ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

# Protein coding gene, *celS2* shows congruency with 16S rRNA phylogeny in highly evolving streptomycetes

**Madison P Munar[1,2*], Young-Geun Eom[1], Yeong-Suk Kim[1] and Tae-Jong Kim[1]**

[1]Department of Forest Products and Biotechnology, College of Forest Science, Kookmin University, 77, Jeongneung-ro, Seongbuk-gu, 136-702, Seoul, South Korea
[2]Department of Biology, College of Science, University of the Philippines Baguio, Governor Pack Road, Baguio City 2600, Philippines

**ABSTRACT**

This paper examined the potential of putative cellulose binding protein gene, *celS2*, in defining the molecular phylogeny of streptomycetes, a highly evolving and important soil-dwelling bacterial species. There have been numerous attempts on the use of molecular markers in establishing the phylogeny of this medically important group. However, cellulase genes have not yet been investigated. Results highlighted the higher evolutionary rate in *celS2* gene (5%) than 16S rRNA (0.5%) making it a good candidate as a molecular chronometer for future evolutionary studies in streptomycetes. Moreover, this study showed the congruency of tree topology using three different building algorithms, Unrooted Neighbor-Joining, Unrooted Maximum-Likelihood and Unrooted Unweighted-Pair Group with Arithmetic Mean of a protein coding gene and 16S rRNA which were analyzed simultaneously using multiple sequence alignment.

**Keywords:** *celS2,* 16S rRNA, *Streptomyces coelicolor* M145, phylogeny, Neighbor-Joining, Maximum-Likelihood

***Corresponding author***

ISSN: 0975-8585

## INTRODUCTION

Streptomycetes are soil-dwelling bacterial species that are reported to produce a plethora of secondary metabolites such as antifungal, antibacterial and anti-cancer drugs. Streptomycetes are also reported to produce industrially important compounds such as cellulase. The high guanine and cytosine bases in the genome of this bacterial species are of great interest because this made streptomycetes a highly evolving bacterial species [1]. At present, there are about 800 unidentified strains of streptomycetes and the growing number of these unidentified strains is becoming a challenge.

Several studies were conducted in investigating the molecular phylogeny of streptomycetes which include the use of molecular markers such as ATP synthase (*atpD*), DNA gyrase subunit B (*gyrB*), DNA repair gene (*recA*), RNA polymerase (*rpoB*), tryptophan synthase (*trpB*) and 16S rRNA [2,3]. Previous researches suggested to look into more robust and highly polymorphic molecular markers that would resolve the confusing taxonomy of this bacterial group. Cellulase genes, however, were not yet fully investigated in streptomycetes. Cellulase genes in the genome of the model actinomycete, *S. coelicolor* A3 (2), were reported to be found in the arms of the linear chromosome of this fungal-like bacterial species [1]. Basic local alignment search tool (BLAST) suggests that there are about 19 species of *Streptomyces* with reported cellulase coding genes. This relatively low turnout of genetically related species indicates that cellulase genes in streptomycetes are not well studied and their importance as agents in carbon cycling appears yet under appreciated.

## MATERIALS AND METHODS

### Control strain

Pure culture strain of *S. coelicolor* M145 was obtained from the Extract Collection of Useful Microorganisms (ECUM), Myongji University in South Korea. This was cultured in R5 minus agar medium which was modified from R5 medium that lacks $KH_2PO_4$, $CaCl_2$ and L-proline [4]. The strain was incubated for five days at 28°C prior to extraction of chromosomal DNA.

### DNA Extraction

Bacterial cells grown in R5 minus agar plate were aseptically transferred in 100 ml of R5 minus broth medium. The inoculated broth medium was incubated at 28°C with 200 rpm for three days in a shaking incubator (Innova 4300, New Brunswick Scientific Co., Inc, Edison, NJ, USA). About five hundred microliters of bacterial cells were collected in a microcentrifuge tube by centrifugation at 13,500 rpm. DNA was extracted using CTAB method [5].

### Primer design

There are 11 cellulase genes in *S. coelicolor* A3 (2) according to the genomic project (http://www.sanger.ac.kr). The DNA sequence of *celS2* gene (SCO1188) was viewed and annotated using the Artemis program. Homologous sequences of *celS2* gene were identified using protein BLAST search in the database of National Center for Biotechnology Information (NCBI). Among the homologous sequences were *S. coelicolor* A3(2) (AL939108.1 and AL939124.1), *Streptomyces* sp. THW31(HQ286612.1), *S. avermitilis* MA-4680 (BA000030.3), *S. hygroscopicus subsp. jinggangensis 5008* (CP003275.1), *S. flavogriseus ATCC 33331* (CP002475.1), *Kitasatospora setae* KM-6054 (AP010968.1), *S. scabiei* 87.22 (FN554889.1), *Streptomyces* sp. SirexAA-E (CP002993.1), *S. violaceusniger* Tu 4113 (CP002994.1). Conserved sequences among the homologous strains were identified and oligonucleotide primers were designed using Primer3 program. The forward sequence with 19 base pairs having 57.89 % GC ratio (CTGTACAACTGGTTCGCCG) and the reverse sequence with 23 base pairs having 47.83% GC ratio (GAGAAGAAGTTTTCCTGGCTGTC) of primers were used during the polymerase chain reaction for the amplification of *celS2* gene.

**Polymerase Chain Reaction**

The optimized PCR mixture consisted of 1x e-*Taq* buffer, 0.2 mM dNTP, 0.8 µM/L each of forward and reverse primers, 0.05 U/µL e-*Taq* polymerase (SolGent Co., Ltd), 1.0 µL of extracted DNA template was used and the total volume was adjusted to 10 µL per reaction volume.

The optimized PCR profile cycle began with pre-denaturation at 94°C for two minutes for one cycle, followed by 40 cycles of denaturation at 94°C for 20 seconds, annealing at 61°C for 40 seconds and extension at 72°C for one minute and then a final extension at 94°C for five minutes for one cycle. PCR gradient machine (BIOER, Genepro) was used in this study.

Five microliters of PCR products with 1.0 µL 6X loading dye (Invitrogen) were loaded into 1% agarose gel for electrophoresis using MUPID-2 plus (Mini gel electrophoresis) at 120V for 20 minutes. One kb ladder (Invitrogen) was used for size estimation in the gel electrophoresis. Ethidium bromide was used for gel staining for 15 minutes with constant shaking and the gel was de-stained with distilled water for another 15 minutes with constant shaking using Orbital shaker (FINEPCR). PCR amplicons were observed using Gel Documentation System (BIORAD). PCR amplicons were excised from gel and were sent to SolGent, Korea for sequencing.

**Phylogenetic analysis**

*celS2* gene and 16S rRNA deduced amino acid sequences were simultaneously analyzed using three different building algorithms which include two distance-based tree methods, Neighbor-Joining (NJ) [6] and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) [7] and one character-based tree method, Maximum-Likelihood (ML) [8]. The logic behind the use of three different building methods is that if a clade or certain grouping of species was supported by different building methods, with all their advantages and disadvantages, the phylogenetic tree obtained is from the data under analysis and is not biased from the method being used [9]. Jones-Taylor-Thornton (JTT) model was used for distance correction [10]. Bootstrap analysis using 1000 replicates was employed throughout the analyses to estimate the confidence limit of each divergence scenario [11].

## RESULTS AND DISCUSSION

NCBI annotated nine strains of *Streptomyces* with reported *celS2* gene which includes *S. coelicolor* A3 (2) (AL939108.1), *S. ambofaciens* ATCC 23877 (AM238663.1), *S. viridosporus* T7A (AF126376.1), *S. griseus* subsp. *griseus* NBRC 13350 (AP009493.1), *S. flavogriseus* ATCC 33331 (CP002475.1), *S. halstedii* (U51222.1), *Streptomyces* sp. THW31 (HQ286612.1), *Streptomyces* sp. SirexAA-E (CP002993.1) and *Kitasatospora setae* KM-6054 (AP010968.1). However, only six strains with available 16S rRNA sequences were used in the phylogenetic analyses (Table 1). DNA sequences of unknown streptomycetes isolated from soil were also included in the phylogenetic analyses.

**Table 1: Strains and GenBank accession numbers of sequences used in the phylogenetic analyses of *celS2* gene and 16S rRNA in streptomycetes.**

| Strain | GenBank accession number |
|---|---|
| *Streptomyces coelicolor* A3(2) | AL939108.1 |
| *Streptomyces sp*. SirexAA-E | CP002993.1 |
| *Streptomyces hygroscopicus subsp. jinggangensis* 5008 | CP003275.1 |
| *Streptomyces griseus subsp. griseus* NBRC 13350 | AP009493.1 |
| *Streptomyces flavogriseus* ATCC 33331 | CP002475.1 |
| *Kitasatospora setae* KM-6054 | AP010968.1 |

Based on *celS2* phylogenetic trees (Fig. 1, 2 and 3), *S. coelicolor* M145 is identical with *S. coelicolor* A3 (2) and this was supported by 99% bootstrap value. The phylogenetic tree also suggests that all of the unidentified strains are unique to those species available in the public database. Notably, strains 119.13, 119.103 and 119.104 were consistently grouped in separate cluster from other strains in NJ, ML and UPGMA analyses (Fig. 1, 2 and 3, respectively). *Streptomyces*. sp. SirexAA-E and *S. flavogriseus* were consistently

ISSN: 0975-8585

grouped in NJ , ML and UPGMA building methods and these were supported by 92% - 95% bootstrap values. Tree topology and clustering of species were congruent in the three building methods.

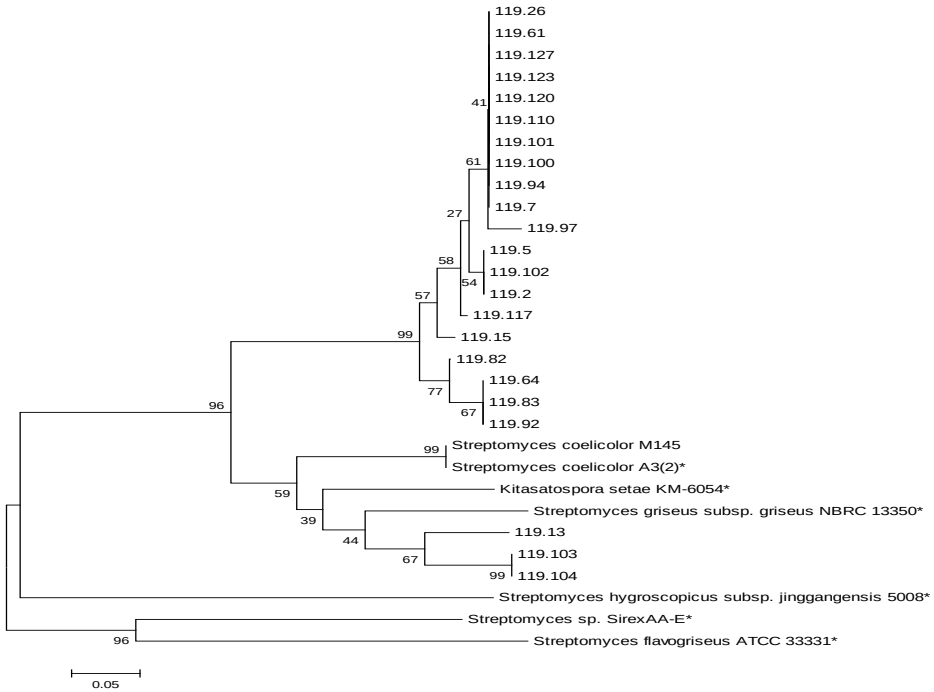


**Figure 1: Unrooted Neighbor-Joining (NJ) tree of *Streptomyces* strains based on *celS2* gene amino acid sequences. Unknown isolates of *Streptomyces* were found different from strains present in NCBI. Numbers at nodes represent bootstrap values from 1000 resampled datasets. Bar indicates sequence divergence. Jones-Taylor-Thornton (JTT) model was used for distance correction. (*) Homologous sequences from GenBank (see Table 1).**

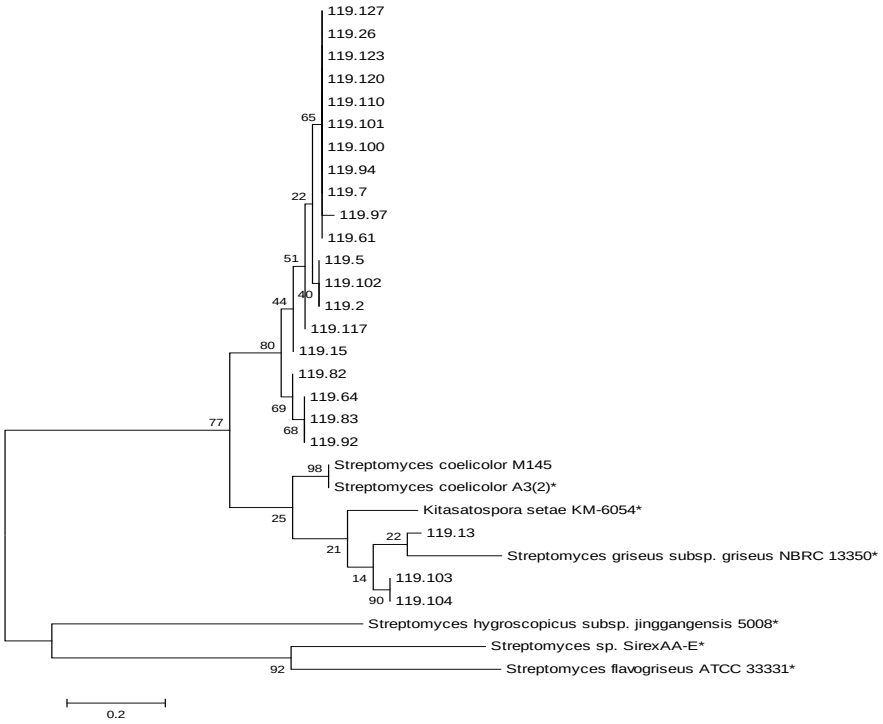


**Figure 2: Unrooted Maximum-Likelihood (ML) tree of *Streptomyces* strains based on *celS2* gene amino acid sequences. Unknown isolates of *Streptomyces* were found different from strains present in NCBI. Numbers at nodes represent bootstrap values from 1000 resampled datasets. Bar indicates sequence divergence. Jones-Taylor-Thornton (JTT) model was used for distance correction. (*) Homologous sequences from GenBank (see Table 1).**

ISSN: 0975-8585

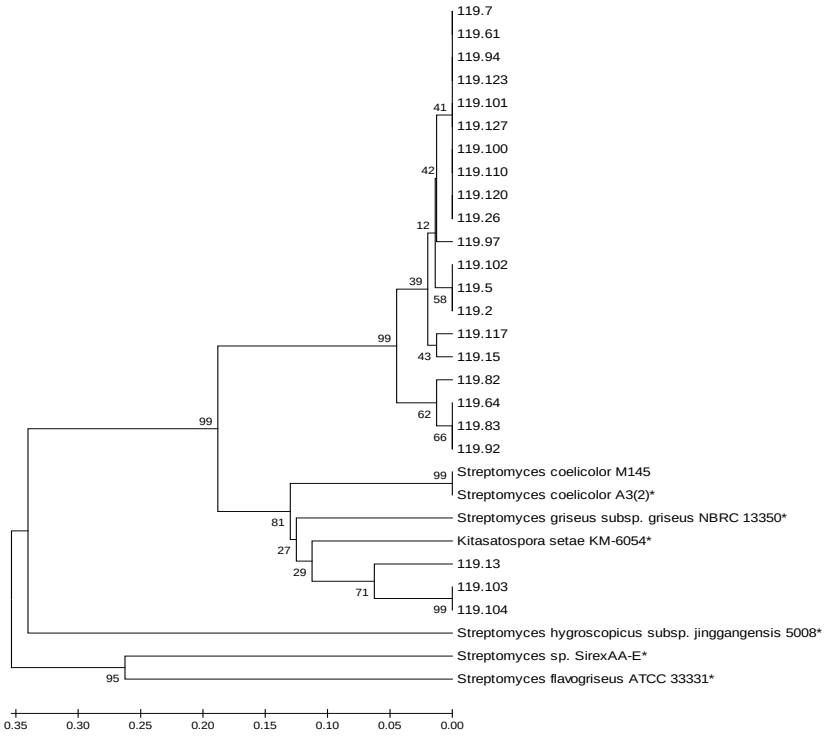


**Figure 3: Unrooted Unweighted Pair Group with Arithmetic Mean (UPGMA) tree of *Streptomyces* strains based on *celS2* gene amino acid sequences. Unknown isolates of *Streptomyces* were found different from strains present in NCBI. Numbers at nodes represent bootstrap values from 1000 resampled datasets. Bar indicates sequence divergence. Jones-Taylor-Thornton (JTT) model was used for distance correction. (*) Homologous sequences from GenBank (see Table 1).**

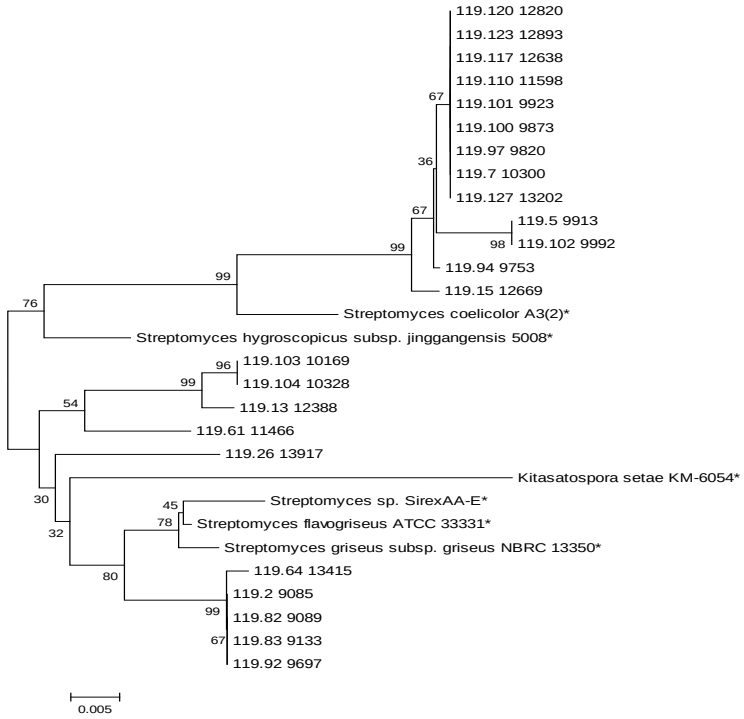


**Figure 4: Unrooted Neighbor-Joining (NJ) tree of *Streptomyces* strains based on 16S rRNA amino acid sequences. Unknown isolates of *Streptomyces* were supported by 16S rRNA phylogenetic tree to be different strains from those strains registered in NCBI. Numbers at nodes represent bootstrap values from 1000 resampled datasets. Bar indicates sequence divergence. Jones-Taylor-Thornton (JTT) model was used for distance correction. (*) Homologous sequences from GenBank (see Table 1). *S. coelicolor* M145 was not included in the phylogenetic analysis due to the unavailability of 16S rRNA sequence. Nonetheless, gene-based phylogenetic tree was observed to be supported by the highly conserved molecular chronometer, 16S rRNA, especially in discriminating unidentified strains.**

ISSN: 0975-8585

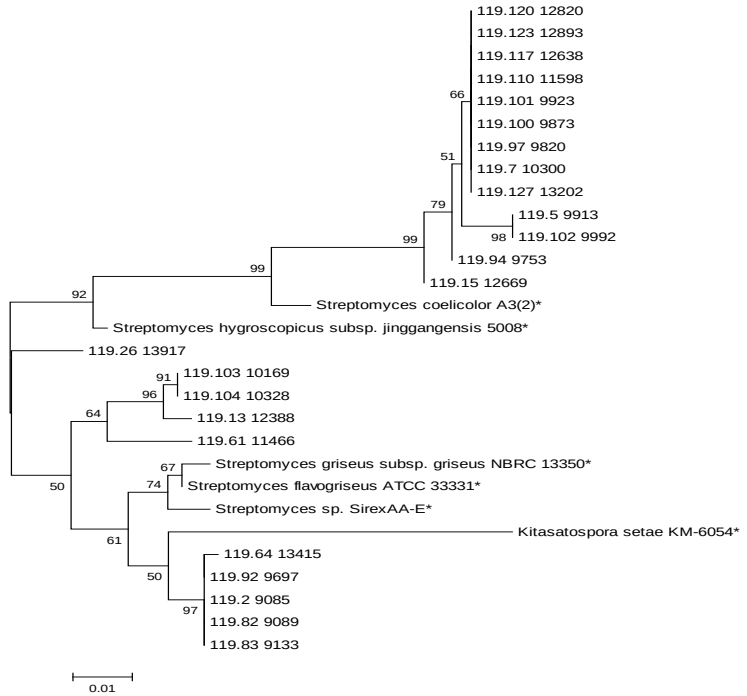


**Figure 5: Unrooted Maximum-Likelihood (ML) tree of *Streptomyces* strains based on 16S rRNA amino acid sequences. Unknown isolates of *Streptomyces* were supported by 16S rRNA phylogenetic tree to be different strains from those strains registered in NCBI. Numbers at nodes represent bootstrap values from 1000 resampled datasets. Bar indicates sequence divergence. Jones-Taylor-Thornton (JTT) model was used for distance correction. (*) Homologous sequences from GenBank (see Table 1). *S. coelicolor* M145 was not included in the phylogenetic analysis due to the unavailability of 16S rRNA sequence. Nonetheless, gene-based phylogenetic tree was observed to be supported by the highly conserved molecular chronometer, 16S rRNA, especially in discriminating unidentified strains.**

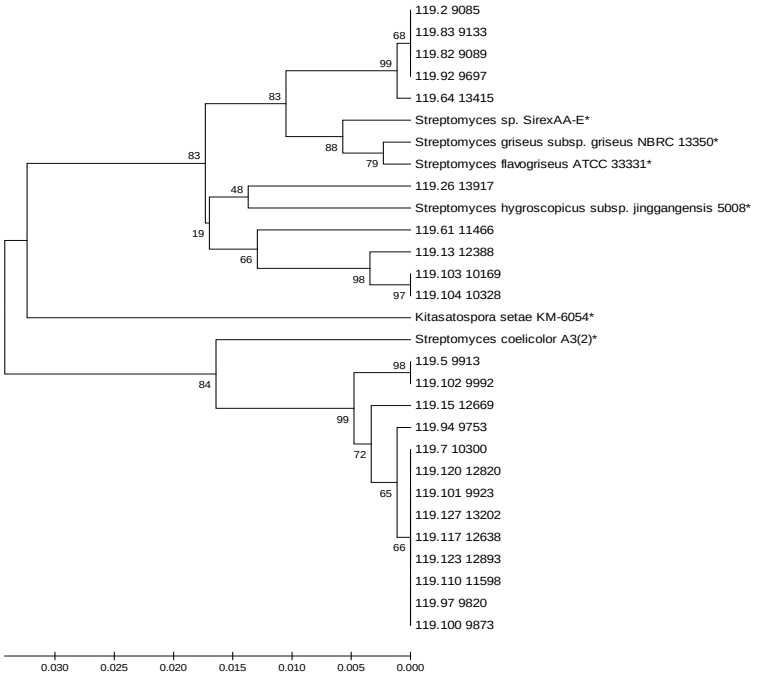


**Figure 6: Unrooted Unweighted Pair Group with Arithmetic Mean (UPGMA) tree of *Streptomyces* strains based on 16S rRNA amino acid sequences. Unknown isolates of *Streptomyces* were supported by 16S rRNA phylogenetic tree to be different strains from those strains registered in NCBI. Numbers at nodes represent bootstrap values from 1000 resampled datasets. Bar indicates sequence divergence. Jones-Taylor-Thornton (JTT) model was used for distance correction. (*) Homologous sequences from GenBank (see Table 1). *S. coelicolor* M145 was not included in the phylogenetic analysis due to the unavailability of 16S rRNA sequence. Nonetheless, gene-based phylogenetic tree was observed to be supported by the highly conserved molecular chronometer, 16S rRNA, especially in discriminating unidentified strains.**

16S rRNA phylogenetic tree (Fig. 4, 5 and 6) confirmed that unidentified strains are unique to strains available in the public database. The phylogenetic tree generated using 16S rRNA showed a more interesting topology. First, having 0.5% sequence divergence which is the lowest observed among the analyses. Second, there is clearer clustering of unidentified strains as compared to the *celS2* gene-based phylogenetic tree. Unidentified isolates, 119.13, 119.103 and 119.104, were also shown to be in the same cluster as observed in the *celS2* gene-based phylogeny. Other clustering observed in 16S rRNA phylogenetic tree were samples 119.2, 119.83, 199.82, 119.92, 119.64 and 119.5, 119.102, 119.7, 119.120, 119.101, 119.117, 119.123, 119.110, 119.97, 119.100, 199.15, 119.127 and 119.94. These three different clusters formed by unknown and reference strains using 16S rRNA were evident as two distinct clusters in *celS2* gene-based phylogenetic tree. *Streptomyces* sp. SirexAA-E and *S. flavogriseus* were also observed to be consistently grouped in the phylogenetic analyses suggesting a close evolutionary origin between the two species.

Design and optimization of polymerase chain reaction parameters to amplify the target gene was successful. Interestingly, *celS2* gene has higher sequence divergence as compared to 16S rRNA. Having higher sequence divergence than 16S rRNA, *celS2* gene has a potential to be used as a reference molecular chronometer in future phylogenetic studies on *Streptomyces* [12]. Apart from having higher sequence divergence, *celS2* gene was also observed to be conserved in a number of *Streptomyces* species. Notably, tree topology of the *celS2* gene demonstrated two separate clusters whereas the 16S rRNA phylogeny revealed one monophyletic (*S. coelicolor* A3 (2)-like) and two separate clusters of *K. setae*-like phylogeny. The congruency of tree topology and species clustering in 16S rRNA-based phylogeny and *celS2* gene-based phylogeny suggests that *celS2* can also be used in intraspecies discrimination within this bacterial group. In present study, we also found out that *K. setae* KM-6054 to be highly related species to *Streptomyces* as previously reported [13].

Three different building methods, NJ, ML and UPGMA also showed strong agreement on every analyses made using *celS2* gene and 16S rRNA. This study reiterates that intraspecies delineation of *Streptomyces* can be achieved using simultaneous analysis of protein coding genes and 16S rRNA.

## ACKNOWLEDGEMENT

This research was funded by the Next-Generation BioGreen 21 Program (No.: PJ009490), Rural Development Administration, Republic of Korea. Sincere gratitude is extended to Professor Elsie C. Jimenez of University of the Philippines Baguio for editing the manuscript and to the graduate students of Kookmin University, BioResource Laboratory, Youngseok Ham, Hee-Hoon Ahn, Yoon-Hee Kim and Han-Saem Park, for the technical support they extended during the conduct of the study.

## REFERENCES

[1] Bently SD, et al. Nature 2002; 417:141-147.
[2] Kim BJ, et al. Int J Syst Evol Microbiol 2004; 54: 593-598.
[3] Guo Y, et al. Int J Syst Evol Microbiol 2008; 58: 149-159.
[4] Lingzhu M, et al. J Antibiot 2011; 64: 97-101.
[5] William S, et al. Bacterial genomic DNA isolation using CTAB. DOE Joint Genomic Institute 2004.
[6] Saitou N, and Nei M, Mol Biol Evol 1987;4: 406-425.
[7] Murtagh F, Computational Statistics Quarterly 1984; 101–113.
[8] Felsenstein J, PHYLIP (Phylogeny inference package), version 3.5c. Department of Genetics, University of Washington, Seattle, USA 1993.
[9] Chelo IM, Ze-Ze L, Tenreiro R, Int J Syst Evol Microbiol 2007; 57: 276-286.
[10] Jones DT, et al. Comput Appl Biosci 1992; 8: 275-282.
[11] Felsentein J, Evolution, 1985; 39: 783-791.
[12] Yanez MA, et al. Int J Syst Evol Microbiol 2003; 53: 875-883.
[13] Hsiao N, and Kirby R, Antonie van Leeuwenhoek 2008; 93: 1-25.