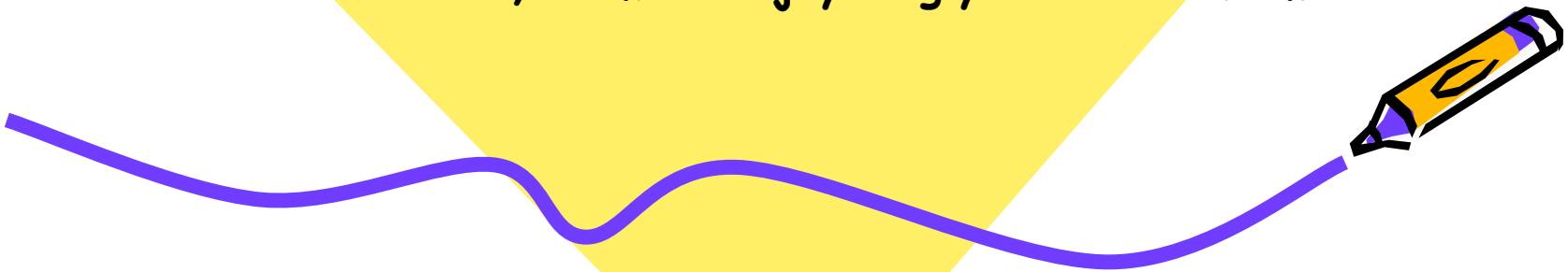
PredictBias: A server for the identification of genomic and pathogenicity islands in prokaryotes

Sachin Pundhir, Hemant Vijaywargiya and Anil Kumar



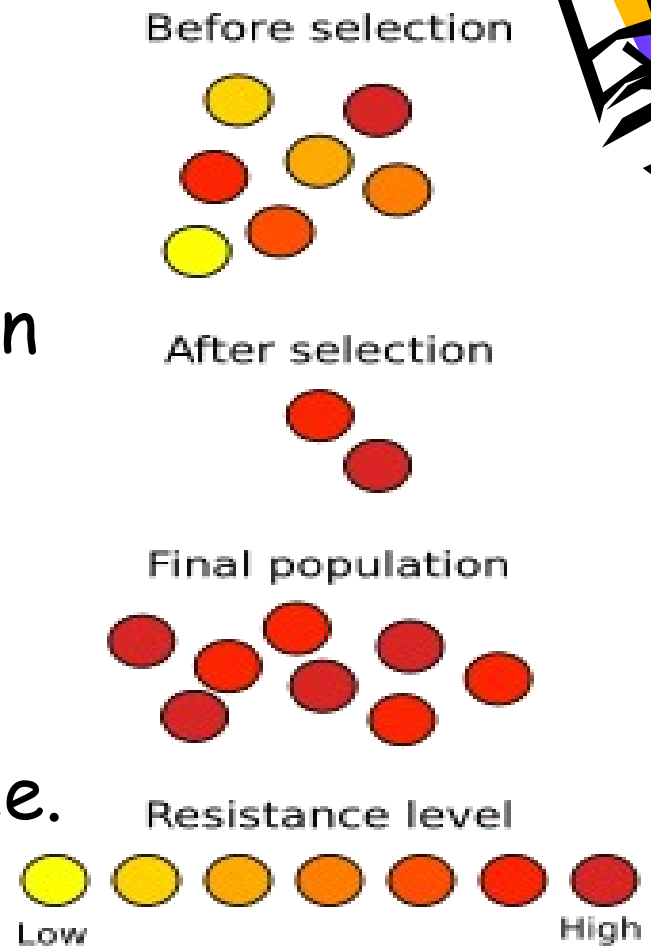
Bioinformatics sub-centre, School of Biotechnology,
Devi Ahilya University, Indore- 452001, INDIA

Antibiotic resistance

- **Antibiotic resistance** is the ability of a micro-organism to withstand the effects of an antibiotic.
- Penicillin, discovered by Alexander Fleming in 1928 crippled many types of disease-causing bacteria.
- In 1980s, the perception was that we had overcome the bacterial infection problem. Drug companies weren't working on new agents. They were concentrating on other areas, such as viral infections.
- As compared to 0.02% of *S. pneumoniae* resistant to antibiotics, today 6.6% of them are resistant.

Why is it happening?

- Consequence of evolution via natural selection.
- The antibiotic action acts as an environmental pressure
- those bacteria which have a mutation allowing them to survive will live on to reproduce.



Measures to combat

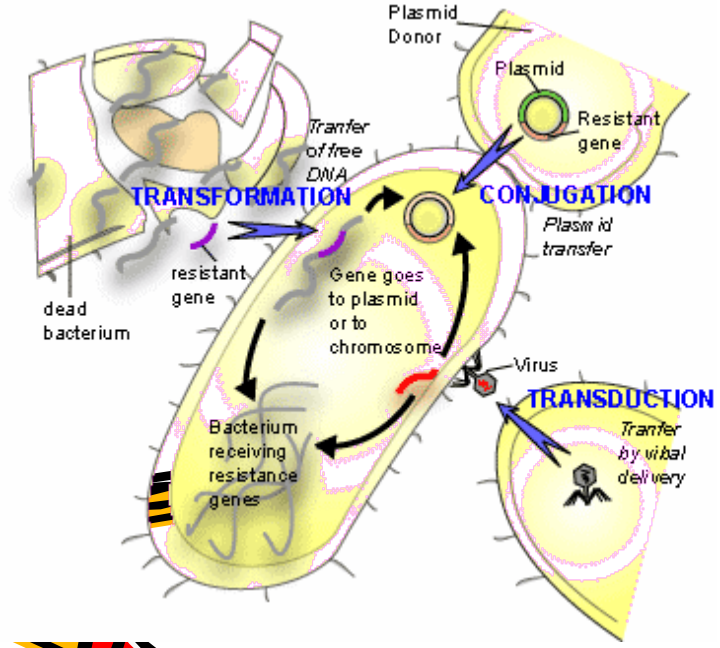
- Antibiotic resistance is inevitable, but there are measures to make it slow.
- Repeated and improper uses of antibiotics are primary causes of the increase in drug-resistant bacteria.
- Do not take an antibiotic for a viral infection like a cold or the flu.
- Do not take expired medicines & do not skip doses
- Identify novel drug targets.

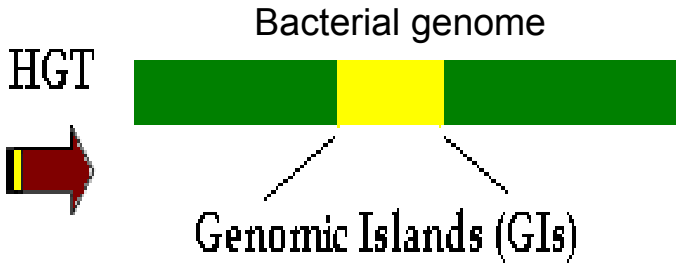
Novel drug targets

- Daunting task considering the huge amount of microbial genomic and proteomic data available in various public repositories like Genbank, TrEMBL and others.
- Whole genome sequence of more than 500 microbial genomes is available at NCBI's genome database and sequencing of more than 1000 microbes is in progress.
- *Escherichia coli CFT073* alone encodes for 5379 proteins.
- Challenge lies in finding subset of proteins encoded in a microbial genome that are vital for the pathogenicity of a pathogenic bacteria.

Horizontal gene transfer (HGT)

- process in which an organism transfers genetic material to another cell that is not its offspring.
- may account for 10 to 50% of all the genes in a bacterial or archaeal genome.





If involved in……

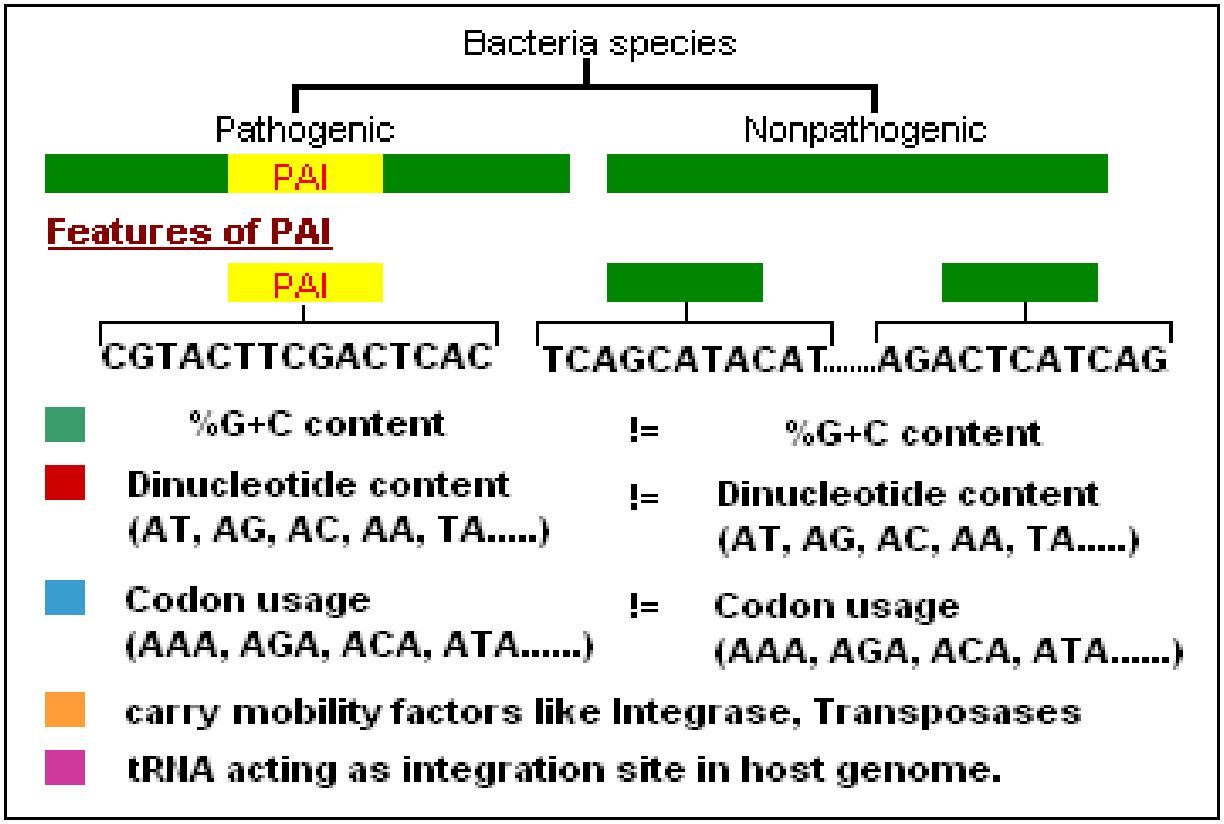
Fitness – Fitness islands

Resistance – resistance islands

Pathogenicity – Pathogenicity islands

What are PAIs?

- Pathogenicity islands are distinct chromosomal regions of pathogenic bacteria that contain genes encoding virulence factors viz. adhesins, toxins and invasions and contribute to the virulence of the respective pathogen.



PAI properties

Significant composition bias

Relative absence in non-pathogenic sp.

Presence of Insertion elements like tRNA, Integrases & Transposase.

Presence of ORFs encoding proteins responsible for virulence (Virulence factors).

Can we identify PAIs computationally?

- Yes, many prediction servers like ISLANDPATH, PAIDB are available online for the prediction of PAIs.

- But how can we relate biological features of PAIs to computational algorithm scheme for PAI identification?

1. Composition bias analyses

- **%GC bias**

  %GC Bias (Cluster) = %GC (Cluster) - %GC (Genome)

- Dinucleotide bias

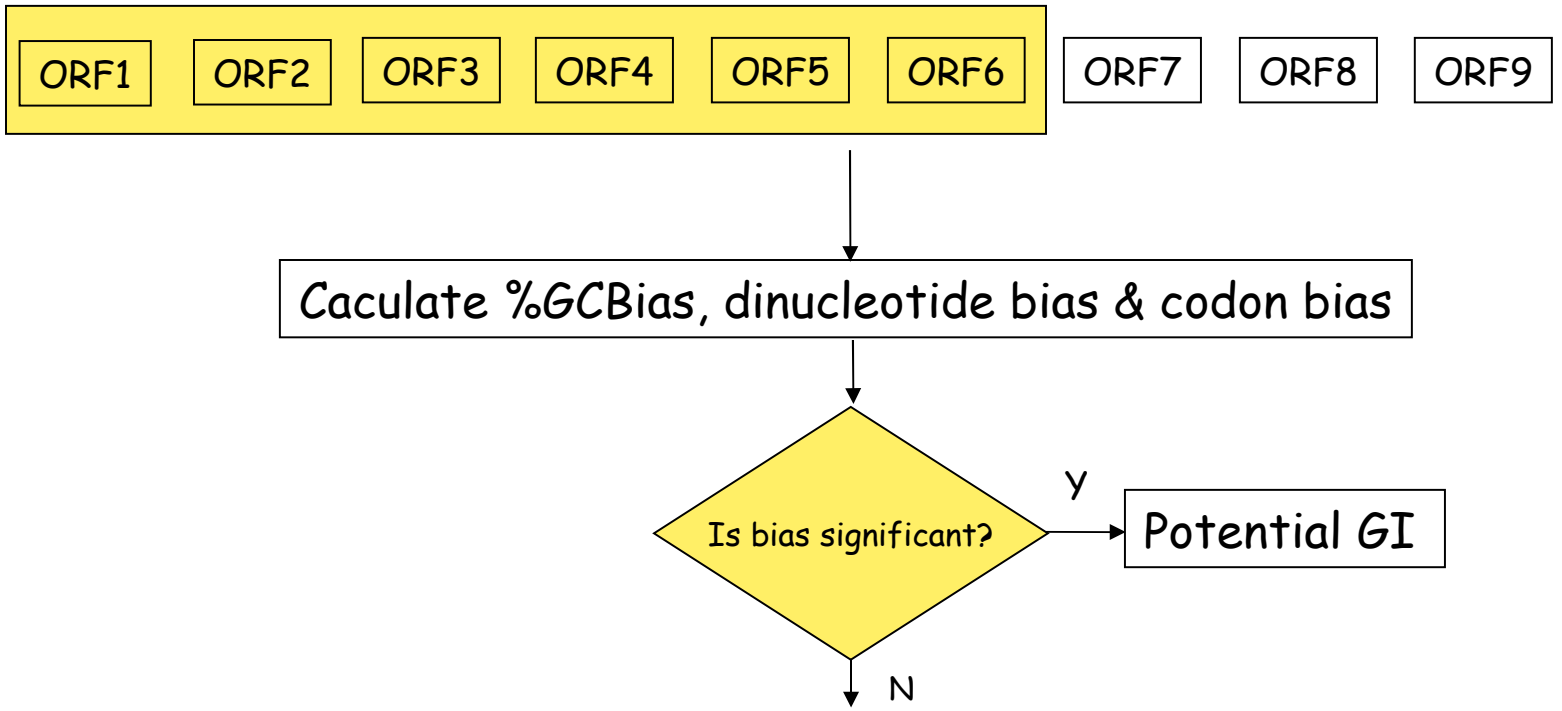
$$\delta^{*}(f,g) = \frac{1}{16}\sum \left| \rho^{*}_{xy}(f) - \rho^{*}_{xy}(g) \right|$$

- Codon bias

$$B(F \mid G) = \sum_{a} p_a(F) \left[\sum_{(x,y,z)=a} \left| f(x,y,z) - g(x,y,z) \right| \right]$$

Reference: Karlin, S. (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. Trends Microbiol. **9**, 335-343.

Sliding window approach



Host specific genome segment

1. Insertion elements

- Can be identified by keyword search of genome file (Genbank or EMBL format) for the presence of tRNA, transposase and integrase or using software like tRNAScan.

```
gene            complement(125063..125186)
                /locus_tag="APEtRNA01"
                /db_xref="GeneID:1445697"
tRNA            join(complement(125063..125101),
                complement(125151..125186))
                /locus_tag="APEtRNA01"
                /db_xref="GeneID:1445697"
                /product="tRNA-Thr"
                /anticodon=(pos:125152..125154,aa:Thr)
                /note="codon recognized: ACG; anticodon:CGT"
intron          complement(125102..125150)
                /locus_tag="APEtRNA01"
gene            125249..125635
                /locus_tag="APE0167"
                /db_xref="GeneID:1445698"
CDS             125249..125635
                /locus_tag="APE0167"
                /protein_id="NP_147016.1"
                /transl_table=11
                /db_xref="GI:14600500"
                /db_xref="GeneID:1445698"
                /codon_start=1
                /product="hypothetical protein"
                /translation="MESGSWRPPFSTGKIVGNYGLLKLYLEVAREKGRRDLVDKALIS
                EDDVDMLRRLSASPGATAEDFVNALEERFVERVDPEVASEALARAGINVDGDTARRMI
                ARILAGWLVEMGEEMKLYRLRRSWED"
```

Available applications……

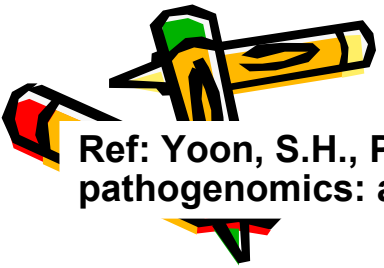
- ISLANDPATH (Composition based):
  - Based on composition bias, &
  - presence of insertion elements.

  Limitation:
  Give much false positive results for PAIs as a region showing distinct nucleotide content may be alien to the host genome (HGT/ GIs) but may not necessarily be involved in pathogenicity

Ref: Hsiao, W., Wan, I., Jones, S. J. and Brinkman, F. S. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. Bioinformatics 19, 418-420.

- PAIDB (Composition and similarity based):
  - Based on composition bias
  - presence of insertion elements, &
  - similarity search against known PAIs.

  Limitation:
  The detected PAIs are limited by the query dataset of known PAIs.

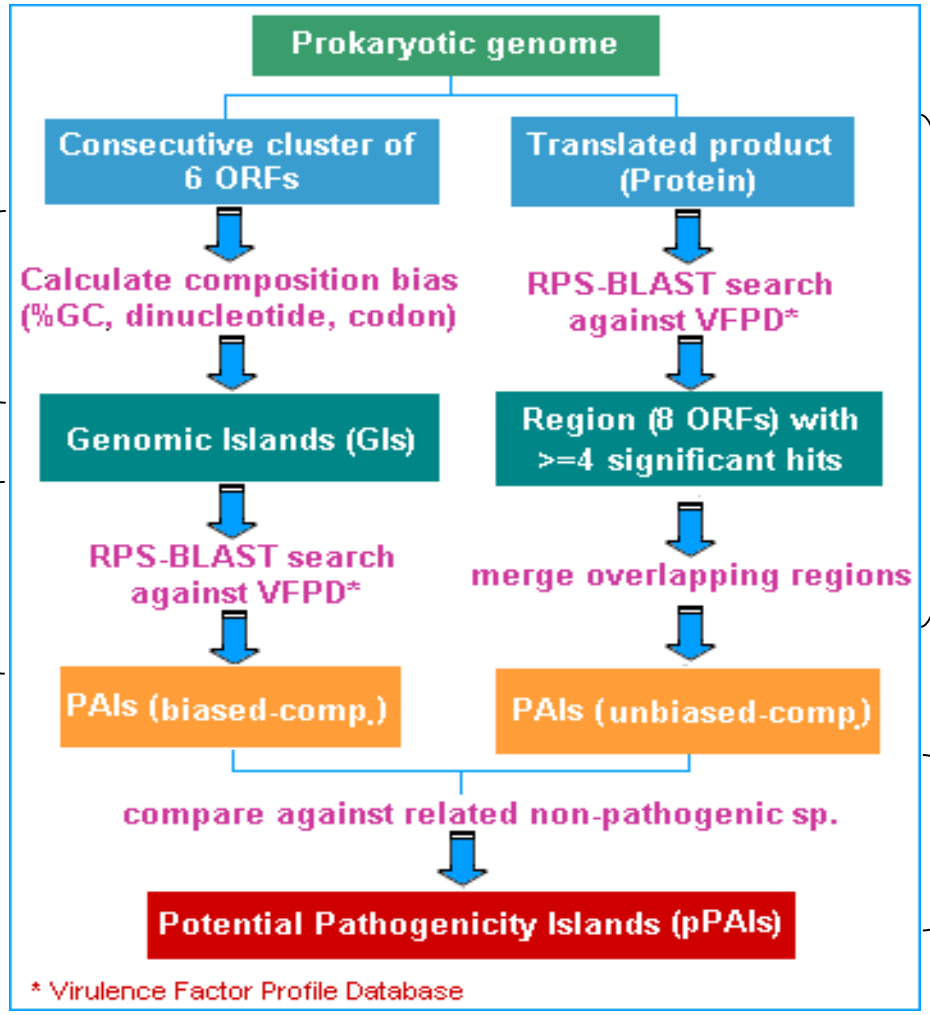


Ref: Yoon, S.H., Park, Y.K., Lee, S., Choi, D., Oh, T. K., Hur, C. G. and Kim, J. F. (2007) Towards pathogenomics: a web-based resource for pathogenicity islands. Nucleic Acids Res. 35, D395-D400.

What can be done?

- Some of the important features that can aid in validating the role of a region in pathogenicity are:
  - its relative absence in related non-pathogenic species
  - presence of virulence factors in the potential PAI region

Our approach!!!



Evaluate regions with significant bias

Identify potential virulence factors in a GI

Identify potential PAIs with insignificant bias

Comparative genome analysis

PredictBias **RESULTS** | Auto analysis | Manual analysis | Help | Genomes Analyzed

Change threshold

Bias analysis

Dinucleotide bias dev.: 2 (default) | Codon bias dev.: 5 (default) | %GC bias: 3.5 (default) | Go

Compare a potential Genomic or Pathogenicity Island in related non-pathogenic sp. (Phylogenetic tree)

Select Organism: Acidiphilium cryptum JF-5 | Compare

Note: The results of genome comparison greatly depends on the phylogenetic distance between the pathogenic and non-pathogenic sp. Smaller the distance, more relevent will be the results.

Potential GIs and PAIs in Escherichia coli 536 (NC_008

S.No	Start	End	Bias	Virulence
1	ECP_0113	ECP_0121	Y	Y
2	ECP_0142	ECP_0147	Y	N

LocusTag	DNBias	CDNBias	%GCBias	Product
ECP_0113	5	11	-8.180293	putative colicin
ECP_0114	6	8	-8.097082	colicin immunity protein
ECP_0115	4	9	-7.618726	uropathogenic specific protein
ECP_0116	3	9	-1.547990	colicin immunity protein
ECP_0117	4	17	0.489055	uropathogenic specific protein
ECP_0118	5	20	0.986028	colicin immunity protein
ECP_0119	4	13	1.985453	transcriptional regulator, GntR family
ECP_0120	3	17	2.312056	hypothetical protein
ECP_0121	3	16	2.509363	pyruvate dehydrogenase E1 component

ORFs having significant similarity with Known Virulence factors

LocusTag	Hits	Score	E-value	Comments
ECP_0113	PYOCINKILLER	182	4e-048	Pyocin S killer protein signature.

```
>PYOCINKILLER#Pyocin S killer protein signature.
          Length = 617

 Score =  182 bits (463), Expect = 4e-048
 Identities = 92/253 (36%), Positives = 125/253 (49%), Gaps = 21/253 (8%)

Query: 343 GEGTPYENVRVANMQWNEQTQRYEFT---PAHDVDGPLITWTPENPEHGNVPGHTGN--D 397
           G   P + V V    +N  T  YE T      +    ++TWTP +P     P  T
Sbjct: 377 GVSVP-KAVPVRMAAYNATTGLYEVTVPSTTAEAPPLILTWTPASPPGNQNPSSTTPVVP 435

Query: 398 RPPLDQPTILVTPIPDGTDTYTTPPFPVPDPKEFNDYILVFPAGSGIKPIYVYLKEDPRK 457
           +P        +TP+       T   +P         D I+ FPA SGIKPIYV   + DPR
Sbjct: 436 KPVPVYEGATLTPV-----KATPETYPGVITLP-EDLIIGFPADSGIKPIYVMFR-DPRD 488

Query: 458 LPGVVTGHGVPLSPGTRWLDMSVSNNGNGAPIPAHIADKLRGREFKTFDEFREALWLEVS 517
           +PG  TG G P+     WL    ++ G GAPIP+ IADKLRG+ FK + +FRE  W+ V+
Sbjct: 489 VPGAATGKGQPV--SGNWLG--AASQGEGAPIPSQIADKLRGKTFKNWRDFREQFWIAVA 544

Query: 518 QDPELIAQFSSGNQTRIKQGLTAKAPIDGWHYGPKDIVKKFQIHHRVAIEYGGSVYDIDN 577
            DPEL  QF+ G+   ++ G            G +    K +IHH+V +  GG VY++ N
Sbjct: 545 NDPELSKQFNPGSLAVMRDGGAPYVRESE-QAGGR---IKIEIHHKVRVADGGGVYNMGN 600

Query: 578 LRIVTPRLHDEIH 590
           L  VTP+ H EIH
Sbjct: 601 LVAVTPKRHIEIH 613
```

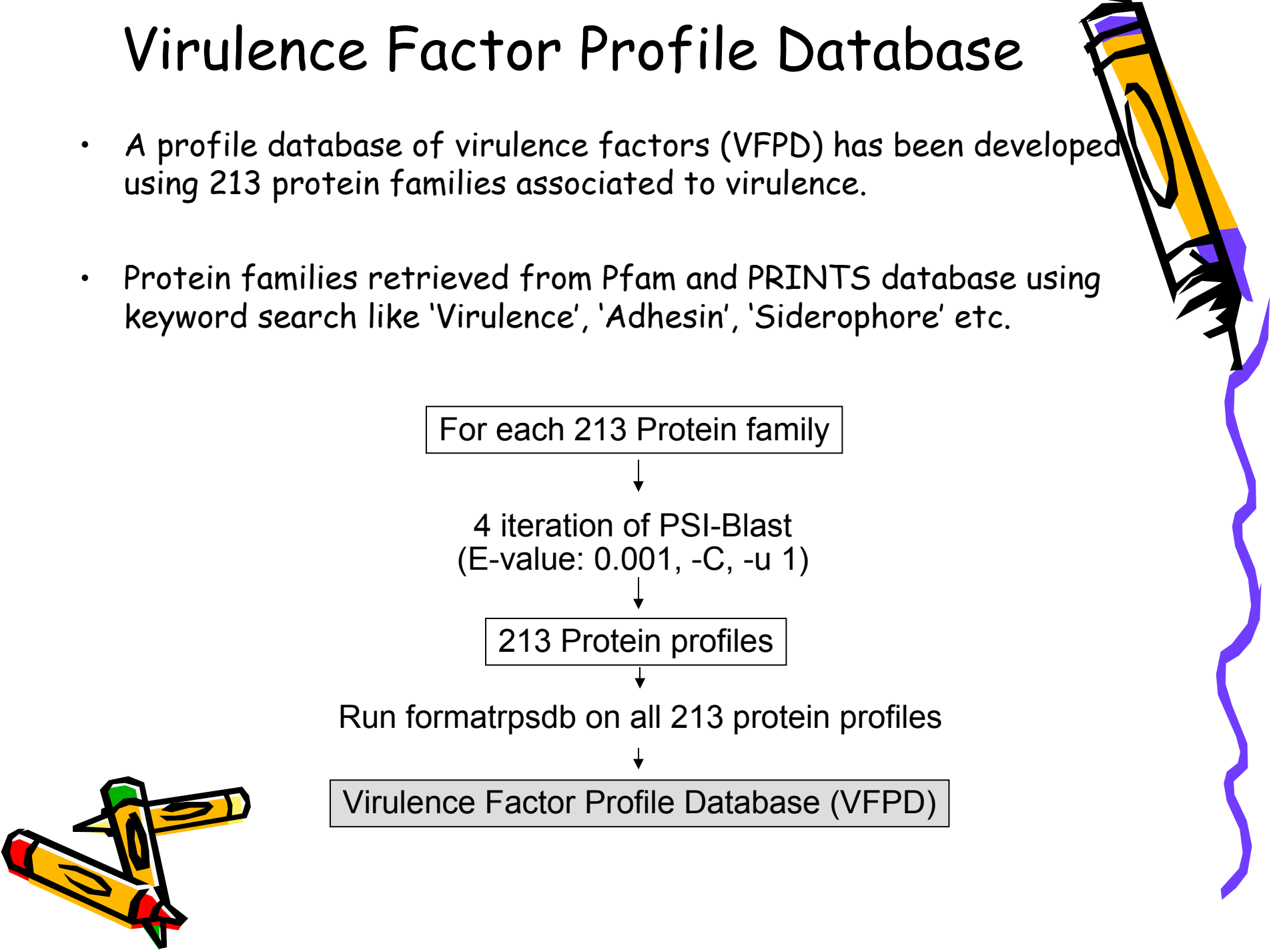
For determining the threshold values, 73 GIs distributed across 29 bacterial organisms and listed in Islander database were studied.

Organism	%GC	Group	Mean $B(F\|G)$ deviation[a§]	Mean $\delta^*(f, g)$ deviation[b§]	Mean %GC bias[c§]
Staphylococcus epidermidis	32	Firmicutes	3.31	2.53	6.49
Lactococcus lactis	35.3	Firmicutes	3.85	1.56	1.02
Streptococcus mutans	36.8	Firmicutes	11.89	2.25	4.74
Enterococcus faecalis	37.4	Firmicutes	3.08	1.09	4.27
Listeria innocua	37.4	Firmicutes	3.87	2.21	1.71
Haemophilus influenzae	38.1	Gammaproteobacteria	4.47	0.53	0.31
Nostoc sp. PCC 7120	41.3	Cyanobacteria	3.21	2.18	2.98
Bacteroides thetaiotaomicron	42.9	Bacteroidetes	8.05	2.64	4.31
Bacillus subtilis	43.5	Firmicutes	10.63	2.04	8.26
Lactobacillus plantarum	44.4	Firmicutes	6.01	2.07	4.45
Vibrio parahaemolyticus	45.4	Gammaproteobacteria	6.46	3.88	3.28
Shewanella oneidensis	45.9	Gammaproteobacteria	3.95	1.64	3.3
Yersinia pestis KIM	47.7	Gammaproteobacteria	15.16	2.95	5.75
Escherichia coli CFT073	50.5	Gammaproteobacteria	8.71	3.4	2.95
Shigella flexneri 2a str. 2457T	50.9	Gammaproteobacteria	9.73	3.15	3.35
Salmonella enterica Typhi Ty2	52.1	Gammaproteobacteria	9.86	4.16	3.19
Xylella fastidiosa	52.6	Gammaproteobacteria	29.64	5.07	12.18
Brucella melitensis	57.2	Alphaproteobacteria	18.15	3.32	5.68
Agrobacterium tumefaciens	59	Alphaproteobacteria	15.87	4.1	4.31
Bifidobacterium longum	60.1	Actinobacteria	7.12	2.02	3.26
Sinorhizobium meliloti	62.2	Alphaproteobacteria	14.13	3.53	4.06
Mesorhizobium loti	62.5	Alphaproteobacteria	11.07	2.62	4.06
Corynebacterium efficiens	63.1	Actinobacteria	13.11	5.98	3.27
Bradyrhizobium japonicum	64.1	Bradyrhizobium japonicum	17.58	2.82	6.28
Xanthomonas campestris	65.1	Xanthomonas campestris	11.63	2.87	3.97
Pseudomonas aeruginosa	66.6	Pseudomonas aeruginosa	20.07	3.47	5.71
Deinococcus radiodurans	66.6	Deinococcus radiodurans	18.67	7.71	2.51
Ralstonia solanacearum	67	Betaproteobacteria	8.42	1.79	2.59
Streptomyces coelicolor	72	Actinobacteria	16.05	1.58	4.18

Virulence Factor Profile Database

- A profile database of virulence factors (VFPD) has been developed using 213 protein families associated to virulence.

- Protein families retrieved from Pfam and PRINTS database using keyword search like 'Virulence', 'Adhesin', 'Siderophore' etc.



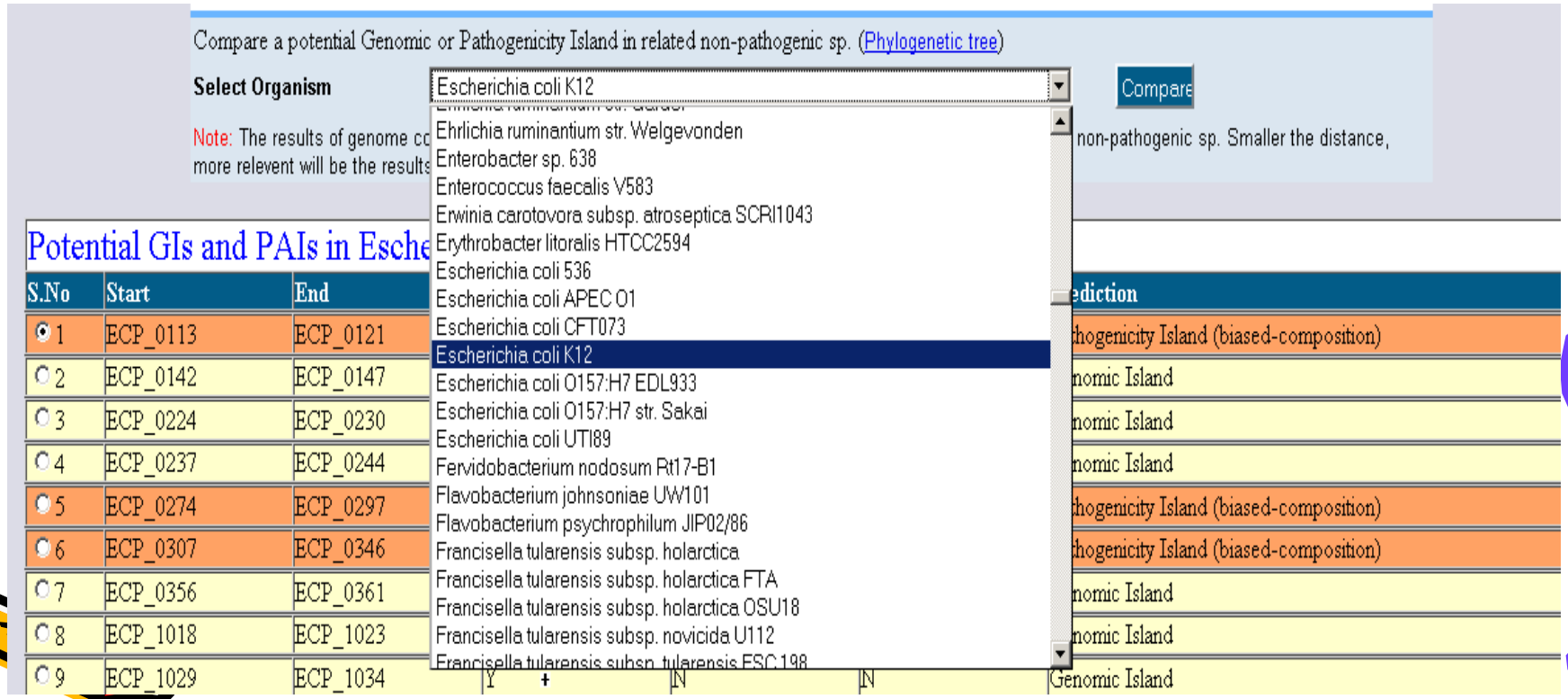
- Each region with significant composition bias is searched against VFPD and marked as potential PAI, if shows significant similarity in VFPD.

P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
1	Y	7 - Y	-3	-5	2	-2	1	1	3	-2	0	-4	0	-3	6	0	-3	-2	-3	1	-4	-4
2	S	8 - T	0	-4	2	-1	0	-2	1	-4	0	1	3	3	-3	-3	-2	-3	0	1	-3	-1
3	C	9 - C	-2	-5	-3	-3	-3	-3	-4	-4	-5	11	-3	-3	-4	-5	-5	-5	-5	-6	-6	-6
4	D	10 - N	-1	-3	-5	-5	-5	-4	-5	-6	-4	-5	-2	-1	-4	4	-2	3	0	-1	6	-1
5	G	11 - E	0	4	1	-2	-1	-3	-4	-4	-4	4	-1	0	0	1	-3	3	-3	-4	-3	0
6	C	12 - C	-2	-5	-3	-3	-3	-3	-4	-4	-5	11	-3	-3	-4	-5	-5	-5	-5	-6	-6	-6
7	L	13 - K	-3	3	-1	2	-1	-2	-4	-4	-4	-4	0	-3	-4	1	2	-3	1	0	-1	-1
8	K	14 - H	-1	1	0	-1	-1	2	-3	-4	-4	-4	1	1	0	1	1	1	2	-2	-3	0
9	P	- -	-1	0	-3	-1	-2	-3	-3	-4	5	-4	-2	0	1	1	-2	2	-2	-3	1	3
10	I	15 - H	-3	-5	6	1	1	-1	2	-4	-5	-3	-1	-3	-2	-1	-4	2	-4	-4	-5	-4
P	C	Master	A	G	I	L	V	M	F	W	P	C	S	T	Y	N	Q	H	K	R	D	E
11	V	16 - V	0	-4	2	-1	1	3	-2	5	1	-3	0	1	2	-1	0	-3	-3	-1	-1	-3
12	G	17 - E	-2	5	-5	-5	-4	-4	-5	-4	2	-4	-2	-3	-4	1	-2	-3	-2	-1	-2	3
13	V	18 - T	-1	-4	0	-1	2	0	0	-4	3	-4	0	0	0	0	-2	1	1	2	-3	1
14	R	19 - R	-3	-5	-3	-3	0	1	1	-3	-4	-5	-3	-3	2	-3	0	-2	0	7	-4	-2
15	Y	20 - W	-4	-5	1	-1	0	1	5	7	-5	-4	-4	-4	6	-5	-4	-2	-4	-4	-6	-5

Compare genome analysis

- Potential PAI regions having significant composition bias and virulence factors are compared in non-pathogenic species using BLAST family of programs (BLASTp).

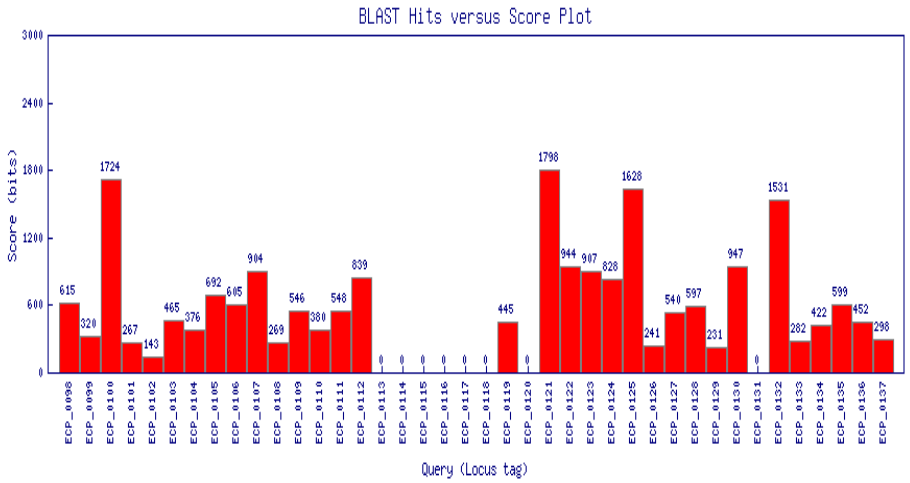
- Regions having no corresponding proteins in non-pathogenic species are marked as PAIs else marked as false positives and thus eliminated from the final results.



A

Compare Genome Results

s feature is based on Blastp version of BLAST family of programs [1] and is meant to investigate the relative arrangement of a query region in non-pathogenic species. It should not be assumed as
ome comparisons.



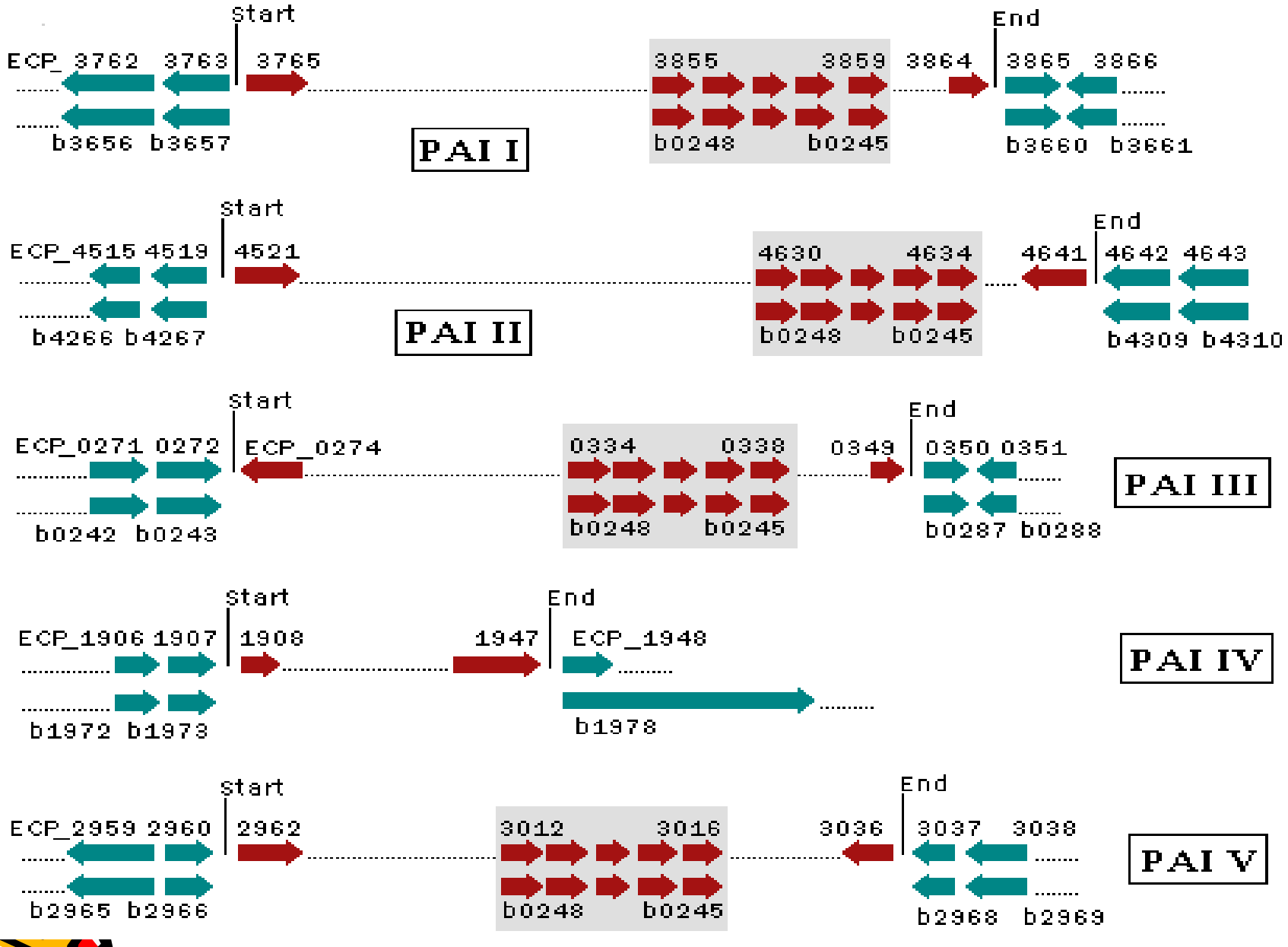
..Previous

Next..

B

S.No.	Query	Hit (1st)
1	ECP_0098	b0096@95
2	ECP_0099	b0097@96
3	ECP_0100	b0098@97
4	ECP_0101	b0099@98
5	ECP_0102	b0101@99
6	ECP_0103	b0102@100
7	ECP_0104	b0103@101
8	ECP_0105	b0104@102
9	ECP_0106	b0106@103
10	ECP_0107	b0107@104
11	ECP_0108	b0108@105
12	ECP_0109	b0109@106
13	ECP_0110	b0110@107
14	ECP_0111	b0111@108
15	ECP_0112	b0112@109
16	ECP_0113	--
17	ECP_0114	--
18	ECP_0115	--
19	ECP_0116	--
20	ECP_0117	--
21	ECP_0118	--
22	ECP_0119	b0113@110
23	ECP_0120	--
24	ECP_0121	b0114@111
25	ECP_0122	b0115@112
26	ECP_0123	b0116@113
27	ECP_0124	b0117@114
28	ECP_0125	b0118@115
29	ECP_0126	b0119@116
30	ECP_0127	b0120@117
31	ECP_0128	b0121@118
32	ECP_0129	b0122@119
33	ECP_0130	b0123@120
34	ECP_0131	--
35	ECP_0132	b0124@121
36	ECP_0133	b0125@122
37	ECP_0134	b0126@123
38	ECP_0135	b0127@124
39	ECP_0136	b0128@125
40	ECP_0137	b0129@126

1. Based on blastp version of BLAST family of programs.
2. Compares potential PAI regions against local BLAST database of microbial proteome


Start
ECP_3762 3763 3765
b3656 b3657
3855 3859 3864
b0248 b0245
End
3865 3866
b3660 b3661
PAI I
Start
ECP_4515 4519 4521
b4266 b4267
4630 4634 4641
b0248 b0245
End
4642 4643
b4309 b4310
PAI II
Start
ECP_0271 0272 ECP_0274
b0242 b0243
0334 0338 0349
b0248 b0245
End
0350 0351
b0287 b0288
PAI III
Start
ECP_1906 1907 1908
b1972 b1973
1947
End
ECP_1948
b1978
PAI IV
Start
ECP_2959 2960 2962
b2965 b2966
3012 3016 3036
b0248 b0245
End
3037 3038
b2968 b2969
PAI V

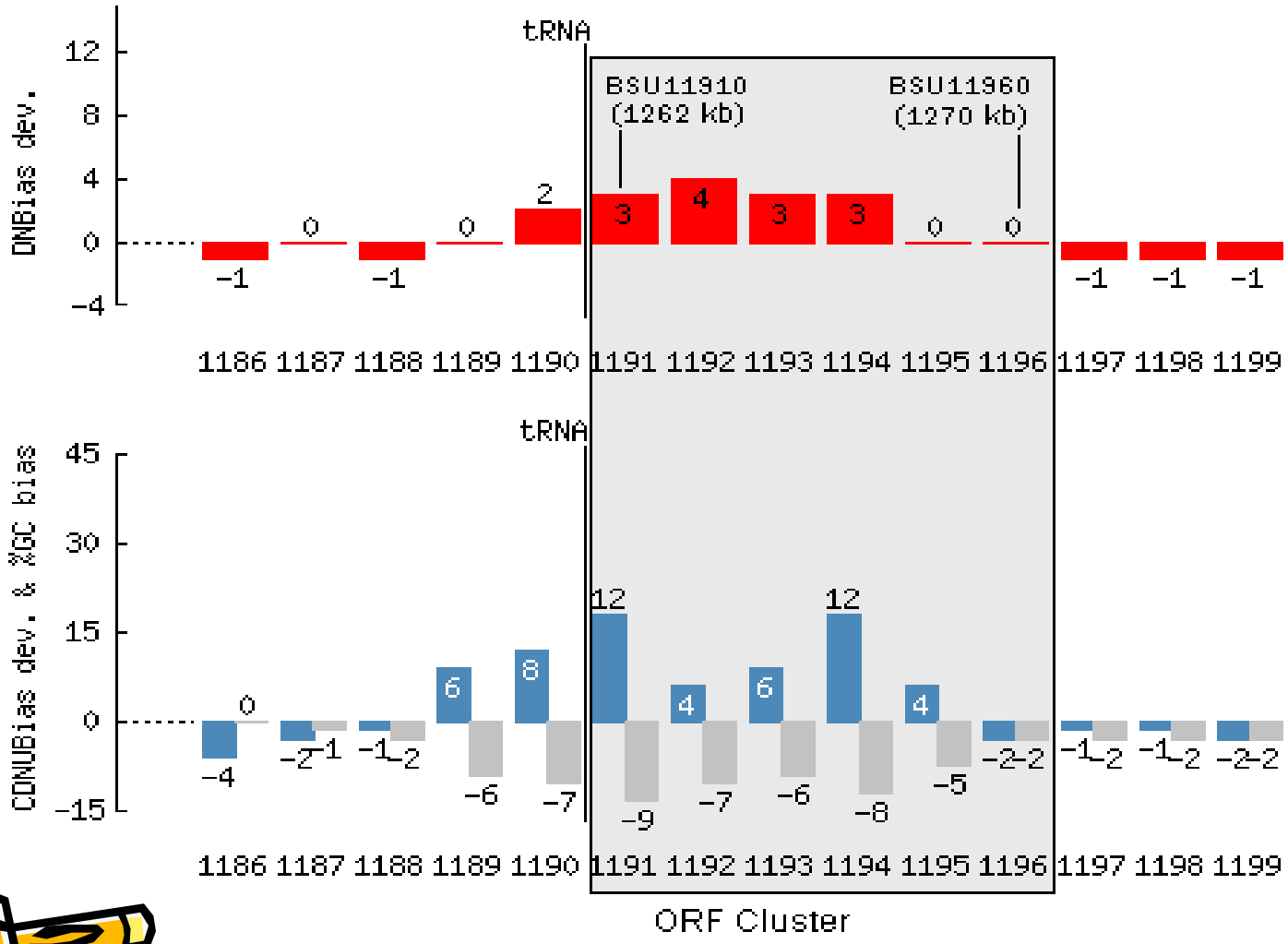
Performance evaluation

- Nicolas et. al. reported the identification of
  - 9 out of 10 prophages integrated in *Bacillus* genome and;
  - 14 DNA segments potentially arising from HGT
- PredictBias results largely in accordance with the findings of Nicolas et. al.
- 8 potential GIs were not identified by PredictBias because of size <6 ORFs.

Ref: Nicolas, P., Bize, L., Muri, F., Hoebeke, M., Rodolphe, F., Ehrlich, S. D., Prum, B. and Bessieres, P. (2002) Mining Bacillus subtilis chromosome heterogeneities using hidden Markov models. Nucleic Acids Res. 30, 1418-1426.

Island	Gene	Nicolas et. al.		PredictBias	
		Repeats	HMM	Position (kb)	Position (ORF)
P1 ‘prophage’	19	202-213	202-220	202-219	BSU01790 - BSU01970
P2 ‘prophage’	39	555-570	529-570	528-564	BSU04790-BSU05170
-	31	-	570-600	-	-
P3 ‘prophage’	12	-	651-664	647-660	BSU05980-BSU06090
Site-specific recombinase	9	-	738-747	-	-
Multidrug-efflux	4	-	818-822	-	-
cotJ operon	24	-	-	752-822	BSU06840-BSU07070
P4 ‘prophage’	10	-	1262-1270	-	-
PBSX ‘prophage’	34	-	-	-	-
P5 ‘prophage’	25	-	1879-1891	1877-1903	BSU17450-BSU17690
P6 ‘prophage’	33	2050-2060	2046-2073	2036-2070	BSU18660-BSU18980
SP propahge	184	-	2151-2286	2151-2283	BSU19810-BSU21640
Skin prophage	53	2654-2701	2652-2701	2652-2666	BSU25750-BSU25930
P7 ‘propahage’	42	2725-2735	2707-2756	2706-2745	BSU26450-BSU26860
Arsenic resistance regul.	6	3462-3469	3463-3467	-	-
-	13	3608-3634	-	3608-3619	BSU35150-BSU35210
Cell wall synthesis	10	3665-3672	3658-3685	3660-3677	BSU35630-BSU35720
ABC transporter	9	-	4123-4134	4122-4131	BSU40120-BSU40200
Streptothricin regul.	6	4189-4190	4184-4190	4182-4186	BSU40710-BSU40760

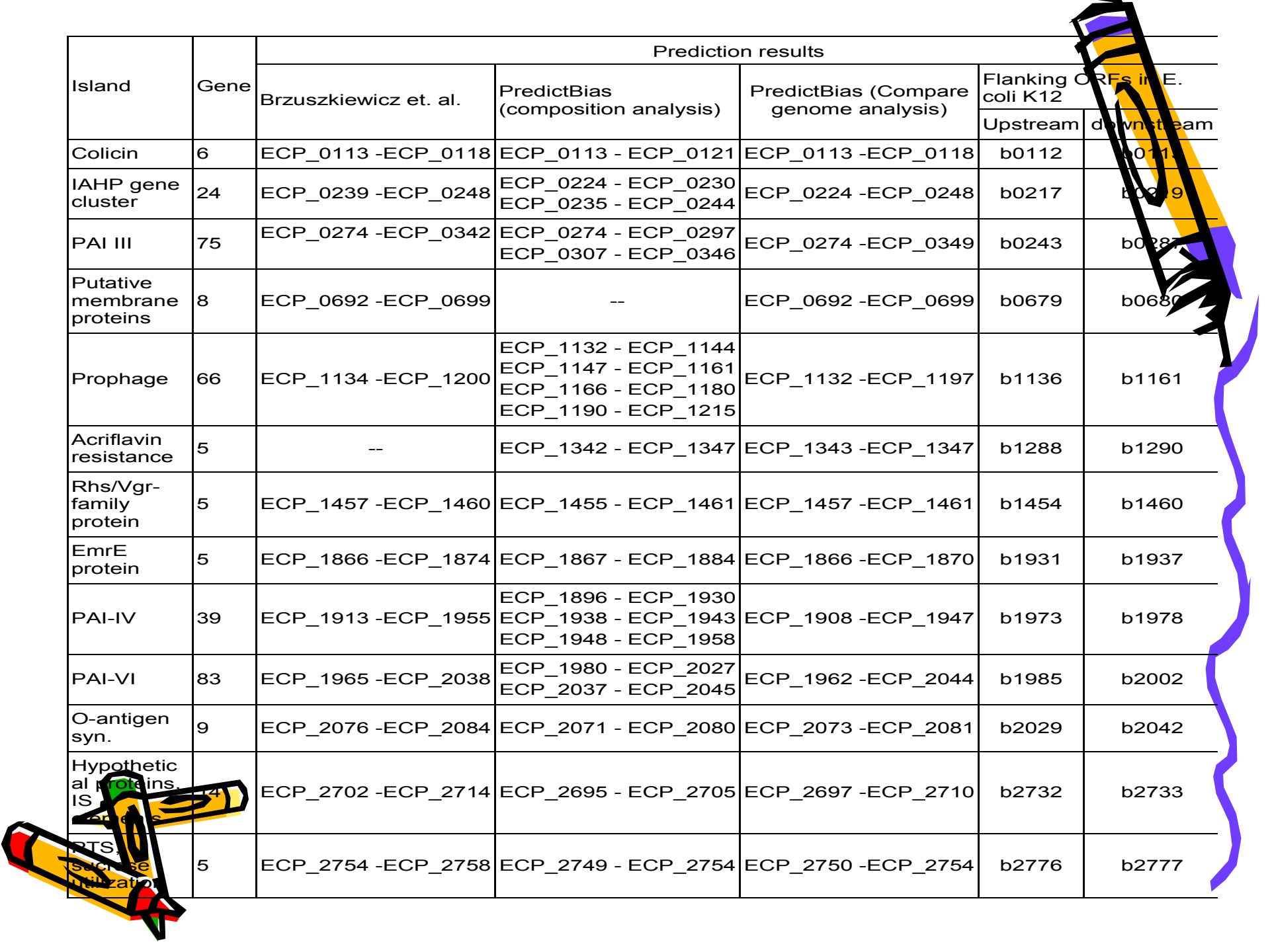
Manual analysis



Bias results for P4 prophage in *Bacillus subtilis*

- Brzuszkiewicz et. al. carried out a comparative genome analysis of uropathogenic *E. Coli* against nonpathogenic *E. Coli* strain K12 and reported the identification of many gene clusters that may contribute to virulence/fitness.
- Sliding window approach significantly aids in the identification of GIs/PAIs but it alone is not sufficient in defining PAI boundaries.
- 'Compare genome feature' of PredictBias significantly aids in defining PAI boundaries.
- On Comparative genome analysis, 18 regions having significant composition bias were observed having corresponding genome segments in *E. coli* K12, thus unlikely to have a role in pathogenicity.

Island	Gene	Prediction results				
		Brzuszkiewicz et. al.	PredictBias (composition analysis)	PredictBias (Compare genome analysis)	Flanking ORFs in E. coli K12	
					Upstream	downstream
Colicin	6	ECP_0113 -ECP_0118	ECP_0113 - ECP_0121	ECP_0113 -ECP_0118	b0112	b0119
IAHP gene cluster	24	ECP_0239 -ECP_0248	ECP_0224 - ECP_0230 ECP_0235 - ECP_0244	ECP_0224 -ECP_0248	b0217	b0249
PAI III	75	ECP_0274 -ECP_0342	ECP_0274 - ECP_0297 ECP_0307 - ECP_0346	ECP_0274 -ECP_0349	b0243	b0287
Putative membrane proteins	8	ECP_0692 -ECP_0699	--	ECP_0692 -ECP_0699	b0679	b0686
Prophage	66	ECP_1134 -ECP_1200	ECP_1132 - ECP_1144 ECP_1147 - ECP_1161 ECP_1166 - ECP_1180 ECP_1190 - ECP_1215	ECP_1132 -ECP_1197	b1136	b1161
Acriflavin resistance	5	--	ECP_1342 - ECP_1347	ECP_1343 -ECP_1347	b1288	b1290
Rhs/Vgr-family protein	5	ECP_1457 -ECP_1460	ECP_1455 - ECP_1461	ECP_1457 -ECP_1461	b1454	b1460
EmrE protein	5	ECP_1866 -ECP_1874	ECP_1867 - ECP_1884	ECP_1866 -ECP_1870	b1931	b1937
PAI-IV	39	ECP_1913 -ECP_1955	ECP_1896 - ECP_1930 ECP_1938 - ECP_1943 ECP_1948 - ECP_1958	ECP_1908 -ECP_1947	b1973	b1978
PAI-VI	83	ECP_1965 -ECP_2038	ECP_1980 - ECP_2027 ECP_2037 - ECP_2045	ECP_1962 -ECP_2044	b1985	b2002
O-antigen syn.	9	ECP_2076 -ECP_2084	ECP_2071 - ECP_2080	ECP_2073 -ECP_2081	b2029	b2042
Hypothetical proteins, IS proteins	14	ECP_2702 -ECP_2714	ECP_2695 - ECP_2705	ECP_2697 -ECP_2710	b2732	b2733
PTS, Sucrose utilization	5	ECP_2754 -ECP_2758	ECP_2749 - ECP_2754	ECP_2750 -ECP_2754	b2776	b2777



metV island, IHAP-like gene cluster	29	ECP_2804 -ECP_2832	--	ECP_2800 -ECP_2828	b2813	b2817
PAI-V, K15 capsule	75	ECP_2962 -ECP_3024	ECP_2960 - ECP_2972 ECP_2975 - ECP_3014 ECP_3023 - ECP_3034	ECP_2962 -ECP_3036	b2966	b2968
Dehydrogenases, putative allatoin degradation	7	ECP_3103 -ECP_3109	--	ECP_3099 -ECP_3105	b4469	b3017
Putative galacticol	8	ECP_3346 -ECP_3353	--	ECP_3342 -ECP_3349	b3256	b3257
Fimbrial proteins	10	ECP_3513 -ECP_3522	ECP_3511 - ECP_3517	ECP_3512 -ECP_3521	b3426	b3428
PTS-dependent fructose utilization	8	ECP_3753 -ECP_3760	--	ECP_3754 -ECP_3761	b3655	b3656
PAI-I	100	ECP_3765 -ECP_3862	ECP_3763 - ECP_3858	ECP_3765 -ECP_3864	b3657	b3660
Hemolysin-coregulated (Hpc)	23	ECP_4024 -ECP_4046	--	ECP_4024 -ECP_4046	b3829	b3832
Sugar utilization	4	ECP_4087 -ECP_4093	--	ECP_4090 -ECP_4093	b3881	b3885
2-oxoglutarate utilization system	9	ECP_4275 -ECP_4282	--	ECP_4274 -ECP_4282	b4054	b4055
Oxidoreductases, regulators	10	ECP_4448 -ECP_4459	ECP_4444 - ECP_4449	ECP_4449 -ECP_4458	b4203	b4205
PAI-II, Fimbrial proteins	121	ECP_4521 -ECP_4641	ECP_4521 - ECP_4573 ECP_4610 - ECP_4650	ECP_4521 -ECP_4641	b4267	b4309

Address http://www.davvbiotech.res.in/PredictBias/ Go Links

PredictBias

Assisting in the identification of Genomic & Pathogenicity islands in prokaryotes

RefSeq Id OR Organism Search

Home | Genomes analyzed | New analysis | Help

Introduction

Although major portion of a prokaryotic genome (70-80%) is of homogenous nucleotide composition, some portion (20-30%) carries segements of DNA having distict nucleotide content known as **Genomic Islands** (GIs). **Pathogenicity Islands** (PAIs) are sub set of GIs that are acquired by horizontal gene transfer and contain genes encoding virulence factors like adhesins, toxins, inavsins and others.

PredictBias is an application which assists in the identification of genomic and pathogenicity islands in prokaryotes. It is based on four important features of PAIs:

- Significant compostion bias (%GC, dincucleotide & codon bias).
- Presence of Insertion elements (Transposase, Integrase, tRNA).
- Presence of genes encoding proteins similar to known virulence factors (adhesins, invasin, toxin & others).
- Absence from same or closely related non-pathogenic sp.

Thank You

With increase in bacterial resistance to conventional antibiotics, there is an urgent need to identify novel drug targets. PAIs encoding genes vital for bacterial pathogenesis may play a significant role in identifying these drug targets.

Reference:

Pundhir, S., Vijaywargiya, H., and Kumar, A. (2008). PredictBias: a server for the identification of genomic and pathogenicity islands in prokaryotes. In Silico Biol. 8, 0019.

How to use?

Search PredictBias:

anism name or *Refeq Id* corresponding to the query genome in the search box (top). To get complete list of analyzed microbial nome at PredictBias Click here.

ew Bias analysis:

as analysis for an organism is not available at PredictBias. Click New analysis and provide the genome file (Genbank format) rresponding to the genome.