IJEMD-BMCR, 1 (2) (2023), 1 – 12 https://doi.org/10.54938/ijemdbmcr.2023.01.2.123

International Journal of Emerging Multidisciplinaries:
Biomedical and Clinical Research

Research Paper
Journal Homepage: www.ojs.ijemd.com
ISSN (print): 2957-5885 ISSN (online): 2960-0731

Identification of a Novel ABC-2 Family Transporter Protein in *Bacillus megaterium* DSM319 Genome using Bioinformatics

Muhammad Shakeeb Sharif [1] and Muhammad Sufian [1*]

[1] Department of Biotechnology, Mirpur University of Science and Technology (MUST), Mirpur, (AJK), Pakistan
** Corresponding Author*

Abstract

Bacillus megaterium is a Gram-positive rod-shaped bacterium that is majorly responsible for the formation of spores aerobically and found throughout the world. It has two strains named as QMB1551 and DSM319 having GC content of 38%. It has a large number of the genes portrayed as hypothetical proteins, with no known function. It has played a promising role in the field of biotechnology as a producer of genetically engineered proteins. Bioinformatics is one of the most advanced field that employed various intelligent tools in order to explore the function, structure, and sequence annotation of proteins. In the study, NCBI BLASTp was utilized for sequence comparison of dataset of hypothetical proteins of *B. megaterium* DSM319 against the subject sequence of dataset of biotechnological important proteins. The best hit obtained from BLASTp results for the hypothetical protein ADF38049.1 was ABC-2 family transporter protein (WP_013081552.1) having known functions. With that about 80% percent identity, a daunorubicin ABC transporter permease (WP_071387832.1) (*Anaerobacillus alkalidiazotrophicus*) having query cover 100% and e-value 1e-153 was obtained. The permease enzyme facilitates the ABC transporter to carry out export of daunorubicin that is produced by Bacillus spp. as a secondary metabolite. Moreover, it was forecasted that *Bacillus megaterium* might be able to produce daunorubicin as a secondary metabolite.

Keywords: ATP-binding cassette; ABC-2 family transporter; *Bacillus megaterium*; Basic Local Alignment Search Tool (BLAST); Daunorubicin; Hypothetical Protein

1. Introduction

Bacillus megaterium is one of the known large-sized Gram positive and rod shaped bacterium. It is greatly responsible for the formation of spores aerobically and found throughout the world [1]. The highly alluring proteins of the respective bacterium are the family of cytochrome P450 monooxygenases

Email addresses: **mshakeebsharif@gmail.com (Muhammad Shakeeb Sharif), msufian@must.edu.pk (Muhammad Sufian)**

[2]. The size of the reference bacterial genome is 5.48 Mb and it contains about 5,858 exons. It has two strains labelled as QMB1551 and DSM319 having 38% of GC content. The bacterium is an ideal organism to be utilized in biotechnology and it has promising significance in various industries due to its intrinsic remarkable features [3]. It is also a potent agent for the creation of recombinant proteins and a variety of enzymes and has ability to mobilize valuable metals from e-waste in a bioleaching process. In *B. megaterium*, the *in silico* study related to the biosynthesis and accumulation of lipids emphasized the need of using appropriate carbon sources for the fermentation process [4]. The respective bacterium has efficacious ability to adapt and survive in a highly acidic environment via various mechanisms with the assistance of genes network. *B. megaterium* has potential activity in order to enhance growth of plants by control the effect of various plant pathogens [5]. Recently, a novel bacteria, *B. megaterium* 1259, has been discovered in chicken ordure and utilized as a probiotic in poultry fodder. Moreover, it has potential to enhance milk productivity and don't cause any side effect on blood metabolites in cows [6].

Hypothetical proteins are those proteins whose existence have been proposed but there is a lack of experimental evidence in order to express them in vivo [7]. With the passage of time, annotation and curation attempts have assisted to control the defiance in order to understand their unique functions. The various researchers have been developed productive techniques for hypothetical proteins in order to decode sequential, structural and functional relationship uniquely in terms of their functional modelling [8]. The annotated hypothetical proteins played an essential function in understanding gene control, target identification and functional analysis. HPs are analyzed in order to find out the operative characteristics about domain and their families. Domains are the conserved areas in proteins, accountable for the unique function and interactions where the protein family assists to classify the protein established on related functions as well as sequence resemblance [9]. There are numerous ways that have been developed to predict the protein function using data from gene expression profiles, protein-protein interactions, phylogenetic profiles, and sequence similarity. Characterization of the putative proteins *in silico* assist in establishing the three-dimensional architectures of the hypothetical proteins, which could show new domains, protein networks, and different routes. In addition to proteins, biochemical pathways are another important component to take into account when developing a shortlist for characterisation. The normal course of the biochemical process could be hampered by a slight alteration to the protein's structure or function [10]. Moreover, the structural as well as annotation of the

hypothesized proteins' functions may also manifest efficacious biomarkers and pharmaceutical target [11].

There is a flourishing need for the hypothetical proteins to undergo automatic annotation by using bioinformatics' tools. The structural genomics drives the extensive structures of various HPs at an enhancing rate [12]. The main objectives of the structural bioinformatics are the construction of novel strategies in order to deal with the biological macromolecules information to solve problems in biology and to generate new data [13]. The several bioinformatics tools have been evolved to assign a unique function to HPs of numerous species. The promising tools are SMART, UniProt & Pfam etc. and these are linked with all the information found in databases utilizing domains and ontological information to assist characterizations of protein function [14]. The imperative database for understanding the purpose of proteins in biological networks is a STRING database [15]. BLAST is a potent technique for locating sequences that are similar to the query sequence in the same or different organisms. The majority of the sequences that go into BLAST are in FASTA and GenBank format, and the output can come in a variety of forms like HTML, XML, or plain text. It is popular in terms of having good balance of sensitivity, reliability and speed [16]. The Protein databases are involving the sequences from a variety of ways, involving translations from annotated coding sections in GenBank, TPA, RefSeq and various data from RCSB PDB and SwissProt. The sequence of proteins are the distinct factors that affect both the structure and function of biological systems [17]. UniProt is a publicly accessible database that contains protein sequences, functional data, and other entries collected from several genome sequencing initiatives. The knowledge base for UniProt is a enlarge agent of protein sequences as well as linked comprehensive annotation. It continues to adapt the gathering, processing and displaying of data in order to enhance the attainability and profitability of protein information for the asset of all [18].

The various studies reported by the researchers about the functional anticipation of hypothetical proteins via *in silico* ways have been victoriously utilized in numerous bacterial species like, *T. pallidum, B. anthracis* , *S. aureus* and *P. marinus* HPs are forecasted by the homologous sequences, but there is deficit of biological as well as chemical evidence [19]. The goal of the ongoing work was to assign one of the HPs found on the *B. megaterium* genome a function for the recognition of a novel protein that might be played an essential role in a biomedical industry.

Materials and Methods

Using regular expressions to define rules or patterns manually is one way to get data. By manually creating patterns that extract information about proteins from biological texts or publications. Using automatically learned pattern-based extraction rules to determine the relation or entity type is another strategy [20]. The complete proteome of *B. megaterium* DSM319 was derived from the NCBI Protein database. (https://www.ncbi.nlm.nih.gov/protein/). The hypothetical proteins were extracted from complete proteome in FASTA format using FaBox tool (https://users-birc.au.dk/~palle/php/fabox/index.php). FaBox is an online toolbox for FASTA sequences and assemblage of straightforward and understandable web services that enable the researchers to complete complex tasks with the sequence data quickly. These proteins are needed to predict their function using bioinformatics tools [21].

A dataset of hundred biotechnological important proteins was made for the current study. The dataset assisted in order to run NCBI BLASTp of hypothetical proteins of *B. megaterium* DSM-319 against it for a sequence similarity search. The NCBI protein database was accessed using the FASTA format. The dataset involved proteins along with their significant industrial uses and proteomic sequences. These proteins play a significance role in various fields of biotechnological industries like food, pharmaceutical, dairy, beverage and others. The dataset was submitted in a centralized BIPs dataset of hundred proteins that have significant role in various fields of biotechnology.

NCBI BLASTp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to perform sequence comparison of query sequence of hypothetical proteins of *B. megaterium* DSM319 dataset against subject sequence of BIPs dataset of hundred proteins. It was done by clicking on align two or more sequences. The algorithm parameters were remain unchanged. It is a heuristic method for creating alignments by optimizing a local similarity metric. Compared to dynamic programming techniques, it compares protein and nucleotide sequences much more quickly [22]. The best hit was searched by using three values: query coverage, e-value, and percent identity. The query of hypothetical proteins should be of large size to avoid partial submissions in NCBI Protein database. Results with good score were checked for pairwise alignments.

The obtained best hit ortholog of known functions and having good query coverage, e-value, and percent identity was searched in the Non-Redundant (nr) database using NCBI BLASTp to determine a

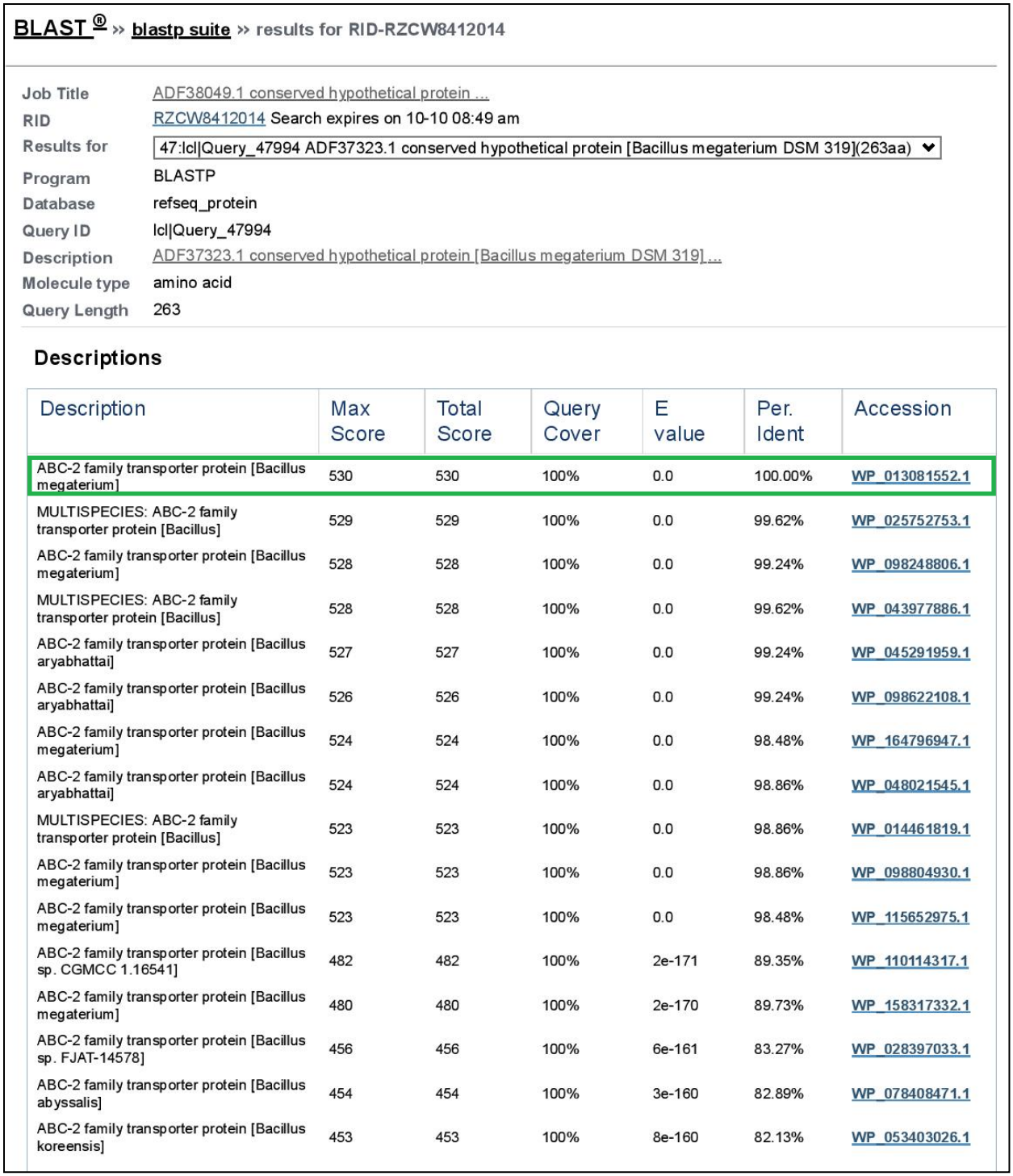
protein's consensus nomenclature and similarity to other proteins. NCBI GenBank CDS translations, PDB, Swiss-Prot, PIR, and PRF sequences that are non-identical are included in NCBI nr database. The BLASTp results of best hit are represented in tabular form and graphical representation of pairwise alignment using dots for similarities.

2. Results and Discussion

Windows 10 Pro, version 1903, OS build 18362.1082, 64-bit AMD E-300 APU with Radeon (tm) HD graphics and 4GB RAM served as the operating system for all computational investigations. Google Chrome, version 85.0.4183.102 (Official Build) (64-bit), was the browser utilized for those calculations.

The bacterium *B. megaterium* was selected for the current study. The DSM319 strain was selected in order to foresee the function of one of its hypothetical protein. Its genome size is about 5.1 Mb having a GC content of 38%. The complete proteome of *B. megaterium* DSM319 was derived from the NCBI Protein database (Accession CP001983.1). It contains 5,100 proteins out of which 1,482 (29%) were hypothetical proteins.

Among the pool of hypothetical proteins of *B. megaterium* DSM319, the best-hit paralog protein to ADF38049.1 was obtained from BLASTp results for sequence similarity is ABC-2 family transporter proteins having known functions (Figure 1). It obtained from *B. megaterium* DSM-319 query ID WP_071387832.1. Its query cover is 100%, e-value 0 and percent identity is 100%. With that about 80% of percent identity a daunorubicin ABC transporter permease WP_071387832.1 (*A. alkalidiazotrophicus*) an ortholog having query cover value 100% and e-value 1e-153 is obtained (Figure 2). Pairwise sequence alignment is illustrated in figure 3. So, it is predicted that hypothetical protein of *B. megaterium* DSM-319 query ID ADF38049.1 has the same function as daunorubicin ABC transport permease. The permease enzyme facilitates the ABC transporter to carries out export of daunorubicin that is produced by Bacillus spp. as a secondary metabolite.



BLAST® » **blastp suite** » results for RID-RZCW8412014

Job Title	ADF38049.1 conserved hypothetical protein ...
RID	RZCW8412014 Search expires on 10-10 08:49 am
Results for	47:lcl|Query_47994 ADF37323.1 conserved hypothetical protein [Bacillus megaterium DSM 319](263aa)
Program	BLASTP
Database	refseq_protein
Query ID	lcl|Query_47994
Description	ADF37323.1 conserved hypothetical protein [Bacillus megaterium DSM 319] ...
Molecule type	amino acid
Query Length	263

Descriptions

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
ABC-2 family transporter protein [Bacillus megaterium]	530	530	100%	0.0	100.00%	WP_013081552.1
MULTISPECIES: ABC-2 family transporter protein [Bacillus]	529	529	100%	0.0	99.62%	WP_025752753.1
ABC-2 family transporter protein [Bacillus megaterium]	528	528	100%	0.0	99.24%	WP_098248806.1
MULTISPECIES: ABC-2 family transporter protein [Bacillus]	528	528	100%	0.0	99.62%	WP_043977886.1
ABC-2 family transporter protein [Bacillus aryabhattai]	527	527	100%	0.0	99.24%	WP_045291959.1
ABC-2 family transporter protein [Bacillus aryabhattai]	526	526	100%	0.0	99.24%	WP_098622108.1
ABC-2 family transporter protein [Bacillus megaterium]	524	524	100%	0.0	98.48%	WP_164796947.1
ABC-2 family transporter protein [Bacillus aryabhattai]	524	524	100%	0.0	98.86%	WP_048021545.1
MULTISPECIES: ABC-2 family transporter protein [Bacillus]	523	523	100%	0.0	98.86%	WP_014461819.1
ABC-2 family transporter protein [Bacillus megaterium]	523	523	100%	0.0	98.86%	WP_098804930.1
ABC-2 family transporter protein [Bacillus megaterium]	523	523	100%	0.0	98.48%	WP_115652975.1
ABC-2 family transporter protein [Bacillus sp. CGMCC 1.16541]	482	482	100%	2e-171	89.35%	WP_110114317.1
ABC-2 family transporter protein [Bacillus megaterium]	480	480	100%	2e-170	89.73%	WP_158317332.1
ABC-2 family transporter protein [Bacillus sp. FJAT-14578]	456	456	100%	6e-161	83.27%	WP_028397033.1
ABC-2 family transporter protein [Bacillus abyssalis]	454	454	100%	3e-160	82.89%	WP_078408471.1
ABC-2 family transporter protein [Bacillus koreensis]	453	453	100%	8e-160	82.13%	WP_053403026.1

Figure 1: BLAST results of hypothetical protein ADF38049.1 against nr database for best-hit paralog protein ABC-2 family transporter protein WP_013081552.1

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: ABC-2 family transporter protein [Bacillus]	450	450	100%	2e-158	85.93%	WP_025909544.1
ABC-2 family transporter protein [Bacillus aryabhattai]	448	448	100%	6e-158	86.31%	WP_094911473.1
ABC-2 family transporter protein [Bacillus sp. CBEL-1]	445	445	100%	1e-156	85.93%	WP_131887143.1
ABC-2 family transporter protein [Fictibacillus arsenicus]	442	442	100%	1e-155	80.61%	WP_077364704.1
ABC-2 family transporter protein [Fictibacillus arsenicus]	442	442	100%	1e-155	80.61%	WP_066286228.1
ABC-2 family transporter protein [Metabacillus litoralis]	442	442	100%	1e-155	79.09%	WP_146950468.1
ABC-2 family transporter protein [Bacillus sp. RO2]	442	442	100%	1e-155	80.99%	WP_169289783.1
ABC-2 family transporter protein [Bacillus horikoshii]	441	441	100%	4e-155	80.99%	WP_148988499.1
ABC-2 family transporter protein [Bacillus sp. CHD6a]	439	439	100%	2e-154	81.37%	WP_060666527.1
ABC-2 family transporter protein [Fictibacillus sp. BK138]	439	439	100%	2e-154	80.23%	WP_130297054.1
ABC-2 family transporter protein [Bacillus horikoshii]	439	439	100%	2e-154	80.61%	WP_064099213.1
daunorubicin ABC transporter permease [Anaerobacillus alkalidiazotrophicus]	439	439	100%	2e-154	79.85%	WP_071387832.1
ABC-2 family transporter protein [Bacillus vietnamensis]	439	439	100%	3e-154	80.23%	WP_148942282.1

Figure 2: BLAST results of WP_013081552.1 against nr database to find its function

daunorubicin ABC transporter permease [Anaerobacillus alkalidiazotrophicus]

Sequence ID: WP_071387832.1 **Length:** 263 **Number of Matches:** 1

Range 1: 1 to 263 GenPept Graphics ▼ Next Match ▲ Pre

Score	Expect	Method	Identities	Positives	Gaps
439 bits(1130)	2e-154	Compositional matrix adjust.	210/263(80%)	236/263(89%)	0/263(0%)

```
Query  1    MDKYIEMIRIRFLMMLAYRTNYYSGILIYAINIGAYYFLWSAIYGGKENIQGLSITQMTT  60
Sbjct  1    .S..L....V...................S..........T.......E.....VI....  60

Query  61   YVAVSWMARAFYFNNIDREMATEIKDGKVAVELIKPYSYLGMKTMQGLGEGIFRLLFFSV  120
Sbjct  61   ....A..............I.A.........I.M.R..N.............V........  120

Query  121  PGMIIVAFLFPVQFSANAATWLYFGLSLIFSFIINTQINLLTGITTFFLFNNDGLIRAKR  180
Sbjct  121  ...F........LN....LT..GL.FI..L....V..........M.......S.......  180

Query  181  VVIDLFSGLLLPISFYPFWAQHIMSYFPFQAISYIPSMIFTNGFKGQEVINALITQAVWS  240
Sbjct  181  ..................I...TV.T.L..............E..V...IYS...L.MI.V  240

Query  241  GLLFIPIACLWNIAKKKMVIQGG  263
Sbjct  241  AI.L...QL..L....QLIV...  263
```

Figure 3: Pairwise alignment of figure 2 using dots for identities

The resultant best hit ortholog daunorubicin ABC transporter permease (*A. alkalidiazotrophicus*) having known functions. ABC transporter stands for ATP-binding cassette. In all organisms these transporters are ubiquitous in nature [23]. The various soil bacterium produced daunorubicin as it is one of the most commonly used anticancer agent. The drrA and drrB open reading frames that lead to carry out the exportation of antibiotics, were predicted to encode ABC transporter permease. Protein that binds to ATB DrrA belonged to the ABC family of transporters and resembled P-glycoprotein, which is prevalent in cancerous cells both structurally and functionally. On the other hand, DrrB may serve as a transporter for the daunorubicin influx. DrrA and DrrB have agreed to work together to create a pump powered by ATP for the purpose of infusing each medication. [24]. Daunorubicin is the prime anthracycline antibiotic isolated from the diverse species of Bacillus and is the most productive anticancer drug. It is widely used for the treatment of lymphocytic leukaemia, both acute and chronic [25]. The exact mechanism for the entry of DNR into the cells is not completely studied but it diffused passively into the plasma membrane from where selectively transported into the nucleus and bind to the proteasomes [26].

All the bacterial ABC transporters mediated the utilization of high affinity binding protein that uptake a broad diversity of substrates like nutrients and osmoprotectants involving small sugar, amino acids and vitamins. The massive compounds utilized energy in order to transport across the exterior membrane via the rich affinity transporters [27]. Daunorubicin is also employed as a CHOP regime in order to treat adult T leukemic cells cognated by Human T-Lymphotopic Virus [28].

Moreover, the daunorubicin is the effective anticancer medication used to treat a variety of diseases of cancerous cells like solid as well as hematological tumors. It controlled the growth of cancerous cells by the process of chimerism between their base pairs and firmly bound to the DNA [29]. Daunorubicin is the high active drug having great ability to distribute indicating a prominent affinity with the tissues. It reached earlier at high peak concentrations intracellularly as compared to that of doxorubicin and its plasma concentration ratio is elevated than for its reductive metabolite daunorubicinol. DNA conjugated drugs have distinctly modify the DNR uptake in the leukemia cells in order to treat it. It is an imperative drug employed for the treatment of myelodysplastic syndromes [30]

3. Conclusions

In current work, the hypothetical protein of *B. megaterium* DSM-319 ADF38049.1 predicted to have the function of ABC-2 family transporter protein. With that about 80% percent identity an ortholog daunorubicin ABC transporter permease (*A. alkalidiazotrophicus*) is obtained. The permease enzyme assists the ABC transporter to carries out export of daunorubicin produced by various soil bacteria as it is one of the most commonly used anticancer agent. *B. megaterium* might be able to produce daunorubicin as a secondary metabolite and daunorubicin ABC transporter permease might be responsible for expelling out daunorubicin in an ATP-dependent manner from the respective bacterium. The current work intensified the industrial significance of *B. megaterium* in the field of pharmaceuticals as the daunorubicin might be proved more effective against cancerous cells.

Acknowledgement

The authors appreciate the support of Researcher for Project number (MUST SRGP Ref. No. ORIC/180/2020), Mirpur University of Science and Technology, MUST, 10250, Pakistan

Competing Interests

None.

4. References

[1] Kalsi, N., *et al*. Whole-Genome Sequence of *Bacillus megaterium* Strain SGAir0080, Isolated from an Indoor Air Sample. *Microbiology resource announcements*, **8(50),** e01249-19. https://doi.org/10.1128/MRA.01249-19 (2019).

[2] Sura, Narendra & Hiremath, Lingayya. Isolation of *Bacillus megaterium* and its Commercial Importance. International Journal of ChemTech Research. **12,** 30-36, 10.20902/IJCTR.2019.120405. https://www.researchgate.net/publication/332892439 (2019).

[3] Eppinger, M., *et al*. Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *Journal of bacteriology*, **193(16),** 4199–4213. https://doi.org/10.1128/JB.00449-11 (2011).

[4] Grage, K., McDermott, P. &Rehm, B.H.A. Engineering *Bacillus megaterium* for production of functional intracellular materials. *Microb Cell Fact.* **16,** 211. https://doi.org/10.1186/s12934-017-0823-5 (2017).

[5] Goswami, G., *et al*. *Bacillus megaterium* adapts to acid stress condition through a network of genes: Insight from a genome-wide transcriptome analysis. *Scientific reports*, **8(1),** 16105. https://doi.org/10.1038/s41598-018-34221-0 (2018).

[6] Deng, B., *et al*. Genomics Analysis of *Bacillus megaterium* 1259 as a Probiotic and Its Effects on Performance in Lactating Dairy Cows. *Animals: an open access journal from MDPI*, **11(2),** 397. https://doi.org/10.3390/ani11020397 (2021).

[7] Sivashankari, S., & Shanmughavel, P. Functional annotation of hypothetical proteins - A review. *Bioinformation*, **1(8),** 335–338. https://doi.org/10.6026/97320630001335 (2006).

[8] Ijaq, J., *et al*. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC bioinformatics*, **20(1),** 14. https://doi.org/10.1186/s12859-018-2554-y (2019).

[9] Thakur, C. J., Saini, S., Notra, A., Chauhan, B., Arya, S., Gupta, R., Thakur, J., & Kumar, V. Deciphering the functional role of hypothetical proteins from *Chloroflexus aurantiacs* J-10-f1 using bioinformatics approach. *Molecular biology research communications*, ***9*(3),** 129–139. https://doi.org/10.22099/mbrc.2020.36894.1495 (2020).

[10] Hassan, S. A., & Haq, I., Revisiting Triangle of Trade-Growth and Pollution for Middle-Income Transitions. International Journal of Emerging Multidisciplinaries: Social Science, **2(1),** 2023.

[11] Rabbi, M. F., Akter, S. A., Hasan, M. J., & Amin, A. In Silico Characterization of a Hypothetical Protein from *Shigella dysenteriae* ATCC 12039 Reveals a Pathogenesis-Related Protein of the Type-VI Secretion System. *Bioinformatics and biology insights*, **15,** 11779322211011140. https://doi.org/10.1177/11779322211011140 (2021).

[12] Mohan, R., & Venugopal, S. Computational structural and functional analysis of hypothetical proteins of *Staphylococcus aureus*. *Bioinformation*, **8(15),** 722–728. https://doi.org/10.6026/97320630008722 (2012).

[13] Shi, M., Gao, J., & Zhang, M. Q. Web3DMol: interactive protein structure visualization based on WebGL. *Nucleic acids research*, **45(W1),** W523–W527. https://doi.org/10.1093/nar/gkx383 (2017).

[14] Da Costa, W. *et al*. Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. *PloS one*, **13(6),** e0198965. https://doi.org/10.1371/journal.pone.0198965 (2018).

[15] Jeong, H., Qian, X., & Yoon, B. J. Effective comparative analysis of protein-protein interaction networks by measuring the steady-state network flow using a Markov model. *BMC bioinformatics*, **17(Suppl 13),** 395. https://doi.org/10.1186/s12859-016-1215-2 (2016).

[16] Shah, N., Altschul, S. F., & Pop, M. Outlier detection in BLAST hits. *Algorithms for molecular biology, AMB*, **13,** 7. https://doi.org/10.1186/s13015-018-0126-3 (2018).

[17] Chen, C., Huang, H., & Wu, C. H. Protein Bioinformatics Databases and Resources. *Methods in molecular biology (Clifton, N.J.)*, **1558,** 3–39. https://doi.org/10.1007/978-1-4939-6783-4_1 (2017).

[18] The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic acids research*, **45(D1),** D158–D169. https://doi.org/10.1093/nar/gkw1099 (2017).

[19] Araújo, C. L. *et al*. In silico functional prediction of hypothetical proteins from the core genome of *Corynebacterium pseudotuberculosis* biovar *ovis*. *PeerJ*, ***8*,** e9643, https://doi.org/10.7717/peerj.9643 (2020).

[20] Asif, S., & Khan, A., An Efficient Hybrid Bi-LSTM Attention Model for Claims Extraction from Research Articles through Deep Learning. International Journal of Emerging Multidisciplinaries: Computer Science & Artificial Intelligence, **1(2),** 84-96 (2022).

[21] Villesen, Palle. FaBox: an online toolbox for FASTA sequences. Retrieved August 15, 2020, from https://users-birc.au.dk/~palle/php/fabox/index.php (2009).

[22] Sorathiya, Hardik and Jadeja, Aishwarya and Georrge, John J. An In Silico Approach for Function Prediction of Hypothetical Proteins of *Mycobacterium laprae* TN. Proceedings of the

National Conference on Innovations in Biological Sciences (NCIBS). Available at SSRN: https://ssrn.com/abstract=3574475 (2020).

[23] Feng, Z., *et al*. A Putative Efflux Transporter of the ABC Family, YbhFSR, in *Escherichia coli* Functions in Tetracycline Efflux and $Na^+(Li^+)/H^+$ Transport. *Frontiers in microbiology*, **11, 556**. https://doi.org/10.3389/fmicb.2020.00556 (2020).

[24] Gandlur, SM., Wei, l., Levine, J., Russell, J., Kaur, P. Membrane topology of the DrrB protein of the doxorubicin transporter of Streptomyces peucetius, **279(26),** 27799-27806. https://europepmc.org/article/med/15090538 (2004).

[25] Saleem, T., & Kasi, A. Daunorubicin. In *StatPearls*. StatPearls Publishing (2022).

[26] Edwardson, D. W. *et al*. Role of Drug Metabolism in the Cytotoxicity and Clinical Efficacy of Anthracyclines. *Current drug metabolism*, **16(6),** 412–426. https://doi.org/10.2174/1389200218881509151120 39 (2015).

[27] Davidson, A. L., & Chen, J. ATP-binding cassette transporters in bacteria. *Annual review of biochemistry*, **73,** 241–268. https://doi.org/10.1146/annurev.biochem.73.011303.073626 (2004).

[28] Mehta-Shah, N., Ratner, L., & Horwitz, S. M. Adult T-Cell Leukemia/Lymphoma. *Journal of oncology practice*, **13(8),** 487–492. https://doi.org/10.1200/JOP.2017.021907 (2017).

[29] Gao, Y., *et al*. Antibiotics for cancer treatment: A double-edged sword. *Journal of Cancer*, **11(17),** 5135–5149. https://doi.org/10.7150/jca.47470 (2020).

[30] Guo, D., *et al*. Daunorubicin-Loaded CdTe QDs Conjugated with Anti-CD123 mAbs: A Novel Delivery System for Myelodysplastic Syndromes Treatment. *International journal of nanomedicine*, **15,** 521–536. https://doi.org/10.2147/IJN.S23339 (2020).