Review Article

Progress in Microbes and Molecular Biology

HH PUBLISHER

Taxonomic and Characterization Methods of *Streptomyces*: A Review

Jodi Woan-Fei Law[1], Kei-Xian Tan[1], Sunny Hei Wong[2], Nurul-Syakima Ab Mutalib[3*], Learn-Han Lee[1,4,5*]

[1]Novel Bacteria and Drug Discovery Research Group, Microbiome and Bioresource Research Strength, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, Selangor Darul Ehsan, Malaysia
[2] Li Ka Shing Institute of Health Sciences, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, Hong Kong
[3]UKM Medical Molecular Biology Institute (UMBI), UKM Medical Centre, University Kebangsaan Malaysia, Kuala Lumpur, Malaysia
[4]Biofunctional Molecule Exploratory Research Group, Biomedicine Research Advancement Centre, School of Pharmacy, Monash University Malaysia, Selangor Darul Ehsan, Malaysia
[5]Center of Health Outcomes Research and Therapeutic Safety (Cohorts), School of Pharmaceutical Sciences, University of Phayao, Phayao, Thailand

Abstract : The 16S ribosomal RNA gene is the gold standard for taxonomic identification and phylogenetic study of most of the bacteria. However, the resolution of 16S rRNA gene is found to be insufficient to distinguish closely related species within the genus *Streptomyces* due to large size of *Streptomyces* member. Thus, it is essential to find alternative phylogenetic gene markers with higher discriminatory power in addition to 16S rRNA gene for the phylogenetic analysis of *Streptomyces* species in order to effectively indicate the evolutionary relationships among *Streptomyces* members at intra- and inter-specific level. This article aims to discuss the resolution power of ribosomal genes (16S rRNA and 23S rRNA gene), of protein-coding genes (*gyrB* and *trpB* gene) and between ribosomal and protein-coding genes in order to identify gene that could provide better resolution for phylogenetic studies of *Streptomyces*.

Keywords: *Streptomyces*; actinobacteria; taxonomic; review; method

*Correspondence to:
Learn-Han Lee, Novel Bacteria and Drug Discovery Research Group, Microbiome and Bioresource Research Strength, Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia; lee.learn.han@monash.edu. Nurul-Syakima Ab Mutalib, UKM Medical Molecular Biology Institute (UMBI), UKM Medical Centre, University Kebangsaan Malaysia, Kuala Lumpur, Malaysia; syakima@ppukm.ukm.edu.my

Received: 8th October 2018
Accepted: 15th November 2018
Published Online: 14th December 2018

***Citation*:** Law JWF, Tan KX, Wong SH, *et al*. Taxonomic and Characterization Methods of Streptomyces: A Review. Prog Microbes Mol Biol 2018; 1(1): a0000009.

INTRODUCTION

Members of *Actinobacteria* consist of different metabolic and physiological characteristics while expressing different morphological characteristics such as coccoid and fragmenting hyphalor branched mycelium forms[1]. *Actinobacteria* possess various lifestyles, for instances, as soil inhabitants (example: *Streptomyces*), gastrointestinal commensals (example: *Bifidobacterium*) and pathogens (example: *Mycobacterium* and *Corynebacterium*)[2]. *Streptomyces* is the largest genus located within the family *Streptomycetaceae*, belonging to the phylum *Actinobacteria* [2]. *Streptomyces* bacteria are biologically active at which they can produce secondary metabolites that have various biological effects[3]. Currently, there are over 10 000 bioactive compounds have been derived from *Streptomyces* [3-5].

Prior to the late 1950s, it was believed that members of *Streptomyces* are eukaryotes instead of prokaryotes due to their distinctive life cycle which is similar to the life cycle of multicellular eukaryotes [6]. As a matter of fact, *Streptomyces* consists of multicellular, Gram positive, filamentous aerobic, and mycelial bacteria that live mainly in soil as saprophytes[7]. These bacteria have special features such as complex life cycle involving the development of substrate mycelium with extensive branches containing LL-diaminopimelic acid as main diamino acid and carries aerial hyphae that are able to differentiate into spores or arthrospores[8,9]. Spores allow extended survival of *Streptomyces* species in an inactive form in the soil because they are known to be resistant to both water and nutrients deficiencies as well as high temperature in unfavourable environments[7]. Another important feature of *Streptomyces* includes

the generation of diverse pigments that form the colour of aerial and substrate mycelia[10]. Interestingly, some *Streptomyces* bacteria exist as marine symbionts, plant root symbionts in rhizophere, growing on gamma-irradiated surfaces or thermal springs[11]. Meanwhile, some *Streptomyces* bacteria are human, animal or plant pathogens such as *Streptomyces scabies* that stimulates potato scab[12].

The genus *Streptomyces* within the order *Actinomycetales* has spectacular diversity in morphology, genomic size, genomic G+C content, and amount of coding sequences. *Streptomyces* bacteria are noted by their large and linearized chromosomes with approximately 8.5-12 Mb in size and high GC content of about 67-78%[3,9,13]. The large genome of *Streptomyces* could account for their ability to produce many secondary metabolites[14]. Several studies have reported the discovery of over 20 biosynthetic gene clusters in the genome of *Streptomyces* which could be associated with the production of secondary metabolites[4]. For examples, the genome of *Streptomyces coelicolor* A3(2) was found to contain more than 20 gene clusters related to numerous metabolites[15]; genome of *Streptomyces avermitilis* was found to contain more than 30 gene clusters related to numerous metabolites[16].

IMPORTANCE OF *Streptomyces*

Being an important soil microbial population, *Streptomyces* has been extensively screened and isolated over the years due to its production of natural biologically active secondary metabolites that are medically and commercially important[4]. It has been an ongoing effort for the screening of novel *Streptomyces* since the discovery of streptomycin and actinomycin from *Streptomyces griseus* and *Streptomyces antibioticus* respectively[17,18]. There are over 75 % of useful antibiotics that are commercially available for the use in both human and veterinary medicines have been produced by various *Streptomyces* species[19]. Many antibiotics belonging to different classes have been derived from *Streptomyces*. For instances: chloramphenicol (chloramphenicol) from *Streptomyces venezuelae*, neomycin (aminoglycoside) from *Streptomyces fradiae*, clavulanic acid (β-lactam) from *Streptomyces clavuligerus*, vancomycin (glycopeptide) from *Streptomyces orientalis*, tetracycline (tetracycline) from *Streptomyces aureofaciens*, and nystatin (polyene-polyol macrolide) from *Streptomyces noursei*[20-22].

Other than the capacity of producing antibiotics, *Streptomyces* bacteria are producers of many other bioactive compounds with high functional and structural diversity such as anti-parasitic, anticancer/antitumour, antifungal, biocatalysts, biopesticides, immunosuppressive, and herbicides biological control agents[23-26]. They are also producers of enzymes which are important in environmental, food biotechnology and other industries[27]. *Streptomyces* produces wood or fungal cell wall hydrolytic enzymes which include hemicellulases, chitinases, glucanases and cellulases to decompose a variety of plant-based polysaccharides, insoluble polymers like chitin and cellulose[28]. They are also involved in biodegradation and bioremediation at which they can degrade lignin and lignin-related aromatic compounds[28,29].

It is important to continue searching for novel *Streptomyces* from different habitats in hope to find novel secondary metabolites such as new antibiotics to combat against pathogens which have gained resistance towards existing antibiotics via chromosomal mutation, vertical or horizontal gene transfer[30]. It is also essential to search for effective antimicrobial compounds to fight against infectious disease agents[31]. Studies indicated that researchers have only discovered a fraction of antibiotics generated by *Streptomyces* species. Hence, it is still possible to isolate novel antibiotics from terrestrial *Streptomyces* despite of being screened continually over the past half century[21]. Nevertheless, exploring novel *Streptomyces* from other areas could lead to the findings of other new/interesting compounds[32]. Many studies have shown the successful isolation of novel *Streptomyces* species from different environments including marine sediments[33], lakes[34], caves[35], mangroves[36], deserts[37], peat swamps[38], and plants[39]. Researchers also found that many novel *Streptomyces* species possess important bioactivities such as antioxidant, anticancer, and antifungal[40,41].

BIOACTIVE METABOLITES FROM *Streptomyces*

Secondary metabolites are different from primary metabolites as they are not essential for growth, reproduction and other metabolisms in the cell[42]. They played an important role in the survival and defense mechanism for their producers to compete in a stressful environment which has a high amount of other metabolically active bacteria[32]. Hence, prolific production of secondary metabolites is an advantage trait of *Streptomyces*.

Streptomyces produces secondary metabolites when encountering nutritional stresses from the surrounding environment and the production of secondary metabolites is referred as physiological differentiation[43]. There is a close association between morphological differentiation and physiological differentiation in the life cycle of *Streptomyces*[44]. It is apparent that morphological differentiation is related with the sensing of nutrient deficiencies and environmental stresses[8]. Therefore, the generation of various secondary metabolites such as antibiotics or antifungals are initiated during the morphological transition of *Streptomyces* from vegetative to aerial growth stage[44]. In most cases, production of antibiotics is depending on the spatial and temporal control of morphogenesis, metabolism, gene expression and metabolites flux[45]. In fact, bioactive secondary metabolites are produced by metabolic pathways that can be linked to primary metabolism; both secondary and primary metabolite pathways share intermediates, also, the intermediate or end products of primary metabolism are usually the precursors used to synthesize secondary metabolites[31,46]. Moreover, the composition of the culture medium will also affect the capacity of *Streptomyces* species in producing bioactive compounds[4,42].

Bioactive secondary metabolites are synthesized extra-

cellularly, which can be isolated and purified with highest purity from fermentation broth by using a combination of different extraction and purification methods. Often, the methods utilized for extraction and purification of secondary metabolites are dependent on their applications. These methods include simple processes such as solvent extraction, chemical precipitation, and processes utilizing advance instruments such as ion-exchange chromatography, high performance liquid chromatography (HPLC) and many more [4,31]. Sometimes, these secondary metabolites may be obtained in the form of crude extracts which can be further subjected to chemical profiling using liquid chromatography-mass spectrometry (LC-MS) or gas chromatography-mass spectrometry (GC-MS) and thus allowing the prediction of compounds present in the extracts prior to purification [47,48].

TAXONOMIC AND CHARACTERIZATION METHODS OF *Streptomyces*

Due to large-scale isolation and screening over the years, there were nearly 3000 species of *Streptomyces* named by year 1970 with most of them identified only on patent literature in the past[49]. Most organisms were given a new name solely based on small variations in cultural and morphological characteristics or for the sake of meeting publications or patent requirements. The issues arisen from this situation include several confusions in the criteria for *Streptomyces* speciation, risks of misclassificaiton as well as concerns on the competency of the individual who describing the new organism as reported by Trejo (1970)[50]. Fortunately, amendments on their taxonomy have been carried out to address these issues [49,50]. To date, *Streptomyces* genus consists of 844 species and 38 subspecies that are validly described (www.bacterio.net).

Taxonomic characterization of *Streptomyces* is undeniably more complicated and challenging as a result of the large amount of described species in the genus *Streptomyces* which is relatively greater as compared to that of other microbial genera[51]. The techniques used for characterization of *Streptomyces* have improved over time; from classical morphological classifications such as spore chain morphology, colour of substrate and aerial mycelia to numerical taxonomic analyses that include phenotypic characterization based on standardized sets and, now, applying molecular and phylogenetic analyses[51,52].

International *Streptomyces* Project (ISP), a traditional approach

International *Streptomyces* Project (ISP) is one of the approaches for the classification of *Streptomyces* strains since 1964 with the objective of producing reliable characterization of existing type strains of both *Streptomyces* and *Streptoverticillium* species. ISP was amended and enacted by the Subcommittee on Taxonomy of *Actinomycetes* of the Committee on Taxonomy of the American Society of Microbiology and the Subcommittee on Taxonomy of *Actinomycetes* of the International Committee on Bacteriological Nomenclature[51,53]. ISP works by sending type and neotype strains of species under these genera to a minimum of three experts in different nations to be investigated under strict standardized experimental and media settings for the identification of morphological characteristics, pigmentation, and carbon utilization profiles of the strains[53,54]. The protocols and outcomes of ISP studies were previously published and the resulting identified type strains were deposited in a few service collections which are internationally recognized[53,55,56].

Although ISP has always been a standard method for the classification and characterization of *Streptomyces*, however, it is not the best approach since it is heavily depending on limited number of selected features or standard phenotypic criteria that are highly based on morphological and pigmentation characteristics[51]. Hence, other methods have been established to further assist in the classification and characterization of bacteria.

Molecular characterization of *Streptomyces*, a modern approach

Identification and characterization methods have evolved to molecular and phylogenetic characterizations with analysis of gene sequences which targets predominantly linear 16S rRNA gene sequences due to the advent of polymerase chain reaction (PCR), DNA-DNA hybridization (DDH) and DNA sequencing approaches[57]. Molecular revolution in the last 30 years allowed not only further insights into molecular cell biology but also brought huge improvement to the study of evolution, conservation and ecology. In other words, genetic techniques enable taxonomic identification of bacteria which can hardly be delineated based on morphological characteristics alone, accompanied with higher efficiency in terms of rapid and high-throughput identification[58].

Molecular characterization - polymerase chain reaction (PCR)

PCR is invented in 1980s by the American biochemist, Kary Mullis, as a revolutionary method[59]. The invention of PCR plays a crucial role in improving the knowledge of DNA evolution and phylogenetic relationships among *Streptomyces* species. An adequate amount of a particular gene or DNA fragment is needed to enable the detection and identification of gene sequence. To achieve this practical amount of DNA, the application of PCR is able to specifically and rapidly amplify any DNA stretch within size limits which is flanked by synthetic oligonucleotide known as primers to produce up to millions or billions of copies of a particular nucleotide sequence in approximately two hours[60]. These primers with approximately 20 base pairs in length will determine the 5'-3' end of the target region of the double-stranded DNA which is the target of amplification. Primers are used to initiate DNA synthesis while heat-stable DNA polymerase such as *Taq* polymerase is used to produce a new DNA strand enzymatically from single-stranded DNA template by adding nucleotides. PCR amplification is operated in a thermal cycler that

consists of repeated cycles of heating and cooling reaction at a series of specific temperature to melt and replicate the target DNA sequence[60,61].

There are three major steps in a PCR amplification process: denaturation, annealing, and extension[61,62]. The PCR reaction begins when high temperature (E.g. 90-97 ºC) is used to melt and separate the double-stranded DNA molecule for one to several minutes to form single-stranded DNA. Next, primers are used to anneal to the DNA template strands at a lower temperature of 50-65ºC. Basically, the annealing temperature is approximately 3-5 ºC lower than the lowest melting temperature of primers pair used. DNA polymerase then extents and synthesizes new complementary DNA strand at the end of the annealed primers at approximately 72 ºC for 2 to 5 minutes. Consequently, the original DNA is duplicated at which each of the new DNA molecules are made up of one new and one old DNA strand. These two new DNA strands can then be used as DNA template to produce more copies in coming cycle. This cycle of denaturing, annealing, extending and synthesizing is repeated up to 30 or 40 times to produce up to billions of identical copies of the original DNA sequence. PCR is widely used in laboratories for many applications including sequencing, genetic engineering and cloning[61]. The application of PCR is certainly useful for the rapid identification of an organism at genus or species level, depending on the primers used. Many studies have been utilizing PCR for the identification of the genus *Streptomyces*, especially via the detection of 16S rRNA gene[63-65].

Molecular characterization - BOX-PCR

BOX-PCR is a repetitive element sequence-based PCR (rep-PCR) fingerprinting technique which possesses high discriminatory power and can be used for typing purpose[66]. It is based on the BOX dispersed-repeat motif, initially found in *Streptococcus pneumonia* but appears to be common in most of the bacterial species. This 154 base pair(bp)-box elements comprising 3 subunits (*boxA* 59bp, *boxB* 45bp, and *boxC* 50bp) represent the naturally occurring, highly conserved and repetitive sequence elements present in multiple copies which are found in most Gram-negative and some Gram-positive bacterial genomes[66,67]. BOX-PCR can simultaneously survey many DNA regions scattered in the genome of bacteria because the BOX repetitive sequences are interspersed throughout the bacterial genome[68]. The primer used in BOX-PCR (BOX A1R primer) targets and amplifies selectively the regions located between BOX elements, thereby resulting in DNA amplification products of different sizes which serve as DNA fingerprints of the bacteria. These amplified fragments will then be resolved using gel electrophoresis to produce BOX-PCR genomic fingerprinting profile for differentiation at species, subspecies, and strain level[67,68]. BOX-PCR is inexpensive, efficient and easy to perform, concurrently offering equivalent or superior discriminatory power as compared to other methods such as restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP)[68].

Molecular characterization - DNA-DNA hybridization (DDH)

DDH is a DNA-based molecular technique used to indicate genetic distance or genetic similarity between species, particularly between those closely related species, through genome-wide comparisons. In the 1960s, DDH has been employed by bacterial taxonomists and it is well known as the gold standard for species delineation and for taxonomic evaluation of strain [69-71]. DNA-DNA reassociation method is often conducted in conjunction with other tests such as morphological, physiological and phylogenetic tests [70]. DDH is a process where two DNA molecules from different biological sources or organisms are mixed and heated to form single-stranded DNA before cooling step. The DNA of one of these sources is label while another is unlabeled. These single-stranded DNA are then allowed to reanneal to reform the hybrid double helices or DNA-DNA hybrids based on contingent sequence homology of the two organisms. Besides that, values of DDH are usually expressed in percentage to quantitatively estimate the genome-wide similarity of the DNA sequence. For instance, two strains that are from same species will have a high value of hybridization which is typically more than 70% while two strains of different species but same genus would have around 30-60% of hybridization value. If the unlabeled DNA is from the same organism as the labelled DNA then the hybridization value would be 100% [72]. DDH analysis is one of the assays included in polyphasic taxonomy which is performed in taxonomic studies involving the delineation of taxa at all levels[73]. The application of DDH can be seen in many recent studies pertaining the discovery of novel *Streptomyces*, for instances, discovery of *Streptomyces mangrovisoli* sp. nov.[74], *Streptomyces humi* sp. nov.[75], and *Streptomyces gilvigriseus* sp. nov.[76]. Henceforth, emphasizing the importance of DDH as a criterion for the delineation of bacterial species despite being time-consuming and labor-intensive[69].

Molecular characterization – next generation sequencing (NGS)

The rapid evolution of new molecular technologies leading to the development of DNA sequencing technologies has revolutionized biological and biomedical related research. These studies assist scientists in many applications including molecular cloning, search of pathogenic genes, comparative and evolutionary studies[77]. In this context, the sequencing of bacterial genome has becoming more common and being routinely applied in many microbial studies. During the early 1990s, DNA sequencing is based on the Sanger's method, also identified as dideoxy chain termination sequencing chemistries. However, the Sanger's "first generation sequencing" method has eventually reached a plateau in terms of technical development, along with several limitations faced by this technique when comes to whole genome sequencing such as time consuming and resource intensive[78-80].

In 2005, the introduction of "next generation sequenc-

ing" (NGS) technologies has offer more advantages than Sanger's method in the aspects of producing massive volume of data at rapid high-throughput and cost-effective manners[77,78]. Besides, the latest NGS technologies operate differently than the Sanger's method as they employ various massively parallel sequencing instruments that are commercially available, for examples, 454 (Roche), Solexa (Illumina), and SOLiD (Applied Biosystems) platforms[81]. All of these instruments have their respective work flows, but they generally share some similar features: (1) simplified initial preparatory steps, (2) library preparation by DNA fragmentation and amplification, and (3) sequencing reaction is conducted followed by automated detection using imaging systems[78,81,82]. With the availability of numerous massively parallel sequencing platforms and the dramatic decrease in the costs for sequencing, it is anticipated that NGS will be routinely used for the whole genome sequencing of bacterial genome, hence, enables the extraction of biologically or clinically useful insights from the genome sequence[32,81].

TARGET GENES FOR *Streptomyces* CHARACTERIZATION

Ribosomal RNA genes

The taxonomic classification of *Streptomyces* species based on morphological, biochemical, nutritional and physiological characterizations are challenging and often problematic [51,83,84]. Therefore, molecular approaches including phylogenetic analysis can be an improvement in evolutionary and diversity studies. The rRNA gene sequences which code for the ribosomal subunits such as 16S, 23S and 5S rRNA are the target of phylogenetic analysis due to the highly conserved rRNA structure in all cells throughout evolution[51]. Majority of the rRNA folding have important functions regardless of the divergence in primary sequence. Precise spatial relationships are needed to produce functional ribosomes, resulting in some regions of rRNA genes which are linked with other components in ribosome are conserved[85,86]. The rRNA gene sequences are universally distributed across distantly related strains, and therefore these sequences can be aligned precisely to make measurement of true differences between them easier [87,88]. However, these rRNA genes also consist of hypervariable regions such as gamma region that diverged over evolutionary time which can be used for species discrimination. Ribosome is defined as large ribonucleoprotein which synthesizes protein and highly conserved among the three life kingdoms. For prokaryotes, ribosome is made up of two subunits which are the small 30S subunit and large 50S subunit. The small 30S subunit consists of a 16S rRNA and 20 proteins while large 50S subunit consists of a 5S rRNA, a 23S rRNA and over 30 proteins. Thus, the conserved regions in rRNA genes are used as tools to study distant phylogenetic relationships while regions between the conserved regions with higher mutation rates are used to discriminate closely related bacteria. It is also important to note that within the highly constrained rRNA genes that are essential for survival of organisms, horizontal gene transfer events are usually implausible to happen[86].

Ribosomal RNA genes - 16S rRNA gene

In general, the 16S rRNA gene is the gold standard in the identification of taxonomic and phylogenetic relationships among different bacteria[83]. The 16S rRNA gene is an essential constituent found in all bacteria, containing highly conserved regions and variable regions with approximately 1600 nucleotides long (bp)[88]. The 16S rRNA gene can clearly distinguish the three main kingdoms which include *Archaea, Bacteria* and *Eukarya*. The hypervariable regions of 16S rRNA gene in a single genome provides sufficient sequence variation for phylogenetic discrimination[88].

Generally, 16S rRNA gene is still sensitive enough for other genera but unfortunately seems to be less sensitive for *Streptomyces*. Previously, many studies attempted to utilize sequences from variable regions of 16S rRNA gene to form taxonomic structure within the S*treptomyces* genus. However, the variation was too low to distinguish between closely related *Streptomyces* species/strains as they might exhibit highly similar or identical 16S rRNA gene sequences, with the addition of overspeciation issue within the genus[89,90]. Alternately, *Streptomyces* strains with highly similar or identical 16S rRNA gene sequences appear to be of different species[49]. Furthermore, there are multiple copies of 16S rRNA gene or heterogeneous rRNA operons within a single genome have been found in some bacteria[87]. This might suggest that there is a potential for the occurrence of horizontal gene transfer events in the rRNA genes[87,90]. Hence, phylogenetic construction based on 16S rRNA gene sequences could not assure well-resolved phylogenetic trees that represent relationship between bacterial species particularly within genus *Streptomyces*[9,12,51]. The 16S rRNA gene sequences provide low resolution as phylogenetic marker for *Streptomyces* species. Furthermore, insignificant phenotypic differences, limitations of DNA fingerprinting and DNA-DNA hybridization appear to ineffectively elaborate taxonomic grouping of *Streptomyces* species within the genus *Streptomyces*[91]. This suggests that it is challenging for *Streptomyces* taxonomic classification, thus, methods with higher resolution are crucial to evaluate taxonomic grouping among closely related *Streptomyces* species.

Ribosomal RNA genes - 23S rRNA gene

The 23S rRNA gene is formerly less preferred to be used for phylogenetic study as compared to 16S rRNA gene due to costing consideration[9]. However, the use of 23S rRNA gene in phylogenetic analysis is eventually being highlighted because of several factors including reduced sequencing costs due to the advancement of technology such as next generation DNA sequencing[92]. The 23S rRNA gene offers similar advantages as 16S rRNA gene with additional properties owing to its longer sequence length of approximately 3000 bp such as better resolution due to greater sequence variation, unique insertions and/or deletions, and more characteristic sequence regions[9,86,92,93]. Recent study conducted by Chaves *et al*. (2018)[94] revealed that the application of 23S rRNA gene can be utilized as an alternative to 16S rRNA gene for phylogenetic study of *Streptomyces* species which also appeared to be more discriminative than the 16S

rRNA gene. In addition, previous study showed that broad-range primers for 23S rRNA genes with a similar universality degree to that of 16S rRNA genes can be generated based on the conserved regions of 23S rRNA genes[92].

Protein-coding genes

The use of protein-coding genes as phylogenetic markers has become more common as they could determine genome relatedness with higher accuracy and might replace DDH for species taxonomy in the future[70,83,95,96]. Currently, multilocus sequence analysis (MLSA) technique that involves the use of a combination of a few protein-coding genes (also referred as housekeeping genes) has been widely applied to deduce bacterial phylogeny and aims to buffer phylogeny distortions caused by recombination[97]. This approach could be an alternative to DDH technique since it is reproducible with comparable resolution to DDH[89]. Studies have reported that there was a high correlation determined between DDH and MLSA at which MLSA evolutionary distance of 0.007 based on five-gene schemes corresponds to DDH value of 70% species cut-off point for species delineation of the genus *Streptomyces*[91,97,98]. Protein-coding genes are proven to improve the resolution and robustness at *Streptomyces* species level and can be used as an acceptable approach for species differentiation. Especially the gyrase B (*gyrB*) and tryptophan B (*trpB*) housekeeping genes can provide greater discriminatory power of closely related strains due to their higher percentages of variable regions and higher molecular evolution rates as compared to 16S rRNA gene[98]. The higher evolutionary rate could be explained by the synonymous substitutions predominantly at the third position of codons in these protein coding genes that occur without affecting the amino acid sequences[9]. Numerous *Streptomyces* phylogenetic analyses were conducted using protein-coding genes due to their higher resolution power, for examples, taxonomic studies of *Streptomyces griseus*[83,98] and *Streptomyces hygroscopicus* [91], and diversity studies on the family *Streptomycetaceae*[9,51].

Protein-coding genes - DNA gyrase subunit B, topoisomerase type II (*gyrB*)

The *gyrB* gene is one of the structural genes of DNA topoisomerases; the DNA topoisomerases are responsible for DNA replication, recombination, transcription and repair, also controlling supercoiling level[99]. The *gyrB* gene encodes for the subunit B protein of bacterial DNA gyrase (DNA topoisomerase type II) that is responsible for DNA replication and involves in direct DNA repair mechanism[96,100]. The specific functions of DNA gyrase include supercoiling relaxed closed circular double-stranded DNA negatively with ATP hydrolysis, whilst relaxing supercoiled DNA molecule in the absence of ATP hydrolysis[100,101]. The *gyrB* gene is found universally distributed among different species of bacteria. As compared to ribosomal 16S rRNA gene, *gyrB* gene has faster molecular evolution rate and also rarely undergo horizontal gene transfer[96,101]. Many studies have proven that the consistency between the results of DDH and *gyrB*-based phylogenetic analysis[70,96,101], whereby approximately 98.5 % *gyrB* gene sequence similarity could corresponds to DNA-DNA relatedness threshold value of 70 % for species delineation[70,102].

Protein-coding genes - tryptophan synthase subunit B (*trpB*)

The *trpB* gene encodes tryptophan synthase subunit B which is responsible for amino acid biosynthesis. It catalyzes the last two steps of tryptophan biosynthesis by causing indole and serine to be condensed irreversibly to produce tryptophan in the pyridoxal phosphate-dependent reaction[100,103]. Protein-coding genes such as *trpB* gene is often used for phylogenetic analysis because it is selectively neutral and is not affected by selection pressure for amino acid changes. This implied that *trpB* gene locus is independent as it underwent evolution[83]. The *trpB* gene has been used individually or in combination with other protein-coding genes for the taxonomic characterization of *Streptomyces* in previous studies and had successfully resolved the phylogenetic relationships between closely related subspecies and species of *Actinobacteria* due to the high genetic variation within these genes as compared to ribosomal genes[9,51,83,89,90,104].

TOOLS AND DATABASES TO FETCH GENES

Basic Local Alignment Search Tool (BLAST)

BLAST stands for 'Basic Local Alignment Search Tool' and it is a sequence similarity search program, accessible as an independent tool or via a web interface. It works by comparing the query sequences either nucleotide or protein sequences to an appropriate nucleotide or protein sequence database. This is to determine type sequences that resemble the query sequence above a particular threshold level. Evolutionary relationships between sequences can then be indicated based on the sequence similarity found between query and type strains[105].

EzBioCloud database

EzBioCloud is the latest improved version of open access database developed by ChunLab Inc., prospers the former versions namely EzTaxon and EzTaxon-e (https://www.ezbiocloud.net/). It is an integrated database that contains comprehensive hierarchical taxonomy of both *Bacteria* and *Archaea* domains from phylum to species level, represented by quality controlled 16S rRNA gene and genome sequences[106]. In the initial version of database, Eztaxon, contains 16S rRNA gene sequences of all validly described prokaryotic type strains which serves as a web-based tool for the analysis of 16S rRNA gene sequences including the determination of pairwise nucleotide similarity values of the gene sequences[107]. Eztaxon was later evolved to the next generation database known as the Eztaxon-e. Extaxon-e database was enhanced by including species with non-validly published names, uncultured phylotypes, and the incorporation of a new function involves the estimation of degree of completeness in sequencing. Besides, sequences of all species and phylotypes within this database have undergone comprehensive phylogenetic analysis based on 16S rRNA gene sequences to produce a complete hierarchical taxonomic system[108]. Despite the identification and comparison of 16S rRNA genes being a reliable method

to determine the identity of a strain of interest, however, it comes with a limitation where it does not guarantee that two strains with nearly identical 16S rRNA gene sequences belong to the same species, alternatively they could be of different species[109]. Hence, analyzing bacteria genome sequences has become a method to overcome this issue. The latter EzBioCloud united database contains information on both 16S rRNA gene and genome sequences as well as several bioinformatics tools such OrthoANI. Furthermore, all genomes in this database were quality filtered where the taxonomic identification at genus, species or subspecies levels was performed according to the algorithm comprising a combination of gene-based search and OrthoANI[106]. Currently, EzBioCloud holds up to 81,151 taxa and 99,418 qualified genomes (available at https://www.ezbiocloud.net/dashboard, accessed on August 2018).

National Center for Biotechnology Information (NCBI) GenBank database

NCBI is a national resource for molecular biology information established in year 1988, which involves in the development of computational biology and software for genome analysis[110]. The NCBI is the host of the GenBank - a comprehensive public online database that contains DNA sequences, supporting bibliographic and biological annotation (http://www.ncbi.nlm.nih.gov/genbank/)[111]. Newly discovered nucleotide sequences such as the non-coding DNA region, a particular gene region or coding region of a DNA sequence, and whole genome shotgun as well as other sequence data will be submitted to this database often by individual laboratories/authors and bulk submissions from high-throughput sequencing projects. Direct submissions to GenBank can be done via a web-based form named as BankIt or stand-alone submission program known as Sequin. GenBank is established as part of the International Nucleotide Sequence Database Collaboration, thus, it contains data from other databases such as the European Molecular Biology Laboratory (EMBL), and DNA DataBank of Japan (DDBJ). Consequently, GenBank holds DNA nucleotide sequences for more than 300,000 validly described species[111,112].

PHYLOGENETIC TREES RECONSTRUCTION

Phylogenetic is the study of evolutionary relationship while phylogenetic analysis is the estimation of these evolutionary relationships. Phylogenetic indicates changes caused by evolution and understand relationships between ancestor and its descendants to estimate divergent time and the evolution in a family[113,114]. Phylogenetic trees can be built by choosing an appropriate phylogenetic information marker (E.g. a particular DNA gene sequence, RNA) of prokaryotes or eukaryotes to determine the degree of relatedness between species, family, genus or order and their hypothetical common ancestors[1,51,115]. Phylogenetic trees consist of nodes and branches, the node is the taxonomic unit or evolutionary event while branch is link between two adjacent nodes. The length of a branch represents the divergence. Branching presents an estimated pedigree of evolutionary relationships among various organisms[115]. Clade is a monophyletic taxon with a set of descendants which are originated from a single most common ancestor, at which the members of a clade are more closely related to one another than members of other clades. Members of the same clade possess a common evolutionary history and share unique features which could not be found in distant ancestors[1,114]. There are two types of phylogenetic trees which are: rooted tree that consists of a common ancestor and unrooted tree which has no common ancestor. An outgroup is the least related operational taxanomic unit to the group of taxa being studied. It is used to root a tree and as a reference to determine evolutionary distance[115].

Bootstrapping analysis

Bootstrapping is one of the statistical methods to determine the reliability of tree branch arrangement/topology, in other words, to estimate the confidence intervals on phylogenies. The phylogenetic tree built is not the actual representation of the evolutionary relationships, instead, it is an estimation of these relationships. Hence, bootstrapping analysis is requested to calculate the reliability of that estimate. Columns in a multiple aligned sequences are resampled during bootstrapping process to produce many new alignments sets and replace the original dataset. The bootstrapping process is continued for at least 100 times while phylogenetic trees are produced from all these alignments sets. The final outcomes determine the number of time a particular branch point is generated out of the total number of the built phylogenetic trees. The branching point will be more valid when the number of occurring time is larger. Bootstrap values are statistically significant if they are ranged between 90 to 100 %. Phylogenetic trees are produced multiple times and the trees produced are not always identical during every production. For instance, a node with a particular cluster descending from it might produce bootstrap value of 90 % when a thousand of trees are built randomly, thereby inferring that node with that identical cluster descending from it appears for nine hundred times out of the one thousand times the tree were rebuilt[115-117].

BioEdit sequence alignment editor software

BioEdit is a free of charge biological sequence alignment editor and sequence analysis software program available online for Windows. BioEdit is commonly used in many molecular biology studies, it provides auto-integration with Clustal W multiple alignments, manipulation, trimming and editing of aligned sequences. The sequences of type strains are aligned with query sequences to determine differences between them. Besides consisting options for sequence analysis, it also provides links to external analysis programs such as BLAST[118,119].

Clustal W alignments

Alignment process is the crucial key for the construction of phylogenetic trees at which poor alignment would result in incorrect phylogenetic tree[120]. An alignment is considered as good if it consists of no gaps. Regions which do not align between sequences should be deleted before building of phylogenetic trees. Clustal W method is a computational multiple alignment methods that

align sequences based on an explicitly phylogenetic criterion or known as a 'guide tree'. The generation of guide tree is based on the initial pairwise sequence alignments while its guide tree is formatted as a PHYLIP tree file. Clustal W tool aligns all the sequences according to the matching identical or similar regions found between these sequences. The 'W' stands for 'Weighted' which refers to the different weights that are applied to parameters and sequences in different regions of the alignment. During the pairwise alignments, distance matrix is calculated among each sequence. Distance is defined as amount of identical matches divided by the length of sequences regardless of gaps[121].

MEGA software

Molecular Evolutionary Genetics Analysis (MEGA) software provides various tools for analyzing DNA and protein sequences according to evolutionary perspective. MEGA is widely utilized for assembling sequence alignments, constructing phylogenetic trees, mining online databases, testing selection, estimating molecular evolution rate and divergence times to study molecular evolutionary histories of genes, genomes and species[122,123]. Several versions of MEGA have been developed, where the initial version MEGA 1 had made available since early 1990s[124]. MEGA version 1 has been further evolved up till MEGA version 6, in which the progress involved multiple upgrades for the use in Microsoft Windows including improved computational algorithms and statistical methods[122,123]. MEGA 7 was later introduced as the latest 64-bit version with greater computational power that capable of handling memory-intensive analysis of large datasets[125]. Recently, MEGA X has been developed as cross-platform version that is able to run natively on Linux and Microsoft Windows with additional upgrade includes the application of multiple computing cores for various molecular evolutionary analyses[123].

TREE-BUILDING METHODS

Tree building methods can be divided into two different methods: (1) distance-based, and (2) character-based. Character-based methods generate trees that optimize the distribution of actual data for each character. Examples of character-based methods are maximum likelihood and maximum parsimony algorithms. Distance methods calculate pairwise distances based on certain measure, followed by elimination of the actual character data while only utilize the fixed distances to obtain trees. Hence, pairwise distances are not fixed because they are identified by the tree topology. Neighbour-joining is one of the commonly used algorithm in distance-based methods[114,126].

Character-based method

This method works by using character data in every step of phylogenetic analysis at which aligned sequences such as DNA sequences are used directly during tree analysis. It also enables the reliability of the position of each base in an alignment to be assessed based on the positions of all other bases. Character-based methods examine every alignment column individually and determine the tree which best accommodates all of the information[114,126].

Character-based method - maximum likelihood (ML) algorithm

ML algorithm builds a tree based on mathematical models - the probability models to deduce the evolutionary distances. It determines the evolutionary model and tree which consist of the highest likelihood of generating the observed data. It searches the most likely tree from all the trees of the given dataset with the help of a tree model for nucleotide substitutions. ML is calculated for each base position in the multiple alignment by determining the likelihood that a certain pattern of variation at a particular site would be generated by a particular substitution reaction when a particular tree and the overall observed base frequencies are provided. Likelihood is the total of the probability of each particular variation pattern produced by a particular substitution process. A tree that best accounts for the high number of variations of the dataset given based on likelihood calculations will be generated[1,114,127].

Character-based method - maximum parsimony (MP) algorithm

The MP algorithm builds shortest phylogenetic tree that minimizes the amount of evolutionary change or consists of the fewest amount of evolutionary changes such as nucleotide substitution to explain the evolutionary differences found among different taxa. The MP algorithm-tree contain of the least parallel changes as compared to NJ and ML algorithms. Ancestral relationship can be determined as the MP algorithm uses all the evolutionary information in every nucleotide bases[1,114,127,128].

Distance-based method

This method build tree according to the evolutionary distance or total of dissimilarity between two aligned sequences. A true tree will be generated based on this method if all the event of genetic divergence in the sequence were recorded accurately. The sequence data will first be converted into pairwise distance in order to build tree using the matrix of pairwise distances between gene sequences in an alignment regardless of the character data. Moreover, the tree is generated according to the distance values and clustering algorithm[128].

Distance-based method - neighbour joining (NJ) algorithm

NJ algorithm is a commonly used distance-based method for the reconstruction of phylogenetic trees[113]. It calculates the amount of differences between two sequences and the actual tree is then generated based on the matrix of distance values[113,127]. The tree construction begins when the most similar sequences or pairs of closest neighbours are joined. Then, a node (common ancestor) is successively inserted between them before next closest neighbours are chosen and added with node. The process is repeated until a tree with a combination of nodes that gives the smallest total branch lengths is produced. NJ algorithm works by trying to locate all neighbours in correct positions before correcting the length of branches[129]. It is frequently used by users due

to its fast computation speed and high accuracy which is equivalent or better than other computationally intensive algorithms[114].

Substitution model

Substitution model is used to determine the probability or rate and how one nucleotide replaces or substitutes another. They calculate evolutionary distance between sequences based on the observed differences found between sequences. This evolutionary model works based on the measurement of normalized distance which is the mean degree of change per length of aligned DNA sequences. Thus, calculations of distance between sequences are improved. Selection of substitution model is important as it influences the alignment and tree building and thus the right substitution model could result in most accurate phylogenetic relationship[1,114].

Substitution model - Kimura-2 parameter

Kimura-2 parameter is used to calculate evolutionary distances and similarity values in NJ tree-building algorithm. It is an evolutionary model which states that transitions which is a change between pyrimidines or between purines (C to T or A to G) occur more frequently as compared to transversions which is a changed between purines to pyrimidines or vice versa (A to T or C to G) when the substitution rate is unequal. Hence, the rate of transitional nucleotide substitutions is greater than rate of transversional nucleotide substitutions[130,131].

Substitution model - Tamura-Nei

This is used to discriminate the rates of two types of transitional substitutions, at which transitional changes between A↔G (purine ↔purine) may occur at different rate as compared to that between C↔T (pyrimidine ↔ pyrimidine) while transversions happen at same rate and this rate can be different from both transitional substitutions rates. Therefore, two different types of transitional substitutions rates and transversional substitutions rate are estimated individually with the consideration of base composition bias which refers to unequal frequency of all four nucleotides[132,133].

GENERAL DISCUSSION

The identification and taxonomic characterization of *Streptomyces* species could be relatively challenging mainly because of the large amount of species present in this genus. Several techniques are often required to be conducted simultaneously to produce a more reliable result for classification of bacterial species. Fortunately, the rapid progress in molecular techniques has greatly improve the efficiency of bacterial species identification and characterization. Traditional phenotypic approach may not be sufficient for describing or differentiating taxa as certain phenotypes changes could be caused by horizontal gene transfer event[57,73]. The incorporation of both molecular genotyping and phenotyping methods could assist in determining the correlations between both traits and thus lead to a better understanding of the *Streptomyces* taxonomy[57]. The application of PCR, DDH, and NGS have govern modern taxonomic studies and these techniques are important to generate genotypic data for classification and allocation of taxa on phylogenetic tree[73].

Ribosomal RNA genes and protein-coding genes are commonly used molecular markers for phylogenetic studies of *Streptomyces* (Figure 1). The rRNA genes are typically considered as best phylogenetic markers as they are universally distributed across living organisms and they made up of a combination of highly conserved, variable, and hypervariable regions that evolve at varying rates, thereby facilitate the differentiation among distinct bacterial groups[73,134,135]. The rRNA genes have slower evolutionary rate as compared to the protein coding genes, hence, rRNA genes are important phylogenetic markers for analyzing the phylogenies of distantly related species[134].

Among the rRNA genes, 16S rRNA gene acts as a very powerful discriminatory tool in resolving phylogenetic relationships and this gene has been the most sequence phylogenetic marker[135]. However, it is arguable whether the gene sequence of 16S rRNA gene should be the gold standard for *Streptomyces* phylogeny. Many studies have suggested that 16S rRNA gene sequence alone is not sufficient enough to discriminate between closely related species, notably species within *Streptomyces* genus due to its low resolution at species level[51,98]. This issue is because of many species in the genus *Streptomyces* share identical or highly similar 16S rRNA gene sequence in addition to highly similar phenotypes[9,51]. Henceforth, the implication of other molecular targets with higher resolution in addition to the 16S rRNA gene is recommended for the phylogenetic analysis within the genus *Streptomyces* to observe their actual relationships. As an alternative, the 23S rRNA gene is employed in taxonomic classification of *Streptomyces* species. The 23S rRNA gene offers better resolution than 16S rRNA gene as it has more variable regions. In the early days, it appears that 23S rRNA gene has somewhat lost favor to 16S rRNA gene as phylogenetic marker in taxonomic classification which might due to difficulties in developing broad-range sequencing primers and sequencing of larger genes. Though the use of 23S rRNA gene has now eventually increased with time and advances in sequencing technology[86].

Nevertheless, it is important to acknowledge that intragenomic or intraspecific diversity of rRNA genes is a common phenomenon in *Streptomyces* that could lead to false identification regarding phylogenetic relatedness and ancestry when there is only a single gene sequence used[86,95,136]. Besides, rRNA genes may not be able to provide high resolution when it comes to distinguishing *Streptomyces* species as the functional constraints on both 16S and 23S rRNA genes for them to be highly conserved to maintain the precise spatial relationships for producing functional ribosomes. Conserved regions are suitable for investigating organisms with distant relationships, whilst variable regions with faster evolutionary rate are suitable for differentiating organisms that are closely related[86,91]. Ribosomal genes were demonstrated to be insufficient to discriminate between

closely related *Streptomyces* species by a number of studies despite of being a powerful tool in elucidating inter- and intrageneric evolutionary relationships[83,91,137].

Protein-coding genes (housekeeping genes) such as *gyrB* and *trpB* genes have exhibit potential for taxonomic classification of *Streptomyces* at species level and have been shown to be effective discriminatory tools for phylogenetic analysis. Also, the utilization of multiple protein-coding genes in MLSA has been conducted in several biosystematics studies of *Streptomyces* species[51,83,91]. Protein-coding genes produce higher resolution in discriminating *Streptomyces* at species level than the rRNA genes because of the presence of more variable regions in the gene sequences[9,91]. Additionally, protein-coding genes such as *gyrB* gene tend to exhibit higher number of base substitutions as compared to rRNA genes. Ochman *et al*. (1987)[138] reported that 16S rRNA gene had average base substitution rate of 1 % in 50 million years whereas the average base substitution rate of *gyrB* gene at the synonymous sites was found to be about 0.7 % to 0.8 % in one million years. As a result, the higher percentage of base substitution rate accounts for higher rate of evolution of protein-coding genes than rRNA genes and thus improve their resolution as phylogenetic markers that capable of determining the relationships between closely related *Streptomyces* species with the generation of stable phylogenetic trees[9,91].

The advent of NGS technology has increase the availability of whole genome sequences of many prokaryotes, along with the rapid progression of bioinformatic tools which are important for analyzing these sequences. The inclusion of whole genome sequencing data could further enrich the prokaryotic systematics by overcoming the limitations of current approaches (E.g. DDH, 16S rRNA gene sequences *etc*.)[139]. There is an increasing evidence that phylogenetic and taxogenomic approach offers robust classification as it is capable of resolving complex taxa below family level (intrageneric and intrafamily)[12,139-141]. Therefore, providing reliable information on the relationship and differentiation of prokaryotes at species and genera level. The taxonomic classification of prokaryotes utilizing genomic data is likely becoming the norm in near future.

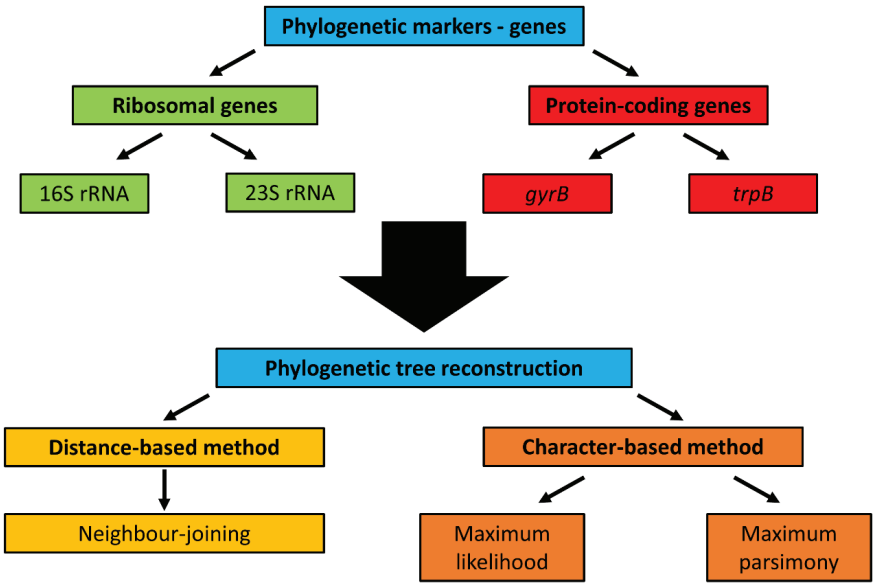

Figure 1. Framework of ribosomal genes and protein-coding genes as molecular markers for phylogenetic analysis based on different algorithms.

CONCLUSION

Streptomyces bacteria have been a focus in systematic research due to the various beneficial properties they can offer towards mankind. However, there are taxonomic chaos and overspeciation problem occurred within the genus *Streptomyces*. Thus, it is important to search for solutions that could ease the taxonomic classification this complex taxon. Several important molecular makers including ribosomal genes and protein-coding genes where each has their own advantages and limitations in determining the phylogenetic relationships of *Streptomyces* species. The availability of various molecular techniques, bioinformatic tools, databases, sequence editing and phylogenetic tree reconstruction software have greatly assist in taxonomic studies of *Streptomyces*. Advances of NGS and associated bioinformatic tools will revolutionize taxonomic practice through the integration of genomic approach. It is anticipated that genome-based phylogenetic analysis will further improve the identification and taxonomic classification of members in the genus *Streptomyces*.

Authors Contributions

The literature review and manuscript writing were performed by JW-FL, K-XT, SHW, N-SAM, and L-HL. L-HL founded the research project.

Conflict of Interest

The authors declare that the research was conducted in

the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

This work was supported by the MOSTI eScience Fund (02-02-10-SF0215) and External Industry Grants from Biotek Abadi Sdn Bhd [vote no. GBA-808138 and GBA-808813] awarded to L-HL.

Reference

1. Reddy BN. Basics for the construction of phylogenetic trees. Web Central Biology 2011; 2(12).
2. Sadeghi A, Soltani BM, Jouzani GS, *et al.* Taxonomic study of a salt tolerant *Streptomyces* sp. strain C-2012 and the effect of salt and ectoine on lon expression level. Microbiol Res 2014; 169(2-3): 232-238.
3. Bérdy J. Bioactive microbial metabolites. J Antibiot 2005; 58(1): 1.
4. Ser H-L, Tan LT-H, Law JW-F, *et al.* Focused review: cytotoxic and antioxidant potentials of mangrove-derived *Streptomyces*. Front Microbiol 2017; 8: 2065.
5. Ser H-L, Yin W-F, Chan K-G, *et al.* Antioxidant and cytotoxic potentials of *Streptomyces gilvigriseus* MUSC 26^{T} isolated from mangrove soil in Malaysia. Prog Microbes Mol Biol 2018; 1(1): a0000002.
6. Hopwood DA. Soil to genomics: the *Streptomyces* chromosome. Annu Rev Genet 2006; 40: 1-23.
7. Samac DA, Willert AM, McBride MJ, *et al.* Effects of antibiotic-producing *Streptomyces* on nodulation and leaf spot in alfalfa. Applied Soil Ecol 2003; 22(1): 55-66.
8. Elliot MA, Buttner MJ, and Nodwell JR. Multicellular development in *Streptomyces*. Myxobacteria: Multicellularity and Differentiation 2008: 419-438.
9. Han J-H, Cho M-H, and Kim SB. Ribosomal and protein coding gene based multigene phylogeny on the family *Streptomycetaceae.* Syst Appl Microbiol 2012; 35(1): 1-6.
10. Dastager S, Li W-J, Dayanand A, *et al.* Seperation, identification and analysis of pigment (melanin) production in *Streptomyces*. Afr J Biotechnol 2006; 5(11).
11. Seipke RF, Kaltenpoth M, and Hutchings MI. *Streptomyces* as symbionts: an emerging and widespread theme? FEMS Microbiol Rev 2012; 36(4): 862-876.
12. Alam MT, Merlo ME, Takano E, *et al.* Genome-based phylogenetic analysis of *Streptomyces* and its relatives. Mol Phylogen Evol 2010; 54(3): 763-772.
13. Law JW-F, Ser H-L, Duangjai A, *et al. Streptomyces colonosanans* sp. nov., a novel actinobacterium isolated from Malaysia mangrove soil exhibiting antioxidative activity and cytotoxic potential against human colon cancer cell lines. Front Microbiol 2017; 8: 877.
14. Kemung HM, Tan LT-H, Khan TM, *et al*. *Streptomyces* as a prominent resource of future anti-MRSA drugs. Front Microbiol 2018; 9: 2221.
15. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature 2002; 417(6885): 141.
16. Ikeda H, Ishikawa J, Hanamoto A, *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 2003; 21(5): 526.
17. Sakula A. Selman Waksman (1888–1973), discoverer of streptomycin: a centenary review. Brit J Dis Chest 1988; 82: 23-31.
18. Fawaz F and Jones GH. Actinomycin synthesis in *Streptomyces antibioticus*. Purification and properties of a 3-hydroxyanthranilate 4-methyltransferase. J Biol Chem 1988; 263(10): 4602-4606.
19. Bhattacharjee K and Joshi S. Phylogenetic rearrangement of *Streptomyces* spp. on the basis of internal transcribed spacer (ITS) region using molecular morphometrics approach. 2013.
20. de Lima Procópio RE, da Silva IR, Martins MK, *et al.* Antibiotics produced by *Streptomyces*. Braz J Infect Dis 2012; 16(5): 466-471.
21. Adegboye M and Babalola O. Phylogenetic characterization of culturable antibiotic producing *Streptomyces* from rhizospheric soils. Mol Biol S 2013; 1: 001.
22. Ser H-L, Law JW-F, Chaiyakunapruk N, *et al.* Fermentation conditions that affect clavulanic acid production in *Streptomyces clavuligerus*: A Systematic Review. Front Microbiol 2016; 7(522).
23. Awad HM, El-Sahed K, and El-Nakkadi A. Isolation, screening and identification of newly isolated soil *Streptomyces* (*Streptomyces* sp. NRC-35) for b-Lactamase inhibitor production. World Appl Sci J 2009; 7(5): 637-646.
24. Law JW-F, Ser H-L, Khan TM, *et al.* The potential of *Streptomyces* as biocontrol agents against the rice blast fungus, *Magnaporthe oryzae* (*Pyricularia oryzae*). Front Microbiol 2017; 8: 3.
25. Ser H-L, Ab Mutalib N-S, Yin W-F, *et al.* Genome sequence of *Streptomyces antioxidans* MUSC 164^{T} isolated from mangrove forest. Prog Microbes Mol Biol 2018; 1(1): a0000001.
26. Ser HL, Chan KG, Tan WS, *et al*. Complete genome of mangrove-derived anti-MRSA streptomycete, *Streptomyces pluripotens* MUSC 135^{T}. Prog Microbes Mol Biol 2018; 1(1): a0000004.
27. Van Wezel GP, Krabben P, Traag BA, *et al.* Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering. Appl Environ Microbiol 2006; 72(8): 5283-5288.
28. Dharni S, Alam M, Samad A, *et al.* Phylogenetic analysis and biological activity of *Streptomyces* sp. CIMAP-A 2 isolated from industrially polluted soil. Indian J Biotechnol 2012; 11: 438-444.
29. Nishimura M, Ooi O, and Davies J. Isolation and characterization of *Streptomyces* sp. NL15-2K capable of degrading lignin-related aromatic compounds. J Biosci Bioeng 2006; 102(2): 124-127.
30. Arasu MV, Duraipandiyan V, Agastian P, *et al.* In vitro antimicrobial activity of *Streptomyces* spp. ERI-3 isolated from Western Ghats rock soil (India). Journal de Mycologie Médicale 2009; 19(1): 22-28.
31. Mehdi RBA, Sioud S, Fguira LFB, *et al.* Purification and structure determination of four bioactive molecules from a newly isolated *Streptomyces* sp. TN97 strain. Process Biochem 2006; 41(7): 1506-1513.
32. Lee L-H, Chan K-G, Stach J, *et al.* Editorial: The search for biological active agent(s) from *Actinobacteria.* Front Microbiol 2018; 9(824).
33. Phongsopitanun W, Kudo T, Ohkuma M, *et al. Streptomyces verrucosisporus* sp. nov., isolated from marine sediments. Int J Syst Evol Microbiol 2016; 66(9): 3607-3613.
34. Sharma TK, Mawlankar R, Sonalkar VV, *et al. Streptomyces lonarensis* sp. nov., isolated from Lonar Lake, a meteorite salt water lake in India. Antonie van Leeuwenhoek 2016; 109(2): 225-235.
35. Maciejewska M, Pessi IS, Arguelles-Arias A, *et al. Streptomyces lunaelactis* sp. nov., a novel ferroverdin A-producing *Streptomyces* species isolated from a moonmilk speleothem. Antonie van Leeuwenhoek 2015; 107(2): 519-531.
36. Lee L-H, Zainal N, Azman A-S, *et al. Streptomyces pluripotens* sp. nov., a bacteriocin-producing streptomycete that inhibits meticillin-resistant *Staphylococcus aureus*. Int J Syst Evol Microbiol 2014; 64(9): 3297-3306.
37. Santhanam R, Okoro CK, Rong X, *et al. Streptomyces atacamensis* sp. nov., isolated from an extreme hyper-arid soil of the Atacama Desert, Chile. Int J Syst Evol Microbiol 2012; 62(11): 2680-2684.
38. Tanasupawat S, Phongsopitanun W, Suwanborirux K, *et al. Streptomyces actinomycinicus* sp. nov., isolated from soil of a peat swamp forest. Int J Syst Evol Microbiol 2016; 66(1): 290-295.
39. Mingma R, Duangmal K, Thamchaipenet A, *et al. Streptomyces oryzae* sp. nov., an endophytic actinomycete isolated from stems of rice plant. J Antibiot 2015; 68(6): 368.
40. Nimaichand S, Tamrihao K, Yang L-L, *et al. Streptomyces hundungensis* sp. nov., a novel actinomycete with antifungal activity and plant growth promoting traits. J Antibiot 2013; 66(4): 205.
41. Ser H-L, Palanisamy UD, Yin W-F, *et al. Streptomyces malaysiense* sp. nov.: a novel Malaysian mangrove soil actinobacterium with antioxidative activity and cytotoxic potential against human cancer cell lines. Sci Rep 2016; 6: 24247.
42. Hasani A, Kariminik A, and Issazadeh K. Streptomycetes: characteristics and their antimicrobial activities. IJABBR 2014; 2(1): 63-75.
43. Horinouchi S. A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in *Streptomyces griseus*. Front Biosci 2002; 7: 2045-2057.
44. Worrall JA and Vijgenboom E. Copper mining in *Streptomyces*: enzymes, natural products and development. Nat Prod Res 2010; 27(5): 742-756.
45. Hopwood DA. *Streptomyces* in nature and medicine: the antibiotic makers. In. New York: Oxford University Press; 2007.
46. Drew SW and Demain AL. Effect of primary metabolites on secondary metabolism. Annu Rev Microbiol 1977; 31(1): 343-356.
47. Halket JM, Waterman D, Przyborowska AM, *et al.* Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 2004; 56(410): 219-243.
48. Tan LT-H, Ser H-L, Yin W-F, *et al.* Investigation of antioxidative and anticancer potentials of *Streptomyces* sp. MUM256 isolated from Malaysia mangrove soil. Front Microbiol 2015; 6: 1316.
49. Antony-Babu S, Stien D, Eparvier V, *et al.* Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. Sci Rep 2017; 7(1): 11089.
50. Trejo W. Section of Microbiology: An evaluation of some concepts and criteria used in the speciation of streptomycetes. Transactions of the New York Academy of Sciences 1970; 32(8 Series II): 989-997.
51. Labeda D, Goodfellow M, Brown R, *et al.* Phylogenetic study of the species within the family *Streptomycetaceae.* Antonie Van Leeuwenhoek 2012; 101(1): 73-104.
52. Williams S, Goodfellow M, Alderson G, *et al.* Numerical classification of *Streptomyces* and related genera. Microbiology 1983; 129(6): 1743-1813.
53. Shirling EB and Gottlieb D. Cooperative description of type cultures of *Streptomyces* III. Additional species descriptions from first and second studies. Int J Syst Evol Microbiol 1968; 18(4): 279-392.
54. Shirling ET and Gottlieb D. Methods for characterization of *Streptomyces* species. Int J Syst Bacteriol 1966; 16(3): 313-340.
55. Shirling EB. Cooperative description of type cultures of *Streptomyces*. II. Sperative descriptions from first study. Int J Syst Bacteriol 1968; 18: 69-189.
56. Shirling EB and Gottlieb D. Cooperative description of type cultures of *Streptomyces*. IV. Species descriptions from the second, third and fourth studies. Int J Syst Evol Microbiol 1969; 19(4): 391-512.

57. Anderson AS and Wellington E. The taxonomy of *Streptomyces* and related genera. Int J Syst Evol Microbiol 2001; 51(3): 797-814.
58. Emerson D, Agulto L, Liu H, *et al.* Identifying and characterizing bacteria in an era of genomics and proteomics. Bioscience 2008; 58(10): 925-936.
59. Mullis KB. The unusual origin of the polymerase chain reaction. Scientific American 1990; 262(4): 56-65.
60. Garibyan L and Avashia N. Research techniques made simple: polymerase chain reaction (PCR). J Invest Dermatol 2013; 133(3): e6.
61. Joshi M and Deshpande J. Polymerase chain reaction: methods, principles and application. Int J Biomed Res 2010; 2(1): 81-97.
62. Law JW-F, Ab Mutalib N-S, Chan K-G, *et al.* Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations. Front Microbiol 2015; 5: 770.
63. Tan LT-H, Chan K-G, Khan TM, *et al. Streptomyces* sp. MUM212 as a source of antioxidants with radical scavenging and metal chelating properties. Front Pharmacol 2017; 8: 276.
64. Ser H-L, Tan LT-H, Palanisamy UD, *et al. Streptomyces antioxidans* sp. nov., a novel mangrove soil actinobacterium with antioxidative and neuroprotective potentials. Front Microbiol 2016; 7: 899.
65. Lee L-H, Zainal N, Azman A-S, *et al.* Diversity and antimicrobial activities of actinobacteria isolated from tropical mangrove sediments in Malaysia. Sci, World J 2014; 2014.
66. Lee A and Wong E. Optimization and the robustness of BOX A1R PCR for DNA fingerprinting using Trout Lake E. coli isolates. J Exp Microbiol Immunol 2009; 13: 104-113.
67. Koeuth T, Versalovic J, and Lupski JR. Differential subsequence conservation of interspersed repetitive *Streptococcus pneumoniae* BOX elements in diverse bacteria. Genome Res 1995; 5(4): 408-418.
68. Brusetti L, Malkhazova I, Gtari M, *et al.* Fluorescent-BOX-PCR for resolving bacterial genetic diversity, endemism and biogeography. BMC Microbiol 2008; 8(1): 220.
69. Janda JM and Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol 2007; 45(9): 2761-2764.
70. Hatano K, Nishii T, and Kasai H. Taxonomic re-evaluation of whorl-forming *Streptomyces* (formerly *Streptoverticillium*) species by using phenotypes, DNA–DNA hybridization and sequences of gyrB, and proposal of *Streptomyces luteireticuli* (ex Katoh and Arai 1957) corrig., sp. nov., nom. rev. Int J Syst Evol Microbiol 2003; 53(5): 1519-1529.
71. Goris J, Konstantinidis KT, Klappenbach JA, *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 2007; 57(1): 81-91.
72. Barbu V. Molecular hybridization techniques of nucleic acids. Innov Rom Food Biotechnol 2007; 1: 1.
73. Vandamme P, Pot B, Gillis M, *et al.* Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev 1996; 60(2): 407-438.
74. Ser H-L, Palanisamy UD, Yin W-F, *et al.* Presence of antioxidative agent, Pyrrolo [1, 2-a] pyrazine-1, 4-dione, hexahydro-in newly isolated *Streptomyces mangrovisoli* sp. nov. Front Microbiol 2015; 6: 854.
75. Zainal N, Ser H-L, Yin W-F, *et al. Streptomyces humi* sp. nov., an actinobacterium isolated from soil of a mangrove forest. Antonie van Leeuwenhoek 2016; 109(3): 467-474.
76. Ser H-L, Zainal N, Palanisamy UD, *et al. Streptomyces gilvigriseus* sp. nov., a novel actinobacterium isolated from mangrove forest soil. Antonie van Leeuwenhoek 2015; 107(6): 1369-1378.
77. Liu L, Li Y, Li S, *et al.* Comparison of next-generation sequencing systems. Biomed Res Int 2012; 2012.
78. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet 2010; 11(1): 31.
79. Snyder LA, Loman N, Pallen MJ, *et al.* Next-generation sequencing—the promise and perils of charting the great microbial unknown. Microb Ecol 2009; 57(1): 1-3.
80. Law JW-F, Ab Mutalib N-S, Chan K-G, *et al.* An insight into the isolation, enumeration, and molecular detection of *Listeria monocytogenes* in food. Front Microbiol 2015; 6: 1227.
81. Shendure J and Ji H. Next-generation DNA sequencing. Nat Biotechnol 2008; 26(10): 1135.
82. Xiong M, Zhao Z, Arnold J, *et al.* Next-generation sequencing. J Biomed Biotechnol 2011; 2010: 370710.
83. Guo Y, Zheng W, Rong X, *et al.* A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. Int J Syst Evol Microbiol 2008; 58(1): 149-159.
84. Singh V, Praveen V, Khan F, *et al.* Phylogenetics of an antibiotic producing *Streptomyces* strain isolated from soil. Bioinformation 2009; 4(2): 53.
85. Bhattacharjee K and Joshi S. Phylogenetic rearrangement of *Streptomyces* spp. on the basis of internal transcribed spacer (ITS) region using molecular morphometrics approach. Indian J Biotechnol 2013.
86. Pei A, Nossa CW, Chokshi P, *et al.* Diversity of 23S rRNA genes within individual prokaryotic genomes. PloS One 2009; 4(5): e5437.
87. Acinas SG, Marcelino LA, Klepac-Ceraj V, *et al.* Divergence and redundancy of 16S rRNA sequences in genomes with multiple rrn operons. J Bacteriol 2004; 186(9): 2629-2635.
88. Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004; 17(4): 840-862.
89. Rong X, Guo Y, and Huang Y. Proposal to reclassify the *Streptomyces albidoflavus* clade on the basis of multilocus sequence analysis and DNA–DNA hybridization, and taxonomic elucidation of *Streptomyces griseus* subsp. solvifaciens. Syst Appl Microbiol 2009; 32(5): 314-322.
90. Rong X, Doroghazi JR, Cheng K, *et al.* Classification of *Streptomyces* phylogroup pratensis (Doroghazi and Buckley, 2010) based on genetic and phenotypic evidence, and proposal of *Streptomyces pratensis* sp. nov. Syst Appl Microbiol 2013; 36(6): 401-407.
91. Rong X and Huang Y. Taxonomic evaluation of the *Streptomyces hygroscopicus* clade using multilocus sequence analysis and DNA–DNA hybridization, validating the MLSA scheme for systematics of the whole genus. Syst Appl Microbiol 2012; 35(1): 7-18.
92. Hunt DE, Klepac-Ceraj V, Acinas SG, *et al.* Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. Appl Environ Microbiol 2006; 72(3): 2221-2225.
93. Sallen B, Rajoharison A, Desvarenne S, *et al.* Comparative analysis of 16S and 23S rRNA sequences of *Listeria* species. Int J Syst Evol Microbiol 1996; 46(3): 669-674.
94. Chaves JV, Ojeda CPO, da Silva IR, *et al.* Identification and phylogeny of *Streptomyces* based on gene sequences. Research Journal of Microbiology 2018; 13: 13-20.
95. Kim B-J, Kim C-J, Chun J, *et al.* Phylogenetic analysis of the genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase β-subunit gene (rpoB) sequences. Int J Syst Evol Microbiol 2004; 54(2): 593-598.
96. Wang L-T, Lee F-L, Tai C-J, *et al.* Comparison of gyrB gene sequences, 16S rRNA gene sequences and DNA–DNA hybridization in the *Bacillus subtilis* group. Int J Syst Evol Microbiol 2007; 57(8): 1846-1850.
97. Glaeser SP and Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Syst Appl Microbiol 2015; 38(4): 237-245.
98. Rong X and Huang Y. Taxonomic evaluation of the *Streptomyces griseus* clade using multilocus sequence analysis and DNA–DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. Int J Syst Evol Microbiol 2010; 60(3): 696-703.
99. Wang JC. Recent studies of DNA topoisomerases. Biochim Biophys Acta, Gene Struct Expression 1987; 909(1): 1-9.
100. Mohan V and Stevenson M. Molecular data analysis of selected housekeeping and informational genes from nineteen *Campylobacter jejuni* genomes. F1000Research 2013; 2.
101. Kasai H, Watanabe K, Gasteiger E, *et al.* Construction of the gyrB database for the identification and classification of bacteria. Genome Inform 1998; 9: 13-21.
102. Kasai H, Tamura T, and Harayama S. Intrageneric relationships among *Micromonospora* species deduced from gyrB-based phylogeny and DNA relatedness. Int J Syst Evol Microbiol 2000; 50(1): 127-134.
103. Buller AR, van Roye P, Cahn JK, *et al.* Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. J Am Chem Soc 2018; 140(23): 7256-7266.
104. Meintanis C, Chalkou KI, Kormas KA, *et al.* The use of trpB gene in resolving phylogenetic diversity within the group of *Streptomyces*. Microbiology 2009; 5.
105. Johnson M, Zaretskaya I, Raytselis Y, *et al.* NCBI BLAST: a better web interface. Nucleic Acids Res 2008; 36(suppl_2): W5-W9.
106. Yoon S-H, Ha S-M, Kwon S, *et al.* Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 2017; 67(5): 1613-1617.
107. Chun J, Lee J-H, Jung Y, *et al.* EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. Int J Syst Evol Microbiol 2007; 57(10): 2259-2261.
108. Kim O-S, Cho Y-J, Lee K, *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(3): 716-721.
109. Fox GE, Wisotzkey JD, and Jurtshuk JR P. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Evol Microbiol 1992; 42(1): 166-170.
110. Coordinators NR. Database resources of the national center for biotechnology information. Nucleic Acids Res 2013; 41(Database issue): D8.
111. Benson DA, Clark K, Karsch-Mizrachi I, *et al.* GenBank. Nucleic Acids Res 2015; 43(Database issue): D30.
112. Mizrachi I. GenBank: the nucleotide sequence database. The NCBI Handbook [Internet], updated 2007; 22.
113. Kumar S and Gadagkar SR. Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. J Mol Evol 2000; 51(6): 544-553.
114. Brinkman FS and Leipe DD. Phylogenetic analysis. 2nd ed In: Baxevanis AD and Ouellette BFF. Bioinformatics: a practical guide to the analysis of genes and proteins. USA: John Wiley & Sons; 2001.
115. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 2013; 30(5): 1229-1235.
116. Efron B, Halloran E, and Holmes S. Bootstrap confidence levels for phylogenetic trees. Proc Natl Acad Sci 1996; 93(23): 13429-13429.
117. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985; 39(4): 783-791.
118. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. in Nucleic Acids Symposium Series. 1999. [London]: Information Retrieval Ltd., c1979-c2000.
119. Hall T, Biosciences I, and Carlsbad C. BioEdit: an important software for molecular biology. GERF Bull Biosci 2011; 2(1): 60-61.
120. Thompson JD, Higgins DG, and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22(22): 4673-4680.
121. Chenna R, Sugawara H, Koike T, *et al.* Multiple sequence align-

ment with the Clustal series of programs. Nucleic Acids Res 2003; 31(13): 3497-3500.
122. Tamura K, Stecher G, Peterson D*, et al.* MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 2013; 30(12): 2725-2729.
123. Kumar S, Stecher G, Li M*, et al.* MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 2018; 35(6): 1547-1549.
124. Kumar S, Tamura K, and Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. Bioinformatics 1994; 10(2): 189-191.
125. Kumar S, Stecher G, and Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 2016; 33(7): 1870-1874.
126. Du Z, Lin F, and Roshan UW. Reconstruction of large phylogenetic trees: a parallel approach. Comput Biol Chem 2005; 29(4): 273-280.
127. Rizzo J and Rouchka EC. Review of phylogenetic tree construction. University of Louisville Bioinformatics Laboratory Technical Report Series 2007: 2-7.
128. Lakshmi P and Rao AA. Application of new distance matrix to phylogenetic tree construction. IJCSS; 1(5): 1.
129. Tamura K, Nei M, and Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci 2004; 101(30): 11030-11035.
130. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980; 16(2): 111-120.
131. Gascuel O and Steel M. Neighbor-joining revealed. Mol Biol Evol 2006; 23(11): 1997-2000.
132. Tamura K, Dudley J, Nei M*, et al.* MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 2007; 24(8): 1596-1599.
133. Tamura K and Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 1993; 10(3): 512-526.
134. Patwardhan A, Ray S, and Roy A. Molecular markers in phylogenetic studies-a review. J Phylogenetics Evol Biol 2014; 2014.
135. Yarza P, Yilmaz P, Pruesse E*, et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 2014; 12(9): 635.
136. Pei AY, Oberdorf WE, Nossa CW*, et al.* Diversity of 16S rRNA genes within individual prokaryotic genomes. Appl Environ Microbiol 2010; 76(12): 3886-3897.
137. Song J, Lee S-C, Kang J-W*, et al.* Phylogenetic analysis of *Streptomyces* spp. isolated from potato scab lesions in Korea on the basis of 16S rRNA gene and 16S–23S rDNA internally transcribed spacer sequences. Int J Syst Evol Microbiol 2004; 54(1): 203-209.
138. Ochman H and Wilson AC. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 1987; 26(1-2): 74-86.
139. Sangal V, Goodfellow M, Jones AL*, et al.* Next-generation systematics: An innovative approach to resolve the structure of complex prokaryotic taxa. Sci Rep 2016; 6: 38392.
140. Harrison N and Kidner CA. Next-generation sequencing and systematics: what can a billion base pairs of DNA sequence data do for you? Taxon 2011; 60(6): 1552-1566.
141. Pérez-Cataluña A, Collado L, Salgado OA*, et al.* A polyphasic and taxogenomic evaluation uncovers *Arcobacter cryaerophilus* as a species complex that embraces four genomovars. Front Microbiol 2018; 9: 805.