METABOLOMIC AND GENOMIC INSIGHTS INTO THE TAXOL PRODUCING BACTERIA FROM MARINE MACROALAGE

Synopsis submitted to Madurai Kamaraj University for the degree of

DOCTOR OF PHILOSOPHY in MICROBIOLOGY

by

S. MUGESH

Reg. No: F9067

UNDER THE GUIDANCE OF

Dr. M. Murugan

Assistant Professor and Head-in-charge

Department of Microbial Technology

School of Biological sciences

Madurai Kamaraj university

Madurai – 21.

Metabolomic and genomic insights into the taxol producing bacteria from marine macroalgae

All living organisms produce biomolecules as metabolites, which are categorized as primary and secondary metabolite (Mann, 1987). The marine environment comprises half of the earth's biodiversity and the largest biomolecule reservoir (Kolanjinathan *et al.,* 2014). Marine macroalgae are also known as seaweeds, which are the most abundant biomass producers in the marine environment (Bhadury and Wright, 2004). Marine macroalgae are present in the benthic habitats often and throughout the marine milieu as phytoplankton (Garson, 1989; Kadam *et al.*, 2013).

In the marine environment, all living and non-living surfaces are harbor diverse bacterial assemblages, and small molecules produced by bacteria which determine the selectivity of encrusting on eukaryotic surfaces or in tissues (Puglisi *et al.,* 2019). Despite the epiphytes of marine macroalgae, bacteria also inhabitant in the inside of the algal cell. Marine macroalgae and the bacterial association have been known a long time before through the algal plastids origination from endosymbiotic cyanobacteria (Margulis, 1998). Microbial association of algae, coral, and sponges have been exhibited through 16S rRNA gene approach (Olsen *et al.,* 1986; Olson and Kellogg, 2010; Geng and Belas, 2010). The ecological impact of bacterial assemblages not only limited to its host but also influential to the other eukaryotes of the same ecosystem (Hadfield, 2011). The ecological coherence of bacterial association with algae was defined as highly host-specific, which appealing evolutionary perspectives (Philippot *et al.,* 2010). The recent advancement of molecular techniques could exhibit both physiological and functional relationships of bacteria associated with algae (McFall-Ngai, 2008; Philippot *et al.,* 2010).

Metabolomics is the systematic analysis of metabolites produced from the biological system and has three principal approaches, namely metabolite profiling, metabolic fingerprinting, and metabonomics (Oliver *et al.,* 1998; Hall, 2006; Barchet, 2007;). Genomics is the comprehensive study of genomic DNA in the biological system which maps the genome.

Taxol is a diterpenoid pseudoalkaloid compound and known as an anti-microtubule agent used as a chemotherapeutic drug for cancer (Rowinsky *et al*., 1990). In the early 1960s, the

National Cancer Institution (NCI) and the US Department of Agriculture (USDA) initiated the plant screening program for anticancer substances (Zubrod *et al.,* 1966). Among several plant extracts, the bark extract of *Taxus brevifolia* Nutt. (Pacific Yew) showed potent activity against cancer cells like Melanoma B16 and the active constituent purified, characterized as Taxol (Fig. 1.3.) (Wani *et al.,* 1971). Taxol has been further studied for their mode of action, like preventing depolymerization in cancer cells lead to cell death (Schiff *et al.,* 1979). However, the Taxol has been approved by the Food and Drug Administration (FDA) after 21 years of its discovery for the treatment of refractory ovarian and metastatic breast cancer (Suffness and Wall, 1995). More than 400 taxanes were reported from the *Taxus* species (Baloglu and Kingston, 1999; Itokawa, 2003; Shi and Kiyota, 2005), and it has been found in the vascular cambial region of the stem, needles and root tissue of *Taxus* (Strobel *et al.,* 1993). Taxol as a plant immune compound in *Taxus* spp., which is a potent inhibitor against the full range of plant fungal pathogens (Young *et al.*, 1992; Elmer *et al.,* 1994), woodworms (Daniewski *et al.,* 1998) and is toxic to mammals (Odgen, 1988).

Taxol and its analogs are the leading drugs for the treatment of cancer, and its sales cost five million US dollars (Croteau *et al*., 2006; Onrubia *et al.,* 2013). Overexploitation of *T. brevifolia* and the cost of drug influence, further, discovery of taxol biosynthesis has implicated and developed to meet the demand (Arbuck and Blaylock, 1995; Goodman and Walsh, 2001). In this regard, we need to understand the entire taxol biosynthesis pathway in order to enhance productivity and it could be achieved by the metabolically engineered bacterial host (McElroy and Jennewein, 2018). In *Taxus* spp., 19 enzymatic steps have been characterized from the geranylgeranyl pyrophosphate (GGPP) to taxol in a long end Methylerythritol Phosphate (MEP) pathway (Jennewein *et al.*, 2004). Other than *Taxus* spp., microbial endophytes of *Taxus* spp. were reported as taxol producers (Flores-Bustamante *et al.,* 2010). In the case of bacteria, *Bacillus cereus*, *Bacillus megaterium*, *Bacillus subtilis*, *Pantoea* sp., *Curtobacterium* sp. and *Sphingomonas* sp. were patented and no subsequent scientific report for taxane/taxol production (Page *et al*., 2000). On the other hand, fungi are well characterized for taxol biosynthesis and identified respective genes like *Taxus* spp. (Hao *et al*., 2013).

The present work aims to explore the taxol-producing endophytic bacteria from marine macroalgae. The main objectives of the present work include,

- Collection and isolation of endophytic bacteria from marine macroalgae
- Screening of taxol production in bacterial isolates from marine macroalgae
- Genome sequencing and draft genome assembly of taxol-producing bacteria
- In silico analysis of taxol biosynthesis pathway in the genome of taxol-producing endophytic bacteria
- Profiling of other secondary metabolites produced by taxol-producing bacteria

Taxol producing endophytes are known for several decades and attained the industrial concern for its high reproducibility. There are several reports elucidated on endophytic fungi for taxol production, even though obtained from non-taxol producing plants. However, in the case of bacteria, only a few have been reported for taxol production, which was patented for mass production. The patented bacteria were isolated from *Taxus* plant and remain anonymous. After a long time, our study unveils the marine endophytic bacteria and theri taxol production. The marine environment is plausibly known for its diverse chemical ecology, which was the reason relayed for this research. Nine different marine macroalgae were collected and isolated 32 endophytic bacteria, which were screened for taxol production. Among the bacterial isolates, 3 (DMTMMB08, DMTMMB10, and DMTMMB24) were showed taxol production which was confirmed by chromatography and mass spectrometry analysis along with standard taxol. In this study, taxol-producing bacteria isolated from *Sargassum polycystum* and *Acanthaphora specifera*, both are a distinct class of algae (brown and red algae). The 16S rRNA sequencing of the newly isolated taxol-producing bacterial isolates were identified as *B. flexus* DMTMMB08, *B. licheniformis* DMTMMB10, and *O. picturae* DMTMMB24. These taxol-producing bacteria have propelled to the forefront in investigating other secondary metabolites using GC-MS analysis. The GC-MS analysis highlighted the various cyclic dipeptide profile and exhibited the phthalates biosynthesis, which could strengthen the discovery in bacteria.

The inadequate research on the bacterial biosynthesis of taxol aims to address it, and the genome analysis delves into it. The taxol-producing bacterial genome has been sequenced, assembled, annotate, and deposited in the NCBI database. A total of 3,638 (3,517 protein-coding genes), 4,469 (4,300 protein-coding genes) and 3,678 (3,563 protein-coding genes) were predicted in *B. flexus* strain DMTMMB08, *B. licheniformis* strain DMTMMB10 and *O. picturae* strain DMTMMB24,

respectively. Comparative genome analysis performed, and the results revealed that 994 core genes shared by the three bacterial genomes (Chaudhari *et* al., 2016). In which, accessory genes and unique genes of the across the three bacterial genomes have been noted for considerable common and unique secondary metabolites from these organisms.

Followed by the comparative genome analysis. Biosynthetic gene cluster (BGC) analysis revealed the signature of known and putative BGCs like type 3 polyketides (T3PKs), bacteriocins, siderophore synthases, terpenes, non-ribosomal peptides (NRPs), polysaccharides and fatty acids within it. In *B. flexus* strain DMTMMB08, the novel hydroxamate siderophore signature has been noted in predicted BGC. Notably, the xenotetrapeptide gene cluster match revealed the potential game X peptide production in *O. picture* DMTMMB242. Draft genome assembly and biosynthetic gene cluster analysis does not show any genomic relevance to the taxol biosynthesis of bacteria. The enzymes involved in the taxol biosynthesis has challenge still for its uncovered steps in it. The preliminary step of isoprene biosynthesis occurred through the MEP and MVA pathway, which determined by the precursor availability from the nutrient source. Most of the diterpenoid production achieved through the MEP pathway instead of MVA pathway where occurs in the plastid of the plant. In bacteria, the MEP pathway is the predominant one for the biosynthesis of any terpenes. On the contrary, few bacteria accomplish the isoprene production through the MVA pathway and with the same end product as the MEP pathway. The KAAS analysis of the taxol-producing bacterial genome revealed that the *B.* flexus DMTMMB08 and *B. licheniformis* DMTMMB10 follows the MEP pathway, whereas *O. picturae* DMTMMB24 follows the MVA pathway.

Genome-wide PSI-BLASTP analysis has been performed using *Bacillus flexus* strain DMTMMB08, *Bacillus licheniformis* strain DMTMMB10, and *Oceanobacillus picturae* strain DMTMMB24 genomes against *Taxus* spp. Non-Redundant protein database and taxol biosynthesis candidate genes with a 1e-05 error threshold value. We have found the important taxol pathway genes cytochrome P450 monooxygenases, acetyltransferases, phenylalanyl-CoA-ligase, and phenylanalnyl aminomutase sequences match in this BLASTP analysis. However, we did not find any TXS signature sequence matches. Hence, the terpene synthase profile was curated from the genome of *B. flexus* DMTMMB08, *B. licheniformis* DMTMMB10, and *O. picturae*

DMTMMB24. We have found that both tetraprenyl beta curcumene synthase and Squalene-hopene cyclase enzymes from the terpene synthase family in *B. flexus* strain DMTMMB08 and *B. licheniformis* strain DMTMMB10.

On the other hand, two genes of phytoene/squalene synthase present along with tetraprenyl-beta-curcumene synthase in the genome of *O. picturae* DMTMMB24. Therefore, seven terpene synthase three-dimensional structures were exhibited the domain architecture of terpene synthases of taxol-producing bacteria. This terpene profile of these taxol-producing bacteria contains class I, class IB, and class II motifs, which may lead to the cognizance of terpene synthase evolution further. Among the terpene synthases, tetraprenyl-curcumene synthases have been reported for multi-functionality for the synthesis of sesquiterpenes, triterpenes, and sesquiterpenes and even reported the activity with GGPP as a non-natural substrate.

To analyze the evolutionary relationship of terpene synthase (TPS) in taxol-producing bacteria, TPSs divided into four groups, namely, tetraprenyl beta curcumene synthase group, squalene-hopene cyclase group, phytoene/squalene synthase group-I and phytoene/squalene synthase group-II based on its motif class and motif signature. Tetraprenyl beta curcumene synthase from *B. flexus* DMTMMB08, *B. licheniformis* DMTMMB10, and *O. picturae* DMTMMB24 taken as a query for PSI-BLASTP and 1340 unique sequences were collected for tetraprenyl beta curcumene synthase group. In the same way, sequences were collected for the squalene-hopene cyclase group (4624 unique sequences), phytoene/squalene synthase group-I (10,780 unique sequences), and phytoene/squalene synthase group-II (2454 unique sequences). The sequences were processed and analyzed for phylogenetic tree construction and exhibited that TPSs could be an evolutionary model from lower-level organisms to higher-level organisms. However, the isolation source should determine the evolutionary event/distance. Henceforth, the evolutionary analysis TPSs strongly attributes the marine to terrestrial evolutionary theory. Moreover, we did not find this pattern or organization of terpene synthase in and among the collected sequences of this study.

Furthermore, the *in silico* analysis of Opi-TS showed homodimer in nature, and the molecular docking simulations explored the binding residues of Opi-TS with the GGPP substrate. *In vitro* studies on functional analysis of Opi-TS with GGPP revealed the production of β-springene and

its taxol-producing ability to either relate or differentiate the diterpene biosynthesis by investigating GGPP with the terpene synthases of taxol-producing bacteria in the future.

In conclusion, this study revealed that the bacterial strains, namely *Bacillus flexus* strain DMTMMB08, *Bacillus licheniformis* strain DMTMMB10, and *Oceanobacillus picturae* strain DMTMMB24 isolated from marine macroalgae are the taxol producers. The terpene synthase profile in taxol-producing bacterial genomes would be a great interest in global research on terpene biosynthesis and its evolution. The promiscuity of terpene synthase exploring the diversity and complexity of terpenoid production, which is useful for designing approaches to improve their catalytic mechanism. Moreover, this investigation can expand the boundaries to understand the taxol production in the bacterial system as well as to study the terpene synthase evolution and distribution further.