muller V. Osmoadaptation in bacteria and archaea: common principles and differences. Environ Microbiol 2001;3:743–54.

[48] Paul MJ, Primavesi LF, Jhurreea D, Zhang Y. Trehalose metabolism and signaling. Annu Rev Plant Biol 2008;59:417–41.

[49] Tzvetkov M, Klopprogge C, Zelder O, Liebl W. Genetic dissection of trehalose biosynthesis in *Corynebacterium glutamicum*: inactivation of trehalose production leads to impaired growth and an altered cell wall lipid composition. Microbiology 2003;149:1659–73.

[50] Belocopitow E, Marechal LR. Trehalose phosphorylase from *Euglena gracilis*. Biochim Biophys Acta 1970;198:151–4.

[51] Avonce N, Mendoza-Vargas A, Morett E, Iturriaga G. Insights on the evolution of trehalose biosynthesis. BMC Evol Biol 2006;6:109.

[52] Ruhal R, Kataria R, Choudhury B. Trends in bacterial trehalose metabolism and significant nodes of metabolic pathway in the direction of trehalose accumulation. Microb Biotechnol 2013;6:493–502.

[53] Marchandin H, Teyssier C, de Buochberg MS, Jean-Pierre H, Carriere C, Jumas-Bilak E. Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: implications for phylogeny and taxonomy. Microbiology 2003;149:1493–501.

[54] Engene N, Gerwick WH. Intra-genomic 16S rRNA gene heterogeneity in cyanobacterial genomes. Fottea 2011;11:17–24.

[55] 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 periodontal abscesses. Plos One 2015;10.

[56] Munoz R, Yarza P, Ludwig W, Euzeby J, Amann R, Schleifer KH, et al. Release LTPs104 of the all-species living tree. Syst Appl Microbiol 2011;34:169–70.

[57] Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41:D590–6.

[58] Tang SK, Wang Y, Guan TW, Lee JC, Kim CJ, Li WJ. *Amycolatopsis halophila* sp. nov., a halophilic actinomycete isolated from a salt lake. Int J Syst Evol Microbiol 2010;60:1073–8.

[59] Guan TW, Xia ZF, Tang SK, Wu N, Chen ZJ, Huang Y, et al. *Amycolatopsis salitolerans* sp. nov., a filamentous actinomycete isolated from a hypersaline habitat. Int J Syst Evol Microbiol 2012;62:23–7.

[60] Xie F, Dai SW, Shen JZ, Ren B, Huang P, Wang QS, et al. A new salicylate synthase AmS is identified for siderophores biosynthesis in *Amycolatopsis methanolica* 239^T. Appl Microbiol Biotechnol 2015;99:5895–905.

[61] Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, et al. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 2015;43:W237–43.

[62] van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic Acids Res 2013;41:W448–53.

[63] Challis GL. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. Chembiochem 2005;6:601–11.

[64] Davis JR, Goodwin LA, Woyke T, Teshima H, Bruce D, Detter C, et al. Genome sequence of *Amycolatopsis* sp. strain ATCC 39116, a plant biomass-degrading actinomycete. J Bacteriol 2012;194:2396–7.

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

[66] Kaur N, Kumar S, Bala M, Raghava GP, Mayilraj S. Draft genome sequence of *Amycolatopsis decaplanina* strain DSM 44594^T. Genome Announc 2013;1:e0013813.

[67] Kwun MJ, Hong HJ. Draft genome sequence of *Amycolatopsis lurida* NRRL 2430, producer of the glycopeptide family antibiotic ristocetin. Genome Announc 2014;2.

[68] Garcia-Estepa R, Argandona M, Reina-Bueno M, Capote N, Iglesias-Guerra F, Nieto JJ, et al. The ectD 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;188:3774–84.

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

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

[71] Caspi R, Altman T, Billington 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:D459–71.

[72] Van Ophem PW, Van Beeumen J, Duine JA. Nicotinoprotein [NAD(P)-containing] alcohol/aldehyde oxidoreductases. Purification and characterization of a novel type from *Amycolatopsis methanolica*. Eur J Biochem 1993;212:819–26.

[73] Bystrykh LV, Vonck J, van Bruggen EF, van Beeumen J, Samyn B, Govorukhina NI, et al. Electron microscopic analysis and structural characterization of novel NADP(H)-containing methanol: N,N'-dimethyl-4-nitrosoaniline oxidoreductases from the gram-positive methylotrophic bacteria *Amycolatopsis methanolica* and *Mycobacterium gastri* MB19. J Bacteriol 1993;175:1814–22.

[74] Norin A, Van Ophem PW, Piersma SR, Persson B, Duine JA, Jornvall H. Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaryotic alcohol dehydrogenases–primary structure, conformational modelling and functional correlations. Eur J Biochem 1997;248:282–9.

[75] Van Ophem PW, Bystrykh LV, Duine JA. Dye-linked dehydrogenase activities for formate and formate esters in *Amycolatopsis methanolica*. Characterization of a molybdoprotein enzyme active with formate esters and aldehydes. Eur J Biochem 1992;206:519–25.

[76] Vorholt JA. Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. Arch Microbiol 2002;178:239–49.

[77] Newton GL, Arnold K, Price MS, Sherrill C, Delcardayre SB, Aharonowitz Y, et al. Distribution of thiols in microorganisms: mycothiol is a major thiol in most actinomycetes. J Bacteriol 1996;178:1990–5.

[78] Newton GL, Buchmeier N, Fahey RC. Biosynthesis and functions of mycothiol, the unique protective thiol of actinobacteria. Microbiol Mol Biol Rev 2008;72:471–94.

[79] Jez JM, Cahoon RE. Kinetic mechanism of glutathione synthetase from *Arabidopsis thaliana*. J Biol Chem 2004;279:42726–31.

[80] Mongodin EF, Shapir N, Daugherty SC, DeBoy RT, Emerson JB, Shvartzbeyn A, et al. Secrets of soil survival revealed by the genome sequence of *Arthrobacter aurescens* TC1. PLoS Genet 2006;2:e214.

[81] McLeod MP, Warren RL, Hsiao WW, Araki N, Myhre M, Fernandes C, et al. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci U. S. A 2006;103:15582–7.

[82] Jones AL, Davies J, Fukuda M, Brown R, Lim J, Goodfellow M. *Rhodococcus jostii*: a home for *Rhodococcus* strain RHA1. Ant Van Leeuwenhoek 2013;104:435–40.

[83] Deng Y, Zhu Y, Wang P, Zhu L, Zheng J, Li R, et al. Complete genome sequence of *Bacillus subtilis* BSn5, an endophytic bacterium of *Amorphophallus konjac* with antimicrobial activity for the plant pathogen *Erwinia carotovora* subsp. carotovora. J Bacteriol 2011;193:2070–1.

[84] Alves AM, Euverink GJ, Santos H, Dijkhuizen L. Different physiological roles

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;14:289.
[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]:Unit11 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 2005;33:W451–4.
[93] Song K. Recognition of prokaryotic promoters based on a novel variable-window Z-curve method. Nucleic Acids Res 2012;40:963–71.
[94] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 2006;34:D354–7.
[95] Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 2005;33:W686–9.
[96] 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.
[97] Ghai R, Hain T, Chakraborty T. GenomeViz: visualizing microbial genomes. BMC Bioinforma 2004;5:198.
[98] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[99] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.
[100] Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 2010;26:2334–5.
[101] 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.
[102] Elizabeth Cha I, Rouchka EC. Comparison of current BLAST software on nucleotide sequences. IPDPS 2005;19:8.
[103] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.
[104] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–9.
[105] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–25.
[106] Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 2007;56:564–77.
[107] 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.
[108] Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006;22:2688–90.
[109] KIM S. Genus I. *Thermobispora* Wang, Zhang and Ruan 1996, 937^{VP}. 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. 1967.