







Biomedical Informatics Division,

Rajendra Memorial Research Institute for Medical Sciences (I.C.M.R)

Patna, India.

Table of Content

- 1. Introduction to the Tool:
 - a. About
 - b. Requirement
 - c. Installation
- 2. Using tool for operon prediction
 - a. Input files
 - b. Genetic Parameters
 - c. Fitness function
 - d. Start Prediction
 - e. Output Visualization
- 3. Algorithm
- 4. Evaluation
- 5. Reference

1. INTRODUCTION TO THE TOOL

1.1 What is GAOPP:

GAOPP is standalone GUI tool for operon prediction. It uses unsupervised method Genetic Algorithm for identifying promoters in annotated prokaryotic species. It uses biological features like intergenic distance, Cluster of Gene Ontology and pathway involvement of each gene pair and clusters them in to operons. There are several computational methods are available for this purpose but none of them are GUI based. They need heavy data preparation, also. To meet these requirements GAOPP has been created.

It has three different evaluating functions to evaluate the fitness of each putative operon structure, can be found in literatures. These functions use biological properties like intergenic distance, involvement in metabolic pathway, and functionality from Clusters of Gene ontology (COG) gene functional families. This need needs the protein table file found at *National Centre for Biotechnology Information (NCBI)* FTP (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/). For Pathway information *KyotoEncyclopedia of Genes and Genome (KEGG)* pathway database can be used. A track of experimental promoters in the target species can be used to predict promoters. Terminators can be predicted using *TranTerm* and the output file may be used to provide terminator coordinates in the genome. Windows version of the tool is currently available to download. Binaries for Linux platform will be released soon.

1.2 Installation on windows :

- 1. Download the zipped installation file and extract it.
- 2. To install the tool, simply double click on install.bat file.
- It prompts you to enter installation directory. To accept default destination C:\GAOPP\
 press *y*. Wait until the prompt closes. Double click on the shortcut icon at Desktop.

- 4. To run it from source code, it requires PERL5.8 above and Tkx module. Active perl can be used instead.
- 5. To uninstall the program, simply go to the folder you installed and delete GAOPP directory. Remove the Desktop shortcut.

🔤 :::::::GAOPP INSTALLATION:::::::::::	
**************************************	CXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

	Enter path for Installation :C:\GAOPP\
GI ::::::GAOI	OPP INSTALLATION::::::::::::::::::::::::::::::::::::
** ** Ge ******	GAOPP INSTALLATION ** Genetic Algorithm for Operon Prediction in Prokaryote ** **********************************
Enter path GAOPP-GUI. GAOPP-path GAOPP-prot GAOPP-prot GAOPP-imag GAOPP-imag GAOPP-imag GAOPP-imag GAOPP-imag GAOPP-imag GAOPP-imag GAOPP-imag	th for Installation :C:\GAOPP\ .exe nnual.pdf thway.dat m_align_forr.txt ptEin.ptt r.txt ages\gaopp.gif ages\icon_t.gif ages\RMRI2.gif) copied stalled at C:\GAOPP\ successfully. cut created at Desktop y key to continue
Avira C	ra Control emailGmail - recomend . Center Shortcut (2)
Cra Car	razyTalk GAOPP roc Instal am Sui

GAOPP: Genetic Algorithm for Operon Prediction in Prokaryotes

1.3 System Requirement:

- 1. Operating system Windows 2000/XP/Vista/7, Linux* (available soon)
- 2. To run from source code it requires perl5.8 or above and Tkx installed.
- 3. To run larger genome sequences it may require higher configuration.
- 4. Additional software like PDF reader and Post Script Viewer may be required.









GAOPP: Genetic Algorithm for Operon Prediction in Prokaryotes

2. Working with GAOPP:

2.1 Input files:

Download the required files like .ptt file and pathway file. Note down the KEGG organism code if you are planning to use pathway data, organism code has to be specified. Check that the .ppt file and pathway file are in the following format:

Escherichia	coli	str.	K-12 substr.	MG1655	, comp.	lete ga	enome -	- 14	639675	5							-
4132 protein	ns																
Location	Strar	nd	Length	PID	Gene	Synon	ym	Code	COG	Prod	uct						
190255	+	21	16127995	thrL	b0001	-	-	thr c	peron	leade	r pept	ide					
3372799	+	820	16127996	thrA	b0002	-	COG046	50E,CC	G0527E	: fuse	d aspa	rtokin	ase :	I and	homose	erine	
28013733	+	310	16127997	thrB	b0003	-	COGOUE	33E	homos	erine	kinas	e					
37345020	+	428	16127998	thru	2000d	-	CUG049	98E 	three	nine	synths 	se					
54345530	+	259	1612/999	yaan	50003 50006	_	-	preui	cted p	rued	n nrotei	n					
6529 7959	_	476	16128001	yaax vaaJ	b0000	_	COG111	15F	nredi	cted	transr	orter					
82389191	+	317	16128002	talB	b0008	-	COG017	76G	trans	aldol	ase B	OFOCE					
93069893	+	195	16128003	mog	b0009	-	COG052	21H	predi	lcted	molybd	ochela	tase				
992810494	-	188	16128004	yaaH	b0010	-	COG158	34S	conse	erved	inner	membra	ne pi	rotein	1 assoc	ciated	ł
106431135	6	-	237 16128	3005	vaa⊍	b0011	-	COG47	355	cons	erved	proteix	n				
																	_
path:eco00010	eco:	b0114	eco:aceE ko	:K00163	ec:1.2.	.4.1											
path:eco00010	eco:	b0115	eco:aceF ko	:K00627	ec:2.3.	.1.12											
path:eco00010	eco:	b0116	eco:lpd ko:	K00382 e	c:1.8.1	1.4											
path:eco00010	eco:	b0356	eco:frmA ko	:K00121	ec:1.1.	.1.1 ec	:1.1.1.	284									
path:eco00010	eco:	b0688	eco:pgm ko:	K01835 e	c:5.4.2	2.2											
path:eco00010	eco:	b0755	eco:gpmA ko	:K01834	ec:5.4.	.2.1											
path:eco00010	eco:	b0756	eco:galM ko	:K01785	ec:5.1.	.3.3											
path:eco00010	eco:	b1002	eco:agp ko:	K01085 e	c:3.1.3	3.10											
·								-									

For promoter prediction, a promoter training set need to be specified. A Perl script provided with the program may be used to extract the promoter and non promoter training sets. Simply run script specifying your input files and sequence file. The input files have same .ptt file format. To generate the positive input file, edit the .ptt file keeping only those genes which contains upstream promoter signals, and delete others. Similarly, for negative input file only those genes not having upstream promoter sequence. Run extractProm.pl :

Perl extraxtProm.pl -pos <positive.ptt> -neg <negetive.ptt> -seq <nucl.fna> -out <output.txt>

extra	ictProm.pl 📔 prom	n_neg2.txt	prom_pos2.txt					
1	Escherichia	coli st	r. K-12 subst	r. MG1655,	complete	gen	ome - 1463	9675
2	4132 proteir	15						
3	Location	Strand	Length PID	Gene Syn	onym Code	2	COG Product	
4	190255	+ 21	16127995	thrL b00	01 -	-	thr operon	leader peptide
5	52345530	+ 98	16127999	yaaX b00	05 -	-	predicted p	rotein
6	56836459	- 258	16128000	yaaA b00	06 -	COG	30225 cor	served protein
7	65297959	- 476	16128001	yaaJ b00	07 -	COG	1115E pre	dicted transporter
8	1138211786	5 -	134 16128007	7 yaaI	b0013	-	 predict 	ed protein
9	1675116903	3 -	50 49175991	l hokC	b4412	-	- toxic n	embrane protein, small
10	1748918655	5 +	388 16128013	3 nhaA	b0019	-	COG3004P	sodium-proton antiporter
11	2023320508	3 -	91 16128016	5 insA	b0022	-	COG3677L	KpLE2 phage-like element; IS1 repressor protein InsA
12	2140722348	3 +	313 16128019	9 ribF	b0025	-	COG0196H	bifunctional riboflavin kinase/FAD synthetase
13	2837429195	5 +	273 16128025	5 dapB	b0031	-	COG0289E	dihydrodipicolinate reductase
14	2965130799	ə +	382 16128026	5 carA	b0032	-	COG0505EF	carbamoyl phosphate synthetase small subunit, glutamine amidotrans
15	3430034695	5 +	131 90111079) caiF	b0034	-	- DNA-bir	ding transcriptional activator
16	4041741931	L –	504 16128034	e caiT	b0040	-	COG1292M	predicted transporter
17	4240343173	3 +	256 90111081	l fixA	b0041	-	COG2086C	predicted electron transfer flavoprotein subunit, ETFP adenine nuc
18	5475557109	9 -	784 16128048	3 imp b00	54 -	COG	1452M exp	orted protein required for envelope biosynthesis and integrity
19	5736458179	ə +	271 16128049) djlA	b0055	-	COG10760	DnaJ-like protein, membrane anchored
20	6342965780	- 0	783 16128054	f polB	b0060	-	COG0417L	DNA polymerase II
21	6834870048	3 -	566 16128057	7 araB	b0063	-	COG1069C	L-ribulokinase
22	7038771265	5 +	292 16128058	B araC	b0064	-	COG2207K	DNA-binding transcriptional dual regulator
23	7564477299	9 –	551 16128063	8 sgrR	b0069	-	COG4533R	DNA-binding transcriptional regulator
24	8362283708	3 -	28 16128069) leuL	b0075	-	- leu ope	ron leader peptide
25	8436885312	2 +	314 90111083	3 leuO	b0076	-	COG0583K	DNA-binding transcriptional activator

Fig: Positive promoter input file

📒 ex	tractProm.pl 📙 rev_com	p2.pl	📙 readFFF.pl 🗎 p	rom_neg2.txt				
1	Escherichia co	li si	tr. K-12 subst	r. MG1655,	complete	gen	nome - 14639	9675
2	4132 proteins							
3	Location St	rand	Length PID	Gene Syn	onym Cod	e	COG Product	
4	3372799 +	820	0 16127996	thrA b00	02 -	COG	0460E,COG0527	'E fused aspartokinase I and homoserine dehydrogenase I
5	28013733 +	31(0 16127997	thrB b00	03 -	COG	0083E homo	oserine kinase
6	37345020 +	42	8 16127998	thrC b00	04 -	COG	0498E thre	eonine synthase
7	992810494 -	18	8 16128004	yaaH b00	10 -	COG	1584S cons	verved inner membrane protein associated with acetate transport
8	1064311356	-	237 16128005	yaaW	b0011	-	COG47355	conserved protein
9	1216314079	+	638 16128008	dnaK	b0014	-	COG04430	chaperone Hsp70, co-chaperone with DnaJ
10	1416815298	+	376 16128009	dnaJ	b0015	-	COG04840	chaperone Hsp40, co-chaperone with DnaK
11	1544516557	+	370 16128010	insL	b0016	-	COG3385L	IS186/IS421 transposase
12	1675116960	-	69 16128012	mokC	b0018	-	 regulato 	ory protein for HokC, overlaps CDS of hokC
13	1871519620	+	301 16128014	nhaR	b0020	-	COG0583K	DNA-binding transcriptional activator
14	1981120314	-	167 16128015	insB	b0021	-	COG1662L	IS1 transposase InsAB'
15	2081521078	-	87 16128017	rpsT	b0023	-	COG0268J	30S ribosomal subunit protein S20
16	2118121399	+	72 16128018	yaaY	b0024	-	 predicte 	d protein
17	2239125207	+	938 16128020	ileS	b0026	-	COG0060J	isoleucyl-tRNA synthetase
18	2520725701	+	164 16128021	lspA	b0027	-	COG0597MU	prolipoprotein signal peptidase (signal peptidase II)
19	2582626275	+	149 16128022	fkpB	b0028	-	COG10470	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)
20	2627727227	+	316 16128023	ispH	b0029	-	COG0761IM	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase, 4Fe-4S
21	2729328207	+	304 16128024	rihC	b0030	-	COG1957F	ribonucleoside hydrolase 3
22	3081734038	+	1073 1612	8027 car	B b00	33	 COG0458E 	CF carbamoyl-phosphate synthase large subunit
23	3430034695	+	131 90111079	caiF	b0034	-	- DNA-bind	ling transcriptional activator
24	3478135371	-	196 90111080	caiE	b0035	-	COG0663R	predicted acyl transferase
25	3537736270	-	297 16128030	caiD	b0036	-	COG1024I	crotonobetainyl CoA hydratase
26	3627137839	-	522 49175993	caiC	b0037	-	COG0318IQ	predicted crotonobetaine CoA ligase:carnitine CoA ligase
27	3789839115	-	405 16128032	caiB	b0038	-	COG1804C	crotonobetainyl CoA:carnitine CoA transferase
28	3924440386	-	380 16128033	CalA	b0039	-	COG19601	crotonobetaine reductase subunit II, FAD-binding
29	4318844129	+	313 16128036	fixB	b0042	-	COG2025C	predicted electron transfer flavoprotein, NAD/FAD-binding domain
30	4418045466	+	428 16128037	TIXC	60043	-	COG0644C	predicted oxidoreductase with FAD/NAD(P)-binding domain
31	4546345750	+	95 16128038	f1xX	00044	-	COG2440C	predicted 4re-45 ferredoxin-type protein
32	4580747138	+	443 16128039	yaaU haɗR	DUU45	-	COGU477GEPR	predicted transporter
133	4/24b4777b	+	1/6 16128040	KETF	DUU46	-	UU1-2249R	TIAVODTOLEIN SUDUNIE FOR ENE KETC DOLASSIUM ETTIUX SVSLEM

Fig: Negative promoter inputfile

0					
F:\Operon\current\prom	_align_forr.txt				
Courier New	▼ 11 ▼ B 43 total sequ	ences			
de: Select / Slide 💌	Selection:0 Position: 13: promoter_7956_8057	Sequence Mask: None Numbering Mask: None	Start ruler at: 1		
IDID 🖥		te CAT CAT	Scroll Scroll		
	10 20	40 50		90 90 100	110 120 13
anatan 92 GAGT	10 20 .	30 40 30	00 10	ATTICACIONACIONACIONACIONACIONACIONACIONACI	110 120 13
omoter 513	AGAGATTAGGATTGCGGAGAATAAC			CARTGATAAAAGGAGTAACTTGTG	
omoter 173	CGAATTATTCCTACACTATAATCTG		TRAAATAGTARAAACGATCTATT	CACCIGAAAGAGAAATAAA AGTG	
omoter 213 AAAA	GCCACTTCTTTCTGATTTCGGTACT(TTTCACTGTTTTGAGCCAG. CATG	
omoter 282	TAATTTCTAATTATCAGCGTTTTTG		TTACTCTGAAAACGGTCTATGCA	AATTAACAAAAGAGAATAG TATG	
moter 295 TGCC			ATTAATATGCAAATAAAGTGAGT	GAATATTCTCTGGAGGGTG TTTG	
moter 342 SCAD	ACTATCGATATATCCACAATTTTAA		AACGAGTCATAGCCAGACTTTTA	ATTTGTGAAACTGGAGTTCGTATG	
omoter 423 TGGT	GATTAAAGTTTTATTTCAAAATTAA			TTCATGATTTCTGGAGATG AATG	
omoter 572 GGCG	ATGTCCCACAATTGACCGCAGCCGG			GTGTGTCGATTTGGGGAAT TATG	
omoter 702 GGTT			ATGGACAATTGGTTTCTTCTCTG	AATGGTGGGAGTATGAAAA TATG	
omoter 842 TTTC	TTATTTTATATGCATGATAAATCAT		TTCCAATAAGGGAAAGGGAGTTA	AGTGTGACAGTGGAGTTAAGTATG	
omoter 645 CGCG	TGGGATGATGTTTCGCAGGAGTAAT		CATTTTTGCTAAAGTCGGCATA	AATTTCCTGCAAGGACTGG. TATG	
omoter 795 ChGA	AGTTATCAAGTACCTOGTAGOGTATI			CCGGCAATAAGAGGGATAT	
omoter 117 TAAT	AAAAAGAAAAGATTACGTGCCTGAA			CTTCCCACTAAGAGAATCC TATG	100m - Although
omoter 169 GCCC	CGAGTGATATTTTACCATCAACCOG		GTAGAGTGCCTCTTACTGACCGT	AAGGTCAAGGAGAAGAGAG	13: promoter_7956_805
omoter 205 TCCC	ccacaaagaat a tggatattgtgati		CAACTTACTGATTTAGTGTATGA	TGGTGTTTTTGAGGTGCTC AGTG	21
omoter 419 GCGT	CCACGAGGTTAATAATAATTATT!		CGGGAACGCAAAATAAATATTCG	TTTTCACAGTGGAATTAAC CATG	
omoter 571 GGAC	aggattaacactagcgtagagatga		TGTCACGCGCAACGTTACCGATG	atggaacaataa a atcaac <mark>statg</mark>	
omoter 657 CGAA				GTTTTTTGATGGATTTTCAGCGTG	
omoter 700 GCAT	TTTTATCCATAAGATTAGOGGATCC'		ototactgtttctccatacccgt	TTTTTTGGATGGAGTGAAA GATG	
omoter 772 CTGA				TTGCACAGGAGTTCCCCTT TATG	
omoter 837 CICA	CGCTGGCGGGGCGGTTTTTATTGCGT		CCAGTAACTCTAAAAGCATATCG	CATTCATCIGGAGCIGATI AAIG	
npromoter TAAC	GGTGCGGGCTGACGCGTACAGGAAA(AAAGGTAACGAGGTAACAACATG	
promoter TTGG	TACTGCGCGGGATATGGTGCGGGCAA!			CATGGAAGITAGGAGICICICATG	
promoter MGCC				GCACGAGTACTGGAAAACTHAATG	
npromoter []GCA	RAAAAATGAAATTGGGCAGTTGAAA			ATATATAGTGGAGACGTTT.GATG	
npromoter AAAA	AATAATCGCCCTATAAACGGGTAAT!			TCTAGGGGCAATTTAAAAAAGATG	
npromoter TTGA				CTGATATGATTTAACGTGC GATG	
[1] [1] [2] [2] [2] [2]			CACLACCOMPLELACCOMPCC/	WCATWGTWATCAGGGAGAG AATG	1

Your file is ready if all the sequences must have A(G)TG at the right side.

In order to generate the terminator coordinates, we have provided a compiled transferm binary executable and *expterm.dat* file. This will run **only on Linux** platform (see transTerm usage file) Run the following command on Linux Terminal:

transterm -p expterm.dat seq.fasta annotation.ptt > output.tt

Remember to keep name of .ptt file and FASTA identifier in sequence file, exactly the same. And provide the sequence file earlier than .ptt file as the command line argument. The output file is written after '>'.

To load the input files click on the respective buttons and click on browse to load the files. Providing incorrect files causes anonymous error or result may be ambiguous.

GAOPP::Operon Predictor				
le Help				
Input files	SPECIFY INPUT FILES			
1	Load .PTT file	:	F:/Operon/new/GUI/protein	ptt
GA Paramtetrs	Load Pathway file	:		
	KEGG Organism code	:	(e.g eco for E.coli)	12/2
Fitness Function	Whole Genome(.fna in F	ASTA)	4	
	Promoter Training set(0	ntional)		
	i tomoter maning set(o	ptionaly		
4 Start	TransTerm output file(O	ptional)		
	(the own			
E Aleruit	Tr Open			
	Look in:	🔒 GUI	· · · · · · · · · · · · · · · · · · ·	G 🤌 📂 🖽 🕶
Graphical	Q.	Name	*	Date modified
		IA 🔒		4/26/2012 1:10 PM
10	Recent Places	ECO_result		5/18/2012 12:54 PM
is and Progress :		🔒 figgl		5/19/2012 11:46 AM
WELCOME TO GAOPP***		FigW		5/19/2012 11:47 AM
zy rules Ready	Desktop	GAOPP		5/18/2012 12:49 PM
		GUI linux		5/18/2012 10:09 PM
		Icons		5/13/2012 4:58 PM
	Libraries	images		5/12/2012 6:39 PM
		i result		5/17/2012 12:07 PM
		14 x 6 in. (2)		5/18/2012 5:23 PM
	Commuter	5 x 7 in (2)		5/18/2012 5:23 PM

2.2 Genetic Algorithm Parameters:

Clicking on GA parameters button opens the parameter panel:

	Crossing Over Probability Mutation Probability	:	0.95
	Selection :	Roulet -	0.00
	Inn. Population Size :	10 -	
	Early termination if	0 -	top Indv. have same Score. 📃
	No of Iteration	: 4	
Ê			
- F			

2.2.a Operator Probability:

To implement genetic algorithm operators like Mutation and crossing over user need to set the probability. The probability indicates how often the operon has to be implemented. Generally a high cross over probability and low mutation probability combination gives optimized result. Use the sliders to adjust the probability.

2.2.b Selection:

A selection procedure selects an individual solution to be act as a parent for crossing over and generate offspring for next generation. There are two options for selecting the parents i. Roulette Wheel Selection ii. Best Individual selection.

i. Roulette Wheel Selection:

It selects an individual stochastically form the current generation by simulating rotation of a wheel with an objective to select the fittest individual. During the process individuals having higher fitness score has higher probability to get selected in comparison to less fit individuals.



ii. Best Individual: This method selects only the best individual from the generation.When user opts for this option, a higher mutation probability is advisable.

2.2.c Early Termination:

On attaining the best plausible solution, all the individuals will look much alike and mutation and crossing over does not make any change to the population. Hence continuing the process is worthless. Click on this the check box if user wants to terminate the evolution process when specified number of individuals in the current generation has same score.



Initial Population Number must be higher than the number of individuals checked for early termination.

2.2. No. of Iterations:

This option explicitly specifies how many generations are to be evolved to find the best possible solution. Set this option as per your convenience. Until and unless early termination is not defined the program will run until the specified generation.

2.3 Fitness Function:

Click on Fitness Function Button to change the fitness function. Selecting a fitness function gives the literature reference used for calculating the score.

Fuzzy Fitness Finder (Jacob et.al) function takes a long run about 10-12 hrs for whole genome. Remember to set early termination option when FFF is used.

Rule based Fitness function is a heuristic one and can be used for quicker evaluation and doesn't guarantee better prediction.

2.4 Result Visualization:

Optimization process starts when start button is clicked. Like most standard GA software average fitness score in each generation plotted. This shows a uprising curve for successful optimization process. If the cure is not reliable (not uprising) user need to adjust the probabilities and run the program again.

Click on export button to save the plot in postscript format (.ps) to view it later in any post script viewer like **ghostviwer**. Otherwise the *progress.xls* file can be open after the run and select the two columns and plot using XY sctter.





GAOPP: Genetic Algorithm for Operon Prediction in Prokaryotes

Operon clusters along with their corresponding scores are displayed in the result panel when Result in Text Button is clicked. Result exported to hard disc.

A Graphical viewer has been designed to represent individual operon clusters along with the promoter and terminator signals. The list of operons is displayed on the top. Selecting an cluster displays its total score at the bottom of list. Double clicking on a particular entry loads the entire operon map with terminator and promoter signals. Map in postscript format can be exported.

				The Operons	of the organism in cont	ext are as follows:		
Input files	****	*******	*PREDICT	OR RESUL	[************			- A
GA Paramtetrs	OPER	ON ID		GENES				
Fitness Function	1	b0001	b0002	b0003	b0004	4.422 1.872	2e-008	
Start	2	b0005		0				E
Text Result	3	b0006 b0008	b0007 b0009		-1.287 -0.8224	1 2e-008 24 2e-008	-2.10939998 -3.13605998	
Graphical	5	b0010	b0011		-3.60768 1.872	2e-008	-1.73567998	
tatus and Progress :	6	b0013		0				
rossing 9 and 9	. 7	b0014	b0015		-1.11774 1.872	2e-008	0.75426002	
rossing 9 and 9 rossing 9 and 9 rossing 9 and 9 inished	E	b0016		0				+ Save

le Help	
Imput files	state chiracture chira
GA Paramitetrs	-talB-mog -yaaH
Fitness Function	
Predict i	The Score of -thrt-thrA-thrB-thrC(1) is 50 22 0.5 Graphics Zone
Text Result	
Status and Progress: Crossing 7 and 5 Crossing 0 and 0 Mutating Finished Termination file processed	Operan Score = 50 Terminator found Promoter found Promoter sign: tho-depender
	Export

Fig: Output panel: Result in Text form (A) and Result in Graphical (B). Graphical Result Shows visualizes regulatory signals.

3. Algorithm:



GAOPP: Genetic Algorithm for Operon Prediction in Prokaryotes



GAOPP: Genetic Algorithm for Operon Prediction in Prokaryotes

Evaluation:

We used GAOPP for available test sets like *Escherichia coli* K-12 substr-MG1655and *Bacillus subtilis*. We created positive and negative gene pairs from available experimental data. The predicted operons were compared with these available test set. From these observations we constructed Receiver operating curve.





Reference: