Doctoral Program in Life Science Innovation
Student ID: 201736055
Student Name: Nurul Syahirah Binti Shamsol Anuar

1. Title of the Thesis: Genome Mining in Tropical Freshwater *Streptomyces* spp.: Identification and Expression of Geosmin Synthase Gene in Response to Temperature

2. Summary

Streptomyces is one of the most promising producers of secondary metabolites including antibiotics, immunosuppressant, extracellular enzymes and terpenoid metabolites. Geosmin is a widely known terpenoid compound that contributes to the unpleasant odour in freshwater systems. This odour issue is most prominent during cyanobacterial blooms in a warmer month in temperate regions. However, geosmin production by *Streptomyces* in tropical Asia has not yet been elucidated in detail. The research presented in this thesis was conducted in response to the worldwide taste and odor (T&O) problems occurring in water bodies particularly related to the regulatory mechanism of geosmin production which is still unclear until now. In this thesis, we have explored the context of geosmin synthase gene from the draft genome sequence of *Streptomyces* sp. S1 and *Streptomyces* sp. S5, analyzed the temperature stimulus on the expression of geosmin synthase and investigated if other genes are functionally related with geosmin synthase or not in order to better understand the synthesis pathways of geosmin in this genus.

Previous study successfully isolated two odoriferous *Streptomyces* spp. (S1 and S5) from tropical freshwater of Southeast Asia. Both isolates showed different behavior for the production of geosmin as a same compound under different temperature. S1 showed a clear induction of geosmin at 30°C and none at 20°C contradict with S5 which produced higher amount of geosmin at 20°C compared to 30°C. We believe that they might have a completely different regulatory mechanism even for the same compound production such as geosmin. There are two objectives involved in this study. First, identification and phylogenetic analysis of geosmin synthase gene from the draft genome sequences of tropical *Streptomyces* spp. and second, expression of *geoS1* and *geoS5* in response to temperature together with geosmin operon prediction.

Generally, in the first objective, this study wants to investigate the context of geosmin synthase in the genome of S1 and S5. To better understand the mechanism of geosmin production from these isolates, the genome of S1 and S5 were sent for Next Generation Sequencing (NGS). Bioinformatic approach was used to do some annotation, searching for the homology against the experimentally verified enzymes, identification of geosmin synthase conserved domain and secondary metabolite biosynthetic gene clusters. Based on the homology search and conserved domain against 76 germacradienol/geosmin synthase of *Streptomyces* species, we speculated that *geoS1* and *geoS5* are potential candidate germacradienol/geosmin synthase similar to already characterized *S. coelicolor* A3(2), *S. avermitilis* and *S. albus*. Comparison of the tree topologies between *geoS1*, *geoS5* and whole genome *Streptomyces* indicate the involvement of horizontal gene transfer occurred during the evolutionary history.

Recently, the discovery of geosmin synthase gene in *Streptomyces* spp. and cyanobacteria provide a fundamental knowledge required to begin investigations into growth conditions affecting the expression of the geosmin synthase gene. In order to elucidate the mechanism of geosmin production in *Streptomyces*, this study investigates the expression of *geoS1* and *geoS5* and the relative production of geosmin in response to temperature. To investigate any changes in geosmin synthase gene expression in response to temperature, three different temperatures were chosen based on data of the previous study: 20°C indicate low temperature, 30°C indicate high temperature and control culture was maintained at 25°C. S1 showed no correlation between geosmin production and *geoS1* expression while S5 showed the same trend of geosmin production and *geoS5* expression at different temperatures. This raised an important question, "Is geosmin production under different temperature is based on the regulation of expression or not?" We predicted that both *Streptomyces* species may have different regulation mechanisms for the same compound production, in this case, Geosmin. This is why it shows a different response against

temperature. Two main questions arise from this prediction, (1) How these isolates adapt or response to temperature changes? and (2) How temperature transfer signal to the level of gene expression? Probably the answer to these questions lies in clusters of coregulated genes called operon. Therefore, we attempted to identify geosmin operon structure to evaluate the connection of mRNA between each ORF with *geoS1* and *geoS5*.

Bioinformatic analysis revealed that *geoS1* and *geoS5* have different geosmin synthase operon arrangement. Primers were design covering the flanking region between adjacent genes and geosmin synthase to identify whether adjacent genes of *geoS1* and *geoS5* are likely to be cotranscribed together. RT-PCR showed amplification product of the expected sizes of ~325 bp and ~1.3 kbp in *geoS1* operon while product sizes of ~780 bp and ~978 bp in *geoS5* operon. Therefore, it is possible that adjacent genes can be co-transcribed together and probably form an operon structure to produce geosmin due to the same transcription orientation. This is the first evidence of geosmin synthase is transcribed in polycistronic in *Streptomyces* species. Furthermore, different operon organization in *geoS1* and *geoS5* imparted that adjacent genes may have some functions for geosmin production that requires further investigation. Following this, expression of geosmin synthase and estimation of operon structure of *geoS1* and *geoS5* may provide some estimation on the location of promoter region. Potential promoter region was identified in *geoS5* but none in *geoS1*. Thus, it further reinforcing our assumption that *geoS5* is regulated during transcription.

This study enhances the understanding of exploiting geosmin synthases information, phylogenetic relatedness and gene regulation in *Streptomyces*. By combining the genome mining approach and expression regulation study, I am trying to develop a general knowledge regarding the regulatory mechanism of secondary metabolite production. The idea is "even for synthesis of secondary metabolite production of the same compound in *Streptomyces* species, there may have different regulatory mechanisms". This study not just illustrate the idea but also point out that temperature may not be the key factor for several Streptomyces species due to difference in promoter region and also operon structure. Since *Streptomyces* is a major producer of secondary metabolites of economic importance, this idea could provide useful information on other metabolites, especially antibiotics and other bioactive compounds. Following the outcome of this research, this is a new finding to estimate the regulation of geosmin production as no research may focus on this area. In this case, the regulatory mechanism sense by temperature is different by confirmation of geosmin production and expression of geosmin synthase together with operon structure prediction. Thus, the aim "production of the same compound but have a different regulatory mechanism" might open an exciting research area to be focused on.

(1048 words)