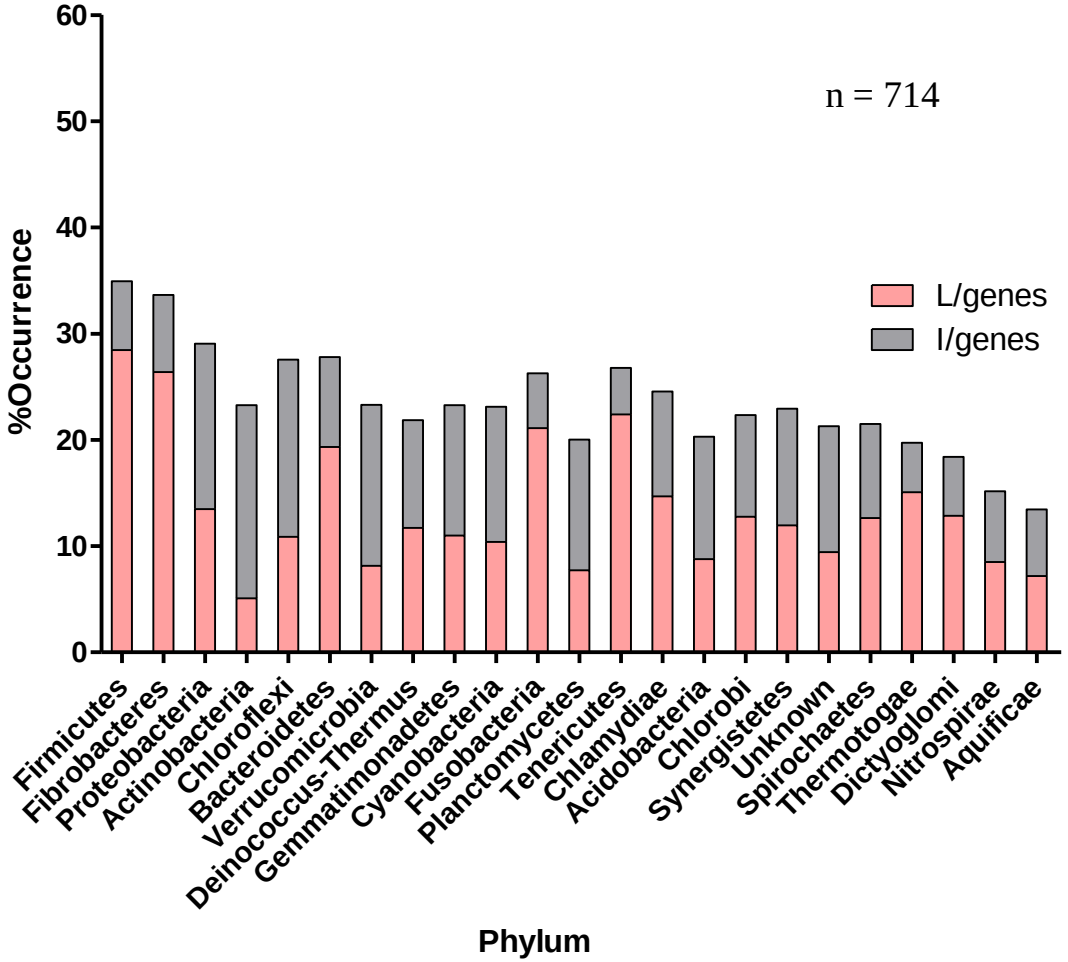
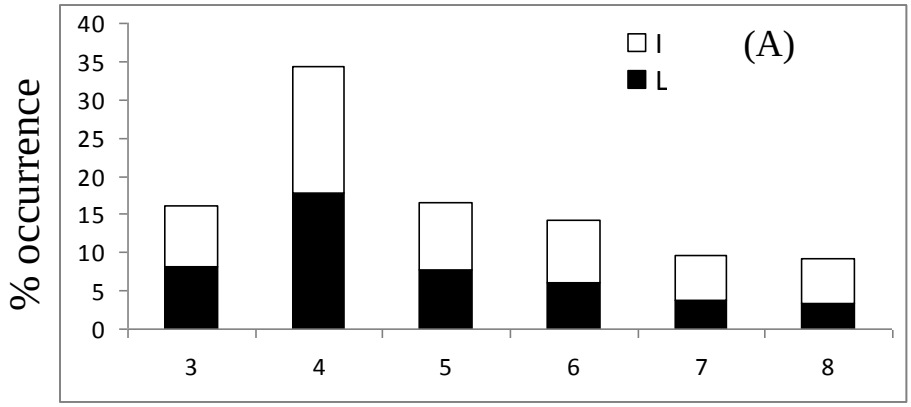
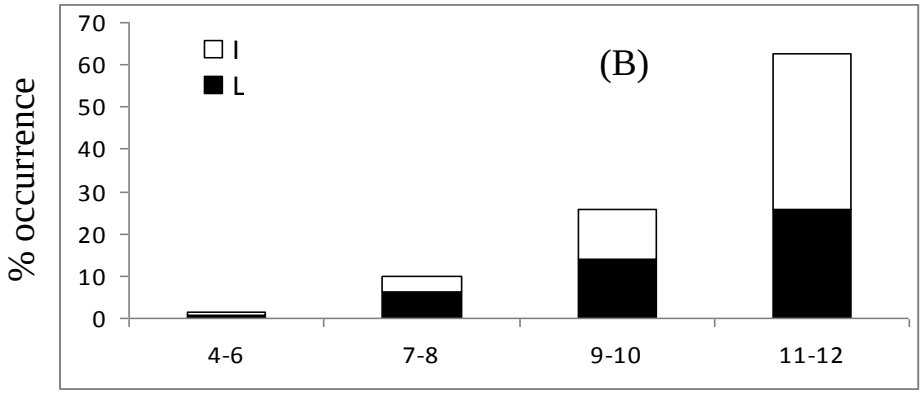
Bacterial genomes	714
Genes	2,211,011
“All” terminators	776,261
“Best” terminators	623,216
Best/Genes	28.1%
L-shaped terminators	319080
I-shaped terminators	304136
%L	51.2%
%I	48.8%
U (tandem terminators)	61275
V	942
XΔG	36157

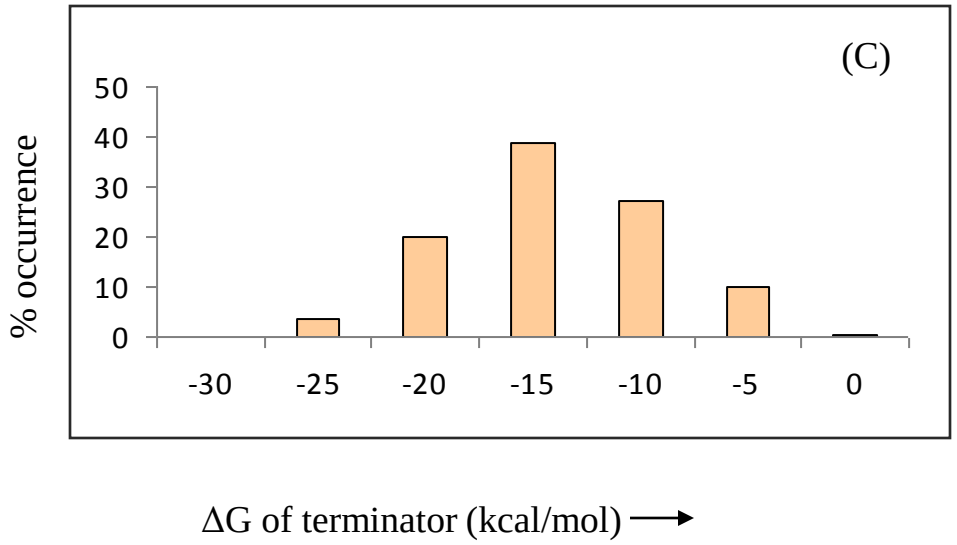
- **Terminator profile from a sample of 714 genomes from WebGeSTer DB.**
- **“Best” = L-shaped + I-shaped terminators.**
- **Individual U, V and X terminators are formed by a combination of L and/or I-shaped terminators.**



- **Prevalence of intrinsic terminators in different bacterial phyla.**
- **Some phyla show high prevalence for L-shaped terminators while others have more of I-shaped terminators.**





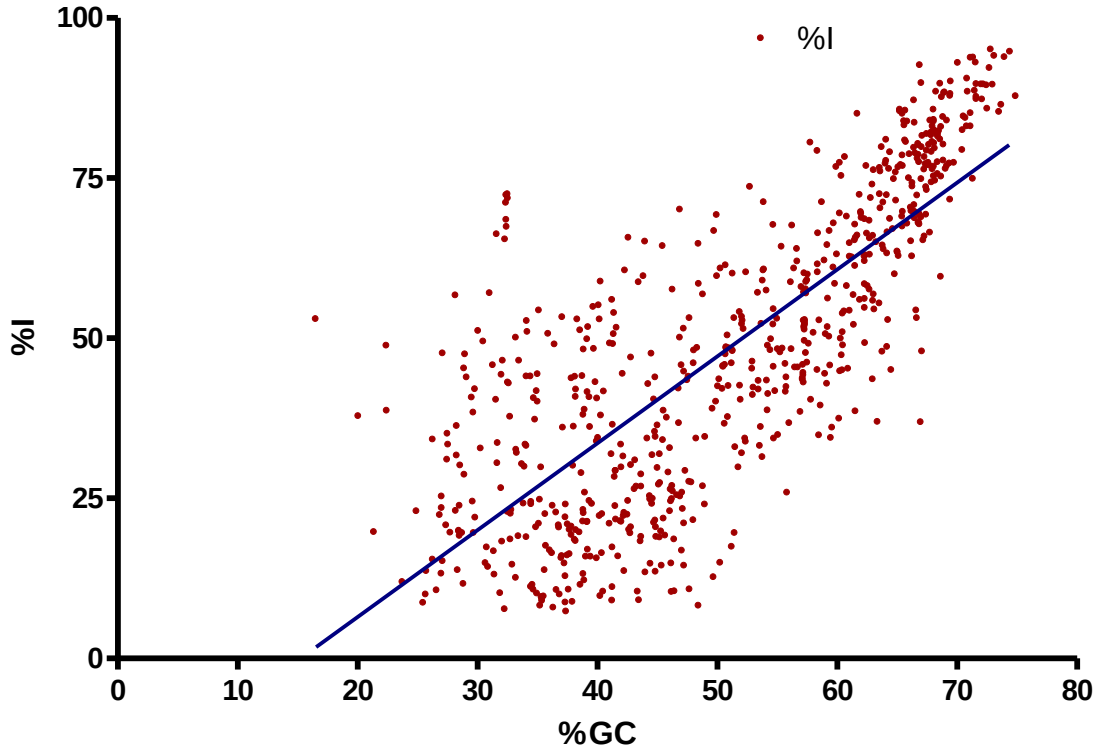


n = 714

➢**Analysis of structural parameters of terminators.**

(A) 65% of terminators have a loop size of 4-6nt; (B) 88% of terminators have a stem-length between 9-12bp. (C) 86% have a ΔG between -10 and -25kcal/mol.

n = 714



- **Correlation(0.86) between prevalence of I-shaped terminators and genomic GC content. The dependence on non-canonical I-shaped terminators increases as GC content rises.**

TABLE S1: Experimentally-determined terminators identified by WebGeSTer.

No.	gene	Terminator present	Classifications	
1	ampC	yes	TP	
2	lpd	no		FN
3	araD(araBAD operon)	yes(araD)	TP	
4	aroG	yes	TP	
5	aspA	yes	TP	
6	tyrA (aroF-tyrA)	yes	TP	
7	birA	yes	TP	
8	carB (car operon)	yes	TP	
9	cdh	yes	TP	
10	sbp	yes	TP	
11	pfkA	yes	TP	
12	cheW	yes	TP	
13	tsr	yes	TP	
14	crp	no		FN
15	cysB(ecoli)	yes	TP	
16	dapD	yes	TP	
17	deoD	yes	TP	
18	dld	yes	TP	
19	dye/arcA	yes	TP	
20	envA/lpxC	no		FN
21	fdhF	yes	TP	
22	accD	yes	TP	
23	fpg/mutM	yes	TP	
24	fumA	yes	TP	
25	fur	yes	TP	
26	guaB	yes	TP	
27	guaA	yes	TP	
28	ihfA	no		FN
29	rpsA	yes	TP(-)	
30	hisL	yes	TP	
31	ilvL	yes	TP	
32	ilvBN	yes	TP	
33	ilvL	yes	TP	
34	infA	yes	TP	
35	leuL	yes	TP	

36	lexA	no		FN	
37	malM	yes	TP		
38	map	yes	TP		
39	metL	yes	TP		
40	flhC	yes	TP		
41	rnhA	yes	TP		
42	rimP	no		FN	
43	infB	yes	TP(-)		
44	nusB	no	TN		true negative 'cos as per literature, there is no terminator either
45	envZ(omp operon)	yes	TP		
46	ompC	no	TP(-)		
47	ompF	no	TP(-)		
48	pheL	yes	TP		
49	pheA	yes	TP		
50	pheV	yes	TP		
51	creD	yes	TP		
52	phoR	no		FN	
53	polA	yes	TP		
54	ponA/mrcA	yes	TP		
55	ponB/mrcB	yes	TP		
56	proC	yes	TP		
57	prs	no		FN	
58	crr	yes	TP		
59	pyrI	yes	TP		
60	pyrE	yes	TP		
61	yceB	yes	TP		
62	recA	yes	TP		
63	thyA	yes	TP		
64	gyrB	yes	TP		
65	rho	yes	TP		
66	rplQ	no	TP		
67	rpoC	yes	TP		
68	rplL	yes	TP		
69	glgP	yes	TP		
70	glnA	yes	TP(-)		
71	glnS	yes	TP		
72	nagE	yes,	TP		

73	gltA	yes	TP	
74	glyS	yes	TP	
75	glyA	yes	TP	
76	gpt	yes	TP	
77	serS	no		FN
78	sodA	yes	TP	
79	sodB	yes	TP	
80	sppA	yes	TP	
81	speD	yes	TP	
82	glyW	yes	TP(-)	
83	trxB	yes	TP	
84	uvrD	yes	TP	
85	valS	yes	TP	
86	aspU	yes	TP	
87	rrfG	yes	TP	
88	rrfF	yes	TP	
89	rrfA	yes	TP	
90	rrfE	yes	TP	
91	trpT	yes	TP	
92	rrfB	yes	TP	
93	thrL	yes	TP	
94	thrC	yes	TP	
95	trpL	yes	TP	
96	trpA	yes	TP	
97	tonB	yes	TP	
98	aspV	yes	TP	
99	xylE	yes	TP	
100	tuf(Mtb)	yes	TP	
101	Rv1324(Mtb)	yes	TP	

Results:

101 experimental genes for which there was sequence information about downstream region of gene

True Positives (TP) = 85 (experimental terminator unambiguously identified in WebGeSTer DB)

False Negatives (FN) = 9 (experimental terminator not identified)

(-) = 6 (the structure predicted in WebGeSTer DB does not match the structure predicted in literature)

True Negative (TN) = (terminators not detected downstream of these genes neither by WebGeSTer DB nor literature)

TN = nusB (44) and >24 genes (which occur internal to the many operons present in the list

Evaluation

The Receiver-operator-characteristic (ROC) plot describes the propability of detection (pD) versus the probability of false alarm (pFA) for various inputs of threshold ΔG values to the WebGeSTer algorithm. The pD(threshold) was defined as the fraction of all the putative true terminators identified by WebGeSTer whose absolute ΔG value was greater than (i.e. more negative) the threshold value. Similarly, we defined pFA(threshold) as the fraction of false positive hits that did not represent terminators but had higher absolute value than the threshold value. To generate the curve, the minimum of the Δ G values of the terminators were first identified and an interval of 1 kcal/mol starting from the minimum was used as the interval to calculate the threshold values for pD and pFA. Both the pD and pFA values climb from 0 for the 'highest' to 1 for the 'lowest' negative threshold.

In the ROC plot, each threshold is indicated by a tick on the curve. The formulation of ROC is according to a previous prescription by Lesnik et al. (2001), except that for pFA calculation we did not divide the false positive counts by the total number of bases in the genome, but by the total count of false positives.

To look at the ROC plot for individual genome, type the URL:
http://pallab.serc.iisc.ernet.in/gester/<directory_code>/roc.pdf

Reference

Lesnik EA, Sampath R, Levene HB, Henderson TJ, McNeil JA, Ecker DJ. (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*. Nucleic Acids Research, 29: 3583-3594.