Supporting information

Hyun-Ju Kim *et al* : Multiple transcription initiation from *LEE1* promoters in enteropathogenic *Escherichia coli* (EPEC)

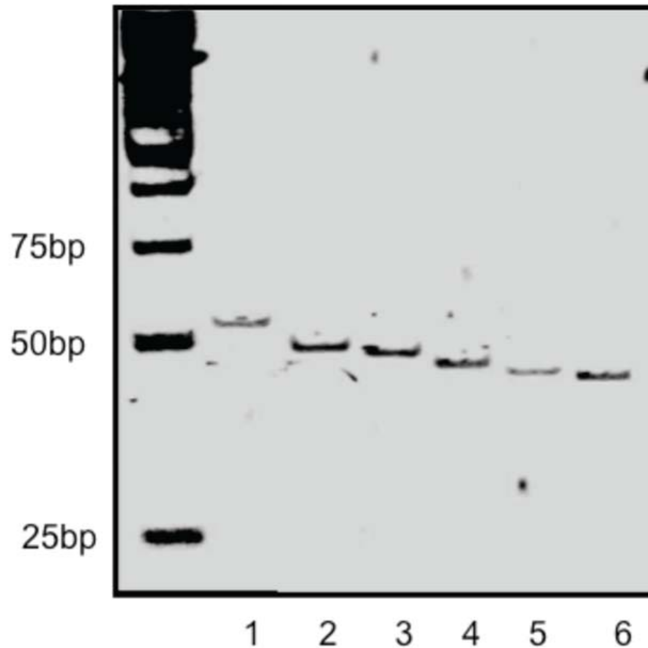


Fig. S1. Identification of *tsp* of *LEE1* P1. Transcription was carried out as described in the Experimental procedure using pSA508::*LEE1* P1 as a template. Transcripts were visualized in an 8 M urea/10% polyacrylamide sequencing gel. Each transcript of *LEE1* P1 origin was excised and the RNA species was extracted with Trizol following the conventional method (Ausubel, 1987) . Subsequently, The 5` cloning linker (5` Linker: 5`-TGGAATrUrCrUrCrGrGrGrCrArCrCrArArGrGrU-3`) was ligated to the purified RNA species, using miRCatTMmiRNA Cloning Kit (Integrated DNA Technologies, USA). Reverse transcription of the linkered RNA species was carried out followed by PCR amplification using forward primer, 5'-TGGAATTCTCG**GGCACC**-3' and reverse primer, 5`-GGTT**CTGCAG**CATCAAACAACCACCTTA-3, each carrying *BanI* and *Pst1* site (bold), respectively. The PCR products were cloned into T vector using TOPcloner^{TM} TA kit(Enzynomics, South Korea). Individual cloned plamids were linearized by *EcoR1* and electrophoresed in an 8 M urea/8% polyacrylamide sequencing gel. The figure shows the DNA obtained from the transcripts presumably initiated from $\underline{A}^{+1}$, $\underline{A}^{+8}$, $\underline{A}^{+9}$, $\underline{A}^{+10}$, $\underline{A}^{+11}$, and $\underline{A}^{+13}$ for the lanes from 1 to 6. Far left lane is E-gel 25 bp DNA ladder (Invitrogen, USA). Subsequent DNA sequencing revealed that each clone carried the expected DNA fragment derived from the *LEE1* P1 RNA (data not shown).

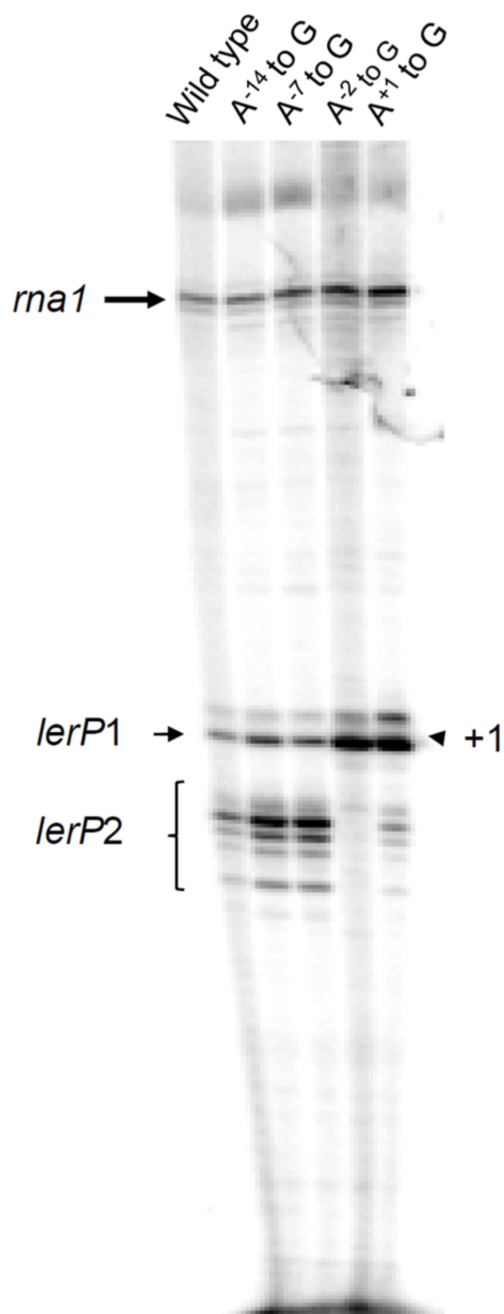
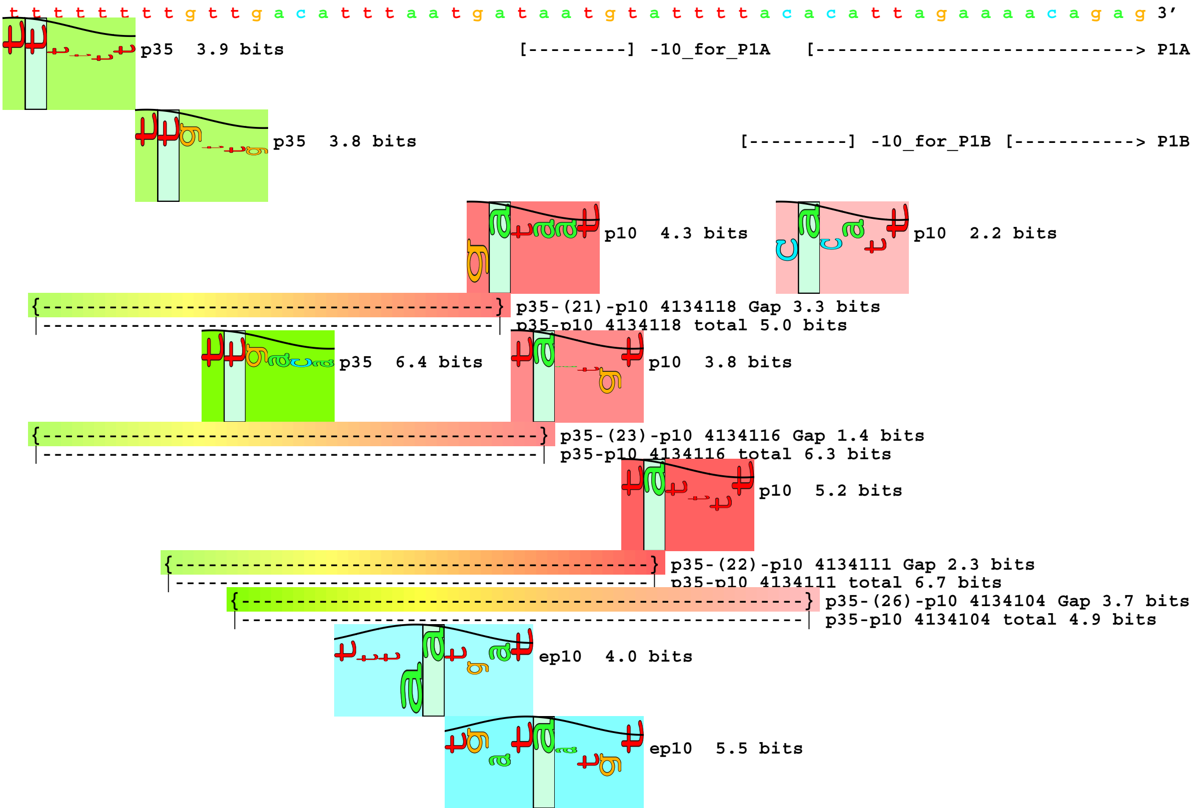
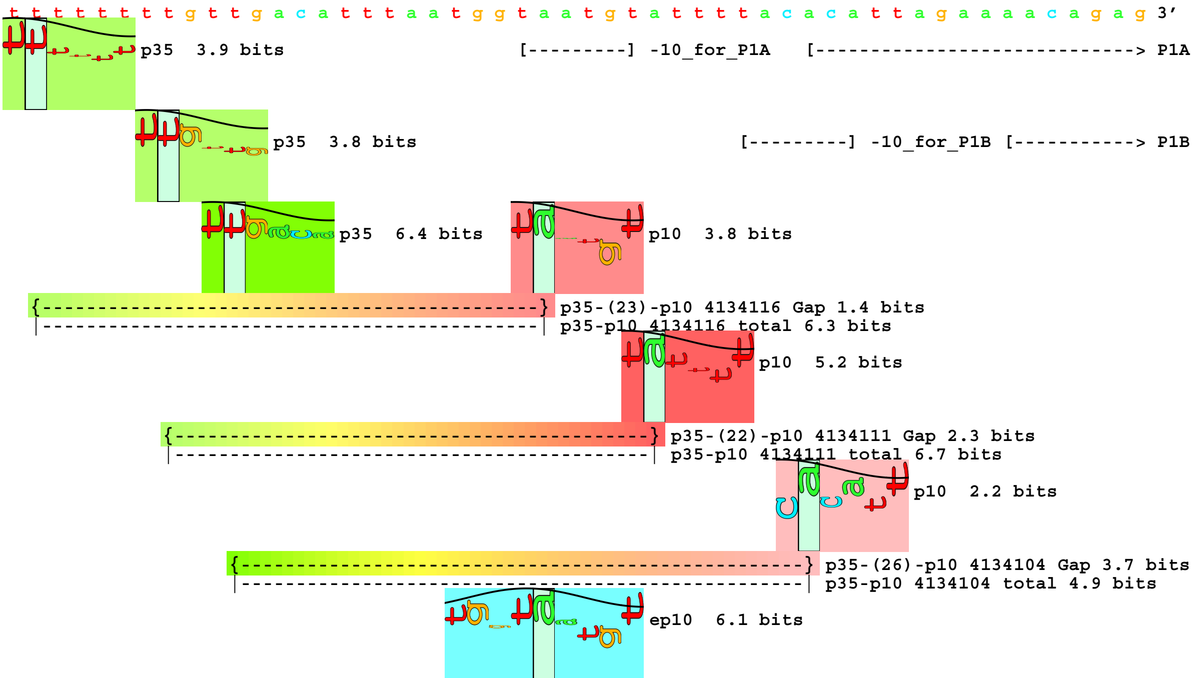


Fig. S2. ***in vitro*** **transcription assay with** ***LEE1*** **P1 variants.** To identify the promoter element responsible for each transcript from *LEE1* P1 DNA, the As near *tsp* were serially substituted and tested by *in vitro* transcription assay. Constructions of the *LEE1* P1 mutants were described in the main text except the substitution of A^{-14} and A^{-7}, which were done with the primers listed in the Supplementary Table 1. These promoter DNAs (-200~+20) were cloned between the *EcoR*I and *Pst*I sites in pSA508(Choy & Adhya, 1993). The assay was carried out with these variant pSA508:: *LEE1* P1 (-200 to +20), and the reaction products were analyzed by 8 M urea/10% polyacrylamide gel electrophoresis. Note that substitution of A_{-14} and $A_{-7}$with G did not change the pattern of transcripts generated from *LEE1* P1 DNA. In contrast, substitution of A^{+1} reduced the intensity of the transcripts initiated from $A^{+8}A^{+9}A^{+10}A^{+11}$ and A^{+13}, and substitution of A^{-2} completely blocked the transcript generation as described in the text (Fig. 3).

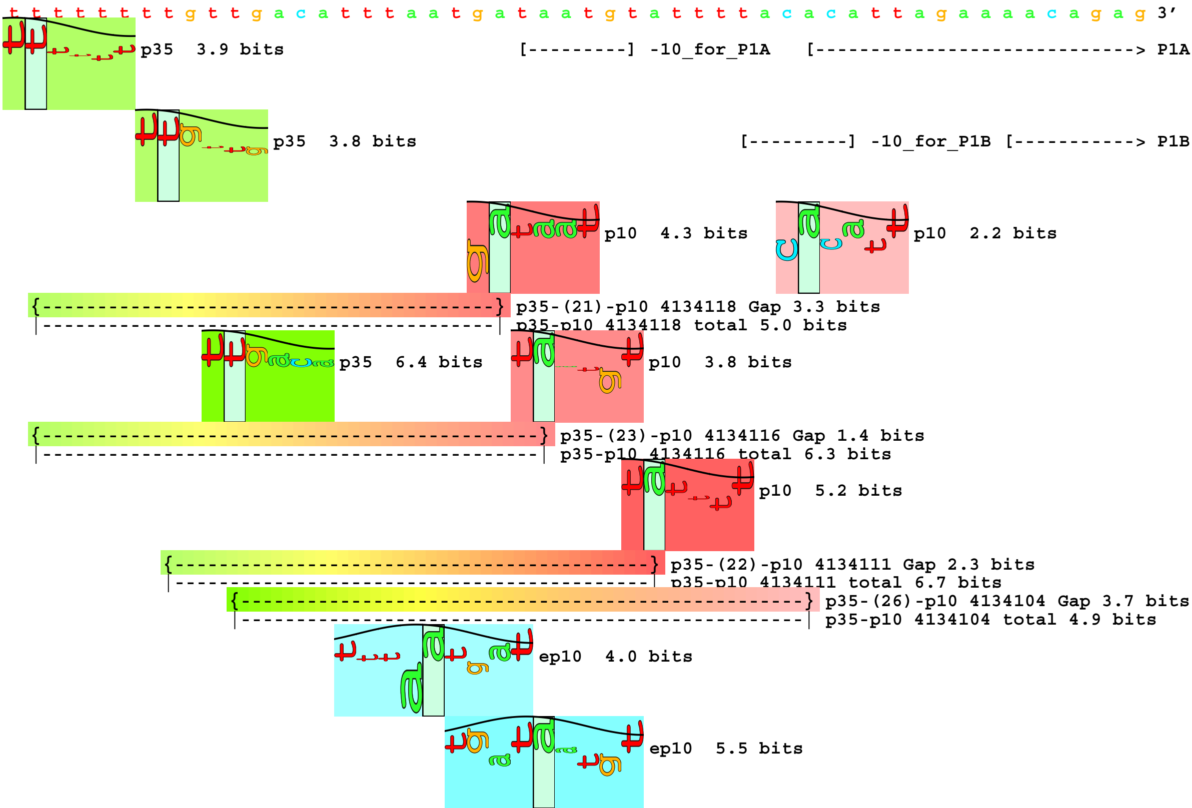
NC_011601.1



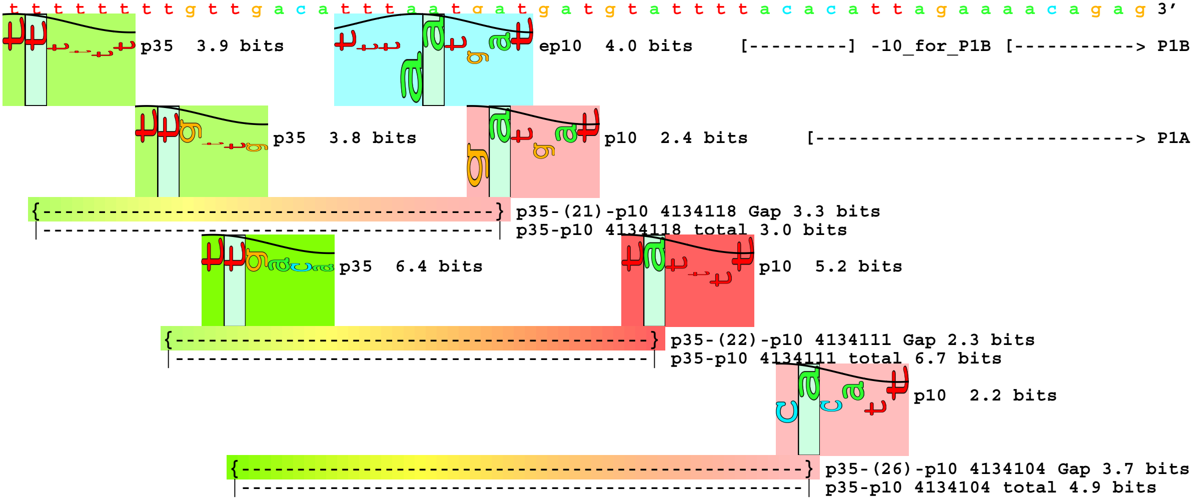
NC_011601.1.t4134118c, A-14G



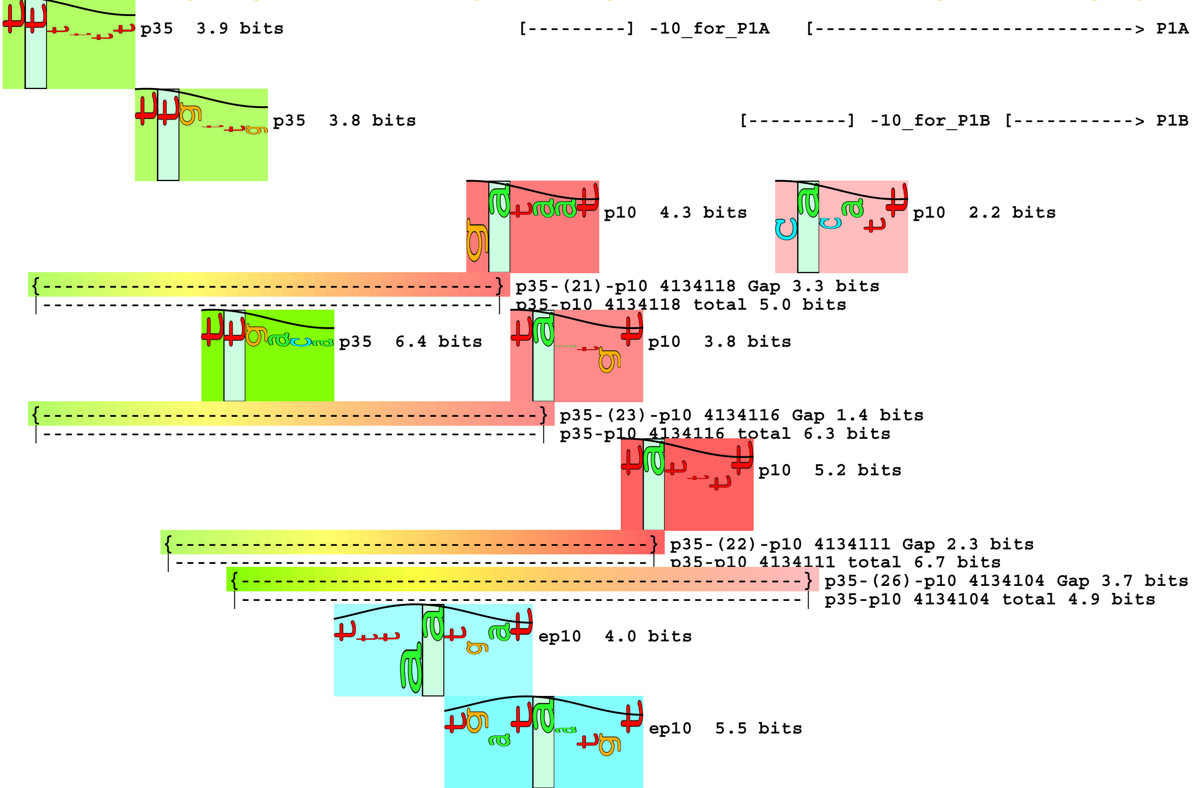
NC_011601.1



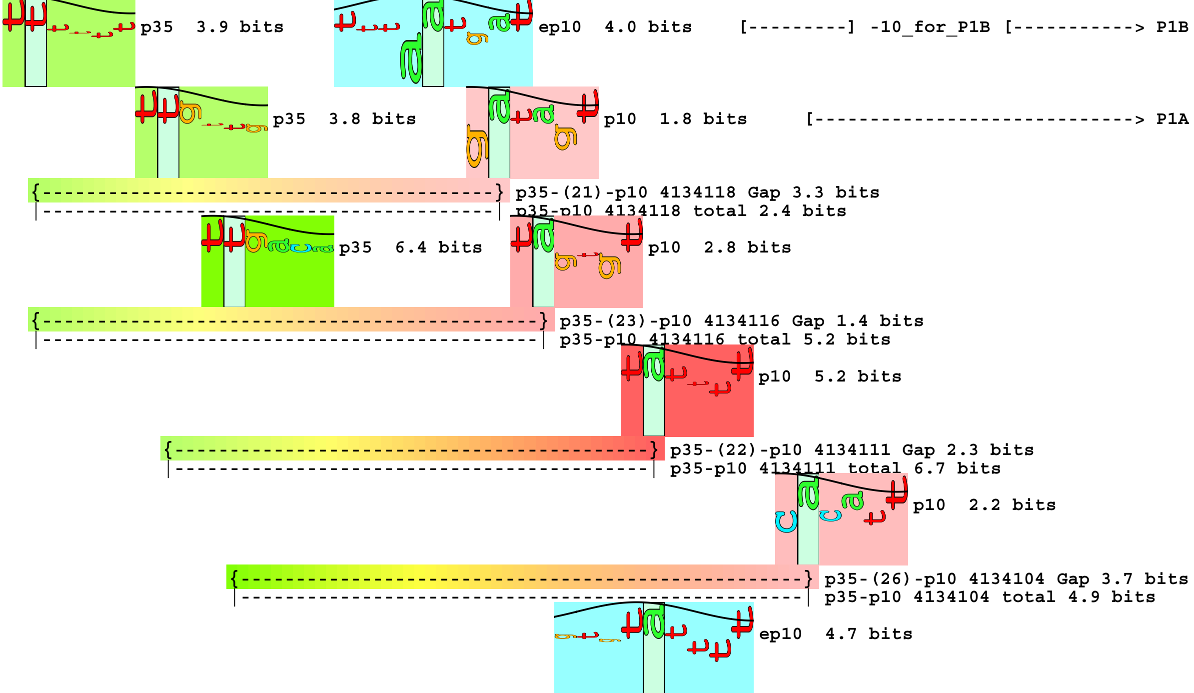
NC_011601.1.t4134116c, A-12G



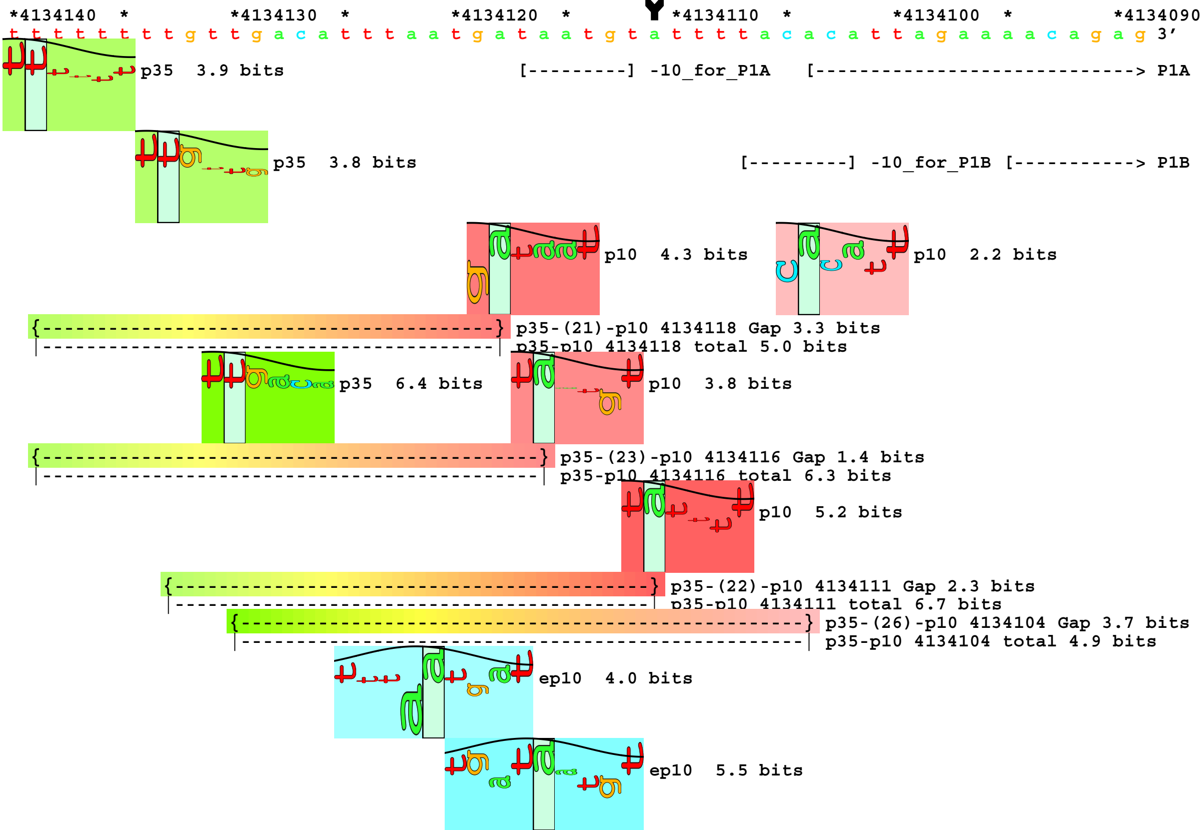
NC_011601.1



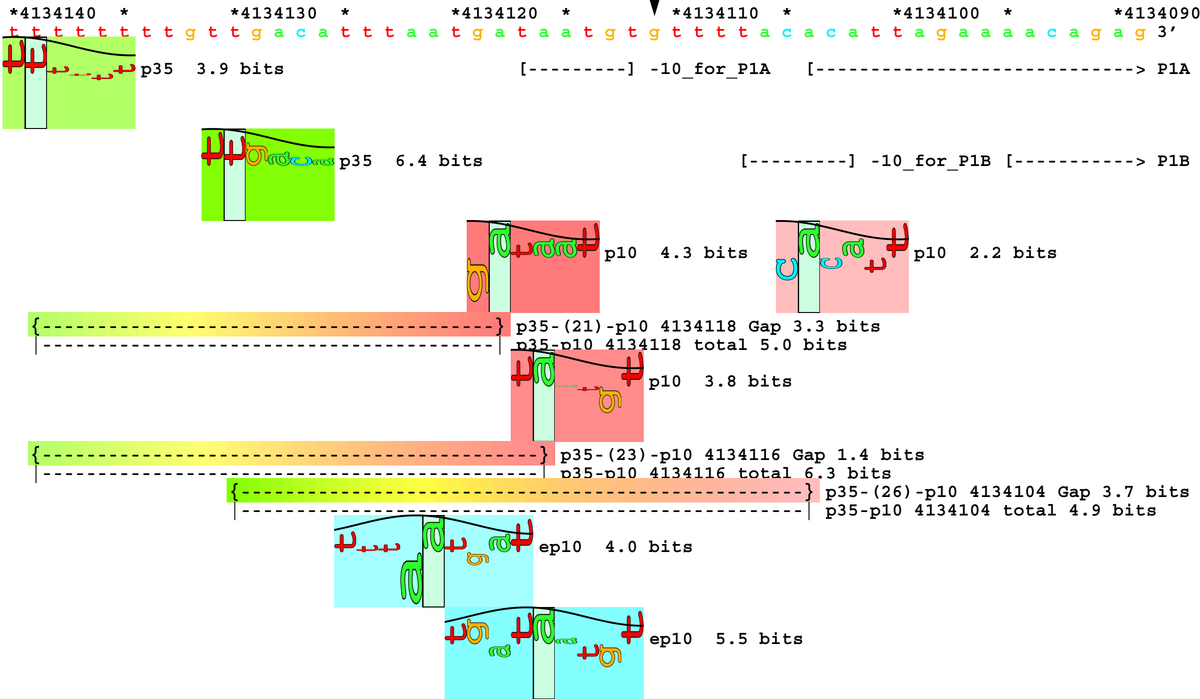
NC_011601.1.t4134115c, A-11G



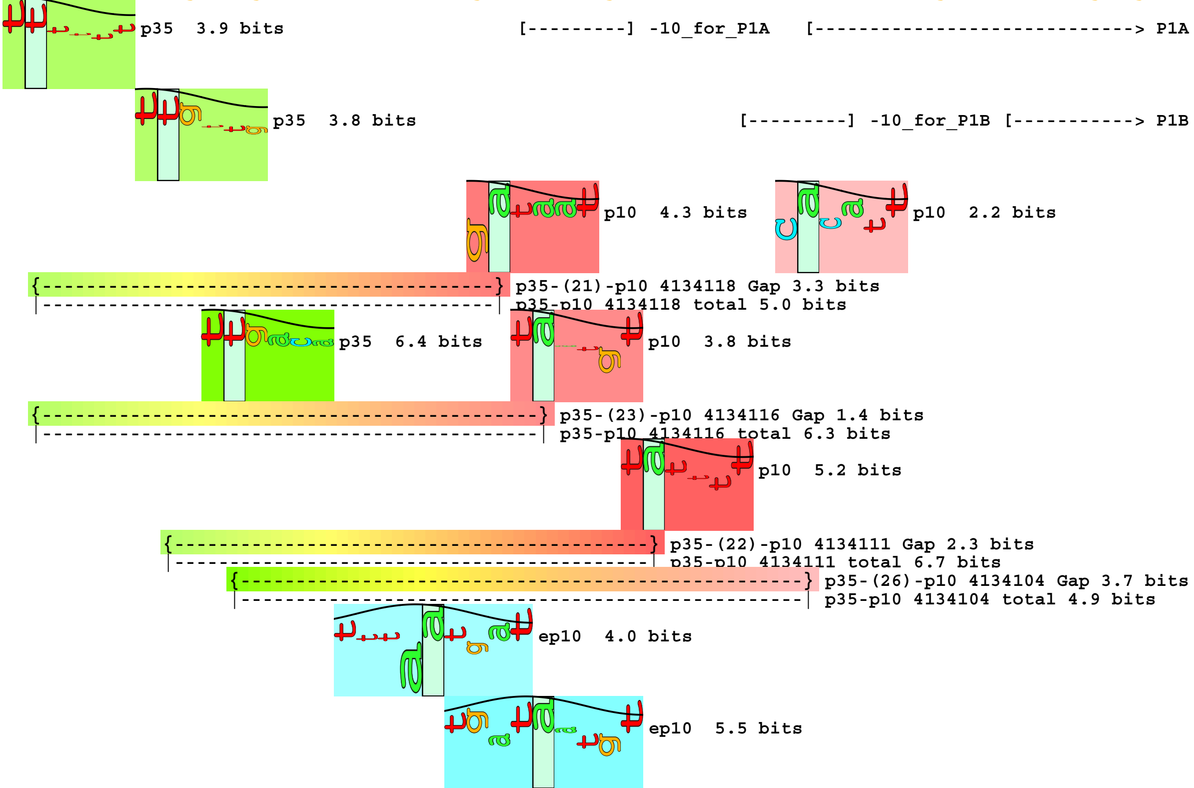
NC_011601.1



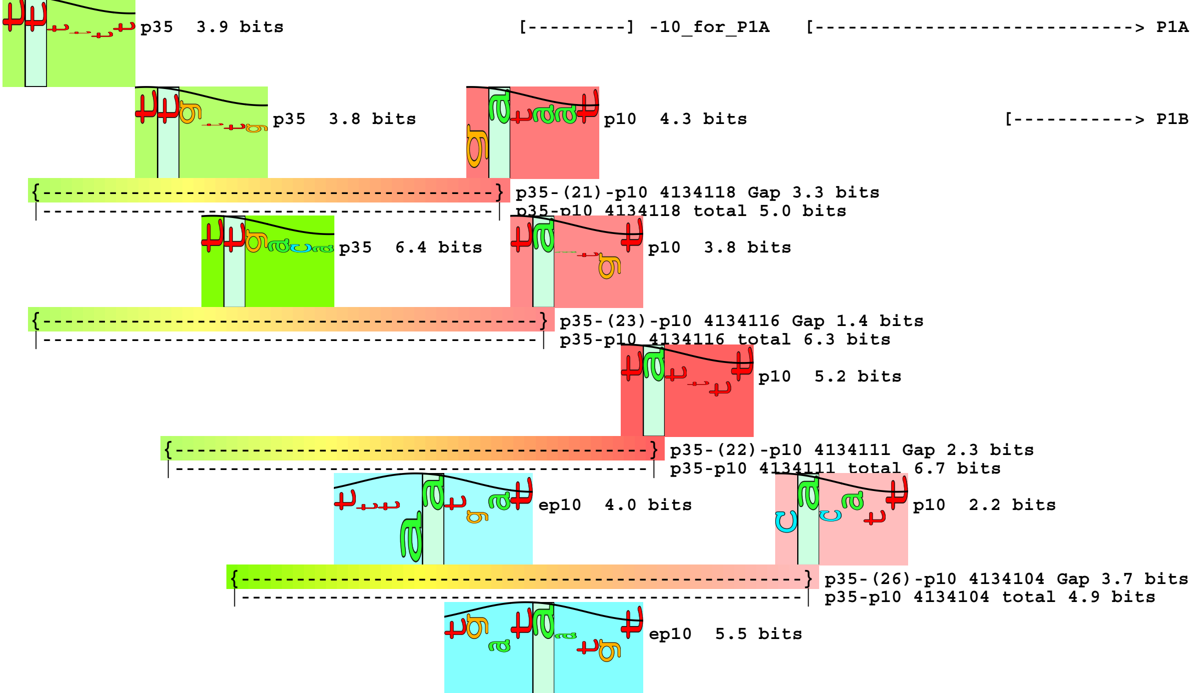
NC_011601.1.t4134111c, A-7G



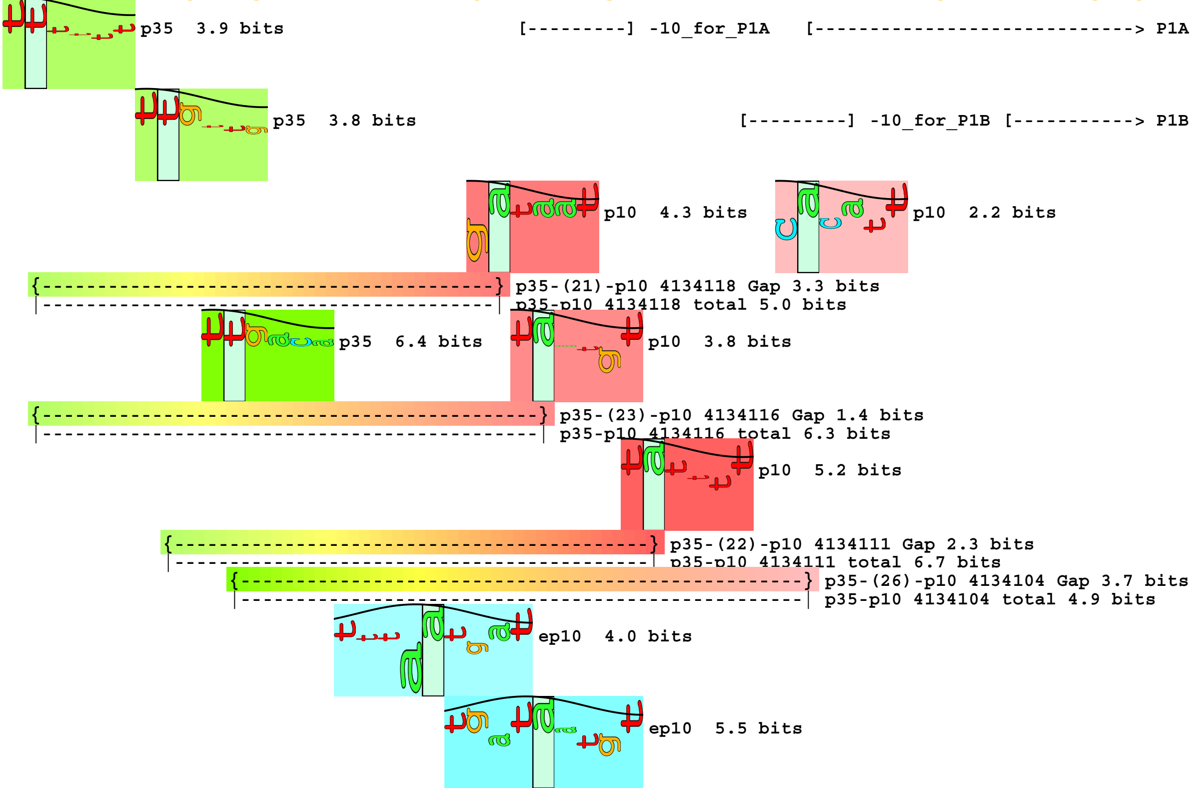
NC_011601.1



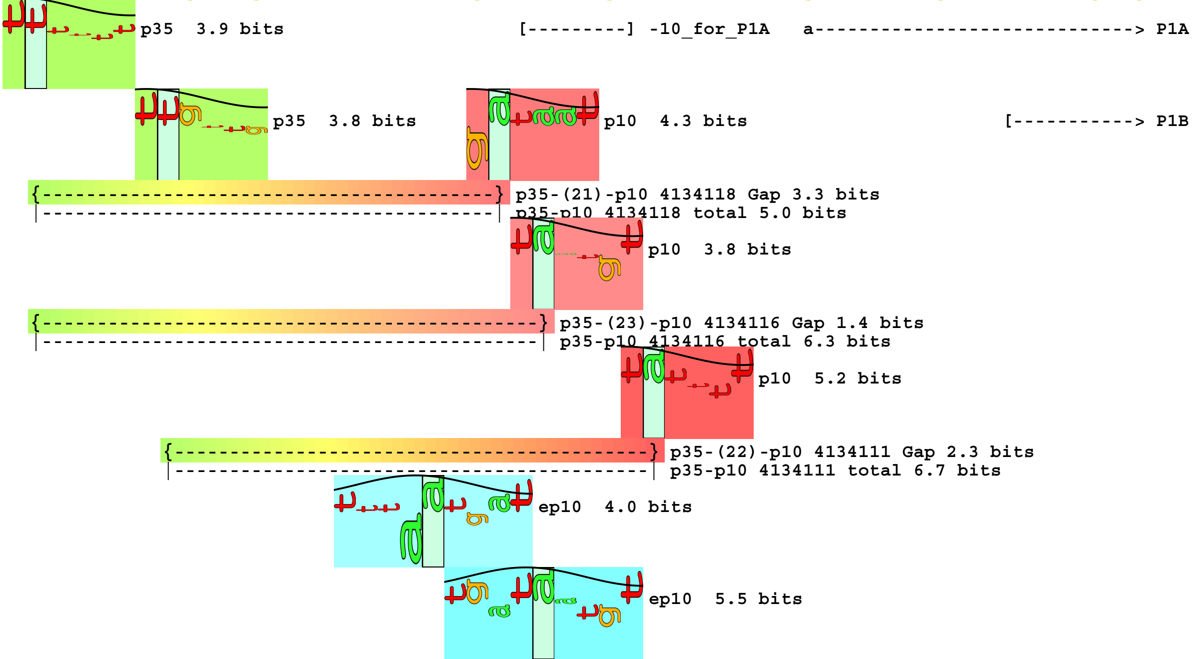
NC_011601.1.t4134106c, A-2G



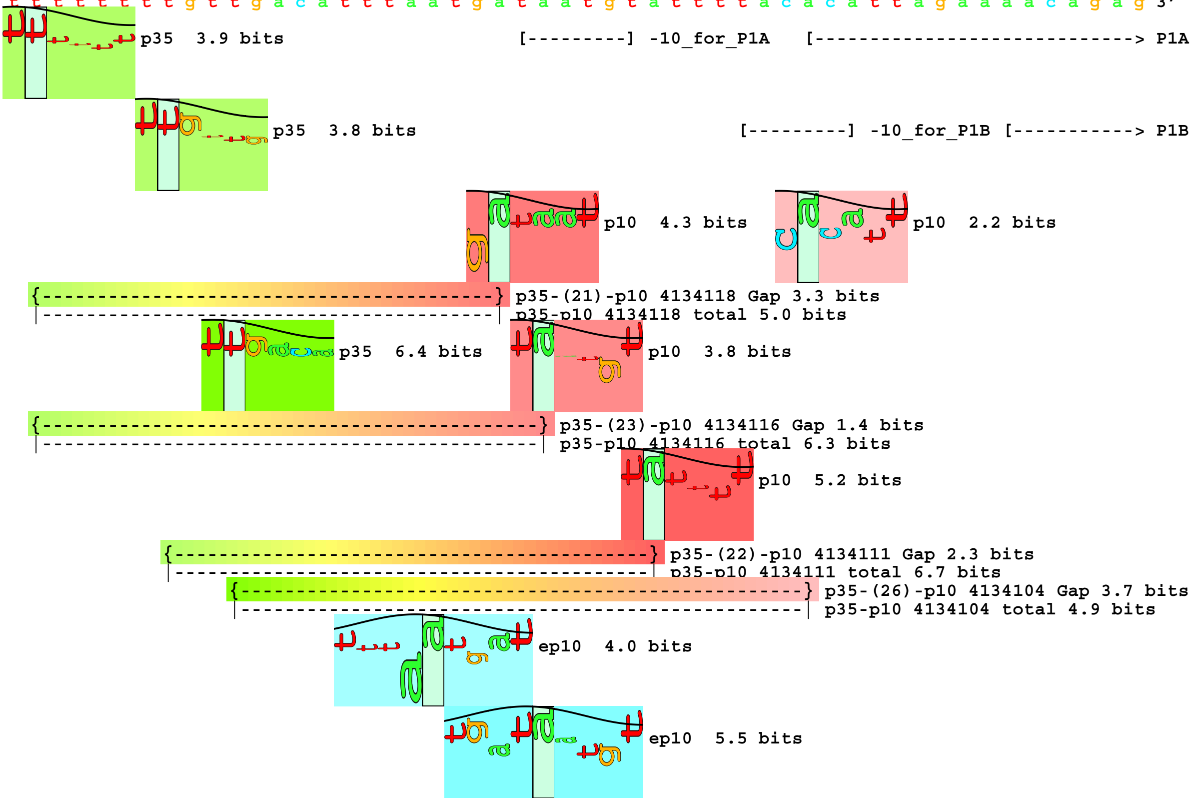
NC_011601.1



NC_011601.1.t4134104c, A+1G



NC_011601.1



NC_011601.1.t4134102c, A+3G

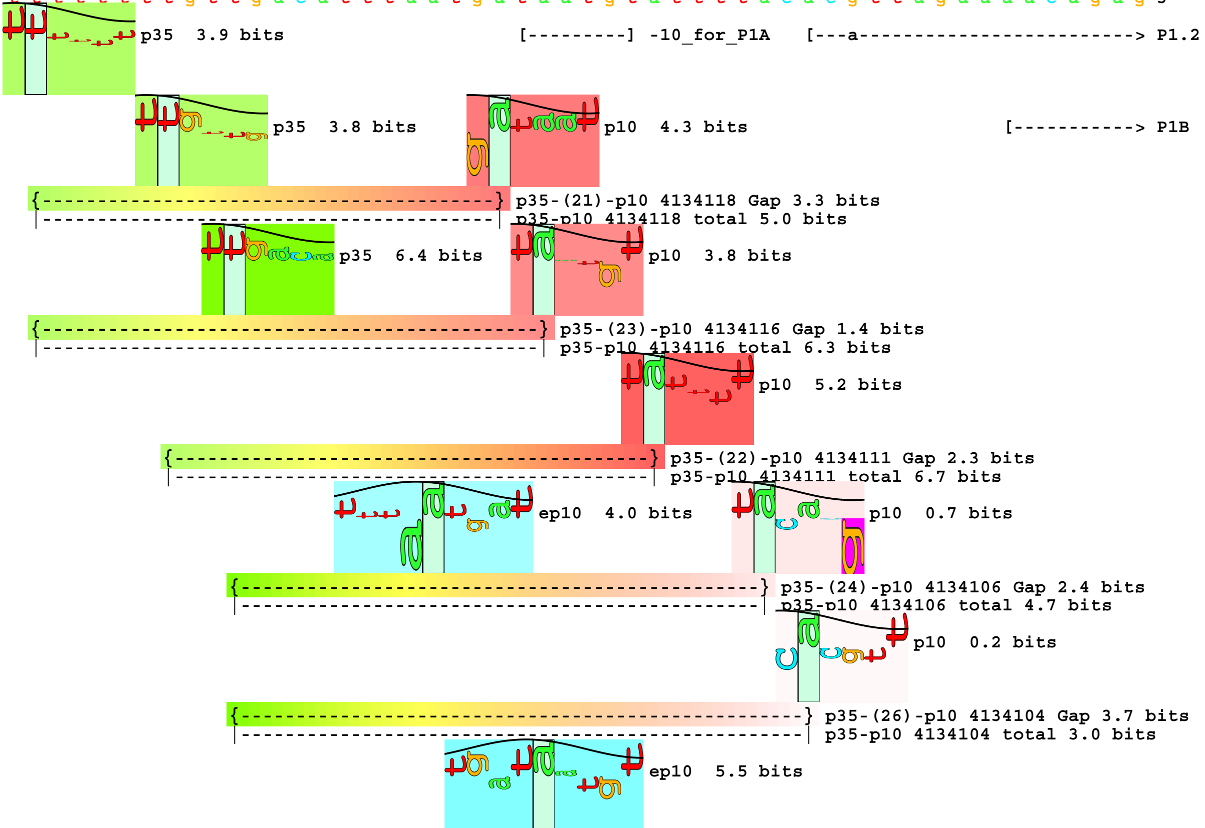


Fig. S3. Analysis of *LEE1* P1 variants using information theory

In order to analyze individual mutant *LEE1* P1 sequences for promoter elements, we used sequence walker technology (Schneider 1997; Schneider 1997; Schneider 1999). *LEE1* P1 of complementary sequence coordinates from 4134142 to 4134089 (Accession and version number: NC_011601.1) and those with A to G substitution at the nucleotide position -14, -12, -11, -7, -2, +1 and +3 were analyzed. Colored rectangles ('petals') behind the walkers identify the kind of site (by hue) and the strength of the site (by saturation)(Chen *et al* 2006), (http://alum.mit.edu/www/toms/papers/t7island/). The connecting bar between parts of a flexible site transitions linearly between the corresponding colors. The σ70 binding sites were located using flexible binding models (Shultzaberger *et al.* 2001), in that there are variable distances between binding components. The horizontal dashed line underneath the -10 walker (p10) and the -35 walker (p35) that is labeled 'Gap', gives the gap surprisal for whatever distance separates the two components, as well as the coordinate of the downstream component. The dashed line that is labeled 'total' gives the total information for the flexible site. The transcription start points for P1A (+1) and P1B (+10) were marked with black arrows. The sequences and coordinates on the map are from GenBank Accession and version number: NC_011601.1 (Friedberg *et al.* 1999). Pages 1 to 7 show the analysis of wild type *LEE1* P1, and those with A to G substitutions at nucleotide position -14, -12, -11, -7, -2, +1 and +3 in sequence. Four overlapping -10 hexamers (red) were indentified with the wild type *LEE1* P1 DNA that have corresponding -35 hexamers (green) and good spacing. The predicted σ70 sites have totals of 5.0 bits (G^{-15}ATAAT), 6.3 bits (T^{-13}AATGT), 6.7 bits (T^{-8}ATTT) and 4.9 bits (C^{-2}ACATT at coordinate positions 4134118, 4134116, 4134111 and 4134104, respectively, with cutoff at 0 bits (top line). These were verified by substituting As in the hexamer site by G and the mutated *LEE1* P1 sequences were analyzed and shown with sequence walkers. The substitution of A^{-14} destroyed the 5.0 bit site while keeping the downstream sites intact. The substition of A^{-12} destroyed the 6.3 bit site and weakened the upstream 5.0 bit site to 3.0 bits. The substitution of A^{-11} reduced both upstream sites: 5.0 bit site to 2.8 bits and 6.3 site to 5.2 bits. The substitution of A^{-7} destroyed the downstream 6.7 bit site with no other effect. The substitution at A^{-2} changed nothing. The weak 4.9 bit site was affected by the $A^{+1}$and A^{+3} changes. Thus, these analyses only partially agreed with *in vitro* transcription assay results (Fig. 3 in main text and Fig. S1). The 6.3 bit site would be responsible for the transcription initiated from +1. Although *in vitro* transcription assay results suggested that the -10 hexamer, $T^{-3}ACACA^{+3}$, would govern the multiple transcription initiation from $A^{+8}A^{+9}A^{+10}A^{+11}$ and A^{+13}, the sequence walker failed to detect this site suggesting that this -10 hexamer should be beyond recognition for conventional σ70-dependent promoter elements by computer program. It has been suggested that prokaryotic transcription start points contain no significant information (Shultzaberger *et al.* 2007). Analysis with an extended -10 model (blue) shows that the 6.3 bit σ70 promoter could be driven entirely from an extended -10. This is consistent with the unusual -35, which is entirely Ts.

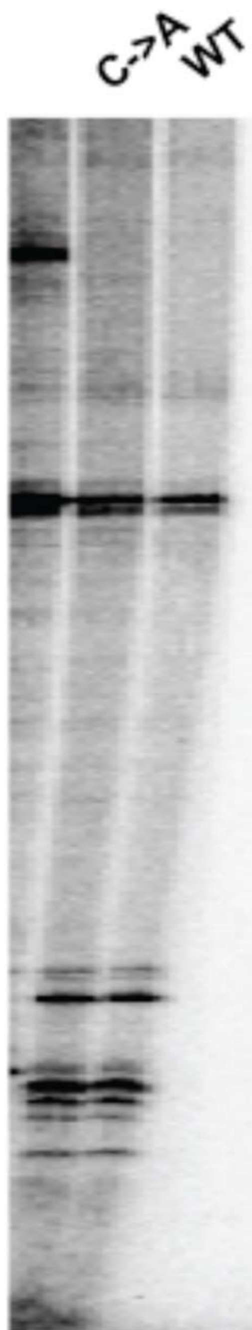


Fig. S4. ***in vitro*** **transcription assay using pSA508 carrying wild type and a mutant** ***LEE1*** **P1 (C^{+12} to A substitution)(-200 to +20)**. The reaction products were analyzed by 8 M urea/10% polyacrylamide gel electrophoresis.

Table S1. Primers used for PCR amplification of various *lerP* DNA*

Primer	Sequence	Product
-200 *EcoRI*	5'-CGGAATTCCAGCTTGGTTTTTATTCTG-3'	pHJ12, pHJ36
+20 *pstI*	5'-GGTTCTGCAGATGTTATTATTCTCTGTTT-3'	pHJ36
A^{+1}->G(R)	5'-GGTTCTGCAGCTCTGTTTTCTAATGCGTAAAATACATTATC-3'	pHJ57
A^{-2}->G(R)	5'-GGTTCTGCAGCTCTGTTTTCTAATGTGCAAAATACATTATC-3'	pHJ58
A^{-7}->G(R)	5'-GGTTCTGCAGCTCTGTTTTCTAATGTGTAAAACACATTATC-3'	pHJ75
A^{-14}->G(R)	5'-GGTTCTGCAGCTCTGTTTTCTAATGTGTAAAATAGATTACC-3'	pHJ76
C^{-+12}->A(R)	5'-GGTTCTGCAGCTCTTTTTTCTAATGCGTAAAATACATTATC-3'	pHJ80

* Mutants carrying altered *LEE1* P1 sequences were generated by cloning the synthetic DNA oligomers in pSA508 (see main text). The primers used for PCR amplification of respective DNA are shown.

References

Ausubel, F. M., (1987) *Current protocols in molecular biology*. Greene Pub. Associates ;
J. Wiley, order fulfillment, Brooklyn, N.Y. Media, Pa.

Chen, Z. & T. D. Schneider, (2006) Comparative analysis of tandem T7-like promoter containing regions in enterobacterial genomes reveals a novel group of genetic islands. *Nucleic Acids Res* **34**: 1133-1147.

Choy, H. E. & S. Adhya, (1993) RNA polymerase idling and clearance in gal promoters: use of supercoiled minicircle DNA template made in vivo. *Proc Natl Acad Sci U S A* **90**: 472-476.

Friedberg, D., T. Umanski, Y. Fang & I. Rosenshine, (1999) Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic Escherichia coli. *Mol Microbiol* **34**: 941-952.

Schneider, T. D., (1997a) Information content of individual genetic sequences. *J Theor Biol* **189**: 427-441.

Schneider, T. D., (1997b) Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucleic Acids Res* **25**: 4408-4415.

Schneider, T. D. a. R., P.K., (1999) Computational analysis of nucleic and acid
information defines binding sites. In: United States Patent. United States, pp.

Shultzaberger, R. K., R. E. Bucheimer, K. E. Rudd & T. D. Schneider, (2001) Anatomy of Escherichia coli ribosome binding sites. *J Mol Biol* **313**: 215-228.

Shultzaberger, R. K., Z. Chen, K. A. Lewis & T. D. Schneider, (2007) Anatomy of Escherichia coli sigma70 promoters. *Nucleic Acids Res* **35**: 771-788.