Journal of Technology and Innovative Knowledge

Journal Home Page: www.jotik.astu.edu.et

Research Article

In silico analysis of regulatory elements of the *bldD* gene of antibiotic-producing Streptomyces species

Sisay Demisie [1], Muktar Ahmed[2] and Ketema Tafess [1,2,*]

[1]Department of Applied Biology, School of Applied Natural Science, Adama Science and Technology University, Adama P.O.Box 1888, Ethiopia
[2]Institutes of Pharmaceutical Sciences, Adama Science and Technology University, Adama P.O.Box 1888, Ethiopia
* Correspondence: ttafess@gmail.com

Citation: Demisie, S *et al.*, : *In silico* analysis of regulatory elements of the *bldD* gene of antiboic-producing Streptomyces species. *Journal of Technology and Innovative Knowledge*, 2024 (1) (1)

Academic Editor: Dr Seifu Juindi

Received: 1 June 2024
Revised: 8 July 2024
Accepted: 15 July 2024
Published: 22 August 2024

Abstract: Streptomyces are known for their ability to produce a great variety of antibiotics and other bioactive compounds. The production of these molecules is temporally and genetically coordinated with the bacterial morphological changes. These changes are regulated by transcriptional regulators which coincide with antibiotics production. The bldD transcriptional regulator BldD protein is identified as one of the key players in the complex morphogenesis and activator of antibiotic production in Streptomyces species. However, much has remained to be explored about the regulatory mechanisms and level of gene expression of *bldD* gene in different antibiotics producing Streptomyces species. We have identified the single transcription start site (TSS) and the promoters in the upstream coding regions of the *bldD* gene of the 13 antibiotic-producing Streptomyces species. All, 13/13 (100%) of *bldD* genes have a TSS flanking the coding regions. Using the MEME algorithm, five motifs (MtS1- MtS5) are identified, of which Motif 1 (MtS1) has the lowest E-value and the most regulatory motifs for *bldD* genes among the five discovered motifs. In additions; using the TOMTOM web program, we identified 13 transcription factors associated with MtS1. The analysis of the CpG Island of the gene indicated the presence of lower CpG islands. In additions, Phylogenetic tree was constructed and the results showed that the *bldD* genes containing streptomyces species are very closely related to other groups of Streptomyces. Inconclusion, *In silco* analysis of gene and gene product is very crucial to get important insight on the gene and mechanisms that trigger streptomyces to produce bioactive compounds. In additions, understanding of regulatory elements of the gene that encodes the key activator of antibiotic biosynthesis in Streptomyces species paves the way to enhance the laboratory-based experiments for the production antibiotics.

Keywords: Streptomyces; *bldD* gene; regulatory elements; transcriptions factors

1. Introduction

Streptomyces are primarily soil-dwelling bacteria that have a complex developmental life stage during the transition stage from vegetative growth to a reproductive phase. The life stage also associated with the production of specialized aerial hyphae that differentiate into exospores for dispersal through a synchronized septation event (Maria A. Schumacher et al., 2017). Aside from their fascinating developmental life stage, streptomyces are also notable as the most abundant source of antibiotics and other natural products used in medicine. Morphological differentiation in streptomyces begins with the development of a spore, which germinates through the outgrowth of vegetative hyphae followed by production of aerial mycelium, on which mature spores develop (Flärdh & Buttner, 2009; Hao Yan et al., 2020). This morphological differentiation has been found to be dependent on two main classes of genes: the *bldD* genes, which are required for the erection of aerial hyphae and the *whiG* genes, which are needed for the formation of mature spores (Hao Yan et al., 2020) .

The *bldD* gene encodes a small highly charged protein with an apparent ability to regulate negatively its own transcription. This apparent auto regulatory function of the gene product suggests that *bldD* may encode a transcription regulator protein (Elliot, Locke, Galibois, & Leskiw, 2003). These regulators are called Bld ("soon") because of mutations in the relevant genes prevent formation of the reproductive aerial hyphae that give wild-type colonies a fuzzy appearance. In additions, studies showed that a Bld regulator is known to be the master regulator of development and it located at the top of the developmental hierarchy and represses the transcription of more than hundreds of sporulation genes during vegetative growth (McCormick & Flärdh, 2012). The bldD transcriptional regulator BldD activity is controlled only when it shown to bind to and be regulated by the second messenger 3, 5-cyclic diguanylic acid (c-di-GMP).

Genome sequencing has led to the discovery of multiple potential gene clusters engaged in secondary metabolite synthesis in Streptomyces strains. A large gene cluster comprising a cluster-situated regulator (CSR) expressing gene is commonly used to biosynthesize each antibiotic. The pleiotropic regulators keep track of developmental status, food availability, and a variety of stressors before sending signals to the CSR genes, which control antibiotic production (G. Liu, Chater Keith, Chandra, Niu, & Tan, 2013). An antibiotic regulatory network has to be elucidated in order to find new approaches to enhance antibiotic production and arouse cryptic antibiotic synthesis. Various degrees of transcriptional regulators tightly govern the commencement of morphological development, which is often associated with antibiotic production, in response to environmental and physiological changes (Choudoir, Pepe-Ranney, & Buckley, 2018; L. Liu et al., 2019; van Wezel & McDowall, 2011).

Understanding the genes and mechanisms that trigger the production of bioactive compounds in streptomyces is therefore significantly important. Extensive genome analyses have revealed that regulatory factors play a key role in the activation and repression of genes involved in drug production (den Hengst et al., 2010; L. Liu et al., 2019). Nowadays, in silico analysis of gene sequences and their products is becoming a common method for identifying gene expression patterns and sequences responsible for the synthesis of novel drugs. This has also led to the identification of numerous new medicinal products. As a result, several computational tools have been developed to assist researchers in this field. The majority of these tools are based on the in silico study of specific genes and gene products. (Ziemert, Alanjary, & Weber, 2016). Even though previous studies on the mutant gene and the key bldD transcriptional regulator have been conducted, the majority of these investigations have focused on Streptomyces coelicolor. Hence, the present study aimed to analyze the promoter regions and regulatory elements of the *whi* in 13 antibiotic-producing Streptomyces species containing the mutant *bldD* gene.

2. Materials and Methods

2.1. Determination of the Promoter Positions/Transcription Start Site (TSS)

Promoter regions are intrinsic DNA elements located upstream of the coding regions of genes and are required for their transcription by RNA polymerase (RNAP). Early approaches to mapping promoters utilized position weight matrices (PWMs) for the -10 and -35 box motifs, taking into account the distribution of spacer lengths between motifs and their distance from transcription start sites (TSSs) (Cassiano & Silva-Rocha, 2020). Accurate identification of promoters is crucial for studying gene expression in bacteria. In this study, promoters are considered as core elements recognized by the sigma subunit of RNAP. This sigma factor recognizes an approximately -35 bp consensus region with two key elements: the -10 box (consensus motif TATAAT) and the -35 box (TTGACA) (Elliot et al., 2003). Besides the core promoter region, other cis-regulatory elements may also play

a significant role in regulating gene expression (Biłas, Szafran, Hnatuszko-Konka, & Kononowicz, 2016).

In this study, sequences of *bldD* genes from 13 Streptomyces species were retrieved in February 2023 from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/gene/?term=bldD). To determine the TSS positions and values, sequences ≥1 kb upstream of the coding regions of the *bldD* gene were analyzed using the NNPP version 2.2 tool. To determine the Linear Discriminant Function (LDF) values, sequences ≥150 bp upstream of the coding regions of the *bldD* gene were analyzed using the BPROM tool. The TSS and distance from the gene start codon were determined using the NNPP toolset version 2.2 with a minimum standard predictive promoter score and a cutoff value of 0.8 (Wang, Ungar, Tseng, & Hannenhalli, 2007). The LDF values were determined using the BPROM algorithm (Brunet, Bernard, & Cascales, 2020).

2.2 Identification of common motifs and transcription factors (TFs)

Analysis of the common motifs for *bldD* genes was conducted using Multiple Em for Motif Elicitation (MEME) software version 3.5.4 (http://meme-suite.org/tools/meme) with the sequence ≥1 kb upstream of the promoter positions or transcription start sites (TSS) (Bailey et al., 2009; Bailey, Johnson, Grant, & Noble, 2015; Jeba Malar et al., 2020). MEME typically identifies the most statistically significant motifs (with low E-values), where the E-value of a motif is determined based on its log likelihood ratio, width, number of sites, background letter frequencies, and the size of the training set. The search results page was linked to the MEME output in HTML format, and the smallest expected value (E-value) was considered for further analysis. The MEME output for each motif was then sent directly to TOMTOM, a web-based motif comparison program that searches against a database of known motifs (Bailey et al., 2009). For this analysis, the CollecTF (Bacterial TF motifs) and EXPREG databases were used as references for binding motifs. The rank of the primary sequences was compared to all 'ab initio' motifs discovered by Sequence Enrichment Analysis (SEA), and the enrichment p-values were used to determine the motifs' rank. Parameters for detecting motif site distribution were set to zero or one site per sequence (ZOOPS), with a maximum of 13 motifs. The motif E-value threshold was unlimited, the minimum motif width was 5, and the maximum motif width was 50.

2.3 Determinations of Transcription Factors Binding Sites (TFBS)

Sequence motifs are important tools in molecular biology and can identify features in DNA, RNA, and protein sequences, such as transcription factor binding sites, splice sites, and protein-protein interaction sites. Several algorithms have been developed to discover motifs and to search databases for matches to given motifs. Gapped Local Alignment of Motifs (GLAM) is one such specialized algorithm for DNA motif discovery and is also important for identifying functional site motifs. As a result, promoter sequences of the *bldD* gene from thirteen antibiotic-producing Streptomyces species were entered into GLAM2. The GLAM2 HTML output was then used to search for transcription factor binding sites (TFBS) (Frith, Saunders, Kobe, & Bailey, 2008).

2.4 Analysis of CpG Islands

The CpG islands of the genes were determined using two algorithms. The first was the offline tool CLC Genomics Workbench Version 8.5 (https://clc-genomicsworkbench.software.informer.com/8.5/) , which was used to search for restriction enzyme MspI cutting sites with fragment sizes between 40 and 220 bp. The second algorithm was the one developed by Kuo et al. (available at http://dbcat.cgm.ntu.edu.tw/), which uses criteria of GC content greater than or equal to 55 percent and an Observed CpG/Expected CpG ratio of 0.65(Kuo et al., 2011).

2.5 Phylogenetic Tree Construction and Analysis

In addition to the 13 Streptomyces species containing *bldD* gene, 20 related Streptomyces species were collected from NCBI by considering the lowest E-values and aligned using Muscle Multiple Alignment Tools (MMAT). The phylogenetic tree was constructed by using UPGMA methods in MEGA 6.0 platform using aligned sequences from prokaryotes (Kumar, Stecher, & Tamura, 2016; Newman, Duffus, & Lee, 2016). With the help of significant aligned sequences from prokaryotes, the phylogenetic relationship of Streptomyces species contains the *bldD* genes was inferred. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) has been shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood Method and are in the units of the number of base substitutions per site. This analysis involved 34 nucleotide sequences including the random anchor (a stretch of 3108 nucleotides) Kitasatospora setae strain KM-6054 23S. All positions containing gaps and missing data were eliminated (complete deletion option).

3. Results

3.1. Determination of the Promoter Positions/Transcription Start Site (TSS)

Determining the location of the transcription start sites and promoter region in a given gene is vital for the study of the mechanism of gene regulation. The core promoter is a minimum promoter region that is capable of initiating transcription in a gene. It contains a transcription start site (TSS) and typically start from −60 to +40 relative to the TSS (V. V. Solovyev & Shahmuradov, 2003). In this study, we included ≥1kb upstream of the coding regions to locate the transcription start site (TSS) using the NNPP tools set and the LDF was determined using BPROM algorithm. Our study analysis indicated that all streptomyces species (containing bldD genes) (13/13 (100%)) have a single TSS. In another way, the TSS of 11/13 (86.61%) is located in less than 100bp upstream of the start codon of the bldD gene (Table 1). This indicated that the transcriptional regulators of the bldD gene in antibiotic-producing Streptomyces species are located closest to the start codon of the genes. The BPROM program utilizes a LDF to make a prediction based on the characteristics in the -200 to +50 bps region of the TSS (V. Solovyev & Salamov, 1997; V. V. Solovyev & Shahmuradov, 2003; Wang et al., 2007; Ziemert et al., 2016) where higher LDF indicates a high probability of expression of the gene. As a result, 150 bp were included to locate the -10 box positions (highly conserved regions) and -35 box positions (less conserved regions) of the gene. Accordingly, the core promoters of the bldD genes of 58470067, 66853606, 61473082, and [69807580 & 69764486] have an LDF threshold of 2.18, 2.14, 2.04, and 2.03, respectively whilst the core promoter genes of [58431103 & 24306276] have the lowest LDF thresholds of 1.18 and 1.47, respectively (Table 1).

Table 1. The TSS and the distance from gene start codon were determined using the NNPP toolset version 2.2 with the minimum standard predictive promoter score and cut off value of 0.8 and Linear Discriminant Function (LDF) values were determined using BPROM.

Gene Name	Gene ID	Chromosome locations	Number of predicted/TSS	TSS position	-10 box at positions	-35 box at positions	Linear Discriminant Function (LDF) value
bldD	6213046	NC_010572	1	97	82	61	1.99
bldD	24306276	NC_013929.1	1	122	107	86	1.47
bldD	15149186	NC_020990.1	1	122	107	86	1.68
bldD	66853606	NZ_CP048261.1	1	98	83	62	2.14
bldD	63978737	NZ_CP070242.1	1	97	82	61	1.99
bldD	61473082	NZ_CP065253.1	1	89	74	53	2.04
bldD	58431103	NZ_KV757141.1	1	97	82	61	1.18
bldD	58470067	NZ_BBQG01000011.1	1	95	80	60	2.18
bldD	69878388	NZ_CP086102.1	1	97	82	61	1.99
bldD	69863271	NZ_CP018074.1	1	98	83	62	1.87
bldD	69807580	NZ_JAGJBY010000001	1	89	74	53	2.03
bldD	69764486	NZ_CP043317.1	1	87	72	51	2.03
bldD	57807597	NZ_JABSUS0100000.1	1	97	82	61	1.99

3.2. Identification of Common Motifs and Transcription Factors (TFs)

Using MEME software; conserved motifs for bldD genes of 13 antibiotics producing Streptomyces species were analyzed. For each promoter region, five candidate motifs were identified (Table 2). The presence of common motifs that serve as binding sites for transcription factors that affect the expression of the gene was determined. The motif which has the least E-value (MtS1) has been submitted to the TOMTOM. Our analysis showed that the sequence of the 5′ promoter regions share equal (100%) common motif binding sites. All of the identified motifs equally shared the binding site distributions (100%); however, they showed variation based on statistical expectation value (E-value). Besides, the MtS2, MtS5, MtS1, MtS3, and MtS4 contain 19, 18, 17, 16, 13, and 11 binding site which could matches with the motif provided in the database respectively.

Table 2. List of discovered motifs, number of promoter-containing motifs, number of binding sites and total number of binding sites matches the bldD gene via motif provided in motif database.

Discovered motifs	Number of promoter containing motifs	E value	Motif width	Number of motifs binding site	Total number of binding site matches in data base
MtS1	13 (100%)	1.0e-216	50	13	16
MtS2	13 (100%)	1.3e-215	50	13	19
MtS3	13 (100%)	7.7e-209	50	13	13

MtS4	13 (100%)	8.1e-202	50	13	11
MtS5	13 (100%)	1.6e-197	50	13	18

In addition, MEME generated thirteen candidate motifs distributed from the position of TSS (+1) to upstream of ≥1 kb. All candidate motifs were distributed in the positive strand with high frequency of binding sites. The distributions and the binding site of MtS1 range from -200 to -700 upstream of the TSS positions and have high binding sites as well as located closest to the TSS positions. While MtS2 lie in the -500 to -1000 range and they are distant from the TSS positions. Besides, MtS3 lie in the -600 to -1000 range and they are distant from the TSS positions. In addition, 53/65 (81.53%) of the identified motifs were found within the range of +1 to -700. From this study, it is possible to suggest that the transcription regulatory factors BldD bind to the motif closest to the TSS positions and activate antibiotic synthesizing genes (Fig. 1)

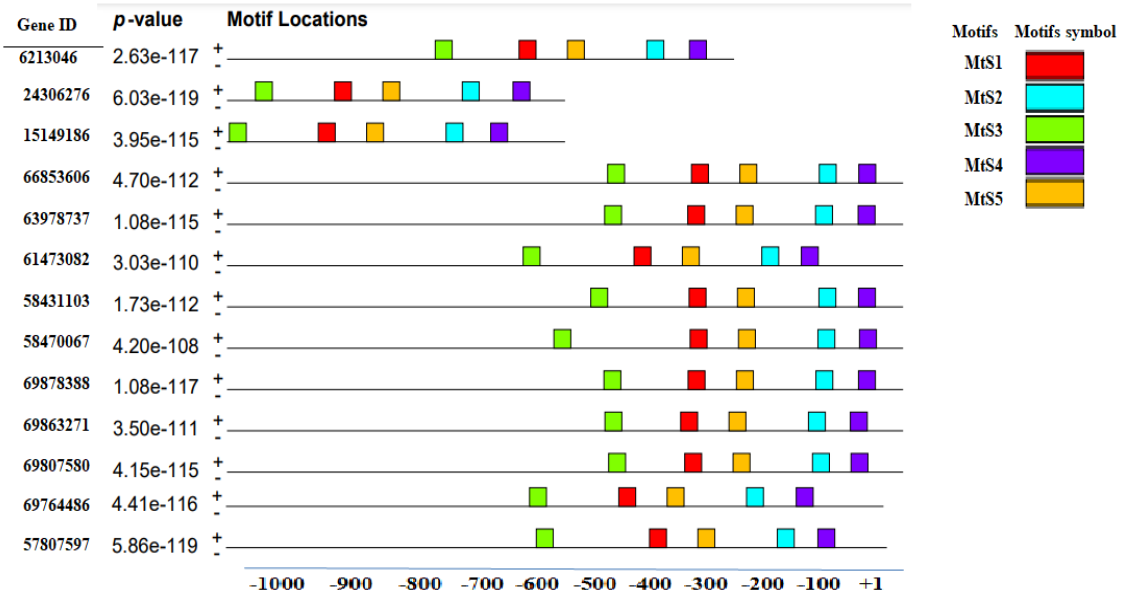

Figure 1. Block diagrams show distribution and location of input promoter sequence of bldD gene candidate motifs.

Transcription factors (TFs) are essential regulatory patterns that control gene expression. Using TOMTOM, we compared the matching MtS1 with the publicly accessible prokaryotic motif database. The analysis results showed numerous matching motifs between MtS1 and the internationally registered motifs. We identified 11 transcription factors associated with MtS1 which includes Putative DNA-binding protein, integrating host factor subunit alpha, RNA polymerase sigma54 factor, positive regulatory protein of alginate biosynthesis, AraC family transcriptional regulator, nucleoid-associated protein EspR, sigma factor PvdS, macrodomain Ter-Protein, RNA polymerase Sigma 70 family protein. The transcription factors play different molecular and biological functions in different groups of organisms (Table 3).

Our study revealed that most of the transcription factors are share a common function in different microorganisms. Notably, the predominant biological function includes DNA-binding transcription activator activity and binding of transcription cis-regulatory region. In addition, positive and negative regulations of transcription and their roles as DNA template are also some of the common feature of transcription factors (Table 3).

Table 3. List of matching candidate transcription factors (TFs) which could bind to common MtS1 and motif GO terms for motif MtS

Organisms Name	**Transcriptions factor/proteins**	**Gene Name**	**Functions**		**E-value**	**Gene expressions database**
			GO – Molecular function	**GO - Biological processes**		
Streptomyces coelicolor A3(2)	Putative DNA-binding protein	SCO1489	DNA-binding transcription repressor activity, Nucleotide binding, Sequence-specific DNA binding & Transcription cis-regulatory region binding	Negative regulation of transcription, DNA-templated	5.84e-02	Collectf/EXPREG_00000fc0
Pseudomonas putida (strain ATCC 47054	Integration host factor subunit alpha	ihfA	DNA-binding transcription activator activity, DNA-binding transcription repressor activity & transcription cis-regulatory region binding	DNA recombination & Regulation of translation	1.20e-01	Collectf/EXPREG_0000006f0
Vibrio cholerae serotype O1 (strain ATCC 39315)	RNA polymerase sigma-54 factor	VC_2529	DNA binding, DNA-binding transcription activator activity, DNA-directed 5'-3' RNA polymerase activity & sigma factor activity	DNA-templated transcription and initiation	2.71e+00	Collectf/EXPREG_0000016e0
Pseudomonas aeruginosa (strain ATCC 15692	Positive alginate biosynthesis regulatory protein	algR	DNA-binding transcription activator activity, DNA-binding transcription repressor activity, Phosphorelay response regulator activity sequence-specific DNA binding & Transcription cis-regulatory region binding	Alginic acid biosynthetic process, Negative regulation of transcription, DNA-templated, Positive regulation of cell motility, Positive regulation of single-species biofilm formation ,Positive regulation of transcription, DNA-templated, Regulation of transcription,	3.44e+00	Collectf/EXPREG_0000009d0
Xanthomonas oryzae pv. oryzae	AraC family transcript	hrpXo	DNA-binding transcription activator	positive regulation of transcription, DNA-templated	4.72e+00	Collectf/EXPREG_0000017f0

	ional regulator		activity & transcription cis-regulatory region binding			
Streptomyces coelicolor (strain ATCC BAA-471	AraC-family transcriptional regulator	SCO2792	DNA-binding transcription factor activity & Sequence-specific DNA binding	Positive regulation of transcription, DNA-templated	4.33e+00	Collectf/EXPREG_000001770
Mycobacterium tuberculosis (strain ATCC 25618 / H37Rv)	Nucleoid-associated protein EspR	espR	DNA binding & Identical protein binding	Regulation of protein secretion , Regulation of transcription, DNA-templated & Response to host immune response	5.32e+00	Collectf/EXPREG_000000c30
Pseudomonas putida (strain ATCC 47054	Integration host factor subunit alpha	ihfA	DNA-binding transcription activator activity, DNA-binding transcription repressor activity & Transcription cis-regulatory region binding	DNA recombination & Regulation of translation	8.00e+00	Collectf/EXPREG_000006f0
Pseudomonas aeruginosa (strain ATCC 15692	Sigma factor PvdS	pvdS	DNA-binding transcription activator activity, Sigma factor activity& Transcription cis-regulatory region binding	Cellular response to iron ion, DNA-templated transcription, initiation, Positive regulation of secondary metabolite biosynthetic process, Positive regulation of transcription, DNA-templated & Regulation of transcription, DNA-templated	8.14e+00	Collectf/EXPREG_000004b0
Escherichia coli (strain K12)	Macrodomain Ter protein	matP	Sequence-specific DNA binding	Chromosome organization, Chromosome segregation & Regulation of transcription, DNA-templated	8.19e+00	Collectf/EXPREG_000007b0
Pseudomonas syringae pv. tomato (strain ATCC BAA-871/ DC3000)	RNA polymerase sigma-70 family protein	PSPTO_2133	DNA-binding transcription activator activity, sigma factor activity & transcription cis-regulatory region binding			

3.3. Determinations of Transcription Factors Binding Sites (TFBS)

Transcription Factor Binding Sites (TFBS) are also crucial for understanding gene expression regulations (Vahed, Vahed, & Garmire, 2022). Thirteen antibiotic-producing Streptomyces species *bldD* gene promoter sequences were entered into GLAM2 and GLAM2 HTML output was buttoned and searched for TFBS. As depicted in Fig.2, 49 nucleotide base pairs TFBS were identified. Alignment was also conducted to check the presence of deletion and insertion among the *bldD* gene of 13 antibiotic-producing Streptomyces. Consequently, the aligned columns have no deletion or insertion and *bldD* genes of antibiotic-producing Streptomyces were ungapped. In addition, GLAM2 analysis showed that 8/13 (61.53%) *bldD* gene indicates a high marginal value, 91.6. These species have a strong motif and better matches to the overall motifs, suggesting a high transcription binding site with the key transcription regulatory factor BldD. In contrast, bldD 69863271 showed a lower marginal value, 75.2, suggesting the species has a weak motif and fewer matches to the overall motifs, as well as signifying the lower gene expression level (Figure 2).

Best Motif Found

Gene ID	Start	SITES	End	strand	MS
6213046	1243	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1292	+	91.6
24306276	697	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	746	+	91.6
15149186	649	aggaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	698	+	86.9
66853606	1753	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1802	+	91.6
63978737	1742	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1791	+	91.6
61473082	1583	agcaccacgtcgtcgggcgcccttggcgcgctcgacgatctcgtgggcgat	1632	+	89.6
58431103	1752	agcaccacgtcgtcgggcgcccttggcgcgctccacgatctcgtgggcgat	1801	+	84.8
58470067	1749	agcaccacgtcgccgggcgcccttggcgcgttcgacgatctcgtgggcgat	1798	+	86.7
69878388	1743	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1792	+	91.6
69863271	1721	agaaccacgtcgtctgcgcccttggcgcgctcgacgatctcgtgtgcgat	1770	+	75.2
69807580	1733	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1782	+	91.6
69764486	1541	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1590	+	91.6
57807597	1632	agcaccacgtcgtcgggcgcccttggcgcgttcgacgatctcgtgggcgat	1681	+	91.6

Score: 904.293

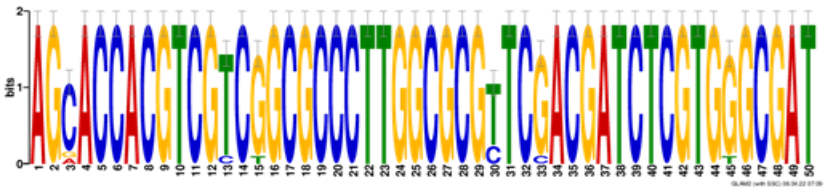

Figure 2. The TFBS and a sample logo for GLAM2 gapped motif respectively.

3.4 Comparisons of the Candidate Motifs to the Database Motifs

All (13) candidate motifs were compared to the motifs in the motif database (Collectf and EXPREG). Our studies showed that all candidate motifs 13/13 (100%) share the same TFBS with VqsM_*P.aeruginosa* and RpoN_*V.cholerae*. Additionally, of the 13 candidate motifs, 9/13 (69.23%) share the same TFBS with PhoP *Y.pestis*, and 52% are activated and 47% are repressed (Table 4).

Table 4. The candidate motifs in the collectf and EXPREG databases match the sequence enriched motif of with E-values ≤ 10.

Alternate name for the motif provided in the motif database file	Regulatory Mode				E value	Number of primary sequences matching the motif	Motif database file
	Activ. (%)	Rep. (%)	Dual (%)	NS (%)			
PhoP_*Y.pestis*	52	47	0	0	1.92e-2	9 / 13 (69.2%)	EXPREG_00000050
IHF_*P.putida*	55	45	0	0	6.92e-2	8 / 13 (61.5%)	EXPREG_00000700
ArgR_*P.aeruginosa*	61	27	0	11	1.01e-1	11 / 13 (84.6%)	EXPREG_00000470
Fur_*V.cholerae*	0	100	0	0	1.01e-1	11 / 13 (84.6%)	EXPREG_000008b0
ToxT_*V.cholerae*	81	18	0	0	2.19e-1	7 / 13 (53.8%)	EXPREG_00000240
OmpR_*Y.pestis*	100	0	0	0	2.19e-1	7 / 13 (53.8%)	EXPREG_00001000
CcpA_*S.suis*	17	25	0	57	2.19e-1	7 / 13 (53.8%)	EXPREG_00001810
CRP_*V.vulnificus*	100	0	0	0	3.03e-1	12 / 13 (92.3%)	EXPREG_00001030
Fur_*N.gonorrhoeae*	60	10	0	30	4.05e-1	10 / 13 (76.9%)	EXPREG_00000ec0
VqsM_*P.aeruginosa*	7	0	0	92	4.38e-1	13 / 13 (100.0%)	EXPREG_00001670
RpoN_*V.cholerae*	100	0	0	0	4.38e-1	13 / 13 (100.0%)	EXPREG_000016e0
Lrp_*E.coli*	1	1	0	97	4.54e-1	12 / 13 (92.3%)	EXPREG_00000840
Zur_*N.meningitidis*	15	84	0	0	1.42e0	10 / 13 (76.9%)	EXPREG_000016a0
IHF_*P.putida*	100	0	0	0	1.93e0	9 / 13 (69.2%)	EXPREG_000006f0
PvdS_*P.aeruginosa*	100	0	0	0	3.44e0	8 / 13 (61.5%)	EXPREG_000004b0
Fur_*A.ferrooxidans*	0	63	0	36	4.02e0	4 / 13 (30.8%)	EXPREG_00000370
Vfr_*P.aeruginosa*	41	11	0	47	4.02e0	4 / 13 (30.8%)	EXPREG_00000b50
PhhR_*P.putida*	90	10	0	0	4.02e0	4 / 13 (30.8%)	EXPREG_00001190
Fur_*P.aeruginosa*	0	100	0	0	5.11e0	11 / 13 (84.6%)	EXPREG_00000c80
CsgD_*E.coli*	33	22	0	44	5.81e0	9 / 13 (69.2%)	EXPREG_00000b00
LexA_*P.difficile*	0	0	0	100	6.02e0	6 / 13 (46.2%)	EXPREG_00000120
CRP_*E.coli*	82	17	0	0	6.02e0	6 / 13 (46.2%)	EXPREG_00000850
H-NS_*V.cholerae*	0	100	0	0	6.38e0	13 / 13 (100.0%)	EXPREG_00001730
CcpA_*C.difficile*	9	36	0	53	6.73e0	5 / 13 (38.5%)	EXPREG_00000d10
LasR_*P.aeruginosa*	98	1	0	0	9.24e0	3 / 13 (23.1%)	EXPREG_000009b0
OxyR_*P.aeruginosa*	3	0	0	96	9.24e0	3 / 13 (23.1%)	EXPREG_00001560
AdpA_S.coelicolor	100	0	0	0	9.40e0	9 / 13 (69.2%)	EXPREG_00001770

Note: Activ: activations, Rep: repression NS: non specified, IHF: integrated host factors, ArgR: arginine responsive regulators, Fur: ferric uptake regulators, OmpR: Outer Membrane Proteins regulators, CcpA: catabolite control protein A, CRP: Cyclic AMP-cAMP receptor protein, VqsM: Virulence and quorum sensing modulator protein, RpoN: RNA polymerase sigma-54 factor, Lrp: leucine-responsive regulatory protein, Zur: Zinc uptake regulator, PvdS: siderophore pyoverdine, Vfr: virulence factor regulator, PhhR: phenylalanine hydroxylase regulators, CsgD: Curlin subunit gene D, H-NS: Histone-Like Nucleoid Structuring Protein, OxyR: oxygen regulators.

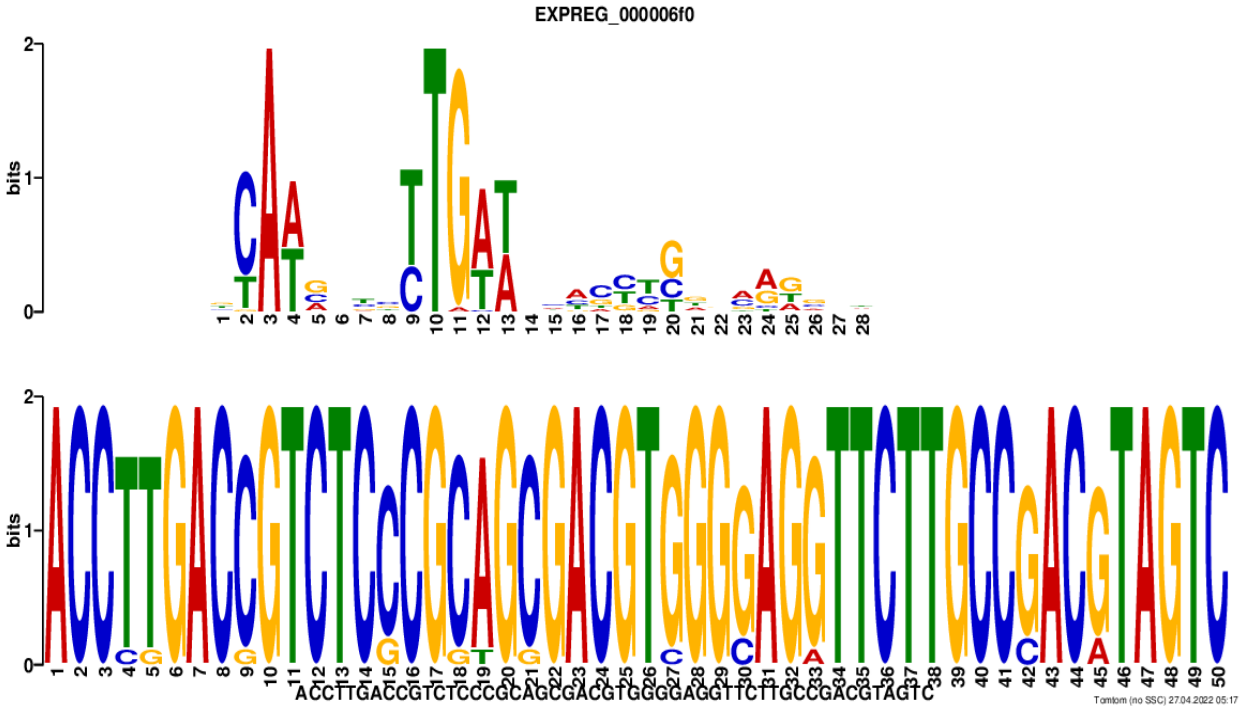

Figure 3. Exemplification logos for the optimal comparison of IHF_*P.putida* sequence and MtS I of the target motifs with the detected motif in the database.

3.5. Analysis of CpG Islands

Two techniques were used for the analysis of the CpG Island: The first is the offline tool CLC Genomics Workbench Version 8.5, with which the restriction enzyme sites MspI with fragment sizes between 40 and 220 bp were searched for parameters. Accordingly, the result revealed that among the 13 Streptomyces species containing *bldD* genes, only one species 1/13 (7.69%) (GI: 69878388) has a single cleavage site; whereas the remains have multiple cleavage sites (Table 5).

Table 5. Identification of MSpI cutting sites and fragment sizes (40 and 220) for bldD gene of streptomyces species.

Gene ID	Nucleotide positions of MspI enzyme cutting sites	Fragment size (40-220)
6213046	Multiple cut (29, 64, 78, 115, 243, 402, 434, 443, 467, 686, 819, 849, 876, 890, 1109, 1121, 1205, 1269, 1278, 1309, 1313, 1322, 1350, 1534, 1569, 1588, 1619, 1670, 1676, 1684, 1713, 1826, 1841)	128,159,219,133,219,84,64,184,51,113
24306276	Multiple cut (19, 55, 100, 112, 384, 719, 729, 810, 840, 854, 867, 881, 1100, 1112, 1206, 1432, 1523, 1577, 1608, 1618, 1659, 1669, 1673, 1702, 1847, 1976, 1987, 1992)	45,81,219,94,91,54,145,129
15149186	Multiple cut (49, 130, 143, 175, 184, 265, 427, 460, 467, 488, 569, 599, 613, 626, 784, 790, 859, 871, 973, 1013, 1035, 1042, 1050, 1057, 1072, 1099, 1136, 1143, 1207, 1282, 1364, 1418, 1432, 1444, 1471, 1500, 1514, 1633, 1645, 1782, 1799, 1804, 1809, 1832, 1861, 1945, 1959, 1987, 1992)	81,81,162,81,158,69,102,40,64,75,82,54,119,137,84
66853606	Multiple cut (12, 29, 53, 68, 343, 368, 627, 660, 667, 688, 799, 813, 826, 1059, 1071, 1155, 1212, 1230, 1253, 1262, 1527, 1612, 1663, 1670, 1674, 1700, 1715, 1825)	111,84,57,85,51

63978737	Multiple cut (17, 53, 95, 110, 382, 727, 747, 808, 838, 852, 865, 879, 1098, 1205, 1433, 1524, 1578, 1609, 1660, 1670, 1674, 1724, 1977, 1987, 1992)	42,61,219,107,91,54,51,50
61473082	Multiple cut (20, 56, 113, 291, 385, 447, 590, 669, 730, 738, 811, 841, 868, 1026, 1032, 1068, 1101, 1113, 1255, 1273, 1278, 1294, 1502, 1556, 1587, 1609, 1638, 1652, 1681, 1794, 1963, 1977)	57,178,94,62,143,79,61,73,158,142, 54,113,169
58431103	Multiple cut (26, 62, 234, 302, 396, 421, 601, 680, 720, 741, 761, 822, 852, 893, 1112, 1124, 1245, 1302, 1423, 1432, 1514, 1568, 1650, 1696, 1928)	172,68,94,180,79,40,61,41,219,57,1 21,82,54,82,46
58470067	Multiple cut (167, 228, 248, 309, 339, 353, 366, 380, 599, 706, 762, 1024, 1078, 1109, 1160, 1170, 1174, 1203)	61,61,219,107,56,54,51
69878388	Single cut (34, 67, 79, 474, 478, 513, 532, 614, 992)	82
69863271	Multiple cut (2, 137, 144, 203, 223, 384, 430, 562, 566, 711, 757, 767, 865, 899, 933, 940, 956, 966, 972, 988)	135,59,161,46,132,145,46,98
69807580	Multiple cut (15, 88, 109, 149, 244, 576, 660, 690, 704, 731, 875, 950, 1132, 1187, 1207, 1299, 1530, 1585, 1637, 1680, 1712, 1761, 1822, 1882)	73,40,95,84,144,75,182,55,92,55,52, 43,49,61,60
69764486	Multiple cut (31, 71, 166, 206, 370, 482, 510, 591, 621, 648, 662, 812, 881, 893, 1029, 1101, 1110, 1231, 1322, 1376, 1458, 1504, 1633, 1849)	40,95,40,164,112,81,150,69,136,72, 121,91,54,82,46,129,216
57807597	Multiple cut (185, 217, 502, 530, 611, 655, 668, 682, 901, 913, 1031, 1065, 1149, 1153, 1183, 1367, 1413, 1448, 1467, 1549, 1563, 1592, 1876, 1881, 1892)	81,44,219,118,84,184,46,82

The other algorithm is Takai and Jones, and the possible CpG island regions and CpG island density are shown in (Fig. 4). Our study revealed that only 1 putative CpG Island was detected for each species gene sequence. However, the percentage of GC content varies among species. Consequently, the GC content of the gene's ranges is from 68 to 73%. In additions, the Streptomyces species GI; 61473082 & 24306276 contains 73 and 68 GC percentage and it revealed the highest and lowest GC percentage respectively (Fig.4).

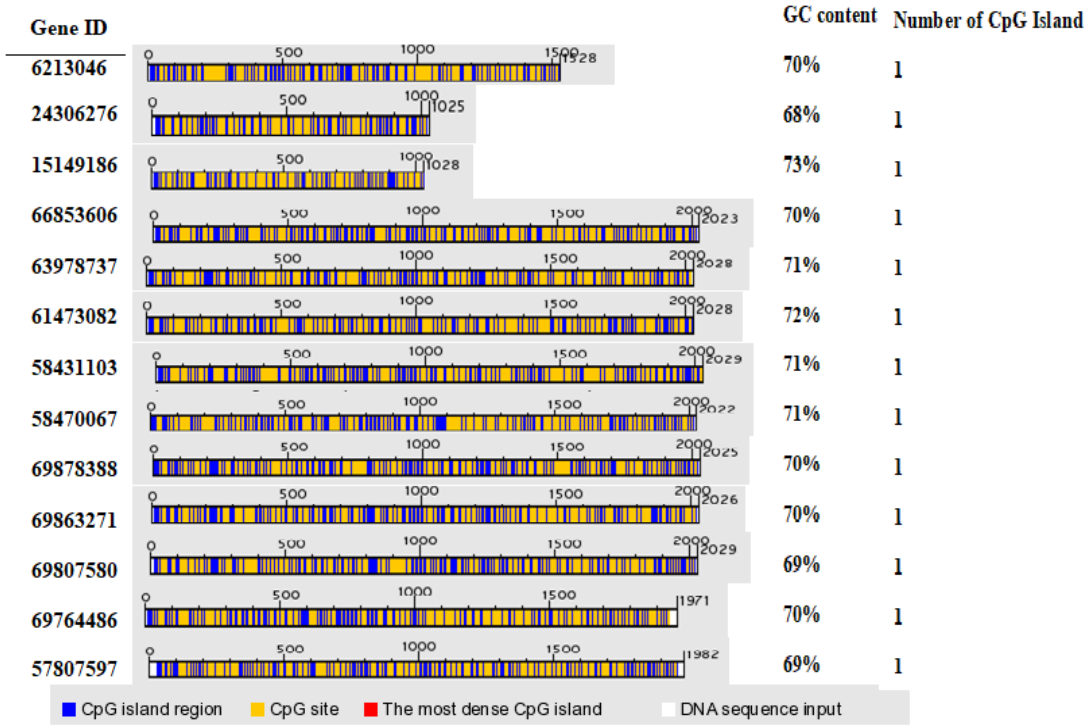

Figure 4. The Possible CpG islands and GC % of the *bldD* gene promoter regions.

3.6. Phylogenetic Tree Construction and Analysis

The nucleotide sequences of the *bldD* gene from 13 antibiotic-producing Streptomyces species and 20 related Streptomyces species were combined, aligned, and used to construct a family tree. Four main criteria were applied for reading, comparing, and interpreting species relationships and divergences: the distance between branch tips, the number of nodes between species, the comparison of time with common ancestors, and the identification of common monophyletic groups. A random anchor, a stretch of 3108 nucleotides from *Kitasatospora setae* strain KM-6054 23S, was used to measure the distance between antibiotic-producing and other Streptomyces species. The combined analysis produced a single significant cladogram, which identified ten clusters and two clades. Consequently, our analysis showed that the Streptomyces species containing *bldD* genes that produce antibiotics fall into clusters II, IV, V, VI, VII, and X (Fig. 5

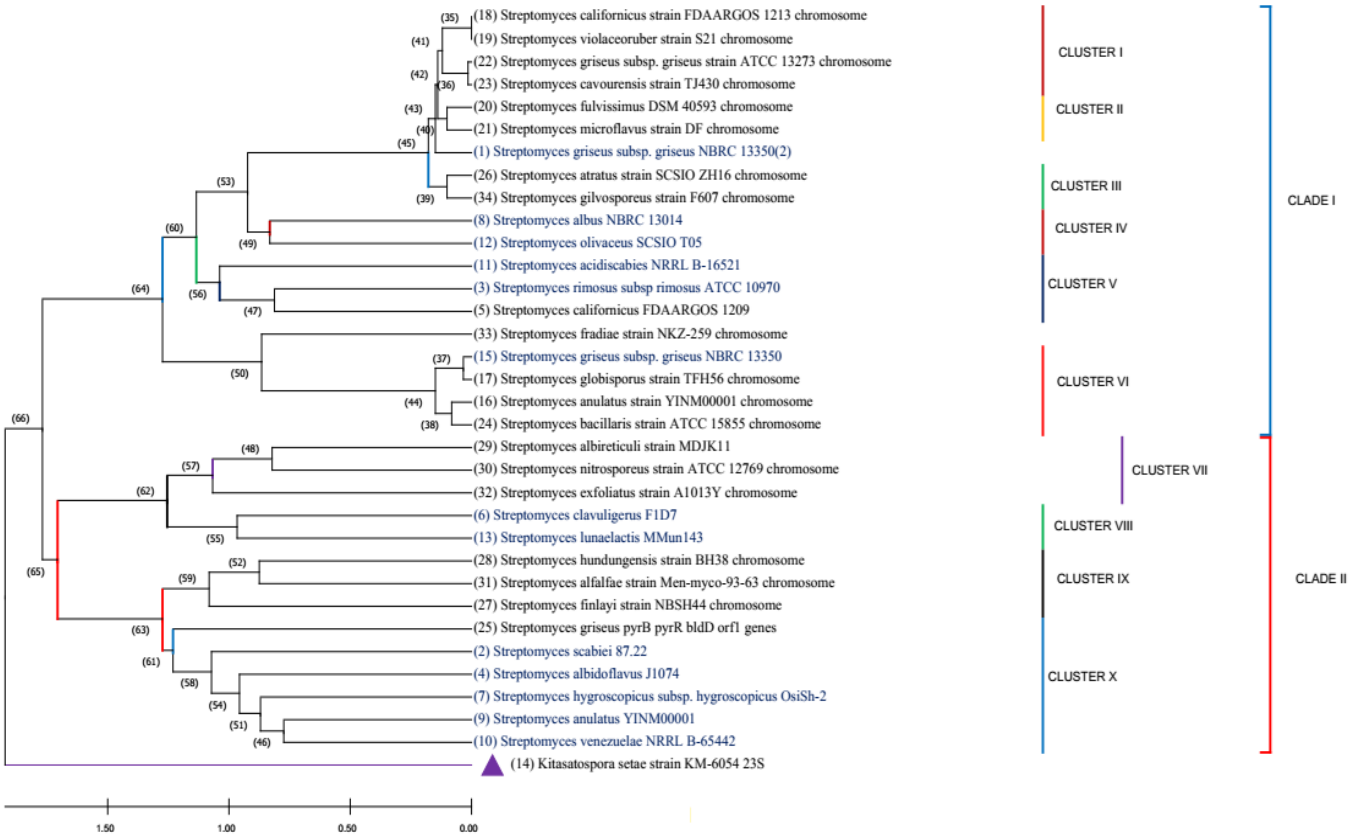

Figure 5. The *bldD* gene family tree of Streptomyces species and other related Streptomyces species.

4. Discussion

The genus *Streptomyces* is a crucial source of bioactive compounds. Several regulatory proteins are critical in activating or repressing antibiotic biosynthetic genes, with BldD functioning at the top of the regulatory cascade that controls antibiotic production in *Streptomyces* bacteria (M. A. Schumacher et al., 2018). Additionally, laboratory-based in vitro studies have highlighted the significant role of BldD in regulating antibiotic production(H. Yan et al., 2020) . In *Streptomyces coelicolor*, BldD is a key transcriptional regulator required for both morphological development and antibiotic synthesis. In silico analysis of the transcriptional regulatory elements of the *bldD* gene in antibiotic-producing *Streptomyces* species can enhance our understanding of drug development and facilitate the implementation of laboratory-based in vitro experiments.

In this study, we used the NNPP and BPROM web-based programs to identify the promoter region and TSS, which are located upstream of the coding regions of the *bldD* gene. Our analysis showed that the *bldD* gene in all antibiotic-producing *Streptomyces* species examined has a single TSS, situated close to the start codon (ATG). This finding aligns with (Lee et al., 2022), who reported that most TSSs in *Streptomyces* are found within 5–100 bp upstream of the start codon. This suggests that the transcriptional regulatory factor BldD regulates the *bldD* gene in antibiotic-producing *Streptomyces* species very close to the start codon (ATG). Our results are also consistent with (Jeong et al., 2016),who found that 68 TSSs mapped to 18 of the 28 secondary metabolic gene clusters in the *S. coelicolor* genome, with an average of 1 TSS for every 2.3 protein-coding genes. TSSs located 500 bp upstream to 150 bp downstream of the annotated start codon of each ORF have been classified as primary (P) or secondary (S) TSS (Lee et al., 2022). The regulation of gene

expression at the transcriptional level is a fundamental process in all biological systems (Romero-Rodríguez, Robledo-Casados, & Sánchez, 2015; Sun, Liu, Zhu, & Liu, 2017) .

Transcription factors are proteins that bind to DNA regulatory sequences (enhancers and silencers) typically located in the 5' upstream region of target genes to modulate gene transcription. This interaction can lead to increased or decreased gene transcription, altered protein synthesis, and subsequent changes in cellular function. In promoter regions, transcription factors attach to short DNA sequence motifs known as binding sites. Position-specific scoring matrices (PSSMs) represent various binding sites recognized by the same transcription factor as a single consensus sequence (Vahed et al., 2022). These matrices indicate the probability of a particular nucleotide at a specific site and can be visualized using logo representations. Bacterial transcriptional regulators are classified into approximately 50 families based on sequence alignment, structural, and functional criteria (Romero-Rodríguez et al., 2015). We identified five significant motifs in the promoter regions of the *bldD* gene. Among these, Motif 1 (MtS1) has the lowest E-value and represents a key regulatory motif for the gene.

Comparative analysis of the motif (MtS1) with known prokaryotic motif databases revealed its association with 11 transcription factors (TFs), including the putative DNA-binding protein, integrating host factor subunit alpha, RNA polymerase sigma 54 factor, positive regulatory protein of alginate biosynthesis, AraC family transcription regulator, nucleoid-associated protein EspR, sigma factor PvdS, macrodomain Ter protein, and RNA polymerase Sigma 70 family protein. These TFs play roles in metabolism, virulence, pathogenesis, replication, and the regulation of various transcriptional processes (Table 3). Similar findings have been reported in other bacterial species (den Hengst et al., 2010; Maria A. Schumacher et al., 2017), suggesting a conserved nature of these TFs across prokaryotes. The identification of TFs associated with metabolism and regulation highlights the potential role of the identified motif in the regulation and activation of secondary metabolites in *Streptomyces*.

Consolidating our results, (Sun, Zhu, Chen, Li, & Wen, 2016) identified a novel AraC-family transcriptional regulator, SAV742, as a global regulator that negatively controls avermectin biosynthesis and cell growth in *Streptomyces avermitilis*, while positively regulating morphological differentiation. Additionally, AraC family members have been reported as key transcription factors in *Streptomyces*, involved in controlling genes related to carbon source utilization, morphological differentiation, secondary metabolism, pathogenesis, and stress responses (Romero-Rodríguez et al., 2015). Fang et al. (Fang et al., 2018) also highlighted the regulatory role of the AraC-family transcriptional regulator BfvR (YPO1737 in strain CO92) in biofilm formation and virulence in *Yersinia pestis* biovar Microtus.

Recently, nucleoid-associated proteins have also been found to influence the expression of specialized metabolic clusters(Gordon et al., 2010; Zhang, Andres, & Elliot, 2021). Leucine-responsive regulatory protein2 (Lsr2) is a small nucleoid-associated protein found throughout the actinobacteria having similarly role to the well-studied Histone-like nucleoid structuring protein (H-NS), it preferentially binds AT-rich sequences and represses gene expression (Chen et al., 2008). In *Streptomyces venezuelae*, Lsr2 represses the expression of many specialized metabolic clusters, including the chloramphenicol antibiotic biosynthetic gene cluster, and deleting lsr2 leads to significant upregulation of chloramphenicol cluster expression. Bacteria, including *Streptomyces*, use protein ADP-ribosylation. Lalić *et al*. (Lalić et al., 2016) characterized the macrodomain protein SCO6735 from *S. coelicolor*, which hydrolyzes PARP-dependent ADP-ribosylation. This protein's

expression is induced by DNA damage, and its deletion leads to increased antibiotic production in S. coelicolor

CpG islands were studied in relation to the *bldD* gene, revealing that these islands are located ≥ 2 kb upstream of the gene's coding regions, as shown in Figure 4. CpG islands are DNA regions longer than 200 bp with at least 50% G+C content and 60% of the expected CpG dinucleotides. They differ from non-CpG island promoters in their transcription regulation. Non-CpG island promoters are repressed by cytosine methylation at CpG dinucleotides, which can directly inhibit transcription by blocking transcription factor binding(Blackledge & Klose, 2011; Yin et al., 2017). According to our study, the bldD gene in each Streptomyces species contains CpG islands at a considerable distance from the start codon. This suggests that the gene has a high potential for antibiotic production. However, the GC content varies among different Streptomyces species containing the bldD gene, ranging from 68% to 73%. Additionally, as shown in the cladogram in Figure 5, the Streptomyces species with the bldD gene are closely related to other groups of Streptomyces

5. Conclusions

Gene mining and in silico analysis are crucial for predicting gene expression patterns and identifying genes responsible for drug synthesis using bioinformatics tools. Our study investigated the *bldD* gene, which has a single TSS near its start codon. We also identified matching transcription factors for the key motif MtS1. Phylogenetic analysis of Streptomyces species with the *bldD* gene, shown in the cladogram, highlights their close evolutionary relationships. This computational study provides a foundation for further wet lab experiments to develop essential antibiotics

Data Availability

The data was extracted from NCBI and can be obtained from the corresponding author

Conflicts of Interest

The authors declare that they have no competing interests.

Funding statement

No funding was received for conducting this research

Acknowledgments

The authors acknowledge research square for a previously published preprint (Preprint Published online May 18th,2022:1-17 available https://assets.researchsquare.com/files/rs-1637072/v1/3ac769d8-6623-4ec8-8043-0633638aca5a.pdf?c=1657196069

Abbreviations

bldD-Bald gene

BPROM-Bacterial promoter

CpG-Cytosine phosphate guanine

NCBI- National Center for Biotechnology Information.

NNPP- Neural Network Promoter Prediction

TFs- Transcriptional factors

TSS-Transcriptional start site

Consent

No consent was necessary

Authors' contributions

SD designed, performed the experiment, analyzed the data, and wrote the manuscript. KT analyzed the data and edited the manuscript. The authors have read and approved the final manuscript.

References

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., . . . Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res, 37*(Web Server issue), W202-208. doi:10.1093/nar/gkp335

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res, 43*(W1), W39-49. doi:10.1093/nar/gkv416

Biłas, R., Szafran, K., Hnatuszko-Konka, K., & Kononowicz, A. K. (2016). Cis-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue and Organ Culture (PCTOC), 127*(2), 269-287. doi:10.1007/s11240-016-1057-7

Blackledge, N. P., & Klose, R. (2011). CpG island chromatin: a platform for gene regulation. *Epigenetics, 6*(2), 147-152. doi:10.4161/epi.6.2.13640

Brunet, Y. R., Bernard, C. S., & Cascales, E. (2020). Fur-Dam Regulatory Interplay at an Internal Promoter of the Enteroaggregative Escherichia coli Type VI Secretion sci1 Gene Cluster. *J Bacteriol, 202*(10). doi:10.1128/jb.00075-20

Cassiano, M. H. A., & Silva-Rocha, R. (2020). Benchmarking Bacterial Promoter Prediction Tools: Potentialities and Limitations. *mSystems, 5*(4). doi:10.1128/mSystems.00439-20

Chen, J. M., Ren, H., Shaw, J. E., Wang, Y. J., Li, M., Leung, A. S., . . . Liu, J. (2008). Lsr2 of Mycobacterium tuberculosis is a DNA-bridging protein. *Nucleic Acids Res, 36*(7), 2123-2135. doi:10.1093/nar/gkm1162

Choudoir, M. J., Pepe-Ranney, C., & Buckley, D. H. (2018). Diversification of Secondary Metabolite Biosynthetic Gene Clusters Coincides with Lineage Divergence in Streptomyces. *Antibiotics, 7*(1), 12.

den Hengst, C. D., Tran, N. T., Bibb, M. J., Chandra, G., Leskiw, B. K., & Buttner, M. J. (2010). Genes essential for morphological development and antibiotic production in Streptomyces coelicolor are targets of BldD during vegetative growth. *Mol Microbiol, 78*(2), 361-379. doi:10.1111/j.1365-2958.2010.07338.x

Elliot, M. A., Locke, T. R., Galibois, C. M., & Leskiw, B. K. (2003). BldD from Streptomyces coelicolor is a non-essential global regulator that binds its own promoter as a dimer. *FEMS Microbiology Letters, 225*(1), 35-40. doi:https://doi.org/10.1016/S0378-1097(03)00474-9

Fang, H., Liu, L., Zhang, Y., Yang, H., Yan, Y., Ding, X., . . . Yang, R. (2018). BfvR, an AraC-Family Regulator, Controls Biofilm Formation and pH6 Antigen Production in Opposite Ways in Yersinia pestis Biovar Microtus. *Front Cell Infect Microbiol, 8*, 347. doi:10.3389/fcimb.2018.00347

Flärdh, K., & Buttner, M. J. (2009). Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology, 7*(1), 36-49. doi:10.1038/nrmicro1968

Frith, M. C., Saunders, N. F., Kobe, B., & Bailey, T. L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol, 4*(4), e1000071. doi:10.1371/journal.pcbi.1000071

Gordon, B. R., Li, Y., Wang, L., Sintsova, A., van Bakel, H., Tian, S., . . . Liu, J. (2010). Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A, 107*(11), 5154-5159. doi:10.1073/pnas.0913551107

Jeba Malar, T. R. J., Antonyswamy, J., Vijayaraghavan, P., Ock Kim, Y., Al-Ghamdi, A. A., Elshikh, M. S., . . . Kim, H. J. (2020). In-vitro phytochemical and pharmacological bio-efficacy studies on Azadirachta indica A. Juss and Melia azedarach Linn for anticancer activity. *Saudi J Biol Sci, 27*(2), 682-688. doi:10.1016/j.sjbs.2019.11.024

Jeong, Y., Kim, J. N., Kim, M. W., Bucca, G., Cho, S., Yoon, Y. J., . . . Cho, B. K. (2016). The dynamic transcriptional and translational landscape of the model antibiotic producer Streptomyces coelicolor A3(2). *Nat Commun, 7*, 11605. doi:10.1038/ncomms11605

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol, 33*(7), 1870-1874. doi:10.1093/molbev/msw054

Kuo, H. C., Lin, P. Y., Chung, T. C., Chao, C. M., Lai, L. C., Tsai, M. H., & Chuang, E. Y. (2011). DBCAT: database of CpG islands and analytical tools for identifying comprehensive methylation profiles in cancer cells. *J Comput Biol, 18*(8), 1013-1017. doi:10.1089/cmb.2010.0038

Lalić, J., Posavec Marjanović, M., Palazzo, L., Perina, D., Sabljić, I., Žaja, R., . . . Ahel, I. (2016). Disruption of Macrodomain Protein SCO6735 Increases Antibiotic Production in Streptomyces coelicolor. *J Biol Chem, 291*(44), 23175-23187. doi:10.1074/jbc.M116.721894

Lee, Y., Lee, N., Hwang, S., Kim, W., Cho, S., Palsson, B. O., & Cho, B.-K. (2022). Genome-scale analysis of genetic regulatory elements in Streptomyces avermitilis MA-4680 using transcript boundary information. *BMC Genomics, 23*(1), 68. doi:10.1186/s12864-022-08314-0

Liu, G., Chater Keith, F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular Regulation of Antibiotic Biosynthesis in Streptomyces. *Microbiology and Molecular Biology Reviews, 77*(1), 112-143. doi:10.1128/mmbr.00054-12

Liu, L., Cheng, Y., Lyu, M., Zhao, X., Wen, Y., Li, J., & Chen, Z. (2019). AveI, an AtrA homolog of Streptomyces avermitilis, controls avermectin and oligomycin production, melanogenesis, and morphological differentiation. *Appl Microbiol Biotechnol, 103*(20), 8459-8472. doi:10.1007/s00253-019-10062-3

McCormick, J. R., & Flärdh, K. (2012). Signals and regulators that govern Streptomyces development. *FEMS Microbiology Reviews, 36*(1), 206-231. doi:10.1111/j.1574-6976.2011.00317.x

Newman, L., Duffus, A. L. J., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher, 78*(7), 608-612.

Romero-Rodríguez, A., Robledo-Casados, I., & Sánchez, S. (2015). An overview on transcriptional regulators in Streptomyces. *Biochim Biophys Acta, 1849*(8), 1017-1039. doi:10.1016/j.bbagrm.2015.06.007

Schumacher, M. A., den Hengst, C. D., Bush, M. J., Le, T. B. K., Tran, N. T., Chandra, G., . . . Buttner, M. J. (2018). The MerR-like protein BldC binds DNA direct repeats as cooperative multimers to regulate Streptomyces development. *Nat Commun, 9*(1), 1139. doi:10.1038/s41467-018-03576-3

Schumacher, M. A., Zeng, W., Findlay, K. C., Buttner, M. J., Brennan, R. G., & Tschowri, N. (2017). The Streptomyces master regulator BldD binds c-di-GMP sequentially to create a functional BldD2-(c-di-GMP)4 complex. *Nucleic Acids Research, 45*(11), 6923-6933. doi:10.1093/nar/gkx287

Solovyev, V., & Salamov, A. (1997). The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proc Int Conf Intell Syst Mol Biol, 5*, 294-302.

Solovyev, V. V., & Shahmuradov, I. A. (2003). PromH: Promoters identification using orthologous genomic sequences. *Nucleic Acids Res, 31*(13), 3540-3545. doi:10.1093/nar/gkg525

Sun, D., Liu, C., Zhu, J., & Liu, W. (2017). Connecting Metabolic Pathways: Sigma Factors in Streptomyces spp. *Front Microbiol, 8*, 2546. doi:10.3389/fmicb.2017.02546

Sun, D., Zhu, J., Chen, Z., Li, J., & Wen, Y. (2016). SAV742, a Novel AraC-Family Regulator from Streptomyces avermitilis, Controls Avermectin Biosynthesis, Cell Growth and Development. *Sci Rep, 6*, 36915. doi:10.1038/srep36915

Vahed, M., Vahed, M., & Garmire, L. X. (2022). BML: a versatile web server for bipartite motif discovery. *Brief Bioinform, 23*(1). doi:10.1093/bib/bbab536

van Wezel, G. P., & McDowall, K. J. (2011). The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. *Nat Prod Rep, 28*(7), 1311-1333. doi:10.1039/c1np00003a

Wang, J., Ungar, L. H., Tseng, H., & Hannenhalli, S. (2007). MetaProm: a neural network based meta-predictor for alternative human promoter prediction. *BMC Genomics, 8*, 374. doi:10.1186/1471-2164-8-374

Yan, H., Lu, X., Sun, D., Zhuang, S., Chen, Q., Chen, Z., . . . Wen, Y. (2020). BldD, a master developmental repressor, activates antibiotic production in two Streptomyces species. *Molecular Microbiology, 113*(1), 123-142. doi:https://doi.org/10.1111/mmi.14405

Yan, H., Lu, X., Sun, D., Zhuang, S., Chen, Q., Chen, Z., . . . Wen, Y. (2020). BldD, a master developmental repressor, activates antibiotic production in two Streptomyces species. *Mol Microbiol, 113*(1), 123-142. doi:10.1111/mmi.14405

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., . . . Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science, 356*(6337), eaaj2239. doi:doi:10.1126/science.aaj2239

Zhang, X., Andres, S. N., & Elliot, M. A. (2021). Interplay between Nucleoid-Associated Proteins and Transcription Factors in Controlling Specialized Metabolism in Streptomyces. *mBio, 12*(4), e0107721. doi:10.1128/mBio.01077-21

Ziemert, N., Alanjary, M., & Weber, T. (2016). The evolution of genome mining in microbes - a review. *Nat Prod Rep, 33*(8), 988-1005. doi:10.1039/c6np00025h