International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(4):1439-1454

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i04.01815

Detection of Metabolic Gene of Interest (MGOI) in WGS of *Microbacterium barkeri* (LMA4)

Krishna Kumar Das, Smaranika Pattnaik*

Department of Biotechnology and Bioinformatics, Laboratory of Medical Microbiology, School of Life Sciences, Sambalpur University, Jyoti Vihar, Burla, India. *Corresponding Author's Email: smaranika2010@suniv.ac.in

Abstract

This communication reports about detection of metabolically active genes of interest cell free extract of actinomycetal isolates from Pond soil. The strain LMA4 previously identified as *Microbacterium barkeri*, based on biotyping as well as 16 s r RNA sequencing, was observed with moderate bioactivity against test strains of *Staphylococcus aureus* and *Escherichia coli*. The Actinomycetal strain was Gram positive, aerial profuse branching bearing rounded spores. The whole genome sequencing was carried using Next Generation Sequencing tools and techniques. The FASTA sequence received from sequencing farm had been subjected to BLASTn analysis and aligned with sequences of databse. The sequence similarity was checked considering percentage of identity, as inferred from online portal. The whole genome sequence of LMA4 had inferred about elution of antibiotic producing genes namely [i] Thiazole synthase protein, [ii] Indole-3-glycerol phosphate synthase Protein, [iii] Antibiotic synthesis MbtH protein, [iv] Antibiotic biosynthesis monooxygenase protein, [v] class II terpene cyclase protein and [vi] YcaO-like family protein in the sequence profile. The genes were identified, ORFs were predicted and more over conserved domain was spotted. The genes were structurally represented using modelling approaches and more so ligand based (target oriented) molecular simulation study had predicted affirmative interactions between receptor and ligand molecule. The receptors were WGS detected gene products, and the ligands were Sortase of *Staphylococcus aureus* and D-ala-ala ligase of *Escherichia coli*. The study may pave a way towards extracting the metabolites of interest and implementing as natural cost effective drugs.

Keywords: Conserved Domain, Homology Modeling, Ligand Based Molecular Simulation, *Microbacterium barkeri,* Metabolic Gene of Interest, Whole Genome Sequencing.

Introduction

Microbacterium barkeri, a freshwater pond soil isolate (1) is a Gram positive, with profuse aerial branching system. The organism is prominent in actinomycetes research field, because of its ability to produce industrially important enzymes (a) polyvinyl biodegradation activity in textile waste (2), cellulase (3), (b) agar-agar (4), (c) low and high density polyethylene film (5), (d) bioconversion of steviol glycosides into steviol (6) and so on. Therefore, it is being enlisted as an organism for the source of industrially important enzymes. For the high throughput production of enzymes/proteins/or compounds, it is mandated to carry out genomics studies as genes are the master molecules of each of the products. However, few authors had contributed their efforts to bring about genes of metabolic activity except a report (7), which had recommended this organism as a biocontrol agent, making a draft genomic study. Hence, it is understood that *Microbacterium barkeri* is a designated biocontrol agent as evidenced from published reports. However, previous reports had not indicated the genes which are detected in LMA4 strain. The genes eluted are annotated as antimicrobial agents specifically terpene cyclase, which carry forward the terpenes or terpenoids synthesis. Therefore, the actinomycetes could have inhibited the bacterial growth, when tested in vitro using qualitative as well as quantitative assays. Therefore, this present pursuit aims at detecting metabolically active genes in real time, using in vitro based study annotated with *in silico* based tools and techniques, using genome sequencing approaches. The DNA sequencing is a vertical process to know the order of nucleotides present in the stretch of DNA. The order of sequencing is carried out considering four numbers of building block units are Adenine, Guanine, Cytosine and Thiamine. The sequencing information can determine the location or position of DNA and also degree of expression, thus researcher can infer

(Received 29th July 2024; Accepted 22nd October 2024; Published 30th October 2024)

about the specific gene of interest. Due to advent of the Next generation sequencing (NGS), researchers although had skewed towards WGS, can precisely defined as the process of determination of whole DNA sequence at a single time. In NGS the read length is the number of base pairs sequence from DNA fragment under study. In sequencing, a definition, contig is used which is derived from the word "contiguous", a set of sequences that has been over lapped in such a way that, a contiguous representation of genomic region can be made. This region provides a physical map of segments of DNA across a genomic region. Thus, contigues provides an actual DNA sequence. In genome sequencing the stretches of these contigues are the generated DNA sequences which are called "Sequence ridge". In addition to these terms another term "Gap Filling" is used in WGS. Gap filling is the reconstruction of gaps / missing sequences between the contiguous sequences (8, 9). Therefore, NGS bioinformatics has made its entry and generating considerable developments due to elevated computational algorithms and software tools by assisting all the steps from raw data processing to finish data analysis and more so detailing the interpretations. In bioinformatics there is a tremendous use of term Open Reading Frame, which is portion of DNA sequence that spans between start and stop codon. In the process of gene finding specifically in Prokaryotes, "ORF finding tools" is used which can detect the start codon A, T, G or its alternatives and terminating codons TAA, TAG and TGA. As the codon is triplicate codon depending on the detection of starting point or nucleotide, there are six possible frames are considered "three on forward strand and three on complementary strands" by translating into amino acid sequence according to the genetic code. Hence these frames are called "Reading Frames". The Coding zone Sequence (Figure 1) refers to nucleotides can be divided into codons, which are translated into amino acids by the Ribosomal translational machinery (10). Detection of specific gene clusters of secondary metabolism importance are possible using various bioinformatics tools (11). The authors had detected the gene cluster using various bioinformatics based gene predication tools, however the searching had involved multi step metabolic pathway assembling, coordinating and finally giving the output.

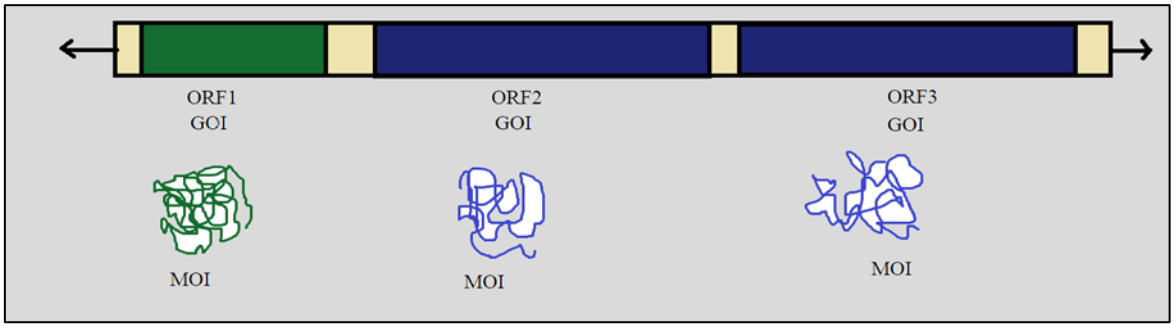

Figure 1: The Coding Zones of Typical Chromosome Harbored with Gene of Interest 1, Gene of Interest 2 and Gene of Interest 3, Expressing Respective Secondary Metabolites

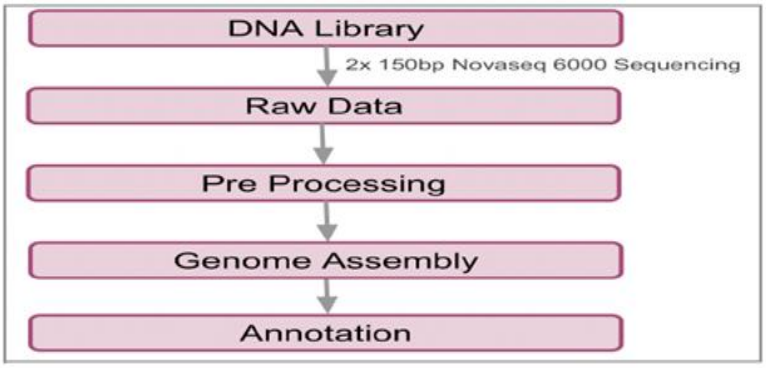

Figure 2: Depicting the Data Analysis Workflow, For WGS of LMA4. The DNA Library was Created, The Raw Data was Assembled, was Preprocessed, The Contigs were Assembled and the Eluted Sequences at A Stretch, were Annotated Structurally as Well as Functionally

With this brief introduction it is understood that the WGS is a directive to detect metabolite of interest (MOI) at gene level in newly isolated actinobacterial strain of *Microbacterium barkeri* (LMA4) using NGS in the eluted sequence profile.

Materials and Methodology

Whole Genome Sequencing (WGS)

LMA4 (*Microbacterium barkeri*), the Pond soil isolate, previously identified with 16s rRNA sequencing (12) was selected for Whole Genome Sequencing using NGS approaches.

Preparation of Pellet

The actinomycetes cells grown on the plates were separated from the surface of agar and suspended in re-suspending buffer and were subjected to centrifugation using a cooling centrifuge. The supernatant was discarded and the settled palette containing cellular biomass was collected. The NGS (Next generation liquid) sequencing platform was Illuming (Novaseq 6000), 150PE.

The Work Flow of Genome Sequencing

The data analysis workflow which was maintained for the sequencing is given in the Figure 2. The DNA library was constructed using 2x150bp Novadeq 6000 Sequencing approach, and the raw data (RD) was assembled. The RD was subjected to Pre-Processing, Genome Assembly and Annotation. The DNA library was created with 2x 150bp Novaseq 6000 sequencing, which had generated the raw data (RD). The RD was preprocessed and the genomic assembly was made for the annotations were completed using OMICS techniques and tools.

Quality Control of Data (QC)

The raw data quality was checked using FastQC and MultiQC softwares. (www.bioinformatics.babraham.ac.uk/projects/fastqc and https://multiqc.info/).The data was checked for base call quality base call quality distribution, % bases above Q30, %GC, and sequencing adapter contamination. The FastQC software, filters and reports about the quality of raw data generated from high throughput NGS data. The basic functions are for (a) making a statistics of raw data, (b) per base sequence quality checking, (c) per base sequence content, (d) GC content analysis, (e) Sequence length distribution, (f) Sequence Duplication levels, (g) over presentation of sequences and more so k-mer contents. While, the MultiQC collects numerical statistics from each module at the top the report, so that the data could be tracked upon. Accordingly, the data was processed for further study.

Summary Data Generated from WGS of LMA4

The summary of data generated from WGS of LMA4 is as follows. There were 44586960 number of 'Reads', data in GB (Giga Byte) was 6.73, Read length was 151, GC% was 58%, Q% was 97.8% respectively. The Mean Sequence Quality is defined by Phred Score. The Phred quality score is used to indicate the measure of base quality in DNA sequencing (6). High consistency of a sequenced base is indicated by greater values of Phred. A Phred Score of 20 indicates the likelihood of finding 1 incorrect base call among 100 bases. In other words, the precision of the base call is 99%. As the Phred Score is 20, there was 1 incorrect base/100bases under screening, a part of 'Base calling'.

Adapter Sequence

The adapter sequence taken for Read 1 and Read 2 are given in Figure 3. Adapters are oligonucleotides ligated during the sequencing library preparation to the 3′ and sometimes 5′ ends of biological sequences. An adapter helps in the specific enrichment of adapter-ligated DNA during the PCR step (7).

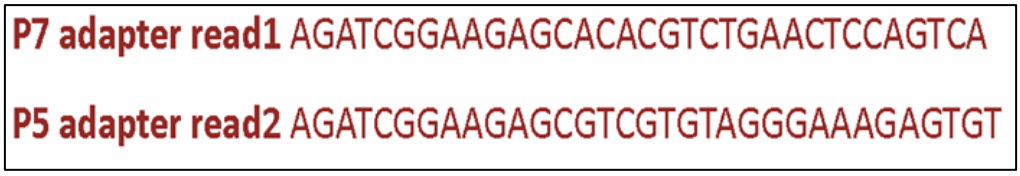

Figure 3: The Adapter Sequence for Read 1 and Read 2 Respectively

The Whole Genome Sequencing

The whole genome sequencing was carried out for LMA4 in the sequencing laboratories of MedGenome Labs Ltd. 3rd Floor, Narayana Netralaya Building, Narayana Health City, # 258/A, Bommasandra, Hosur Road, Bangalore – 560 099. The culture was subjected to De-novo Genome Sequencing, using the Sequencing Platform,

namely Illumina NovaSeq 6000 (PE 150), with Coverage X, 761,Preprocessing using FASTp, assembled by SPAdes v3.15.3 software, Gene annotation by RAST software, whereas, Gene Completeness had been assessed by BUSCO.

Homology Modeling of the Unidentified Genes / Proteins Structure and their Validation by Ramachandran Plot Analysis: From the results inferred from conserved domain search (CDS) of eluted proteins of interest, it was observed that there were six no. of proteins namely [i] Thiazole synthase protein, [ii] Indole-3-glycerol phosphate synthase Protein, [iii] Antibiotic synthesis MbtH protein, [iv] Antibiotic biosynthesis monooxygenase protein, [v] class II terpene cyclase proteinand [vi] YcaO-like family protein sequence in LMA4 could be identified. Keeping the sequences in record the modeling approaches were made for each of the eluted protein sequences using MODELLER software SWISS MODEL (https://swissmodel.expasy.org/) in the Expasy web server. Accordingly the eluted sequences were analyzed for its validity as protein structure using Ramachandran Plotting analysis (RCP) the optimized model of each of the proteins were noted down and a comparative structural analysis were made (13).

Bioactive Enzymes/ Proteins/ Drug Compounds of Interest for Molecular Docking Study: The sequences of enzymes/proteins/drug compounds of interest (Ligands) were captured entrusted with conserved domain search at NCBI portal, which was detailed in previous chapter. The Protein sequences were further hyperlinked with SWISS-MODEL (https://swissmodel.expasy.org/) for homology modeling and the PDB structural files were acquired. The PDB files were visualized in discovery studio and also in the SWISS-MODEL server visualizer. The eluted modeled three dimensional structures (3D structures) are given in the below Figures. While the respective targets of each of ligands, were searched in RSCB portal (www.rcsb.org) and molecular docking was carried out using online docking server for protein-protein complex named HawkDock (A web server for the structural prediction and analysis of protein-protein complex) and also standalone software Autodock4.2.6 (14). The PDB files of targets of interest (Sortase of *Staphylococcus aureus* and D-ala-ala ligase of *Escherichia coli*) were downloaded from RCSB (www.rcsb.org) and the PDBqt files from Mcule data bank (www.mcule.com), are given in Figure 4. The homologically modelled structure of 6 nos. of proteins of *Microbacterium barkeri* (LMA4) having sufficient drug ability were considered as ligand, while D-ala-D-ala ligase (*E.coli*) and sortase (*S. aureus*) were the receptor molecules. The docking simulation study was carried out in the HawkDock server (A web server for the structural prediction and analysis of protein-protein complex) for the protein-protein interaction study. Accordingly, the protein sequences (PDB files) were submitted (both ligand molecules and receptor molecules) in the prescribed window. Then the program was run using the default parameters. Binding energies in the form of docking score were recorded. From the best 10 docked poses the top model having highest negative score was considered as the best docking interactive pose.

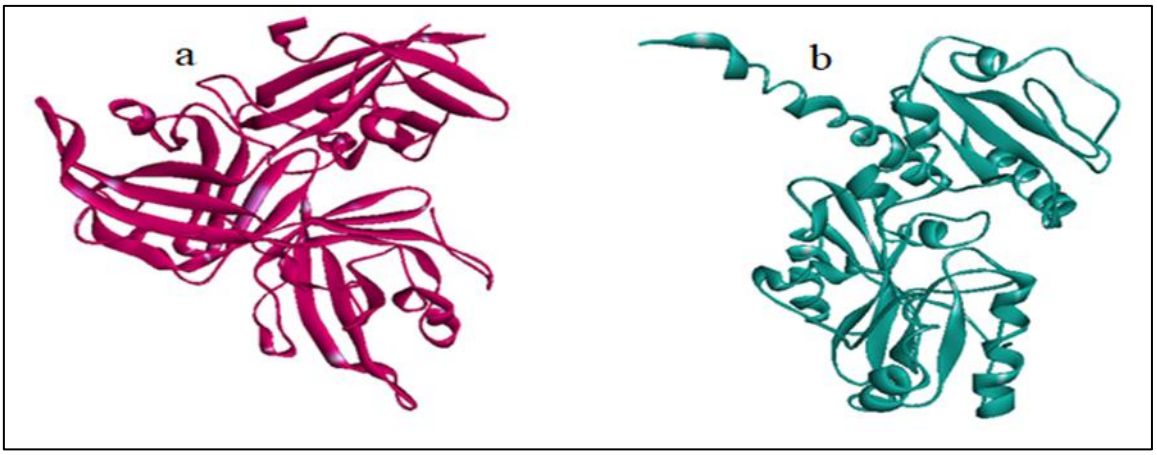

Figure 4: The Receptors of Interest for Molecular Docking Study. A. Sortase of *Staphylococcus Aureus*; B. D-Ala-Ala Ligase (Present in Cell Wall of *Escherichia Coli*)

The dockings were made following the method of the HawkDock server (A web server for the structural prediction and analysis of protein-protein complex). The bonding patterns in the form of Hydrogen bonds, hydrophobic interactions, salt bridges were observed. The molecular dockings made with respective ligands and receptors, were observed and the snapshots were noted down from discovery studio visualization (15).

Results and Discussion

Figure 5 depicts the luxuriant growth of LMA4 on Actinomycetal agar plates, which was subjected to Whole Genome sequencing.

Genome Analysis of LMA4 (*Microbacterium barkeri*)

The WGS of LMA4 had 1291 eluted nodes. Out of which metabolically active genes (a)Thiazole synthase protein, (b) Indole-3-glycerol phosphate synthase Protein, (c) Antibiotic synthesis MbtH protein, (d) Antibiotic biosynthesis monooxygenase protein, (e) class II terpene cyclase protein and (f) YcaO-like family protein sequence with bioactivity potential were chosen and analyzed. The genes with specific bioactivity are provided herewith.

Thiazole Synthase Gene

Thiazole Synthase Gene Sequence

TTGTTGAAAATTGGACCTTATGAATTTAACTCCAG
GCTGCTGCTGGGAACAGGGAAATATCCAAGCTTTG
ATGTGCAAAGGGAAGCTGTAGAAGCTTCGGCAACA
GAGGTGCTTACCTTTGCTGTACGCCGTATGAACATT
TTTGAAGCGGGCCAGCCCAACTTTTTGGAAAAGCT
GGATTTAAAGAAATATAAATTACTCCCTAACACGG
CAGGAGCAAGTACAGCTGAGGAGGCTGTCCGAATA
GCAAGGTTATCCAAAGCTTCCGGGCTCTGCGACATG
ATTAAAGTTGAGGTTATTGGATGCAGCAAAACGCT
TCTTCCTGATCCAATAGAAACAATAAAAGCAGCGG
AGATTCTATTAGACGAGGGTTTTATTGTTCTGCCC
TACACTTCTGATGATGTCGTCCTTGCTCGAAAACTT
GAAGAACTAGGCTGCCATGCGGTCATGCCATGTGC
TTCTCCAATTGGTTCAGGGCAGGGAATCATTAATC
CTATTAATCTTCAATTTATTACAGAGCAGGCAACT
GTGCCGATTATTGTGGATGCTGGAATTGGAAGCCC
GGCAGATGCAGCATTGGCAATGGGGCTGGGAGCAG
ATGGAGTGTTATTAAATACGGCTGTCTCCTCTGCCG
ATGATCCTGTTAAAATGGCAAAAGCAATGAAACTC
GCCATTGAAGCAGGGAGGCTGGGCTATGAAGCAGG
CCGGATGCCCCAAAAACGTTATGCGTCAGCAAGCA
GCCCGGCGGAAGGAATGATTTTTAGTTGA

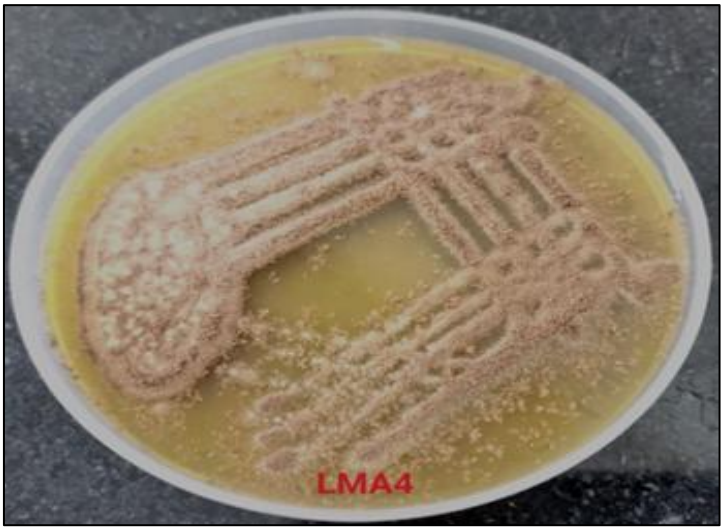

Figure 5: The Typical Actinomycetal Growth on Actinomycetes Isolation Agar Media. LMA4 were Off White, Turquoise and Milky White in Colour Respectively. LMA4 off White in Colour and the Reverse Colony Colour was Straw Yellow for All the Isolates

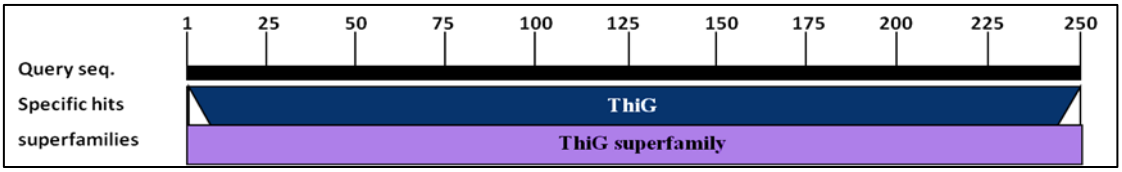

Figure 6: The Conserved Domain of Thiazole Synthase

Thiazole synthase (ThiG) catalyzes the formation of thiazole from 1-deoxy-D-xylulose 5-phosphate (DXP) and dehydroglycine. Catalysis of the reaction: 1-deoxy-D-xylulose 5-phosphate + 2-iminoacetate + thiocarboxy-adenylate-[sulfur-carrier protein ThiS] = 2-[(2R, 5Z)-2-carboxy-4-methylthiazol-5(2H)-ylidene] ethyl phosphate + [sulfur-carrier protein ThiS] + 2 H2O. Thiazole moiety containing compounds are proved to be broad spectrum bioactive agents, being effective against bacteria and fungi, along with antioxidant, anti-inflammatory and anti-proliferative agents (16). The conserved domain of said protein is given in Figure 6.

Indole-3-glycerol Phosphate Synthase

Indole-3-glycerol Phosphate Synthase Gene Sequence

GTGAGTGTGCTCGACGAGATCATCGACGGAGTCCG
TGCCGACCTCGCGGAGCGGCAGGCGCGCGTCAGCCT
CGACGAGCTCAAGGAGCGCGCGGCGAAGGCCCCGGC
CGCCAAGGACGGCGTGGCCGCGCTGCGCGGTGAGGG
CGTCAAGGTCATCTGCGAGGTGAAGCGCTCCAGTCC
CTCCAAGGGCGCGCTGGCCGCGATCGCGGACCCGGC
CGGGCTCGCCGCGGACTACGAGGCGGGCGGCGCGGC
CGTCATCTCGGTCCTGACCGAGCAGCGCCGCTTCGG
CGGCTCGCTGGCCGACCTGGAGGCGGTCCGGGCCCG
GGTGGACATCCCGGTGCTGCGCAAGGACTTCATCGT
CACGTCGTACCAGCTGTGGGAGGCGCGCGCGTACGG
CGCCGACCTGGCGCTGCTGATCGTCGCCGCCCTCGA
CCAGCCGGCCCTGGAGTCGCTGATCGAGCGCGCCGA
GTCCATCGGCCTCACCCCGCTCGTCGAGGTCCACGA
CGAGGACGAGGTCGAGCGCGCGGTGGCCGCGGGCGC
CAAGGTGATCGGCGTCAACGCGCGCAACCTGAAGA
CCCTGGAGGTCGACCGCTCCACCTTCGAGCGGGTCG
CCCCCGAGATCCCCGCCCGCGTCGTCAAGGTCGCCG
AGTCCGGTGTCCGCGGCCCGCACGACCTGATCGCGT
ACGCCAACGCGGGCGCCGACGCGGTCCTGGTCGGCG
AGTCCCTGGTCACGGGCCGCGACCCCAAGTCGGCGG
TGGCCGACCTGGTGGCGGCGGGCGAACACCCCGCGC
TCCGGCACGGCAGGCCCTGA

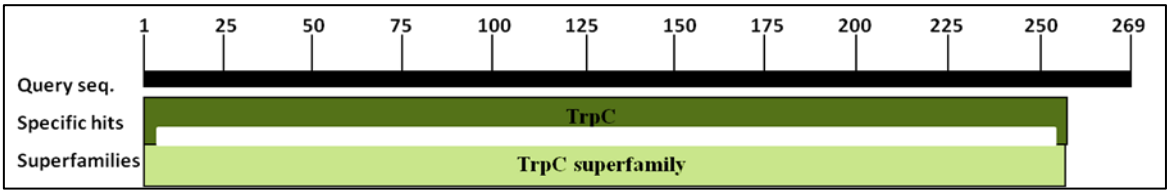

Figure 7: The Conserved Domain of Indole-3-glycerol Phosphate Synthase

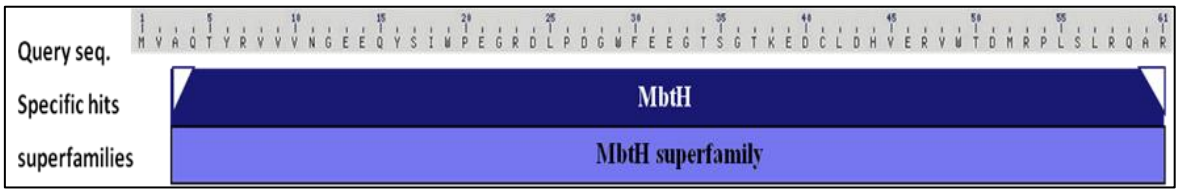

Figure 8: The Conserved Domain of MbtH Protein

The detection of Indole-3-glycerol phosphate synthase in the genome of LMA4 (Microbacterium barkeri), infers about the significance of the compound in context to its bioactivity, as suggested before (17). The conserved domain is given in Figure 7.

Antibiotic Synthesis Protein MbtH

Antibiotic Synthesis MbtH Gene Sequence

ATGGTGGCTCAGACGTACCGCGTGGTGGTGAACGG
CGAGGAGCAGTACTCGATCTGGCCGGAGGGACGGG
ACCTGCCCGACGGCTGGTTCGAGGAGGGCACCTCCG
GCACCAAGGAGGACTGCCTCGACCACGTCGAGCGGG
TCTGGACCGACATGCGGCCGCTCAGCCTCCGCCAGG
CGCGGTGA

MbtH-like protein; this domain is found in the MbtH protein as well as at the N-terminus of the antibiotic synthesis protein NIKP1. MbtH and its homologous were first noted in gene clusters involved in non-ribosomal peptides and other secondary metabolites by Quadri et al. This domain is about 70 amino acids long and contains 3 fully conserved tryptophan residues. The structure of the PA2412 protein shows it adopts a beta-beta-beta-alpha-alpha topology with the short C-terminal helix forming the tip of an overall arrowhead shape. MbtH proteins have been shown to be required for the synthesis of antibiotics, siderophores and glycol peptidolipids. Bacteria MbtH-like Proteins Stimulate Nonribosomal Peptide Synthetase (18). The conserved domain of MntH protein is represented in Figure 8.

Antibiotic Biosynthesis Monooxygenase

Antibiotic Biosynthesis Monooxygenase Gene Sequence

ATGTTTATTCAATTAAAAACGATTACTGTGAAAGA
AGGCCATTCAAATAAGCTAGTCGAGCGCTTTGGAG
GGGAAGGAATAATCGAAGAGCAGCCGGGCTTCATT
GATTTGAATGTTTTAAAGAAGAAGCAGCGCAGCGG
GGACGAAGAAGTGGCTATTATGATTCGCTGGGAAT
CGGAAGGAGCCTGGAAGGCGTGGGAAACAAGTGAT
GTCCATATTGCCGGGCACAAGGCAAACCGCGGCAA
GCCGAAGCCGGAATATATTATTGATAGCCGTCAGG
ATGTCTATCATTCATTG
GGCCAGAAGCAGTACCGTGAGCCAGCTGAAATTAA
ATAA

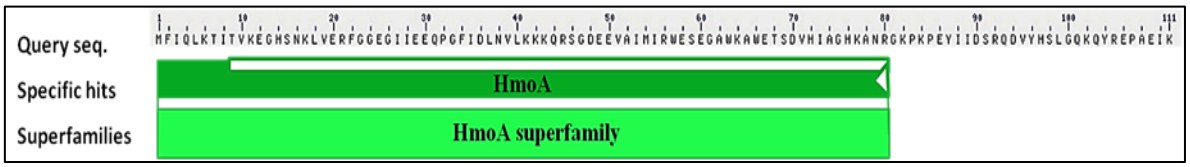

Figure 9: The Conserved Domain of Antibiotic Biosynthesis Monooxygenase Gene

The monooxygenase has hydroxylase activity, forwarding the catalysis of the incorporation of one atom from molecular oxygen into a compound and the reduction of the other atom of oxygen to water. Further, acting on paired donors with incorporation or reduction of molecular oxygen. The oxygen incorporated need not be derived from O_2. The past literature had inferred about the mutation of a single gene in the biosynthetic gene cluster encoding an antibiotic biosynthesis monooxygenase (ABM) superfamily protein leads to a significant increase both in total formicapyridine production and their enrichment relative to the Fasamycins (19). The conserved domain is given in Figure 9.

Class II Terpene Cyclases

Class II Terpene Cyclases Gene Sequence

ATGGCCACTACGGCAACCATGGCAACTACAGAAGCC
AGCAAGCCCCTCGAAGCCCAAGCCAGGACAGCCTTG
ACCAAAGCAACCAACTACGCCTGGGAGATCTTCTCC
AACCGCCACTGGTGCGGCGAGCTCGAGTCTAACGTG
ACCGTGACCTGCGAGCACATCTTCTTTCTTTATGTT
CTCTACCAACACATCGACCCCGGCGAAGGCAGCCAA
TACCGCCAATGGCTGCTCTCGCAGCAAAACTCCGAC
GGCTCCTGGGGCATCGCGCCCAACTACCCCGGCGAC
ATCTCCACAAGCGCAGAGGCATACCTGGCCCTGAGA
ATCATAGGAATGTCCACAGACAGCCCTGAGCTGTA
CCGAGCGCGAACCTTCATCCGGGCCGCCGGCGGGCT
CTCCAAGATGCGCATGTTCACCCGCATCTTCTTCGC
AGAGTTCGGCCTGGTCCCTTGGACTGCCATACCCCA
GCTGCCCGCAGAGTTTATTCTTGTCCCAGCTCACTT
CCCCATCAGCATCTACCGCCTTGCCTCATGGGCACG
CAGCAACGTAGTCCCGCTCCTCATAATCGCCCACCA
CCGACCCCTCTACCCGCTTCCCAACGGGCTACACAA
GCAAAACCCGTTCCTGGATGAGCTCTGGCTGGACCC
GGCCACAAAGCCCCTCCCCTACGGCTCCTCAGACCC
TACCGACCCCGTCGCTTTTGTCTTTACCATACTCGA
CAAAGCCCTCTCTTACCTGGGCGGGCTCCGCCGCAG
CCCTACAAGAGGTTACGCGCGCCGTCGCTGTGTCCA
GTGGATCCTGCAGCACCAAGAGAAAGCCGGCGACT
GGGCAGGTATCATCCCGCCTATGCACGCGGGTATCA
AGGCCCTTTTGCTCGAGGGGTACAAGTTACACGAC
GAGCCTATCCAACTCGGTCTAGCAGCCATCGAACGC
TTCACCTGGGCCGACAACCGTGGCAAGCGTCTCCAA
TGCTGCATCTCCCCCGTCTGGGACACGGTACTCATG
ATCCGCGCATTGCAAGATACTCCCGCCTCCCTGGGC
ATCAAATTGGACCCTCGCATCGCCGACGCCCTGGCC
TGGACGGCCGAGAATCAGCACCGTGGGCCCGAGGGC
GACTGGCGCGTCTACAAGCCCAACATCCCCGTCGGT
GGGTGGGCGTTCGAATACCATAACACGTGGTACCC
GGATATCGATGATACGGCCGCGGCCGTTCTCGCCTT
CTTGACGCATGACCCGGCCACCGCGCGATCGAGACT
CGTGCGCGATGCGGTGCTGTGGATTGTAGGCATGC
AGAATGCCGATGGCGGGTGGGCAGCGTTCGACCAT
GACAATAACCAGCTGTTTCTGAACAAGATCCCGTT
TAGCGATATGGAAAGTCTTTGTGATCCGAGCACGC
CGGATGTGACGGGGCGGACGATTGAGTGTTTGGGC
ATGCTGCGCGATTTGCTGATGCGTCCGGCCGAGAAT
GCCGAAAATGGGGAGAAATATGGATATCCCGATGG
AGAGGGGGATGCGGCAGCCGACGCGCACCTCCTCCA
AATCATCAACACCGCCTGCGCCCGCGCAATTCCCTA
CCTCATCCGCTCTCAGGAAGCAACGGGCACCTGGTA
CGGCCGGTGGGCGGTCAACTATGTCTATGGTACAT
GCCTTGTTCTGTGCGGACTGCAGTATTTCAAGCACG
ACCCGAAGTTCGCGCCGGAAATCCAAGCCATGGCGG
CTCGTGCAGTGAAGTGGCTAAAGCAAGTCCAAAAC
TCCGACGGCGGATGGGGTGAGTCCCTCCTCTCGTAC
CGCGAGCCCTGGCGGGCCGGGTGTGGGCCTTCCACG
CCCTCGCAGACAGCATGGGCGCTTATGGGTATTTTG
ACAGTCTGTGGGGGCGAGGATCGGTCTGTGCAGCG
TGGGGTGAGGCATTTGGTGGATACGCAGGATGATA
CCTTGAGTCAGGGTGACGGTGGGGCGGCGGCGTGG
ACAGAGCGCGAATTCACGAGTACGGGGTTTCCCAA
CCATTTCTACATCTCGTATACGCTGTATCGAGTTTA

TTTTCCTATTACCGCGTTGGGGAGGTATCTTTCTTT
GATCGAAGGAGGCCAGGAGAAGAAGAAGAAAGGTG
GTGGCACTTAA

This group contains class II terpene cyclases, protein prenyltransferases beta subunit, two broadly specific proteinase inhibitors alpha2-macroglobulin (alpha (2)-M) and pregnancy zone protein (PZP) and, the C3 C4 and C5 components of vertebrate complement. Class II terpene cyclases (20,21) include squalene cyclase (SQCY) and 2,3-oxidosqualene cyclase (OSQCY), these integral membrane proteins catalyze a cationic cyclization cascade converting linear triterpenes to fused ring compounds. The protein prenyltransferases include protein farnesyltransferase (FTase) and geranylgeranyltransferase types I and II (GGTase-I and GGTase-II) which catalyze the carboxyl-terminal lipidation of Ras, Rab, and several other cellular signal transduction proteins, facilitating membrane associations and specific protein-protein interactions. Alpha (2)-M is a major carrier protein in serum and involved in the immobilization and entrapment of proteases. PZP is a pregnancy associated protein. Alpha (2)-M and PZP are known to bind to and, may modulate, the activity of placental protein-14 in T-cell growth and cytokine production thereby protecting the allogeneic fetus from attack by the maternal immune system (22). The conserved domain is given in Figure 10.

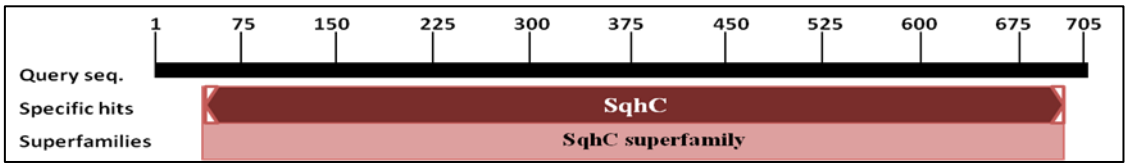

Figure 10: The Conserved Domain of Class II Terpene Cyclase

YcaO-like Family Protein

YcaO-like Gene Sequence

ATGAAGCGCTGTATTTTGGTCGTTGGCGAAGGATT
GATGGCTGAAGGCGCATCACAACAATTATCTGCTG
AATATAACGTTATCCGCAGGTCTGATTTTGAGAAC
AGAATACCGGCGGAGACTTCGTTCGTTCTTGTGCTG
CACGACGCTTGGCGGCCTTTGTTATATCGCGAGGCC
GAAAAAGTGCTAAGGGCGGCAGGGATTCCCTGGCT
TCGCGGGTTTGCATCTTTTGGGGAAGGGGTAATCG
GGCCGCTCGTTTATCCGTCAGAGTCAGGATGTTCAA
AATGCGCCGATATGAGGTGGTATGTCTCCGGCCCCG
ACAGCATGGAGACGCTTGAACTGGAACAAAAAAGC
AATGACATAAACCGCGATCCGTGGGCTTCAAAGAC
CGGCTGCCGCCAGATGGTCAGCCTGATCTGCACAGA
AGTCGACACCATCATGAACGGCGGGCGCGGCCGATT
GGAAAAAAAGCTGTACATCGTCAATTTGAAAACGC
TAGAAAGTTCGCTGCACGCTTTCATACCGGAACCGA
CATGCCCTATATGCGGAAATCTGCCCGAAGACGGAC
CGGAATCCGCGCGGATATCGCTCCGGCCCAGCCTGA
AAACGAACCCGGACAGCTTTCGCGTGCGGAACACG
GACGACTTAAAACAAACGCTCTCCAAAGACTATTT
GGACCAGAGAACGGGGATATTCAACGCCAAAATTT
ATGACGCGATTCTGCCGTTTGCCGATATGATCGTCA
ATATCCCTTTATGGATGGGAAATGAAGGGGTAGGT
GGGCGGAGCCATTCGTATGAAGACAGCGAATTAAC
CGCCATTTTAGAAGGTTTGGAACGCTATTCCGGGA
TTGAACCGCGCGGCAAACGGACGGTCGTACATGAC
ACCTATGAAAATGTAAAAGATCGTGCGCTCGACCC
GGCAAAAGTCGGGCTTCATCAAGCCGAACAATACA
CGCATCCCGATTTTCCTTTTCAGCCGTTTCATCCGG
CCCGCCCGCTCCGCTGGGTATGGGGCTATTCATTTC
TGCAGAAACGGCCGATTCTCGTTCCTGAATCGCTCG
CGTATTACAGCCTGGGGCGGGAAGACAGCTTTGTC
TATGAAACATCGAACGGCTGTGCACTCGGAGGGAG
TTTGGAGGAAGCGATTTTCCACGGCATATTAGAAG
CGGCTGAGCGCGATTCATTTTTGCTTGCCTGGTATG
CCAGGCTTCCACTTCGGCAGCTGGATCTCGACTCTT
CAGGAGATCCGGAATTGCGCCTCATGACAGAGAGG
GTGCGAAGTGTGGCGGGATATGATCTGCATGTCTT
TAATTCAACGATGGAATACGGCATTCCAAGCATTT
GGGCGATCGCCAAAAACAGAAAACGGGAAGGATTG
AATCTGATTTGTTCCGCAGGCGCCCACCCCGATCCC
GTAAAAGCGGTGAAAAGCGCCCTTCATGAGCTGGC
AGGCATGATGTTGAAGCACGACCGCAAATTTGAAT
CTGAAAGGAGCAAATATCACCAGATGCTCCGCAGT
CCTGCTTTAGTGCGGGAAATGGAGGATCACAGCTT
AATGTACGGACTGCCCGAAGCGGAAGAGCGGCTCA
GTTTTTTGCTGAATCAAAGCCGGCCTCTGCAAACAT
TCGACGATGCCTTTAAAAAAAGAAAGAAACACCTT
GATTTAACCGAAGACCTCTGCGACCTGCTTGAAGCT
TTCCGGAGTCTCAACCTTGAAGTGATCGTCGTTGAT
CAGACTTCCCCTGTTATCAAACGGAATGGACTTTAT
TGTGTGAAAGTTCTCATCCCGGGCCTGATTCCGATG
ACATTCGGACAGCGCTTCATCCGCCTGGAAGGGCTT
GAGAGAATTTTCACCGTGCCGAAGAAGCTCGGCTT
TGCGGAAGAGCCGTTAAGACCAGAACAGCTGAACC
TGCACCCGCACCCGTTCCCTTAA

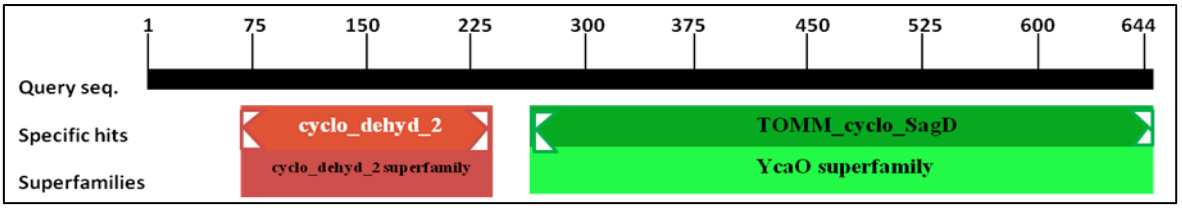

Figure 11: The Conserved Domain of Ycao-Like Family Protein

The YcaO-like family protein may be involved in azoline formation, macroamidine formation, thioamide formation, and/or potentiation of RimO-dependent methylthiolation (23). In addition, it was also noted that the sequence was constitute with thiazole/oxazole-forming peptide maturase, SagD family component related to SagD, previously referred to as a scaffold or docking protein involved in the biosynthesis of streptolysin S in *Streptococcus pyogenes* from the protoxin polypeptide (product of the sagA gene). In later period, it was described an enzymatic activity, an ATP-dependent cyclo dehydration reaction. Further, to put the understanding that this protein family serves as a marker for widely distributed prokaryotic systems for making a general class of heterocycle-containing bacteriocins, with bacteriocin biosynthesis cyclodehydratase, ThiF-like domain of a fusion protein found in clusters associated with the production of TOMMs (thiazole/oxazole-modified microcins), small bacteriocins with characteristic heterocycle modifications (24-26). The conserved domain is shown in Figure 11.

Table 1: The Conserved Domain Properties of the Six Selected Protein from NCBI Database

S.L No	Protein Name	Accession	Description	Interval	E-value	Domain Architecture ID
1	Thiazole synthase	COG2022	Thiazole synthase ThiGH, ThiG subunit (thiamin biosynthesis) [Coenzyme transport and metabolism] ;Thiazole synthase ThiGH, ThiG subunit (thiamin biosynthesis) is part of the Pathway/BioSystem: Thiamine biosynthesis	1-255	4.11e-156	18578739
2	Indole-3-glycerol phosphate synthase	COG0134	Indole-3-glycerol phosphate synthase [Amino acid transport and metabolism] ;Indole-3-glycerol phosphate synthase is part of the Pathway/BioSystem: Aromatic amino acid biosynthesis	2-256	3.18e-132	10000629
3	Antibiotic synthesis protein MbtH	COG3251	MbtH family protein, regulates adenylation domains of NRPSs [Secondary metabolites biosynthesis, transport and catabolism]	3-61	3.81e-31	10007125
4	Antibiotic biosynthesis monooxygenase	COG2329	Heme-degrading monooxygenase HmoA and related ABM domain	1-80	3.98e-16	10006366

			proteins [Coenzyme transport and metabolism]			
5	Class II terpene cyclases	COG1657	Terpene cyclase SqhC [Lipid transport and metabolism]	39-692	0e+00	11447352
6	YcaO-like family protein	TIGR03604	thiazole/oxazole-forming peptide maturase, SagD family component ;Members of this protein family include enzymes related to SagD, previously referred to as a scaffold or docking protein involved in the biosynthesis of streptolysin S in Streptococcus pyogenes from the protoxin polypeptide (product of the sagA gene). Newer evidence describes an enzymatic activity, an ATP-dependent cyclodehydration reaction, previously ascribed to the SagC component. This protein family serves as a marker for widely distributed prokaryotic systems for making a general class of heterocycle-containing bacteriocins.	261-644	1.67e-172	10024029
		TIGR03882	bacteriocin biosynthesis cyclodehydratase domain ;This model describes a ThiF-like domain of a fusion protein found in clusters associated with the production of TOMMs (thiazole/oxazole-modified microcins), small bacteriocins with characteristic heterocycle modifications. This domain is presumed to act as a cyclodehydratase, as do members of the SagC family modeled by TIGR03603	68-232	2.39e-61	10024029

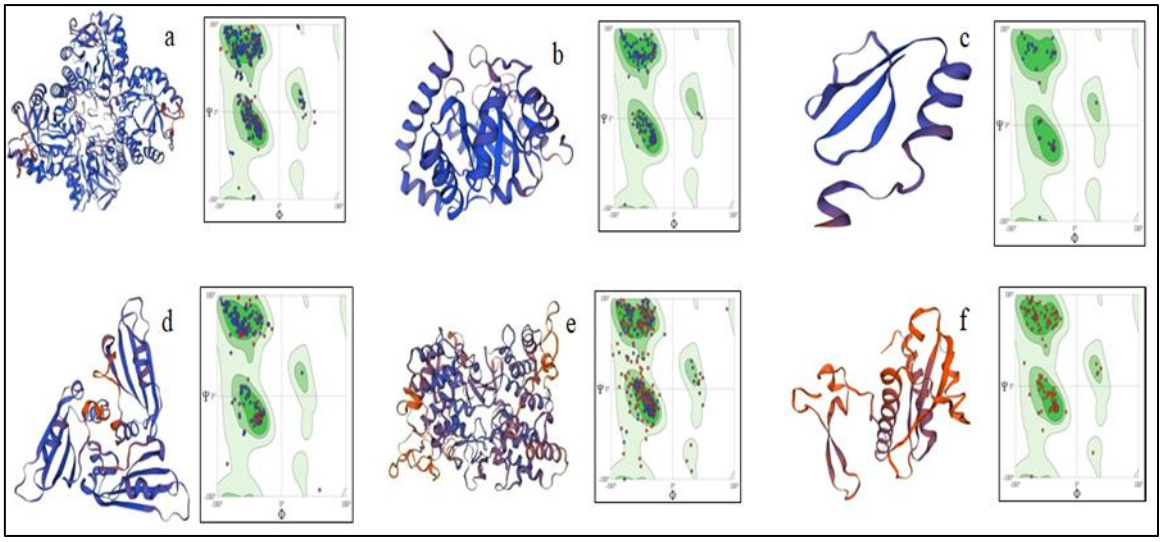

Figure 12: Pictures Showing the Structure of the Selective Identified Protein Structure from the SWISS-MODEL Server Along with its Validation Ramachandran Plot Graph of LMA4 Whole Genome Sequence Results. (a) Thiazole Synthase Protein, with 96.86%; (b) Indole-3-Glycerol Phosphate Synthase Protein with 99.24%; (c) Antibiotic Synthesis Protein Mbth Protein with 100.00%; (d) Antibiotic Biosynthesis Monooxygenase Protein with 92.63%; (e) Class II Terpene Cyclase Protein with 90.49% And (f) Ycao-Like Family Protein with 95.21% Depicted Favourable Regions as Shown in Respective Ramachandan Plots Given in Right Panels Prediction of 3D Structure of the Proteins of LMA4

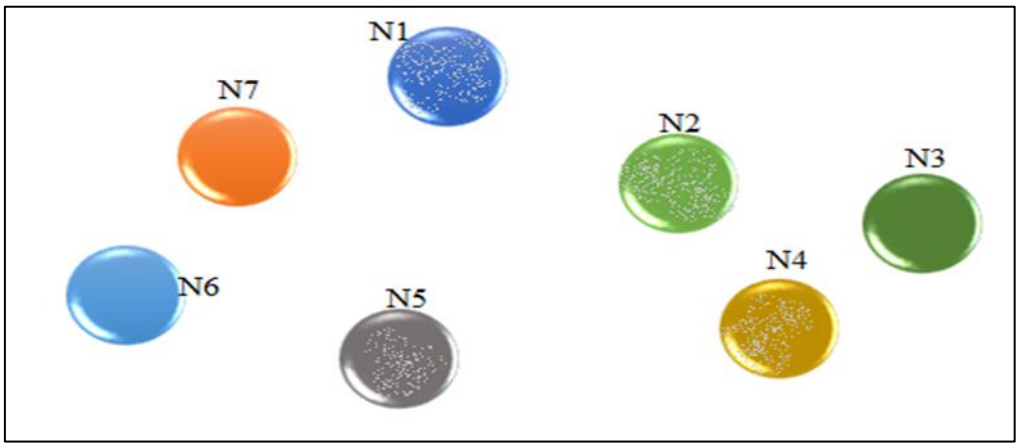

Figure 13: The Snapshot Result Eluted in STRING Program Based Gene-Gene Interaction Study in the Defined Portal. There was Participation of 7 Numbers of Nodes, But No Significant Interaction as Expected

The Whole genome sequence result of LMA4 gives the identified and unidentified genes / proteins / enzymes in the form of secondary metabolites. By taking the genes of interest 6 genes along with their nucleotide and amino acid sequences are mentioned in the above. Due to unavailability of the 3D crystal structure according to their sequence given, the Homology modeling process was done to construct the appropriate structure and their validation in the SWISS MODEL web server which was mentioned in the Figure 12. This is important to note that the threshold value of favoured region depicted in Ramachandran plot is ≥90%. Most of the models generated were observed to be with favored region above 90%. From the above results it was found that Antibiotic synthesis protein MbtH Protein (100.00%) were constituted with absolute Ramachandran favoured region (FR) followed by Indole-3-glycerol phosphate synthase Protein having FR 99.24%.

The Gene-Gene Interaction Study of Detected Gene of Interest: The genes, namely [i] Thiazole synthase protein, [ii] Indole-3-glycerol phosphate synthase Protein, [iii] Antibiotic synthesis MbtH protein, [iv] Antibiotic biosynthesis monooxygenase protein, [v] class II terpene cyclase protein and [vi] YcaO-like family protein sequence, detected in the WGS of LMA4 was

subjected to gene-gene interaction study using STRING (www.string-db.org) online and it was inferred that there was no significant interaction of input proteins in the program as expected. There was participation of 7 nodes, which were found to be not interacting with each other (Figure 13). From this result, it may be inferred that the genes detected in the WGS of LMA4, could have expressed independently, with possibility of each of gene ontology. As actinomycetes are the soil inhabitants, possibilities are more intrusion of other organism genes entering into the host actinomycetes system those are co-inhabiting in the soil niche, and getting intercalated in between the host genes. The genes namely, expressing bacteriocins of Lanthipeptide category, other bacteriocins are reportedly expressed from genes of Bacillus spp. (the Gram positive) bacteria not Actinomycetes. In addition, the expression of Lovastatin had also made a twisted view, as the specific compound produced from fungal genera. Therefore, it is suggested that there was a direct correlation between the co inhabitation and expression of genes of co inhabited organisms of course. The respective binding scores (in Kcal/mol) was noted and graphically presented in Figure 14. While the figures depicting respective docking poses are given in Figure 15 and Figure 16 for *Microbacterium barkeri* (LMA4). From the docking scores, it was observed that Indole synthase had been docked with D-ala-D-ala receptor, with a docking score of -6794.34 followed by terpene cyclase (-6517.83), Thiazole synthase (-5146.34), YcaO like family protein (-5100.29), Antibiotic synthesis MbtH (-5030.22) and Antibiotic biosynthesis monooxygenase (-4748.45).Likewise the docking scores was observed for surface proteins in Gram-positive bacteria (*Staphylococcus aureus*) i.e. sortase protein was docked with YcaO like family protein, with a docking score of -6817.38 followed by Thiazole synthase (-6093.60), Antibiotic synthesis MbtH (-5986.38), Indole 3 glycerol phosphate synthase (-5477.27), ClassII terpene cyclase (-5363.78) and Antibiotic biosynthesis monooxygenase (-4939.34). The figures for the protein- protein docking complexes of LMA4 containing proteins as ligand compound and the sortase / D-ala-D-ala as receptor compound of the target bacterial cell membrane proteins are given here. From the Table 1, it was observed that amongst 6 nos. of proteins of LMA4 Indole 3 glycerol phosphate synthase docked with D-ala-D-ala ligase had highest docking score -6794.34 followed by ClassII terpene cyclase (-6517.83). The other proteins had also appreciable docking score with D-ala-D-ala ligase receptor molecules. It was also noted that YcaO like family protein (-6817.38), had the best docking score with sortase of *S. aureus* followed by and Thiazole synthase (-6093.60). Again, it was observed that the other proteins had also comparable binding energy scores (docking score). This work is embedded with significant findings, on genes which are expressed when subjected to co culture with bacterial of medical importance. The genes are biologically significant as the expression is related with co incubated bacteria, in a definite time period, with optimised growth condition. This fragment of study is dedicated to readers because this study determines the elution of genes of interest which with antimicrobial activity, eluted when co incubated with its competitor. There is sufficient evidence regarding the antimetabolic/ antibiotic property of gene products, which were experimented using a co culture method (27) where both a Gram positive (a strain of *Staphylococcus aureus*) and Gram negative (a strain of *Escherichia coli*), inferring about broad spectrum antibiotic activity.

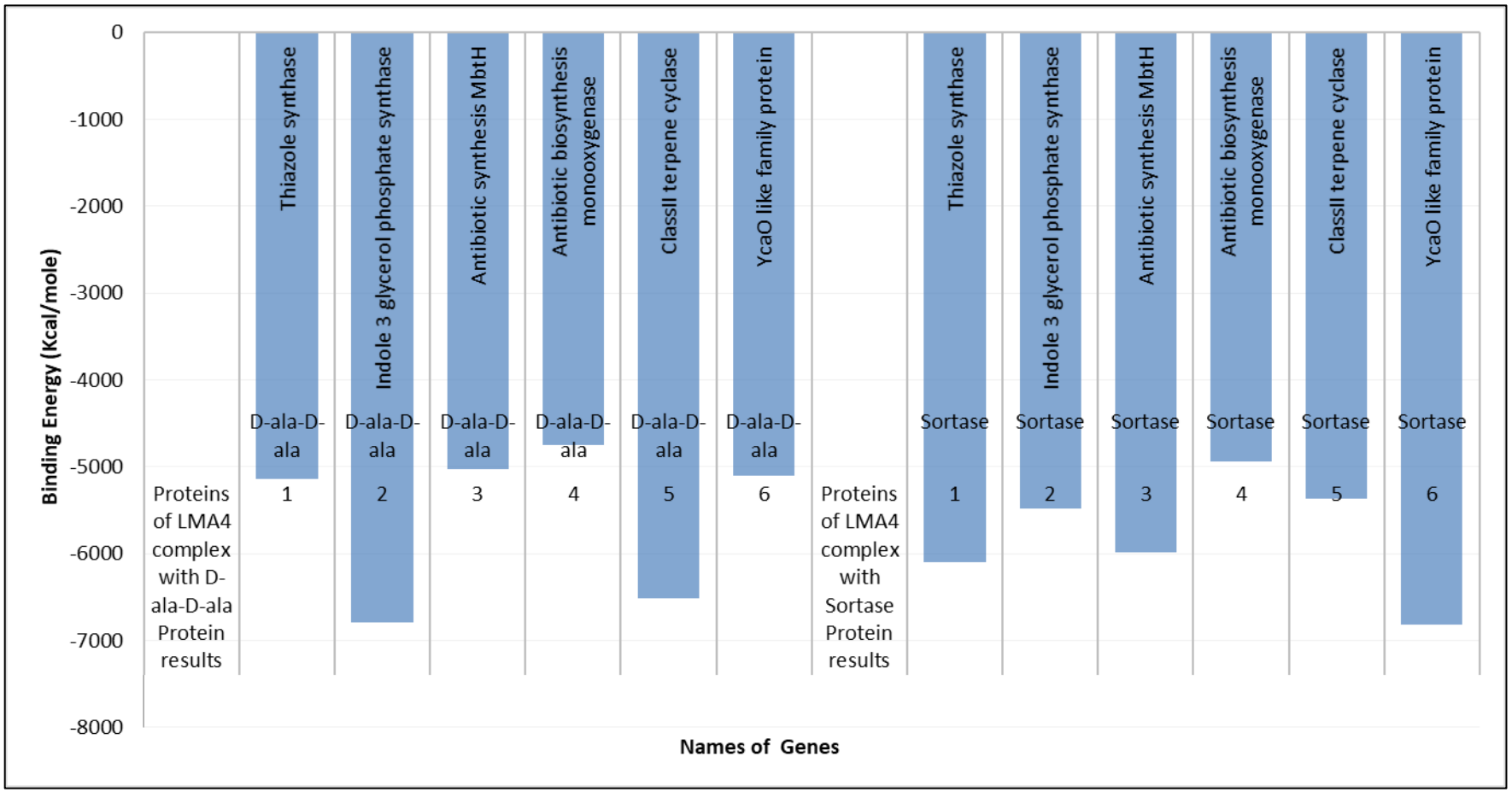

Figure 14: The Graphical Representation of Respective Binding Energy (Kcal/Mole), Graphically Presented as Inverted Bars (-Ve Gibbs Energy) Required for the Molecular Dockings Between (A) Thiazole Synthase Protein(B) Indole-3-Glycerol Phosphate Synthase (C) Antibiotic Synthesis Protein Mbth Protein (D) Antibiotic Biosynthesis Monooxygenase Protein (E) Class II Terpene Cyclase Protein and (F) Ycao-Like Family Protein. The Left Panel Represent Proteins of LMA4 Complexed with D-Ala-D-Ala and the Right Panel Represent Proteins of LMA4 Complexed with Sortase Protein

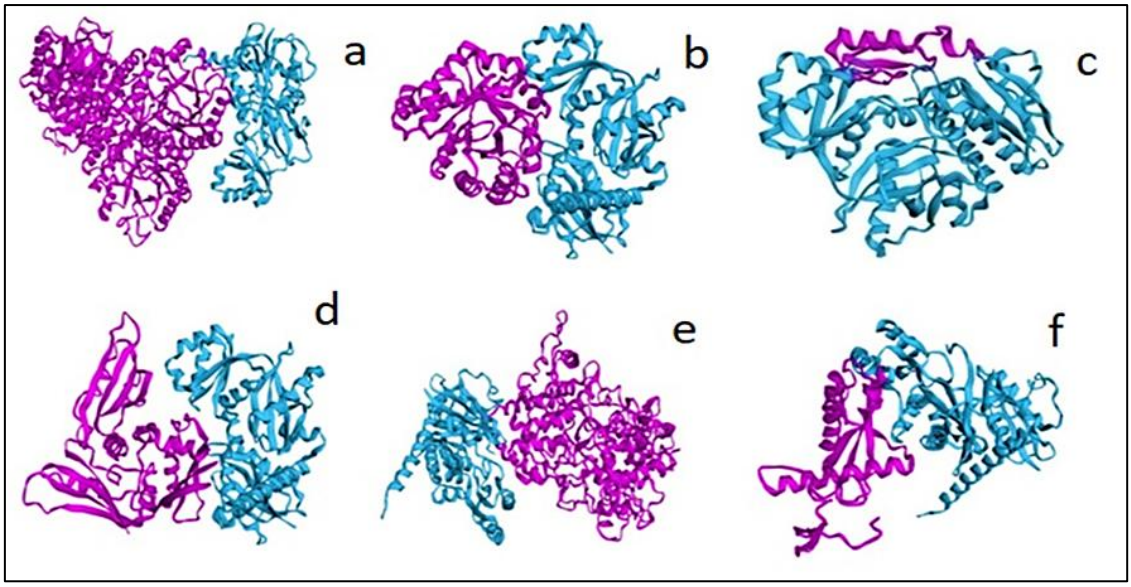

Figure 15: Containing D-Ala-D-Ala-Mediated Cell Wall Surface Protein in Gram-Negative Bacteria (*Escherichia Coli*) I.E. D-Ala-D-Ala Ligase Protein with the Genes/ Proteins Predicted from the WGS of the Actinomycetes Isolate Namely LMA4. The Interactions between the Proteins are Like (a) Thiazole Synthase-D-Ala-D-Ala; (b) Indole 3 Glycerol Phosphate Synthase-D-Ala-D-Ala; (c) Antibiotic Synthesis Mbth-D-Ala-D-Ala; (d) Antibiotic Biosynthesis Monooxygenase-D-Ala-D-Ala; (e) Class-II Terpene Cyclase-D-Ala-D-Ala; and (f) Ycao Like Family Protein-D-Ala-D-Ala

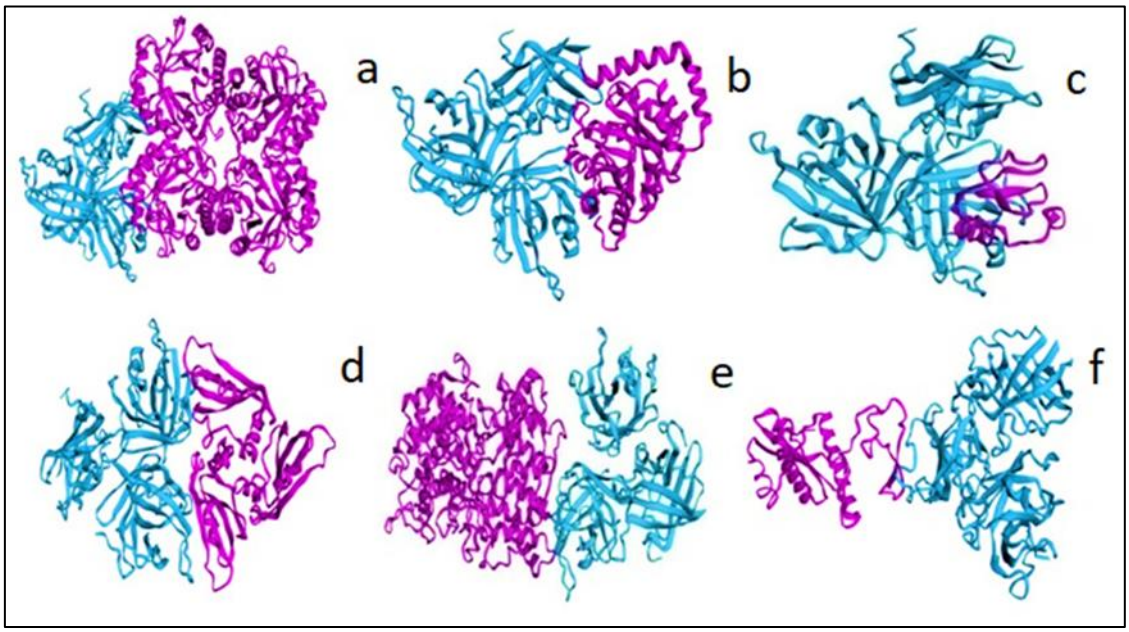

Figure 16: Containing Sortase-Mediated Cell Wall Surface Proteins in Gram-Positive Bacteria (*Staphylococcus aureus*) Sortase Protein with the Genes/ Proteins Predicted from the WGS of the Actinomycetes Isolate Namely LMA4. (a) Thiazole Synthase-Sortase ; (b) Indole 3 Glycerol Phosphate Synthase-Sortase; (c) Antibiotic Synthesis Mbth-Sortase; (d) Antibiotic Biosynthesis Monooxygenase-Sortase; (e) Class II Terpene Cyclase-Sortase and (f) Ycao Like Family Protein-Sortase

Conclusion

The genes with proven bioactivity were also screened for interaction based expression studies, using STRING networking software online. From the modelling *in silico* experimentation, it was found that the proteins of LMA4 [i] Thiazole synthase protein, [ii] Indole-3-glycerol phosphate synthase Protein, [iii] Antibiotic synthesis MbtH protein, [iv] Antibiotic biosynthesis monooxygenase protein, [v] class II terpene cyclase proteinand [vi] YcaO-like family protein

was structurally valid and satisfying the drugability criteria to be get docked with bacterial target molecules (receptors). This is to add here that the metabolic synthesis of polyketide synthases (PKS), non-ribosomal protein synthetases (NRPS) can be considered as a biomarker metabolic profile, while in comparison with other species within the same genus (28). From this study, it is concluded that the docking studies made between the said proteins of LMA4 with D-ala-D-ala ligase and sortase leads to the best docking scores among the other proteins, they are Indole 3 glycerol phosphate synthase with D-ala-D-ala ligase (-6794.34) and YcaO like family protein with sortase (-6817.38) for LMA4 proteins had the best docking scores. This is further to clarify that the genes are all of pharmaceutically validated products and can be used as drug actives against microbial candidates of clinical importance. This is to mention that there is a direct correlation between organism's ecological function specifically in presence of a co inhabitant in a niche and expression of genes assigned with constitutive functions like chemical weapons which are directed towards the inhibition of growth and multiplication of other microbial contender, trying to interfere the life style of actinobacterial strain. This leads to production of antimetabolites which are of immense importance in Biotechnology. However, the genes are considered as novel for the said isolate except terpene cyclases which are produced in plants.

Abbreviation

Nil.

Acknowledgements

This article is part of Thesis work registered under Sambalpur University.

Author Contributions

All the authors contributed equally.

Conflicts of Interest

The authors declare there is no competing interest.

Ethics Approval

Not applicable.

Funding

The research was funded by SURF (Sambalpur University Research Fellowship) vide letter No.01077/RDC/SU, SURF 2022.

References

1. Das KK, Pattnaik S. Bioactivity of *Microbacterium barkeri* (LMA4) *In Vitro* and Candidate Gene Annotation *in Silico*. Applied Biochemistry and Biotechnology. 2024 May;196(5):2839-50.
2. Chung J, Kim S, Choi K, Kim JO. Degradation of polyvinyl alcohol in textile waste water by *Microbacterium barkeri* KCCM 10507 and *Paenibacillus amylolyticus* KCCM 10508. Environmental technology. 2016 Feb 16;37(4):452-8.
3. Tanuja, Phillip J, Molecular Characterization Of Cellulase Producing *Microbacterium barkeri* Isolated From Agricultural Waste. International Journal of Advanced Research (IJAR). 2017;5(8);1344-1349.
4. Parashar S, Kumar N. Isolation and characterization of a novel agar-degrading bacterium, *Microbacterium barkeri* Sp. Sela 4, from soil enriched with laboratory agar. Journal of Microbiology, Biotechnology and Food Sciences. 2021 Jan 6;2021:78-83.
5. Sharma S, Mathur N, Singh A, Agarwal M. Biodegradation of Low-and High-Density Polyethylene Films by *Microbacterium barkeri* SH20. Current World Environment. 2022;17(1):245.
6. Jiang HL, Xuan Y, Zeng Q, Yu QM, Zhang YQ, Chen YR, Luo HB, Huang H, Xu Q. Bioconversion of steviol glycosides into steviol by *Microbacterium barkeri*. Journal of Asian Natural Products Research. 2021 Nov 15;23(11):1057-67.
7. Liu J, Zhou Q, Ibrahim M, Liu H, Jin G, Zhu B, Xie G. Genome sequence of the biocontrol agent *Microbacterium barkeri* strain. J Bacteriol. 2011;194(23):6666-6667.
8. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. Sealer: a scalable gap-closing application for finishing draft genomes. BMC bioinformatics. 2015 Dec;16:1-8.
9. Waterston RH, Lander ES, Sulston JE. More on the sequencing of the human genome. Proceedings of the National Academy of Sciences. 2003 Mar 18;100(6):3022-4.
10. Tuller T, Zur H. Multiple roles of the coding sequence 5′ end in gene expression regulation. Nucleic acids research. 2015 Jan 9;43(1):13-28.
11. Chavali AK, Rhee SY. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. Briefings in bioinformatics. 2018 Sep;19(5):1022-34.
12. Das KK, Pattnaik S, Behera SK. Identification and Genomic study of the biopotent actinomycetes species Microbacterium barkeri (LMA4): a computational approach. IJRAR-International Journal of Research and Analytical Reviews (IJRAR). 2022 Oct;9(4):845-58.
13. Laskowski RA, Furnham N, Thornton JM. The Ramachandran plot and protein structure validation. In Biomolecular forms and functions: a celebration of 50 years of the Ramachandran map. Singapore: World Scientific; 2013. p. 62-75.
14. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, Hou T. HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA. Nucleic acids research. 2019 Jul 2;47(W1):W322-30.

15. Jejurikar BL, Rohane SH. Drug designing in discovery studio, 2021. Mar/Apr 14;(2): 135-138.
16. Lemilemu F, Bitew M, Demissie TB, Eswaramoorthy R, Endale M. Synthesis, antibacterial and antioxidant activities of Thiazole-based Schiff base derivatives: a combined experimental and computational study. BMC chemistry. 2021 Dec 23;15(1):67.
17. Sun H, Sun K, Sun J. Recent advances of marine natural indole products in chemical and biological aspects. Molecules. 2023 Feb 27;28(5):2204.
18. Zwahlen RD, Pohl C, Bovenberg RA, Driessen AJ. Bacterial MbtH-like proteins stimulate nonribosomal peptide synthetase-derived secondary metabolism in filamentous fungi. ACS Synthetic Biology. 2019 Jul 8;8(8):1776-87.
19. Yang L, Li X, Wu P, Xue J, Xu L, Li H, Wei X. Streptovertimycins A–H, new fasamycin-type antibiotics produced by a soil-derived *Streptomyces morookaense* strain. The Journal of Antibiotics. 2020 May;73(5):283-9.
20. Pan X, Rudolf JD, Dong LB. Class II terpene cyclases: structures, mechanisms, and engineering. Natural Product Reports. 2024 Mar 20;41(3):402-433.
21. Moosmann P, Ecker F, Leopold-Messer S, Cahn JK, Dieterich CL, Groll M, Piel J. A monodomain class II terpene cyclase assembles complex isoprenoid scaffolds. Nature chemistry. 2020 Oct;12(10):968-72.
22. Barrios V, Chowen JA, Martín-Rivada Á, Guerra-Cantera S, Pozo J, Yakar S, Rosenfeld RG, Pérez-Jurado LA, Suárez J, Argente J. Pregnancy-Associated Plasma Protein (PAPP)-A2 in Physiology and Disease. Cells. 2021; 10(12):3576.
23. Burkhart BJ, Schwalen CJ, Mann G, Naismith JH, Mitchell DA. YcaO-dependent posttranslational amide activation: biosynthesis, structure, and function. Chemical reviews. 2017 Apr 26;117(8):5389-456.
24. Goto Y, Suga H. In vitro biosynthesis of peptides containing exotic azoline analogues. ChemBioChem. 2020 Jan 15;21(1-2):84-7.
25. Russell AH, Vior NM, Hems ES, Lacret R, Truman AW. Discovery and characterisation of an amidine-containing ribosomally-synthesised peptide that is widely distributed in nature. Chemical Science. 2021;12(35):11769-78.
26. Costa T, Cassin E, Moreirinha C, Mendo S, Caetano TS. Towards the Understanding of the Function of Lanthipeptide and TOMM-Related Genes in *Haloferax mediterranei*. Biology. 2023 Feb;12(2):236.
27. Pattnaik S. Induction Of bioactive compounds in a co-cultured strain of *Micromonosperma echinospora* (DST-4) isolated from secluded Hirakud dyke soil. Indian Journal of Biotechnology. 2018 Oct;17:586-594.
28. Corretto E, Antonielli L, Sessitsch A, Höfer C, Puschenreiter M, Widhalm S, Swarnalakshmi K, Brader G. Comparative genomics of Microbacterium species to reveal diversity, potential for secondary metabolites and heavy metal resistance. Frontiers in microbiology. 2020 Aug 4;11:1869.