Molecular characterization and functional annotation of a hypothetical protein (SCO0618) of *Streptomyces coelicolor* A3(2)

Nadim Ferdous¹, Mahjerin Nasrin Reza¹, Md. Tabassum Hossain Emon¹, Md. Shariful Islam², A. K. M. Mohiuddin¹, Mohammad Uzzal Hossain³*

¹Department of Biotechnology and Genetic Engineering, Faculty of Life Science, Mawlana Bhashani Science and Technology University, Tangail 1902, Bangladesh
²Laboratory of Reproductive and Developmental Biology, Hoikaido University, Sapporo 060-0808, Japan
³Bioinformatics Division, National Institute of Biotechnology, Savar, Dhaka 1349, Bangladesh

*Corresponding author:
E-mail: uzzalbge10044@gmail.com

*Streptomyces coelicolor* is a gram-positive soil bacterium which is well known for the production of several antibiotics used in various biotechnological applications. But numerous proteins from its genome are considered hypothetical proteins. Therefore, the present study aimed to reveal the functions of a hypothetical protein from the genome of *S. coelicolor*. Several bioinformatics tools were employed to predict the structure and function of this protein. Sequence similarity was searched through the available bioinformatics databases to find out the homologous protein. The secondary and tertiary structure were predicted and further validated with quality assessment tools. Furthermore, the active site and the interacting proteins were also explored with the utilization of CASTp and STRING server. The hypothetical protein showed the important biological activity having with two functional domain including POD-like_MBL-fold and rhodanese homology domain. The functional annotation exposed that the selected hypothetical protein could show the hydrolase activity. Furthermore, protein-protein interactions of selected hypothetical protein revealed several functional partners those have the significant role for the bacterial survival. At last, the current study depicts that the annotated hypothetical protein is linked with hydrolase activity which might be of great interest to the further research in bacterial genetics.

**Keywords:** genome, hydrolases, hypothetical protein, modeling, *Streptomyces coelicolor*

---

**Introduction**

*Streptomyces coelicolor* A3(2) is one of the best studied representatives amongst other members of the genus *Streptomyces*. Like the streptomyces genus in general, it lives in soil [2]. It is considered a model organism to study soil bacteria [3], which has been studied genetically for about 60 years [4]. They have the capability to degrade chitin and other compounds that are difficult to degrade which makes them especially important [5]. This bacterium produces a range of secondary metabolites, including actinorhodin, undecylprodigiosin, calcium-dependent antibiotic, methylenomycin A and perimycin [6]. Some of them have antifungal activities also. So, *Streptomyces coelicolor* A3(2) has the potential to make such secondary metabolites, and metagenomic analysis has revealed it has...
revealed it has tremendous quantities of significant biosynthetic gene sets [7,8]. These characteristics have elicited biotechnological interest in this bacterium and have aroused the interest of researchers in the past few years to investigate the different proteins involved in secondary metabolites production. As an example, it is recently found that albaflavone, germicidin A, and chalcone are produced during germination of *Streptomyces coelicolor* [9] and the genes responsible for the biosynthesis of streptomyces secondary metabolites are generally clustered with high expression of regulation [10]. Another research shows that a group of mtbH-like genes in *S. coelicolor* are necessary for some secondary metabolite production [11]. *Streptomyces coelicolor* has three such genes, cloY is one of them [11]. When all three genes were absent, clorobiocin, an antibiotic which inhibits the enzyme DNA gyrase was produced only in very small amounts, but when cloY was restored, clorobiocin was produced at a more significant level [11].

*Streptomyces coelicolor* A3(2) is reported to have 8,667,507 base pair linear chromosome, containing the largest number of genes so far discovered in a bacterium [10]. The genes so far predicted are 7,825 which include more than 20 clusters coding for known or predicted secondary metabolites [10]. However, there are many proteins of this bacterium which are considered hypothetical proteins as their structures and biological functions are not yet known. These proteins can be very important and their annotation can lead to knowledge about new structures, pathways, and functions. Thus, bioinformatics approaches can play an important role in predicting and analyzing various forms of structure of those hypothetical proteins, their biological functions as well as protein-protein interactions.

With the advancement of in-silico analysis, it became easier to annotate function to a hypothetical protein using various bioinformatic tools. Thus, the purpose of this study was to assign structural and biological function to the hypothetical protein SCO0618 (accession No. NP_624929.1) of *S. coelicolor* for an improved understanding of the protein. Subcellular localization, secondary structure, and active site were predicted and protein-protein interaction was analyzed. Further, a good quality model of the SCO0618 was tried to generate using homology modeling techniques.

**Methods**

**Sequence retrieval and similarity identification**

The sequence information of the hypothetical protein (NP_624929.1) was retrieved from the NCBI database. The sequence was then collected as a FASTA format sequence and submitted to several prediction servers for the in-silico characterization (Table 1). To get the initial prediction about the function of the targeted hypothetical protein, similarity search was performed with the NCBI protein Database (https://www.ncbi.nlm.nih.gov/) against non-redundant and SwissProt [12] database to find out the proteins that might have structural similarities with that of the uncharacterized protein by using BLASTp program [13].

**Multiple sequence alignment and phylogeny analysis**

Multiple sequence alignment was performed using MUSCLE server of EBI (https://www.ebi.ac.uk/Tools/msa/muscle/) [14] and visualized using the CLC Sequence Viewer 7.0.2 (http://www.clcbio.com). The phylogeny analysis was done by using the webtool Phylogeny.fr (http://phylogeny.lirmm.fr/) [15].

**Physicochemical properties analysis**

The physical and chemical properties including molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, total number of negatively charged residues (Asp + Glu), total number of positively charged residues (Arg + Lys), instability index, aliphatic index, and grand average of hydropathicity (GRAVY) predictions, etc. were performed by the ProtParam (http://web.expasy.org/protparam/) [16] tool of ExPASy.

**Subcellular localization analysis**

Subcellular localization was predicted by CELLO [17]. Results were also cross-checked with subcellular localization predictions obtained from PSORTb [18], PSLpred [19], and SOSU1GramN [20]. TMHMM [21], HMMPRED [22], and CCTOP [23] were used for the topology prediction.

**Conserved domain, motif, fold, coil, family, and superfamily identification**

Search carried out at conserved domain database (CDD, available at NCBI) [24], for conserved domain. Protein motif search was carried out using Motif (Genome Net) server [25]. Pfam [26] and SuperFamily [27] database searches were done to assign the protein's evolutionary relationships. For the detection of coiled-coil conformation within the protein, the COILS server [28] was employed. Protein sequence analysis and classification server InterProScan [29] was employed for the functional analysis of the protein. For protein folding pattern recognition, PFP-FunD SeqE server [30] was used. And STRING 10.0 [31] search was carried out for the identification of possible functional interaction network of the protein.

**Secondary structure prediction**

PSI-blast based secondary structure Prediction (PSIPRED) [32]
and self-optimized prediction method with alignment (SOPMA) servers were used for the prediction of the proteins’ secondary structure [33].

### Three-dimensional structure prediction

The three-dimensional structure was predicted by HHpred server ([https://toolkit.tuebingen.mpg.de/tools/hhpred](https://toolkit.tuebingen.mpg.de/tools/hhpred)) [34] of the Max Planck Institute for Developmental Biology, Tübingen which is based on the pairwise comparison profile of hidden Markov models (HMMs). For higher accuracy, the 3D structure was predicted on the basis of best scoring template. Later the 3D structure was refined through YASARA energy minimization server [35].

### Model quality assessment

Finally, PROCHECK ([https://servicesn.mbi.ucla.edu/PROCHECK/](https://servicesn.mbi.ucla.edu/PROCHECK/)) [36], Verify3D ([http://nihserver.mbi.ucla.edu/Verify_3D/](http://nihserver.mbi.ucla.edu/Verify_3D/)) [37], and ERRAT Structure Evaluation server ([https://servicesn.mbi.ucla.edu/ERRAT/](https://servicesn.mbi.ucla.edu/ERRAT/)) [38] were used for quality assessment of the predicted three dimensional structure.

### Active site detection

The active site of the protein was determined by the Computed Atlas of Surface Topography of Protein (CASTp) ([http://sts.bioengr.uic.edu/castp/](http://sts.bioengr.uic.edu/castp/)) [39] which provides an online resource for locating, delineating, and measuring concave surface regions on three-dimensional structures of proteins.

### Results and Discussion

The work-flow of the study was shown in Fig. 1.

#### Sequence and similarity information

The BLASTp result against non-redundant and SwissProt database showed homology with other hydrolase and sulfurtransferase proteins (Tables 2 and 3). Multiple sequence alignment (Supplementary Fig. 1) was considered the FASTA sequences of the hypothetical protein (SCO0618) and the homologous annotated proteins. For the confirmation of homology assessment between the proteins, down to the complex and subunit level, phylogenetic analysis was also performed. Phylogenetic tree was constructed based on the alignment and BLAST result which gives the similar concept about the protein (Fig. 2). The distances between branches are also included.

#### Physicochemical features

The protein consist of 461 amino acids, among the most abundant

---

**Table 1.** Tools used for the in-silico characterization of hypothetical protein SCO0618

<table>
<thead>
<tr>
<th>No.</th>
<th>Server name</th>
<th>Reference</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>protBLAST</td>
<td>Altschul et al. (1999) [40]</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MUSCLE</td>
<td>Madeira et al. (2019) [14]</td>
<td>Multiple sequence alignment</td>
</tr>
<tr>
<td>4</td>
<td>ProtParam</td>
<td>Gasteiger et al. (2003) [16]</td>
<td>Physicochemical characterization</td>
</tr>
<tr>
<td>5</td>
<td>PSORTb</td>
<td>Yu et al. (2010) [18]</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>PSLpred</td>
<td>Bhasin et al. (2005) [19]</td>
<td>Subcellular localization prediction</td>
</tr>
<tr>
<td>7</td>
<td>CELLO</td>
<td>Yu et al. (2006) [17]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>SOSUIGramN</td>
<td>Imai et al. (2008) [20]</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>TMHMM</td>
<td>Moller et al. (2001) [21]</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>HHMTOp</td>
<td>Tusnady and Simon (2001) [22]</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CCTOP</td>
<td>Dobson et al. (2015) [23]</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Pfam</td>
<td>Finn et al. (2014) [26]</td>
<td>Family relationship identification</td>
</tr>
<tr>
<td>14</td>
<td>Superfamily</td>
<td>Wilson et al. (2007) [27]</td>
<td>Superfamily search</td>
</tr>
<tr>
<td>15</td>
<td>COILS</td>
<td>Lupas et al. (1991) [28]</td>
<td>Coiled-coil motif identification</td>
</tr>
<tr>
<td>17</td>
<td>InterPro</td>
<td>Hunter et al. (2009) [29]</td>
<td>Functional classification</td>
</tr>
<tr>
<td>18</td>
<td>STRING</td>
<td>Szklarczyk et al. (2015) [31]</td>
<td>Interaction network analysis</td>
</tr>
<tr>
<td>20</td>
<td>SOPMA</td>
<td>Geourjon and Deleage (1995) [33]</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>HHpred</td>
<td>Zimmermann et al. (2018) [34]</td>
<td>Tertiary structure prediction</td>
</tr>
<tr>
<td>22</td>
<td>PROCHECK</td>
<td>Laskowski et al. (1993) [36]</td>
<td>Structure verification</td>
</tr>
<tr>
<td>23</td>
<td>Verify3D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>ERRAT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
were Ala (92) followed by Val (51), Arg (42), Gly (41), Leu (40), Asp (32), Glu (30), Pro (26), Thr (21), Ser (19), His (17), Phe (11), Ile (10), Tyr (8), Trp (6), Asn (5), Gln (4), Met (4), and Cys (2). The calculated molecular weight was 48216.15 Da and theoretical pI was 5.27 indicating the protein to be negatively charged. Total number of positively charged residues (Arg + Lys) and the total number of negatively charged residues (Asp + Glu) were found to be 62 and 42, respectively. The computed instability index was 32.67 classifying the protein as stable one. Aliphatic index was 94.34 which gives an indication of proteins’ stability over a wide temperature range. The GRAVY was 0.053. Positive value of GRAVY indicates that the protein is polar. Protein half-life computed was found to be 30 h in mammalian reticulocytes (in vitro), > 20 hours in yeast (in vivo), > 10 h in Escherichia coli (in vivo). And the molecular formula of protein was identified as C_{2119}H_{3350}N_{636}O_{643}S_{6}.

Functional annotation of hypothetical protein

The conserved domain search tool revealed that this hypothetical protein sequence was found to have two domains, MBL-fold metallo-hydrolase domain (accession No. cd07724) and rhodanese homology domain (RHOD) (accession No. cd00158). The result was also checked by two other domain searching tools namely InterProScan and Pfam. Pfam server predicted the rhodanese like domain at 362–444 amino acid residues with an e-value of 2.3e-05 and metallo-beta-lactamase superfamily domain at 16–171 amino acid residues with an e-value of 4.7e-07. InterproScan server predicted rhodanese like domain at 249–454 amino acid residues with an e-value of 0.00447.

Table 2. Similar protein obtained from non-redundant UniProt KB/SwissProt sequences

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Organism</th>
<th>Protein name</th>
<th>Identity (%)</th>
<th>Score</th>
<th>e-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP_011027250.1</td>
<td>Streptomyces</td>
<td>MULTISPECIES: MBL fold metallo-hydrolase</td>
<td>100</td>
<td>889</td>
<td>0.0</td>
</tr>
<tr>
<td>WP_003978243.1</td>
<td>Streptomyces</td>
<td>MULTISPECIES: MBL fold metallo-hydrolase</td>
<td>99.57</td>
<td>886</td>
<td>0.0</td>
</tr>
<tr>
<td>WP_121713050.1</td>
<td>Streptomyces sp. ESN91</td>
<td>MBL fold metallo-hydrolase</td>
<td>99.13</td>
<td>884</td>
<td>0.0</td>
</tr>
<tr>
<td>WP_016325181.1</td>
<td>Streptomyces lividans</td>
<td>MBL fold metallo-hydrolase</td>
<td>99.35</td>
<td>883</td>
<td>0.0</td>
</tr>
<tr>
<td>WP_093455449.1</td>
<td>Unclassified Streptomyces</td>
<td>MULTISPECIES: MBL fold metallo-hydrolase</td>
<td>99.35</td>
<td>883</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 3. Similar protein obtained from UniProt database

<table>
<thead>
<tr>
<th>Entry name</th>
<th>Organism</th>
<th>Protein name</th>
<th>Identity (%)</th>
<th>Score</th>
<th>e-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q88FF3.1</td>
<td>Pseudomonas putida KT2440</td>
<td>Hydroxyacylglutathione hydrolase</td>
<td>32.97</td>
<td>60.1</td>
<td>6e-09</td>
</tr>
<tr>
<td>B1JBN3.1</td>
<td>Pseudomonas putida W619</td>
<td>Hydroxyacylglutathione hydrolase</td>
<td>31.49</td>
<td>58.2</td>
<td>3e-08</td>
</tr>
<tr>
<td>B0KN02.1</td>
<td>Pseudomonas putida GB-1</td>
<td>Hydroxyacylglutathione hydrolase</td>
<td>30.77</td>
<td>57.8</td>
<td>3e-08</td>
</tr>
<tr>
<td>ASW167.1</td>
<td>Pseudomonas putida F1</td>
<td>Hydroxyacylglutathione hydrolase</td>
<td>31.32</td>
<td>55.8</td>
<td>1e-07</td>
</tr>
<tr>
<td>D3RBP9.1</td>
<td>Allochromatium vinasum DSM 180</td>
<td>Sulfurtransferase</td>
<td>33.33</td>
<td>51.6</td>
<td>3e-07</td>
</tr>
</tbody>
</table>

Fig. 1. A complete workflow of the study.

Fig. 2. Phylogenetic trees with true distance of different hydrolases proteins.
no acid residues and metallo-beta-lactamase domain at 13–180
amino acid residues. Rhodanese like domain, lactamase-B and
MreB-Mbl domains were also found by Motif server. Superfamily
search revealed present of Metallo-hydrolase/oxidoreductase and
rhodanese/cell cycle control phosphatase superfamily. β-Lact-
amases can catalyze the hydrolysis of a wide range of β-lactam an-
tibiotics. Members of the MBL-fold metallohydrolase superfamily
are mainly hydrolytic enzymes which carry out various biological
functions. Both the active and inactive version of the Rhodanese
domain in a variety of proteins including certain protein phospha-
tases, sulfide dehydrogenases, certain stress proteins and sulfuryl
transferases, where they are thought to play a regulatory role in
multidomain proteins (Fig. 3). All these results confirm the pres-
ence of hydrolytic enzyme containing domains in this protein.
Fold pattern recognition by PFP-FunDSeqE tool revealed the
presence of a ‘(TIM)-barrel’ fold within the protein sequence.
(TIM)-barrel structure is generally eight stranded α/β barrel. The
x-axis of the graph represents the position in the protein of amino
acid number (starting at the N-terminus) and the y-axis shows the
coiled coil whereas ‘Window’ refers to the width of the amino acid
‘window’ that is scanned at one time (Fig. 4).

Subcellular localization nature
Subcellular localization analysis was predicted by CELLO and val-
idated by PSORTb, SOSUIGramN, and PSLpred. The subcellular
localization of the hypothetical protein was predicted to be a cyto-
plasmic protein (Table 4). Absent of transmembrane helices pre-
dicted by THMM and HMMTOP also emphasizes the result of
being a cytoplasmic protein. Also, CCTOP server predicted that
the query protein was not a transmembrane protein. All these re-
sults summarize the protein as a cytoplasmic one.

Secondary structure analysis
The SOPMA secondary structure prediction server analysis re-
vealed the proportions of alpha helix, beta turn, extended strand,
and the random coil of protein as 31.89%, 9.11%, 18.87%, and
40.13%, respectively (Supplementary Fig. 2).

Three-dimensional structure analysis
Prediction of 3D structure was done by HHpred server. The server

Fig. 4. Coil depicts the heptads corresponding to the residue
windows 14 (green), 21(blue), and 28 (red).

Fig. 3. Functional annotation of the hypothetical protein.
predicted 3D structure of the protein with 100% identity with the highest scoring template (PDB ID: 3TP9_A) (Fig. 5). 3TP9 is the crystal structure of Alicyclobacillus acidocaldarius protein with β-lactamase and rhodanese domains. This protein is a homo-dimer which has two chains (chain A and chain B) and the chain A was used as template to build the model. Validation of the predicted three-dimensional model was assessed by PROCHECK through Ramachandran plot analysis, where the distribution of φ and ψ angle in the model within the limits are shown (Table 5, Fig. 6). Residues in the most favored regions covered 90.9%, which is the quality of a valid model. Finally, the established model of 3D structure for the target sequence was verified by structure validation server Verify3D and ERRAT. In the Verify3D graph, 92.73% of the residues have averaged 3D-1D score ≥ 0.2 which indicates that the environmental profile of the model is good and the overall quality factor predicted by the ERRAT server was 69.0583 indicates a good model. The 3D structure was later modified by YASARA energy minimization server. The energy calculated before energy minimization was –77,930.2 kJ/mol whereas after energy minimization (through 3 round of steepest descent method), it was changed to far less value of –244,148.6 kJ/mol making the modeled structure more stable one.

**Table 4. Subcellular localization analysis**

<table>
<thead>
<tr>
<th>No.</th>
<th>Analysis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CELLO 2.5</td>
<td>Cytoplasmic localization</td>
</tr>
<tr>
<td>2</td>
<td>PSORTb</td>
<td>Cytoplasmic localization</td>
</tr>
<tr>
<td>3</td>
<td>SOSUIGramN</td>
<td>Cytoplasmic localization</td>
</tr>
<tr>
<td>4</td>
<td>PSIPred</td>
<td>Cytoplasmic protein</td>
</tr>
<tr>
<td>5</td>
<td>TMHMM 2.0</td>
<td>No transmembrane helices present</td>
</tr>
<tr>
<td>6</td>
<td>HMMPred</td>
<td>No transmembrane helices present</td>
</tr>
<tr>
<td>7</td>
<td>CCTOP</td>
<td>Not transmembrane protein</td>
</tr>
</tbody>
</table>

**Table 5. Ramachandran plot statistics of the hypothetical protein**

<table>
<thead>
<tr>
<th>Ramachandran plot statistics</th>
<th>No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in the most favored regions [A, B, L]</td>
<td>351 (90.9)</td>
</tr>
<tr>
<td>Residues in the additional allowed regions [a, b, l, p]</td>
<td>26 (6.7)</td>
</tr>
<tr>
<td>Residues in the generously allowed regions [a, b, l, p]</td>
<td>8 (2.1)</td>
</tr>
<tr>
<td>Residues in the disallowed regions</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>No. of non-glycine and non-proline residues</td>
<td>386</td>
</tr>
<tr>
<td>No. of end-residues (excl. Gly and Pro)</td>
<td>2</td>
</tr>
<tr>
<td>No. of glycine residues (shown in triangles)</td>
<td>41</td>
</tr>
<tr>
<td>No. of proline residues</td>
<td>25</td>
</tr>
<tr>
<td>Total No. of residues</td>
<td>454</td>
</tr>
</tbody>
</table>

**Protein-protein interaction analysis**

STRING 10.0 search was carried out for the identification of possible functional interaction network of the protein [31]. The identified functional partners with scores were; SCO0619 (0.970), SCO0620 (0.743), SCO0621 (0.739), groES (0.568), SCO2899 (0.568), guaA (0.545), SCO6160 (0.520), pheT (0.508), SCO5178 (0.485), polA (0.473). Of them, SCO0619 is a possible membrane protein. The others are two hypothetical proteins, two chaperonins, GMP synthase, multifunctional fusion protein, phenylalanine tRNA ligase β subunit, putative sulfurylase, and DNA polymerase I (Fig. 7).

**Active site of the hypothetical protein**

The predicted active site of the protein found that 42 amino acids are involved in potent active site (Fig. 8). The best active site was...
found in areas with 613.075 and a volume of 608.774 amino acids. The amino acid residues in the active site were shown in Supplementary Fig. 3.

**Conclusion**

The identification of protein functions is fundamental for the understanding of biological processes. So, this study was aimed to determine the structural and biological function of SCO0618, a hypothetical protein of this bacterium through an in-silico approach. The identified protein revealed several characteristics such as cytoplasmic nature, hydrolytic enzymes containing domain presence, ‘(TIM)-barrel’ fold presence, and hydrolase activity emphasize the significance of this protein. These characters of the hypothetical protein will strengthen basic knowledge on *S. coelicolor*. So, extended in-vitro research has to be carried out to experimentally validate the possibilities shown here and to find out the proteins’ role in biotechnology.

**Fig. 7.** String network analysis of the hypothetical protein, indicates as SCO0618.

**Fig. 8.** Active site of the hypothetical protein. Here the red sphere indicates the active site of the protein.

---

**ORCID**

Nadim Ferdous: https://orcid.org/0000-0003-4240-6829  
Mahjerin Nasrin Reza: https://orcid.org/0000-0002-0331-1416  
Md. Tabassum Hossain Emon: https://orcid.org/0000-0001-5521-4565  
Md. Shariful Islam: https://orcid.org/0000-0002-7631-882X  
A. K. M Mohiuddin: https://orcid.org/0000-0003-4188-6592  
Mohammad Uzzal Hossain: https://orcid.org/0000-0002-9957-122X

**Authors’ Contribution**

Conceptualization: NF, MUH. Data curation: NF, MNR. Formal analysis: NF, MNR, MTHE, MUH. Methodology: NF, MNR. Writing - original draft: NF, MNR. Writing - review & editing: MUH, MTHE, MSI, AKMM.

**Conflicts of Interest**

No potential conflict of interest relevant to this article was reported.

**Acknowledgments**

We are grateful to the book of Gobeshonay Bioinformatics-1st Part.

**Supplementary Materials**

Supplementary data can be found with this article online at http://www.genominfo.org.

**References**

6. Hobbs G, Obanye AI, Petty J, Mason JC, Barratt E, Gardner DC,
29326665