Scientific Research Publishing

Advances in Bioscience and Biotechnology, 2019, 10, 150-164
http://www.scirp.org/journal/abb
ISSN Online: 2156-8502
ISSN Print: 2156-8456

Analysis of the Promoter Region, Motif and CpG Islands in AraC Family Transcriptional Regulator ACP92 Genes of *Herbaspirillum seropedicae*

Mihret Yirgu[1], Mulugeta Kebede[2]

[1]College of Agriculture and Natural Resource, Department of Plant Science, Madda Walabu University, Bale, Robe, Ethiopia
[2]Department of Applied Biology, School of Applied Natural Science, Adama Science and Technology University, Adama, Ethiopia
Email: ymihret@gmail.com, mihret.yirgu@mwu.edu.et

How to cite this paper: Yirgu, M. and Kebede, M. (2019) Analysis of the Promoter Region, Motif and CpG Islands in AraC Family Transcriptional Regulator ACP92 Genes of *Herbaspirillum seropedicae*. *Advances in Bioscience and Biotechnology*, **10**, 150-164.
https://doi.org/10.4236/abb.2019.106011

Received: February 10, 2019
Accepted: June 27, 2019
Published: June 30, 2019

Abstract

Identification of promoters and their regulatory elements are the most important phases in bioinformatics. To understand the regulation of gene expression, identification, and analysis of promoters region, motif and CpG islands are the most important steps. The accurate prediction of promoter's is basic for proper interpretation of gene expression patterns, construction and understanding of genetic regulatory system. Therefore, the objective of this study was to analyze the promoter region, motif such as a transcription factor and CpG islands in AraC family transcriptional regulator ACP92 genes of *Herbaspirillum seropedicae*. The analysis was carried out by identifying transcription start sites in ACP92 genome sequences taken from the *H. seropedicae* assembly of NCBI genome browser, and 29 ACP92 genes sequences. Accordingly, transcription start sites (TSS) were identified, and the result indicated that 37.9% had more than one TSS whereas only 62.1% had one TSS. In the analysis, seven motifs were identified from the thought sequences and MV6 was revealed the common promoter motif for all (100%) in *H. seropedicae* ACP92 gene that serves as binding sites for transcription factors which shared a minimum of 48.27%. Based on a common motif MV6 to find out similar motifs using TOMTOM from the databases of prokaryotes DNA, most of them are transcription factors of fur family. The others are bacterial histone-like protein family, matp and sigma-54 factor family also transcription factor families that are binding candidate to MV6. *H. seropedicae* ACP92 genes are CpG Island which implies that the regulation of gene expression plays an important role.

Keywords

Gene Expression, Motifs, Promoter, Transcriptional

1. Introduction

Herbaspirillum seropedicae is a genus of bacteria that found in roots, stems, and leaves in association with economically important species of *Poaceae* family such as maize (*Zea mays*), rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), sugar cane (*Saccharum officinarum*) [1]. It's commonly found in forage grasses such as elephant grass and tropical fruits like pineapple and banana [2]. It is a nitrogen-fixing *Proteobacterium* isolated from the rhizosphere and tissues of several economically important plants species [3]. *H. seropedicae* is well-characterized class of diazotrophic bacteria capable of establishing endophytic associations and promoting plant-growth of important cereals and forage grasses [4]. It was also studied as a model of bacterial entry into host plants and plant growth promotion [5]. Genome of *H. seropedicae*, involved in the nitrogen fixation process and its regulation, the genes potentially involved in the establishment of efficient interaction with the host plant. Several studies have shown that *H. seropedicae* supplies fixed nitrogen to the associated plant and increases grain productivity [4]. The AraC family of transcriptional regulators, present in bacterial species is involved in a variety of cellular processes from carbon metabolism to stress responses and its regulation according to Munson and Scott [6]. Correspondingly, in AraC family transcriptional regulator ACP92 genes of *H. seropedicae* are also a potential transcriptional regulator that involved in a variety of cellular processes, transcriptional control and expression of genes by binding to specific promoter regions both at transcriptional and post-translational levels.

Promoter is a key region that is involved in differential transcription regulation of protein coding and RNA genes [7]. Promoters are functional regions containing complex regulatory elements for determining the transcription initiation of genes [8]. DNA binding sites or motifs refer to short DNA sequences (typically 4 to 30 base pairs long, but up to 200 bp for recombination sites) that are explicitly bound by one or more DNA-binding proteins or protein complexes [9]. It is often associated with specialized proteins known as transcription factors, and is thus linked to transcriptional regulation [10]. Transcription factors are DNA binding proteins interacting with RNA polymerase complex to activate or repress transcription factors bind to the DNA on specific cis-acting regulatory elements (CAREs) and in the regulation of gene expression the initiation of transcription which is one of the most important control points [11]. CpG islands are also reported as important regulatory elements in the promoter regions of genome [12]. CpG refers to the base cytosine (C) linked by a phosphate bond to the base guanine (G) in the DNA nucleotide sequence [13]. A structural feature that has proven useful in the detection of promoters is the so

called CpG islands, *i.e.* regions that are rich in CpGs, which are important because of their strong link with gene regulation [14]. CpG islands are playing an important role in gene regulation through epigenetic changes [15]. Recent studies have shown that CpG methylation correlates with the activation of some genes [16]. DNA methylation has been shown to repress transcription initiation by interfering directly with the binding of transcriptional activators or indirectly by binding proteins [17].

Prokaryotic and eukaryotic promoters use different DNA sequences to regulate gene expression [18]. Promoters in eukaryotic and prokaryotic genomes using CpG islands and transcription factor binding sites (TFBS) have been developed by Anwar *et al.* [19]. Studies on identifying the promoters on 250 bp long regions upstream of gene start in *Escherichia coli* [20], and also have proposed to identify in *E. coli* promoters reported by Gordon *et al.* [21]. Many methods have been proposed to search for binding sites [22]. Explained large subset of motif-finders among which MEME one is the most important tools for binding motif's discovery [23]. Neural Network Promoter Prediction (NNPP version 2.2) is a widely used on-line tool for the recognition of eukaryotic promoters [24]. However, in prokaryotes the Neural Network Promoter search from https://www.fruitfly.org/seq_tools/promoter.html, and promoter prediction tool set was used [25]. Analysis of promoter region, transcription start site and CpG islands are some of the most important issues in gene expression. Conducted for *Herbaspirillum seropedicae* ACP92 to identify and analysis of these elements and were revealed a common motif that serves as binding sites are very crucial. Therefore, the objective of this study was initiated to analyze the promoter region, motif such as transcription factor and CpG islands in *H. seropedicae* in AraC family transcriptional regulator ACP92 genes.

2. Materials and Methods

Genome sequences were taken from *H.seropedicae* assembly of NCBI genome browser. Genome sequences starting by ATG (starting codon) were identified form AraC family transcriptional regulator ACP92 genes of *H. seropedicae* databases. At the beginning sequences containing start codon were identified and coding sequences were used in this analysis. Only twenty-nine AraC family transcriptional regulator ACP92 genes were discovered and the left AraC families are pseudogene with no ATG. Twenty-nine, *H. seropedicae* ACP92 gene sequences were used for analysis to determine their respective TSSs, 1 kb sequences upstream of the start codon were excised from each gene. Promoter regions were defined as 1 kb region upstream of each TSS. The Neural Network Promoter search from https://www.fruitfly.org/seq_tools/promoter.html and prediction tool was used with a minimum standard predictive scores (between 0 and 1) cutoff value of 0.8 for prokaryote [25]. For those regions containing more than one TSS, the highest value of prediction score was considered so as to have a more accurate prediction.

Identification of *H. seropedicae*, ACP92 promoter sequence was analyzed using the MEME (Expectation Maximization algorithm); via web server (http://bioinformatics.ubc.ca/resources) look for common motifs and transcription factors that regulate the expression of ACP92 genes. MEME was many optional inputs to modify its performance. The following possibilities were used: 1) zero or one occurrence per sequence model was chosen, 2) the maximum width of the motifs was 50, and 3) motifs occurrences were on both strands of the input DNA sequences. Statistically, significant motifs in the input sequence set were researching MEME and the E-value which is the probability of finding an equally well-conserved pattern in random sequences. The MEME output is HTML and shows the motifs as local many alignments of the input sequences. The MEME HTML output was allowed one or all the motifs to be forwarded for further enquiry, was better characterizing the identified motifs, by other web-based programs, TOMTOM. TOMTOM web server was selected where various sequence databases were searched for sequences matching the identified motif. TOMTOM shows that the query motif closely resembles the binding motif [26].

To find the CpG islands in *H. seropedicae* ACP92 promoter regions were two algorithms used. The first CLC searching, genomics Workbench ver. 3.6.5 (http://clcbio.com, CLC bio, Aarhus, Denmark) was used for searching the restriction enzyme *MspI* cutting sites (fragment sizes between 40 and 220 bp), and the second algorithm, Takai and Jones algorithm (stringent) search criteria was used in GC content ≥55%, Observed CpG/Expected CpG ratio ≥ 0.65, and length ≥ 500 bp [27]. The CpG island searcher program (CpGi130) available at the web link, http://dbcat.cgm.ntu.edu.tw// was used.

3. Result and Discussion

3.1. Determination of Transcription Start Sites (TSSs) and Promoter Regions

Identification of transcriptional start site and promoter regions is the first step to understand the regulation mechanisms of gene expression and association with genetic variations in the regions [28]. Accordingly, this study was the first identified transcription starts sites for each 29 transcriptional regulator ACP92 genes in *Herbaspirillum seropedicae*. The prediction more reliable for genes containing more than one TSS, TSS of the highest prediction score was considered and identified. The result indicated that three (in ACP92_RS04670 and ACP92_RS13185), four (in ACP92_RS00045) and six (in ACP92_RS19695) was found the highest TSS number while in the remaining genes a lower number of TSSs was obtained. In addition, 37.9% have more than one TSS whereas 62.1% had only one TSS (**Table 1**).

3.2. Common Motifs and Transcription Factors

Based on the promoter region of *H. seropedicae* significant motifs in the input sequence set was searched MEME via the web server and the E-value, the probability

Table 1. Identified TSSs and predictive score value for *H. seropedicae* ACP92s gene.

No	Gene ID	Corresponding promoter region name	Number of TSS identified	Predictive score at cutoff value of 0.8	Distance from start codon (ATG)
1	29390694	pro-ACP92_RS00045	4	0.88, 0.91, 0.97, 0.98	−12, −239, −715, −21
2	29392361	pro-ACP92_RS00855	1	0.94	−4, 297
3	29390588	pro-ACP92_RS01580	1	0.88	−258
4	29392080	pro-ACP92_RS04595	1	0.84	55
5	29393722	pro-ACP92_RS04620	1	0.90	196
6	29390157	pro-ACP92_RS04670	3	0.82, 0.84.0.85	−2671, −2907, 2445
7	29390499	ACP92_RS04680	2	0.84, 0.98	−33, −24
8	29391790	pro-ACP92_RS04870	1	0.85	−7
9	29390103	pro-ACP92_RS05285	2	0.93, 0.97	−232, −73
10	29393242	pro-ACP92_RS08330	1	1.0	−27
11	29393201	pro-ACP92_RS11865	2	0.86, 0.87	−5, 656, −5, 215
12	29392668	pro-ACP92_RS12020	1	0.84	3, 128
13	29389983	pro-ACP92_RS12195	1	0.94	−1, 664
14	29389939	pro-ACP92_RS12565	1	0.82	−2, 526
15	29393219	pro-ACP92_RS12645	2	0.83, 0.85	1, 935, 1312
16	29390362	pro-ACP92_RS13185	3	0.81, 0.84, 0.97	−58, −400, −76
17	29390889	pro-ACP92_RS14465	1	0.91	−733
18	29389718	pro-ACP92_RS15060	1	0.91	1048
19	29391772	pro-ACP92_RS17440	1	0.96	−3909
20	29390087	pro-ACP92_RS17545	2	0.89, 0.94	−200, −534
21	29394066	pro-ACP92_RS18255	1	0.93	−356
22	29391333	pro-ACP92_RS18865	2	0.84, 0.93	−4922, −4902
23	29393361	pro-ACP92_RS19245	1	0.92	−1907
24	29390041	pro-ACP92_RS19695	6	0.88, 0.95, 0.95, 0.96, 0.99, 1.00	259, 248, 170, 180, 198, 219
25	29393726	pro-ACP92_RS19845	2	0.95, 0.99	2124, 2401
26	29391169	pro-ACP92_RS19860	1	0.87	−46
27	29394235	pro-ACP92_RS22515	1	0.94	48
28	29394160	pro-ACP92_RS22560	1	0.93	1347
29	29389690	pro-ACP92_RS23100	1	0.80	1247

NNPP tool prediction result is considered reliable at 0.8 cutoff values for prokaryote organism [26].

of finding a well-conserved pattern in random sequences. MEME output was revealed seven motifs (MV1, MV2, MV3, MV4, MV5, MV6 and MV7) were identified from the thought's sequences. The study indicated that, motif six (MV6) was found the common promoter motif for all (100%) in *H. seropedicae* ACP92 genes that serve as binding sites for transcription factors and shared a minimum of 48.27% (Table 2). Motif MV6 found to serve as binding sites for

transcription factors in the expression and regulation of genes. After the location and distribution of these motifs largely, it was found between -800 and -100 bp of the transcription start sites (TSSs). Relatively, higher distributions of motifs were found also in positive (96) than negative strands (81) *H. seropedicae* ACP92s gene (Figure 1). In a similar manner, sequence logo for MV6 was generated by MEME (Figure 2).

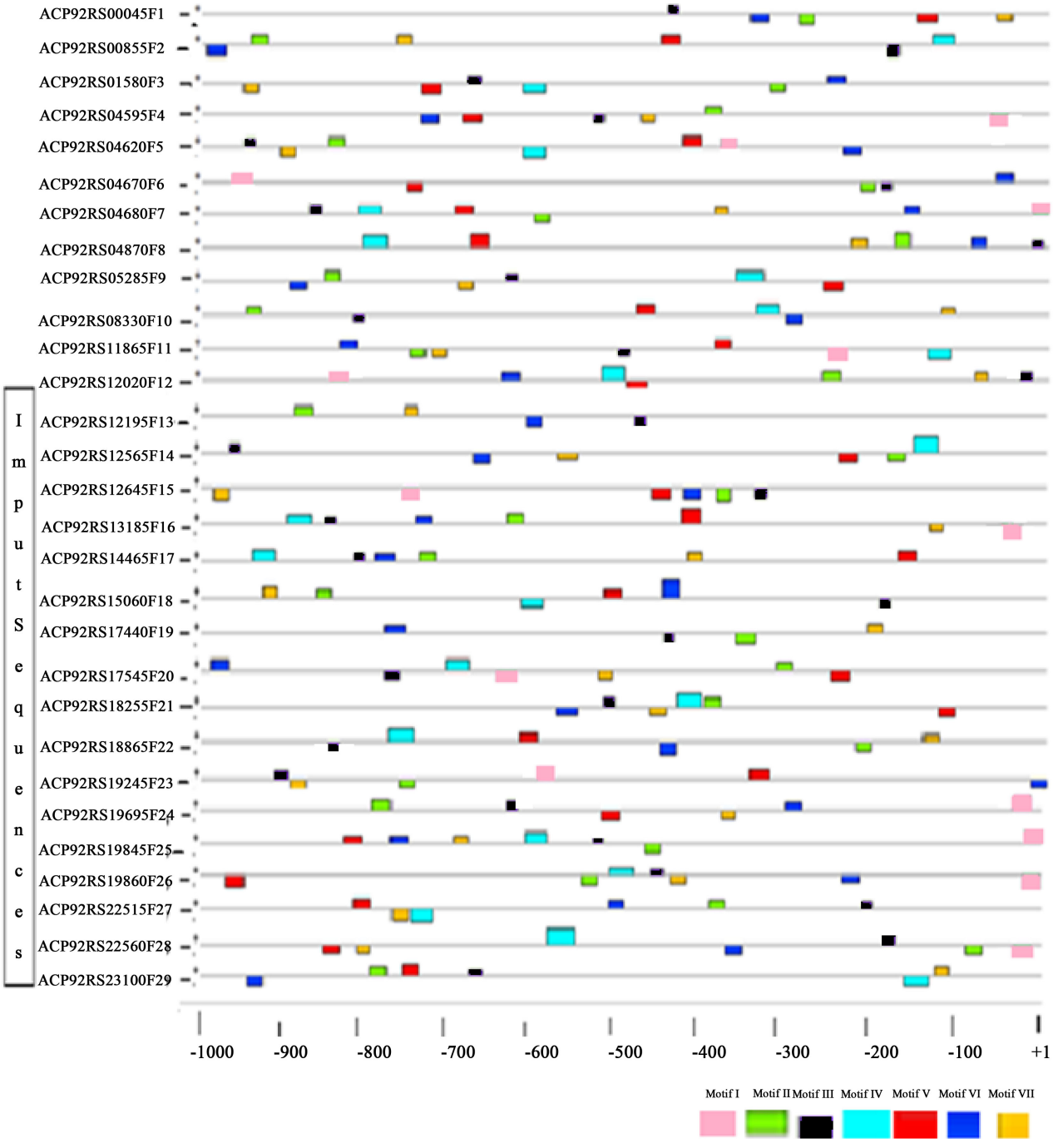


Figure 1. The relative positions of motifs in different ACP92 subfamilies sequences relative to TSSs. Note: the nucleotide positions are specified at the bottom of the graph from +1 (beginning of TSSs) to the upstream 1 kb (−1 kb) bp.

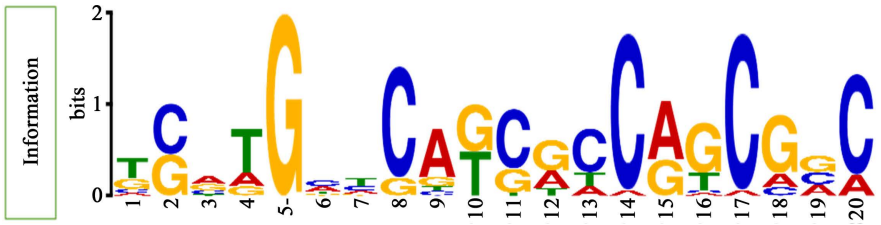


Figure 2. Identified common promoter sequence logos for the motif, MV6 of *H. seropedicae* ACP92 gene.

Table 2. Number of binding site and common motifs in *H. seropedicae* ACP92 gene promoter regions

Discovered motif	Number (%) of ACP92 promoters containing each one of the motifs	E-value*	Motif width	Total no. of binding sites
MV1	14 (48.27%)	3.7e−005	20	14
MV2	29 (100%)	2.5e−002	21	29
MV3	29 (100%)	3.3e−006	14	29
MV4	21 (72.41%)	1.5e−008	29	21
MV5	27 (93.10%)	1.3e−010	23	27
MV6	29 (100%)	2.1e−007	20	29
MV7	28 (96.55%)	1.6e−005	19	28

*Probability of finding an equally well-conserved motif in random sequences.

TOMTOM web server was selected as various sequence databases can easily be searched for sequences matching of the identified motif, based on common motif MV6 to find out similar motifs using TOMTOM from the databases of prokaryotes DNA, collected bacterial transcription factors. The result indicated that, 24 query motifs closely resemble the binding motifs MV6 was identified out of 84 motifs used and found in collected bacterial transcription factors (collectF) prokaryote databases. In addition, only identified four TF families were discovered out of 24 matched motifs and the left query motifs were non-transcription factor families. Four transcription factors families that are binding candidates for MV6 motif was identified namely; bacterial histone-like protein, fur (Ferric uptake regulation protein), matp (Macrodomain Ter protein) and sigma-54 factor (RNA polymerase sigma-54 factor) families (Table 3). Among four families, fur (Ferric uptake regulation protein) is largely matched with the binding motif also known as a transcription factors family for *H. seropedicae* ACP92s gene regulations.

3.3. Determination of CpG Islands in *H.seropedicae* ACP92 Promoter Regions

In this study, CpG islands were determined using twenty-nine in *H. seropedicae* promoter and gene body regions with two algorithms were used to search. CLC searching algorithm was used and identified one possible CpG island in each

Table 3. Classification of transcription factors families which bind to motif MV6 of *H. seropedicae* ACP92s promoter regions from the collect TF database.

Classification of TF families	Candidate transcription factors	Gene	Function
Bacterial histone-like protein family	Integration host factor subunit alpha,	ihfA	– A specific DNA-binding protein that functions in genetic recombination as well as in transcriptional and translational control
Fur family	Ferric uptake regulation protein	fur	– Acts as a repressor, employing Fe^{2+} as a cofactor to bind the operator of the iron transport operon, Involved in exotoxin a regulation, siderophore regulation and manganese susceptibility, transcriptional activator activity and sequence-specific DNA binding
matp family	Macrodomain Ter protein	matP	– Required for spatial organization of the terminus region of the chromosome (Termacrodomain) during the cell cycle. – Prevents early segregation of duplicated Termacrodomains during cell division – Binds specifically to matS, which is a 13 bp signature motif repeated within the Termacrodomain.
sigma-5 factor family	RNA polymerase sigma-54 factor	VC_2529	– Sigma factors are initiation factors that promote the attachment of RNA polymerase to specific initiation sites and are then released

gene; ACP92RS01580, ACP92RS04595, ACP92RS11865, ACP92RS12565, ACP92RS15060, ACP92RS17545, ACP92RS18255, ACP92RS18865, ACP92RS19245, ACP92RS22560, and ACP92RS23100 in promoter regions (Table 4). While, in gene body regions it was identified one possible CpG island in all genes except in gene ACP92RS00045, ACP92RS08330, ACP92RS11865, ACP92RS12565, ACP92RS15060, ACP92RS17440, ACP92RS19845 and ACP92RS19860 (Table 5). The second algorithm using restriction enzyme *MspI* site cutting was used and examined CpG Island has many fragment sizes both in promoter and gene body regions. CpG islands in promoter regions contain several fragments size in all genes except ACP92_RS17545 gene that have only two fragment size (62 and 70 bp) (Table 6).

Similarly, CpG Island was also found in all the gene body regions and contains many fragment sizes except the gene ACP92_RS00045, ACP92_RS11865 and ACP92_RS19695 in AraC family transcriptional regulator ACP92 genes in *H. seropedicae* (Table 7). This event implies that *H. seropedicae* bacteria have CpG Island and an important role the regulation of the gene expression. Also, there were indicating that *H. seropedicae* ACP92 genes are not poor in CpG islands. In contrary to this study in human [29], mouse [30], and pig V1R genes [31] were poor in CpG islands from eukaryotes. Nevertheless, in vertebrates, about 70% of known promoters are CpG islands reported by Deaton and Bird [32].

4. Conclusion

Transcriptional factors modulate gene expression through binding to a specific DNA sequence usually found upstream of the gene, or the genomics region that they control. Gene promoter regions are together with transcription factors binding to regions upstream to the coding sequence. CpG islands are also regulatory elements in the promoter regions of genome and useful in the detection of

Table 4. Possible CpG islands shown in graph using promoter regions.

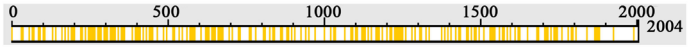
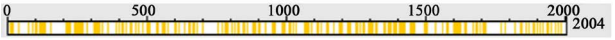
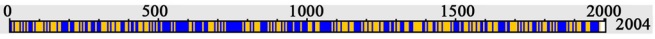
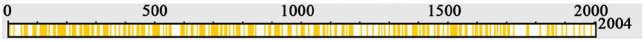
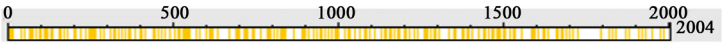
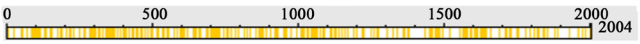
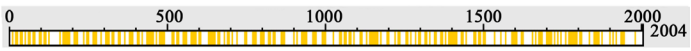
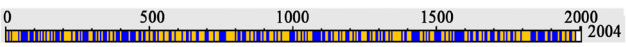
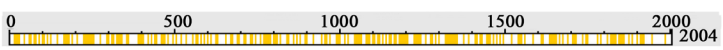
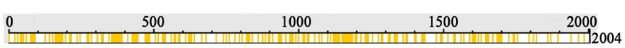
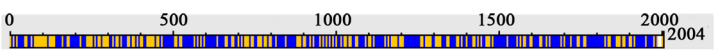
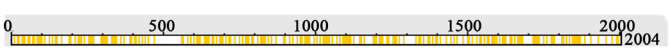
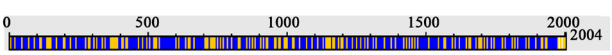
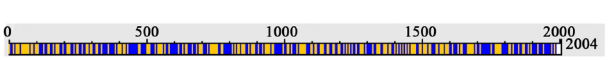
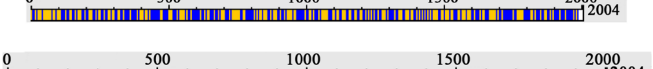
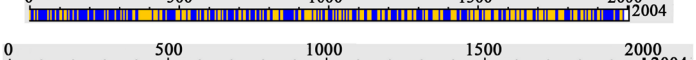
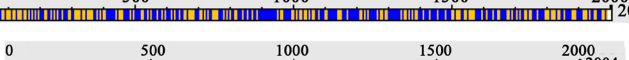
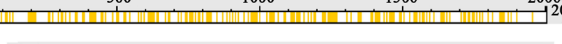
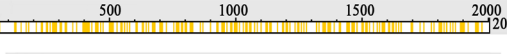
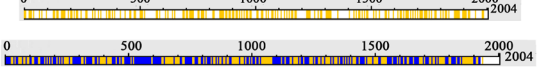
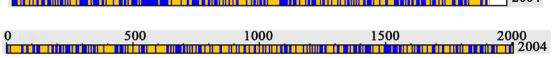
Name of Gene	Possible CpG island(s) found	N. of CpG island (s) found	GC content (%)
ACP92RS00045		-	-
ACP92RS00855		-	-
ACP92RS01580		1	57
ACP92RS04595		1	66
ACP92RS04620		-	-
ACP92RS04670		-	-
ACP92RS04680		-	-
ACP92RS04870		-	-
ACP92RS05285		-	-
ACP92RS08330		-	-
ACP92RS11865		1	100
ACP92RS12020		-	-
ACP92RS12195		-	-
ACP92RS12565		1	53
ACP92RS12645		-	-
ACP92RS13185		-	-
ACP92RS14465		-	-
ACP92RS15060		1	56
ACP92RS17440		-	-
ACP92RS17545		1	53
ACP92RS18255		1	50
ACP92RS18865		1	56
ACP92RS19245		1	50
ACP92RS19695		-	-
ACP92RS19845		-	-
ACP92RS19860		-	-
ACP92RS22515		-	-
ACP92RS22560		1	65
ACP92RS23100		1	60

Table 5. Possible CpG islands shown in graph using gene body regions.

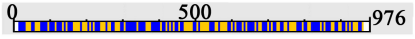
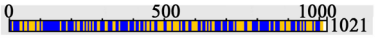
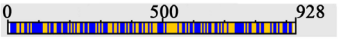
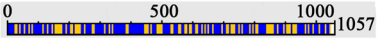
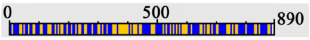
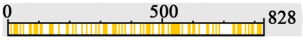
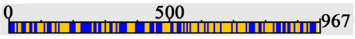
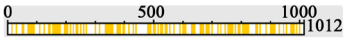
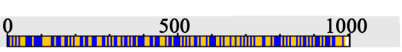
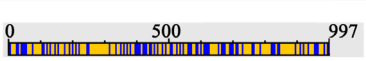
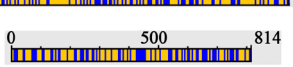
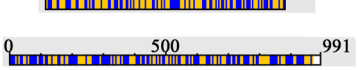
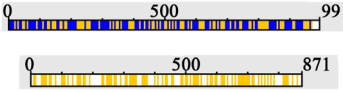
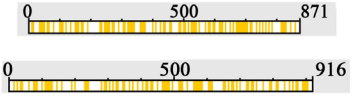
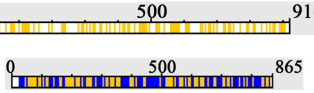
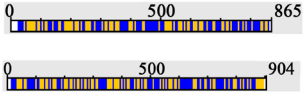
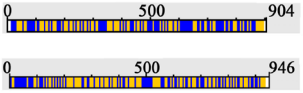
Name of Gene	Possible CpG island(s) found	Number of CpG island(s) found	GC content (%)
ACP92RS00045		-	-
ACP92RS00855		1	69
ACP92RS01580		1	50
ACP92RS04595		1	59
ACP92RS04620		1	61
ACP92RS04670		1	60
ACP92RS04680		1	63
ACP92RS04870		1	75
ACP92RS05285		1	57
ACP92RS08330		-	-
ACP92RS11865		-	-
ACP92RS12020		1	61
ACP92RS12195		1	67
ACP92RS12565		-	-
ACP92RS12645		1	71
ACP92RS13185		1	50
ACP92RS14465		1	68
ACP92RS15060		-	-
ACP92RS17440		-	-
ACP92RS17545		1	50
ACP92RS18255		1	64
ACP92RS18865		1	71
ACP92RS19245		1	72
ACP92RS19695		1	63
ACP92RS19845		-	-
ACP92RS19860		-	-
ACP92RS22515		1	70
ACP92RS22560		1	66
ACP92RS23100		1	62

Table 6. Determination *of MspI* cutting sites and fragment sizes for *H. seropedicae* ACP92s promoter regions analysis results.

Region	Names of corresponding ACP92s	Nucleotide positions of *MspI* sites	Fragment sizes
Promoter region	pro-ACP92_RS00045	Multiple cut (at 26, 80, 177, 219, 231, 572, 839, 1097, 1117, 1213, 1285, 1584, 1734)	42, 54, 72, 96, 97, 150
	pro-ACP92_RS00855	Multiple cut (at 97, 112 , 262, 335, 424, 508, 751, 775, 811, 844, 1033, 1068, 1085, 1192, 1248, 1596, 1617, 1691, 1889, 1947, 1960, 1974)	56, 58, 73, 74, 84, 89, 107, 150, 189, 198
	pro-ACP92_RS01580	Multiple cut (at 142, 154, 245, 260, 323, 419, 457, 667, 679, 782, 825, 917, 1239, 1335, 1724)	63, 91, 92, 96, 96, 210,
	pro-ACP92_RS04595	Multiple cut (at 46, 83, 90, 122, 161, 173, 274, 480, 486, 519, 695, 815, 959, 1049, 1185, 1278, 1307, 1353, 1406, 1480, 1583, 1624, 1640, 1656, 1789, 1820, 1893)	41, 46, 53, 73, 74, 90, 93, 101, 103, 120, 133, 136, 144, 176, 206,
	pro-ACP92_RS04620	Multiple cut (at 243, 357, 410, 439, 579, 584, 728, 1018, 1040, 1132, 1229, 1264, 1272, 1286, 1330, 1356, 1696, 1745, 1902)	44, 49, 53, 92, 97 114, 140, 144, 175
	pro-ACP92_RS04670	Multiple cut (at 460, 673, 798, 883, 1262, 1454, 1974)	85, 192, 213
	ACP92_RS04680	Multiple cut (at 44, 52, 129, 339, 393, 438, 605, 743, 962, 1057, 1350, 1394, 1704, 1720, 1902)	44, 45, 54, 77, 95, 138, 167, 182, 210, 219,
	pro-ACP92_RS04870	Multiple cut (at 189, 312, 336, 510, 694, 837, 859, 1231, 1416, 1416, 1498, 1552)	54, 82, 123, 174, 184, 185,
	pro-ACP92_RS05285	Multiple cut (at 36, 47, 263, 315, 332, 353, 438, 956, 1033, 1109, 1373, 1594, 1663, 1677, 1790, 1876, 1946)	52, 69, 70, 76, 77, 85, 86, 113, 216,
	pro-ACP92_RS08330	Multiple cut (at 21, 53, 65, 104, 189, 227, 389, 535, 597, 680, 701, 721, 752, 832, 922, 1049, 1300, 1465, 1482, 1612, 1639, 1887, 1980)	62, 80, 85, 90, 93, 93, 101, 130, 146, 165, 172, 130,
	pro-ACP92_RS11865	Multiple cut (at 30, 42, 79, 135, 230, 412, 589, 629, 641, 690, 727, 733, 1024, 1069, 1225, 1268, 1528, 1729, 1969)	40, 43, 45, 49, 56, 95, 182, 177, 201, 240
	pro-ACP92_RS12020	Multiple cut (at 275, 379, 391, 506, 569, 666, 863, 920, 1341, 1669, 1733, 1778)	45, 57, 63, 64, 104, 115, 197,
	pro-ACP92_RS12195	Multiple cut (at 39, 45, 171, 379, 638, 694, 740, 968, 998, 1043, 1152, 1163, 1296, 1327, 1364, 1368, 1442, 1442, 1442, 1474, 1591, 1595, 1789, 1918)	45, 46, 56, 74, 109, 117, 126, 129, 133, 208, 129
	pro-ACP92_RS12565	Multiple cut (at 37, 120, 165, 185, 208, 236, 297, 324, 450, 487, 618, 738, 1019, 1024, 1141, 1265, 1383, 1408, 1475, 1632, 1646, 1817, 1884	45, 61, 67, 67, 83, 117, 118, 120, 124, 126, 131, 157, 171,
	pro-ACP92_RS12645	Multiple cut (at 355, 459, 804, 837, 1093, 1361, 1458, 1469, 1509, 1556, 1709, 1743, 1934)	40, 47, 97, 104, 153, 191
	pro-ACP92_RS13185	Multiple cut (at 248, 271, 305, 518, 578, 844, 919, 1013, 1199, 1230, 1295, 1424, 1685, 1701)	60, 65, 75, 94, 129, 186, 213,
	pro-ACP92_RS14465	Multiple cut (at 49, 163, 365, 406, 420, 428, 474, 585, 608, 639, 675, 882, 929, 1017, 1079, 1250, 1306, 1338, 1474, 1744, 1867, 1960, 1981)	41, 46, 47, 56, 62, 88, 93, 111, 114, 123, 136, 171, 202, 207,
	pro-ACP92_RS15060	Multiple cut (at 121, 239, 266, 287, 314, 361, 386, 446, 454, 518, 696, 719, 968, 989, 1270, 1418, 1432, 1640, 1877)	47, 60, 64, 118, 148, 178, 208
	pro-ACP92_RS17440	Multiple cut (at 70, 99, 226, 587, 645, 693, 815, 974, 991, 1097, 1145, 1198)	48, 48, 53, 58, 106, 122, 127, 159
	pro-ACP92_RS17545	Multiple cut (at 166, 236, 529, 910, 1168, 1445, 1507, 1973)	62, 70
	pro-ACP92_RS18255	Multiple cut (at 65, 170, 263, 356, 495, 826, 1098, 1127, 1158, 1222, 1343, 1513, 1574, 1607, 1707, 1827)	93, 93, 64, 105, 100, 120, 121, 139, 170
	pro-ACP92_RS18865	Multiple cut (at 103, 364, 436, 487, 492, 752, 817, 824, 923, 932, 994, 1034, 1282, 1345, 1550, 1733, 1745, 1873)	40, 51, 62, 63, 65, 71, 72, 99, 128, 183, 205
	pro-ACP92_RS19245	Multiple cut (at 1, 20, 47, 65, 100, 119, 181, 265, 337, 351, 401, 445, 580, 593, 682, 725, 797, 964, 1033, 1339, 1450, 1501, 1543, 1562, 1646, 1979)	42, 43, 44, 50, 51, 62, 69 72, 72, 81, 84, 89, 111, 135, 167
	pro-ACP92_RS19695	Multiple cut (at 70, 292, 524, 608, 625, 722, 751, 958, 982, 997, 1063, 1088, 1151, 1342, 1846, 1897)	51, 63, 66, 84, 97, 191, 207, 222,
	pro-ACP92_RS19845	Multiple cut (at 179, 429, 542, 556, 604, 640, 769, 865, 1082, 1096, 1212, 1353, 1443, 1587, 1690, 1740, 1770, 1855)	48, 50, 85, 90, 96, 103, 113, 116, 129, 141, 217

Continued

	pro-ACP92_RS19860	Multiple cut (at 59, 100, 183, 469, 475, 502, 573, 599, 617, 945, 975, 1327, 1359, 1647, 1689, 1794, 1837, 1889, 1944)	41, 42, 43, 52, 55 71, 83, 105
	pro-ACP92_RS22515	Multiple cut (at 29, 375, 451, 939, 960, 976, 1159, 1252, 1278, 1446, 1557, 1677, 1923, 1984)	61, 76, 93, 110, 111, 168, 183
	pro-ACP92_RS22560	Multiple cut (at 73, 260, 446, 516, 813, 845, 1043, 1177, 1397, 1468, 1472, 1625, 1724, 1755)	70, 71, 99, 134, 153, 186, 187, 198, 220
	pro-ACP92_RS23100	Multiple cut (at 16, 20, 46, 131, 156, 189, 203, 262, 374, 386, 422, 473, 799, 820, 1018, 1090, 1351, 1369, 1944, 1996)	51, 52, 59, 72, 85, 198

Pro-promoter.

Table 7. Determination of *MspI* cutting sites and fragment sizes for *H. seropedicae* ACP92 gene body analysis results.

Region	Names of corresponding ACP92s	Nucleotide positions of *MspI* sites	Fragment sizes
Gene body region	pro-ACP92_RS00045	Single cut (at 850)	-
	pro-ACP92_RS00855	Multiple cut (at 33, 187, 240, 518, 523, 598, 609, 740, 890, 921, 932)	53, 131, 150, 154
	pro-ACP92_RS01580	Multiple cut (at 87, 249, 351, 459, 509, 567, 1005)	50, 102, 106, 108, 162
	pro-ACP92_RS04595	Multiple cut (at 135, 293, 413, 461, 504, 603, 651, 685)	48, 48, 99, 158, 120
	pro-ACP92_RS04620	Multiple cut (at 111, 457, 580, 764, 811, 841, 869)	47, 123
	pro-ACP92_RS04670	Multiple cut (at 228, 310, 693, 782, 835)	82, 89
	ACP92_RS04680	Multiple cut (at 255, 312, 380, 537, 572, 622, 795)	50, 57, 68, 157, 173
	pro-ACP92_RS04870	Multiple cut (at 206, 406, 450, 459, 477, 684, 813)	44, 200, 207, 129
	pro-ACP92_RS05285	Multiple cut (at 99, 175, 193, 249, 486, 664, 877)	56, 76, 213
	pro-ACP92_RS08330	Multiple cut (at 72, 85, 407, 465, 486, 549, 616, 763, 786)	58, 63, 67, 147,
	pro-ACP92_RS11865	Multiple cut (at 592, 880, 1041)	161
	pro-ACP92_RS12020	Multiple cut (at 214, 276, 472, 494, 808, 868)	60, 62, 196,
	pro-ACP92_RS12195	Multiple cut (at 111, 174, 213, 324, 333, 489, 580, 645)	63, 65, 91, 111, 156
	pro-ACP92_RS12565	Multiple cut (at 209, 240, 330, 360, 438, 535, 711, 752)	41, 78, 90, 97, 176
	pro-ACP92_RS12645	Multiple cut (at 70, 172, 180, 231, 255, 255, 379, 520, 661, 743, 748, 793, 804, 863, 895, 954, 965, 987, 1024)	45, 51, 59, 82, 102, 124, 141, 141
	pro-ACP92_RS13185	Multiple cut (at 145, 190, 195, 277, 305, 350, 457, 769)	45, 45, 82, 107
	pro-ACP92_RS14465	Multiple cut (at 180, 273, 310, 339, 423, 617, 640)	84, 93, 194
	pro-ACP92_RS15060	Multiple cut (at 54, 109, 138, 240, 299, 398, 437, 529, 719, 788, 901, 1029)	55, 59, 92, 95, 99, 102,
	pro-ACP92_RS17440	Multiple cut (at 141, 214, 268, 341, 373, 497, 526, 538, 633, 822, 880, 972)	54, 58, 73, 92, 95, 124, 189
	pro-ACP92_RS18255	Multiple cut (at 364, 424, 531, 784, 901)	60, 107, 117
	pro-ACP92_RS18865	Multiple cut (at 58, 89, 113, 160, 220, 255, 307, 319, 343, 411, 434, 484, 579, 754, 856, 929)	47, 50, 52, 60, 68, 95, 73, 102, 175
	pro-ACP92_RS19245	Multiple cut (at 64, 127, 175, 199, 222, 300, 727)	48, 63, 78
	pro-ACP92_RS19695	Multiple cut (at 125, 225, 737)	100
	pro-ACP92_RS19845	Multiple cut (at 154, 259, 273, 339, 459, 525, 582, 630, 666, 753, 825)	48, 57, 66, 66, 72, 87, 105, 120
	pro-ACP92_RS19860	Multiple cut (at 72, 157, 187, 237, 340, 417, 484, 574, 715, 831, 845)	50, 67, 85, 90, 103, 116, 141
	pro-ACP92_RS22515	Multiple cut (at 72, 78, 94, 178, 188, 318, 336, 357, 390, 516, 541, 800, 831, 848)	84, 126, 130
	pro-ACP92_RS22560	Multiple cut (at 156, 167, 380, 388, 406, 420, 453, 468, 709, 839, 871)	130, 213
	pro-ACP92_RS23100	Multiple cut (at 88, 104, 301, 313, 466, 478, 574, 759, 857)	96, 98, 153, 197, 185

Searching the CpG Islands using restriction enzyme *MspI* cutting site (*fragment size b/n* 40 *&* 220 bp).

promoters. In this study, we analyzed the promoter region, motif and CpG islands in AraC family transcriptional regulator ACP92 genes of *H. seropedicae*. The result of this analysis helps to understand the transcription factor binding regions and could allow reading of the regulatory genetic code which predicts gene expression of bacterial species in general and *H. seropedicae* in particular. Therefore, knowledge of bioinformatics methods is worthy important to identify gene regulatory regions in the promoter regions and gene body regions could help also to predict gene expression profiles in various bacterial species.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Gyaneshwar, P., James, E.K., Reddy, P.M. and Ladha, J.K. (2002) *Herbaspirillum* Colonization Increases Growth and Nitrogen Accumulation in Al-Tolerant Rice Varieties. *New Phytologist*, **154**, 131-145. https://doi.org/10.1046/j.1469-8137.2002.00371.x

[2] Cruz, L.M., Souza, E.M., Weber, O.B., Baldani, J.I., Dobereiner, J. and Pedrosa, F.O. (2001) 16S Ribosomal DNA Characterization of Nitrogen-Fixing Bacteria Isolated from Banana (*Musa* spp.) and Pine-Apple (*Ananascomosus* (L.) Merril). *Applied and Environmental Microbiology*, **67**, 2375-2379. https://doi.org/10.1128/AEM.67.5.2375-2379.2001

[3] Baldani, J.I., Baldani, V., Seldin, L. and Dobereiner, J. (1986) Characterization of *H. seropedicae* Gen-Nov, Sp-Nov, Root-Associated Nitrogen-Fixing Bacterium. *International Journal of Systematic Bacteriology*, **36**, 86-93. https://doi.org/10.1099/00207713-36-1-86

[4] Pedrosa, F.O. and Elmerich, C. (2007) Regulation of Nitrogen Fixation and Ammonium Assimilation in Associative and Endophyticnitrogen Fixing Bacteria. In: Elmerich, C. and Newton, W.E., Eds., *Associative and Endophytic Nitrogen Fixing Bacteria and Cyanobacterial Associations*, Springer, Berlin, 47-71.

[5] Monteiro, R., Balsanelli, E., Wassem, R., Marin, A. and Brusamarello-Santos, L. (2012) *Herbaspirillum*-Plant Interactions: Microscopical, Histological and Molecular Aspects. *Plant and Soil*, **356**, 175-196. https://doi.org/10.1007/s11104-012-1125-7

[6] Munson, G.P. and Scott, J.R. (1999) Binding Site Recognition by RNS, a Virulence Regulator in the AraC Family. *Journal of Bacteriology*, **181**, 2110-2117.

[7] Solovyev, V., Shahmuradov, I. and Salamov, A. (2010) Identification of Promoter Regions and Regulatory Sites. *Methods in Molecular Biology*, **674**, 57-83. https://doi.org/10.1007/978-1-60761-854-6_5

[8] Abeel, T., Saeys, Y., Rouze, P. and Vande Peer, Y. (2008) Pro SOM: Core Promoter Prediction Based on Unsupervised Clustering of DNA Physical Profiles. *Bioinformatics*, **24**, 24-31. https://doi.org/10.1093/bioinformatics/btn172

[9] Borneman, A.R., Gianoulis, T.A., Zhang, Z.D., Yu, H., Rozowsky, J., Seringhaus, M.R., Wang, L.Y., Gerstein, M. and Snyder, M. (2007) Divergence of Transcription Factor Binding Sites across Related Yeast Species. *Science*, **317**, 815-819. https://doi.org/10.1126/science.1140748

[10] Halford, E.S. and Marko, J.F. (2004) How Do Site-Specific DNA-Binding Proteins Find Their Targets. *Nucleic Acids Research*, **32**, 3040-3052. https://doi.org/10.1093/nar/gkh624

[11] Gidekel, M., Jimenez, B. and Herrera-Estrella, L. (1996) The First Intron of the Arabidopsis Thaliana Gene Coding for Elongation Factor 1 Contains an Enhancer-Like Element. *Gene*, **170**, 201-206. https://doi.org/10.1016/0378-1119(95)00837-3

[12] Rakyan, V.K., *et al.* (2008) An Integrated Resource for Genome-Wide Identification and Analysis of Human Tissue-Specific Differentially Methylated Regions (tDMRs). *Genome Research*, **18**, 1518-1529. https://doi.org/10.1101/gr.077479.108

[13] Lim, D.H. and Maher, E.R. (2010) DNA Methylation: A Form of Epigenetic Control of Gene Expression. *The Obstetrician and Gynaecologist*, **12**, 37-42. https://doi.org/10.1576/toag.12.1.037.27556

[14] Robertson, K.D. (2002) DNA Methylation and Chromatin: Unraveling the Tangled Web. *Oncogene*, **21**, 5361-5379. https://doi.org/10.1038/sj.onc.1205609

[15] Du, X., *et al.* (2012) Features of Methylation and Gene Expression in the Promoter-Associated CpG Islands Using Human Methylome Data. *Comparative and Functional Genomics*, **2012**, Article ID: 598987. https://doi.org/10.1155/2012/598987

[16] Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., *et al.* (2008) MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. *Science*, **320**, 1224-1229. https://doi.org/10.1126/science.1153252

[17] Meyer, P., Niedenh, I. and Ten Lohuis, M. (1994) Evidence for Cytosine Methylation of Non-Symmetrical Sequences in Transgenic Petunia Hybrida. *The EMBO Journal*, **13**, 2084-2088. https://doi.org/10.1002/j.1460-2075.1994.tb06483.x

[18] Hawley, D.K. and McClure, W.R. (1983) Compilation and Analysis of *E. coli* Promoter DNA Sequences. *Nucleic Acids Research*, **11**, 2237-2255. https://doi.org/10.1093/nar/11.8.2237

[19] Anwar, F., Baker, S.M., Jabid, T., Mehedi, H.M., Shoyaib, M., Khan, H. and Walshe, R. (2008) Pol II Promoter Prediction Using Characteristic 4-mer Motifs: A Machine Learning Approach. *BMC Bioinformatics*, **9**, 414. https://doi.org/10.1186/1471-2105-9-414

[20] Huerta, A.M. and Collado-Vides, J. (2003) Sigma 70 Promoters in *Escherichia coli*: Specific Transcription in Dense Regions of Overlapping Promoter-Like Signals. *Journal of Molecular Biology*, **333**, 261-278. https://doi.org/10.1016/j.jmb.2003.07.017

[21] Gordon, L., Chervonenkis, A.Y., Gammerman, A.J., Shahmuradov, L.A. and Solovyev, V.V. (2003) Sequence Alignment Kernel for Recognition of Promoter Regions. *Bioinformatics*, **19**, 1964-1971. https://doi.org/10.1093/bioinformatics/btg265

[22] Arnosti, D.N. and Kulkarni, M.M. (2005) Transcriptional Enhancers: Intelligent Enhanceosomes or Flexible Billboards. *Journal of Cellular Biochemistry*, **94**, 890-898. https://doi.org/10.1002/jcb.20352

[23] Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006) MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Research*, **34**, W369-W373. https://doi.org/10.1093/nar/gkl198

[24] Reese, M.G. (2001) Application of a Time-Delay Neural Network to Promoter Annotation in the Drosophila Melanogaster Genome. *Computers & Chemistry*, **26**, 51-56. https://doi.org/10.1016/S0097-8485(01)00099-7

[25] Reese, M.G., Harris, N.L. and Eeckman, F.H. (1996) Large Scale Sequencing Specific

Neural Networks for Promoter and Splice Site Recognition. *Bio-Computing*: *Proceedings of the* 1996 *Pacific Symposium*, Singapore, 2-7 January 1996. http://www.fruitfly.org/seq_tools/promoter.html

[26] Gupta, S., Stamatoyannopolous, J.A., Timothy, B. and William, S.N. (2007) Quantifying Similarity between Motifs. *Genome Biology*, **8**, R24. https://doi.org/10.1186/gb-2007-8-2-r24

[27] Takai, D. and Jones, P.A. (2002) Comprehensive Analysis of CpG Islands in Human Chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States*, **99**, 3740-3745. https://doi.org/10.1073/pnas.052410099

[28] Rani, T.S., Bhavani, S.D. and Bapi, R.S. (2007) Analysis of *E. coli* Promoter Recognition Problem in Di-Nucleotide Feature Space. *Bioinformatics*, **23**, 582-588. https://doi.org/10.1093/bioinformatics/btl670

[29] Jiang, C., Han, L., Su, B., Li, W.H. and Zhao, Z. (2007) Features and Trend of Loss of Promoter-Associated CpG Islands in the Human and Mouse Genomes. *Molecular Biology and Evolution*, **24**, 1991-2000. https://doi.org/10.1093/molbev/msm128

[30] Sharif, J., Endo, T.A., Toyoda, T. and Koseki, H. (2010) Divergence of CpG Island Promoters: A Consequence or Cause of Evolution? *Development*, *Growth & Differentiation*, **52**, 545-554. https://doi.org/10.1111/j.1440-169X.2010.01193.x

[31] Hunduma, D. and Minh, T.L. (2017) Analysis of Pig Vomeronasal Receptor Type 1 (V1R) Promoter Region Reveals a Common Promoter Motif but Poor CpG Islands. *Animal Biotechnology*, **29**, 293-300.

[32] Deaton, A.M. and Bird, A. (2011) CpG Islands and the Regulation of Transcription. *Genes & Development*, **25**, 1010-1022. https://doi.org/10.1101/gad.2037511