Supplementary Information

Systematic molecular analyses reveal extremely long and complex 5'UTR of a major cold shock protein CsdA in *Escherichia coli*

Soma Jana[1,2] and Partha P. Datta[1*]

[1]Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur, WB, PIN 741246, India

[2] Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA.

[*]Corresponding author: partha_datta@iiserkol.ac.in

Table S1: Primers used to clone *cspA* promoter region in modified pTurboGFP-B vector

Primers	Sequences	Restriction Sites
cspA promoter forward primer	GAAGATCTACTTTTATCCACTTTATTG	BglII
cspA promoter reverse primer	ACGCGTCGACTGATGTGCATTAAGCC	SalI

Table S2: Primers used to clone fragments up to 700 bases upstream of the *deaD* coding region

Primers	Sequences	Restriction Sites
Reverse primer for Cod 103bpU, Cod 208bpU,	ACGCGTCGACGTAGTACGTGTGCCTCAAAATTAATGGC	SalI

Cod 320bpU, Cod 414bpU, Cod 540bpU, Cod 700bpU construct		
Cod 103bpU Forward Primer	GAAGATCTCACCTACTTTGAGCCGGTTCACAC	BglII
Cod 208bpU Forward Primer	GAAGATCTATGACCTGGCAGAATCGGACC	BglII
Cod 320bpU Forward Primer	GAAGATCTGATTTGGACAGCGCCACG	BglII
Cod 414bpU Forward Primer	GAAGATCTAAGGCTCAAGGCGGACG	BglII
Cod 540bpU Forward Primer	GAAGATCTGAAGCTCGATGAGAAGCAGGC	BglII
Cod 700bpU Forward Primer	GAAGATCTTACTTGAGCTTGATCCAACTTACAACTACG	BglII

Table S3: Primers used to clone fragments 700 bases upstream of the *deaD* coding region

Primers	Sequences	Restriction Sites
Reverse primer for 705ntU 53bp, 705ntU 100bp, 705ntU 150bp, 705ntU 200bp construct	ACGCGTCGACTCAAACGCTTCATAGGCAGCATC	SalI

705ntU 53bp Forward primer	GAAGATCTGGCATATATTTAACGCAGGCAGGC	BglII
705ntU 100bp Forward primer	GAAGATCTAAGCGCTGGCAATCCGAC	BglII
705ntU 150bp Forward primer	GAAGATCTGTTGTATGATAGTCTCGGTCTGAGGGC	BglII
705ntU 200bp Forward primer	GAAGATCTCGGGCTTTAACCGATGACGAACG	BglII
EMSA experiment negative control *deaD* gene coding region fragment Forward primer	CGCGGATCCCATGGCTGAATT	-------
EMSA experiment negative control *deaD* gene coding region fragment Reverse primer	AACAGAGGTAAAGAGAATGCTGCAGTTTTTCCGCTCCCC	-------

Figure S1: Schematic representation of the original pTurboGFP–B vector.

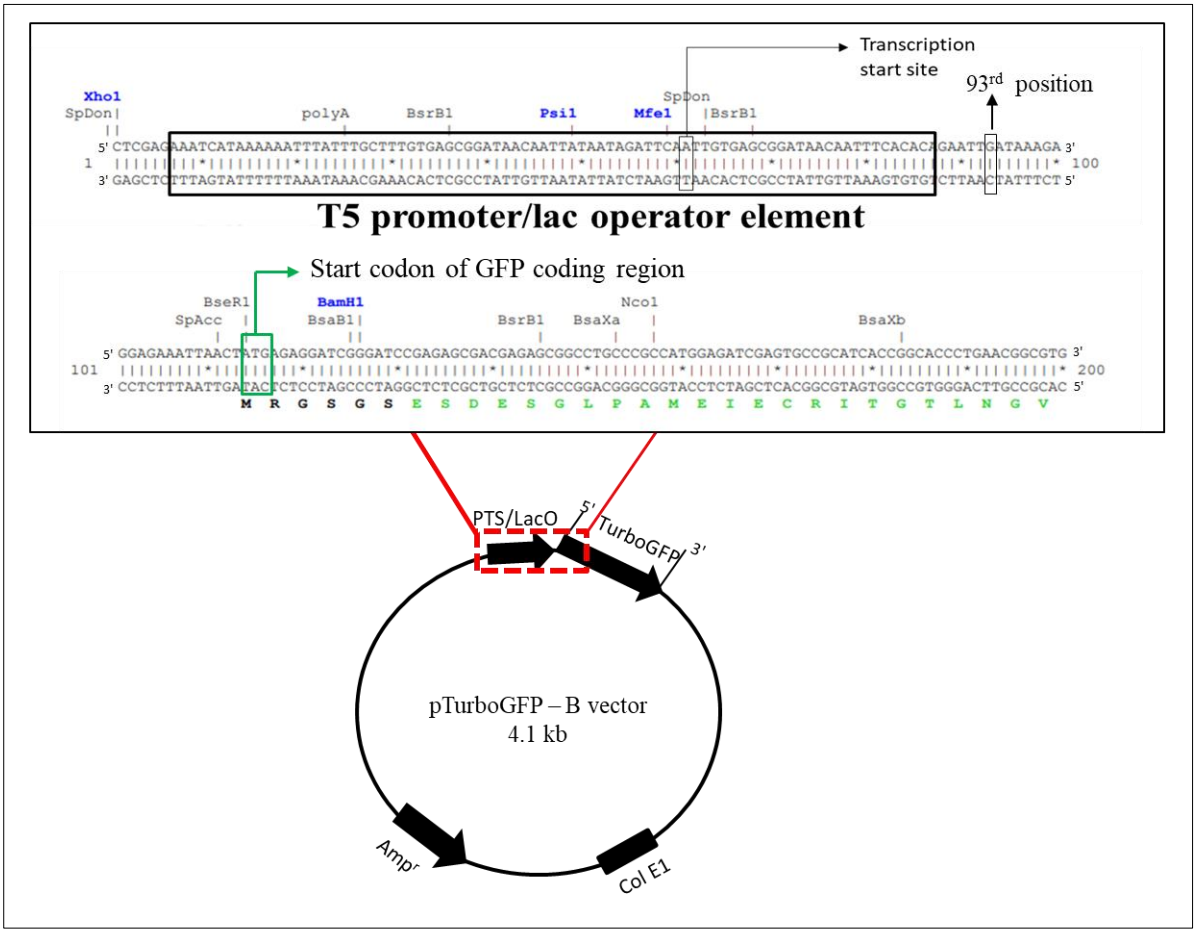
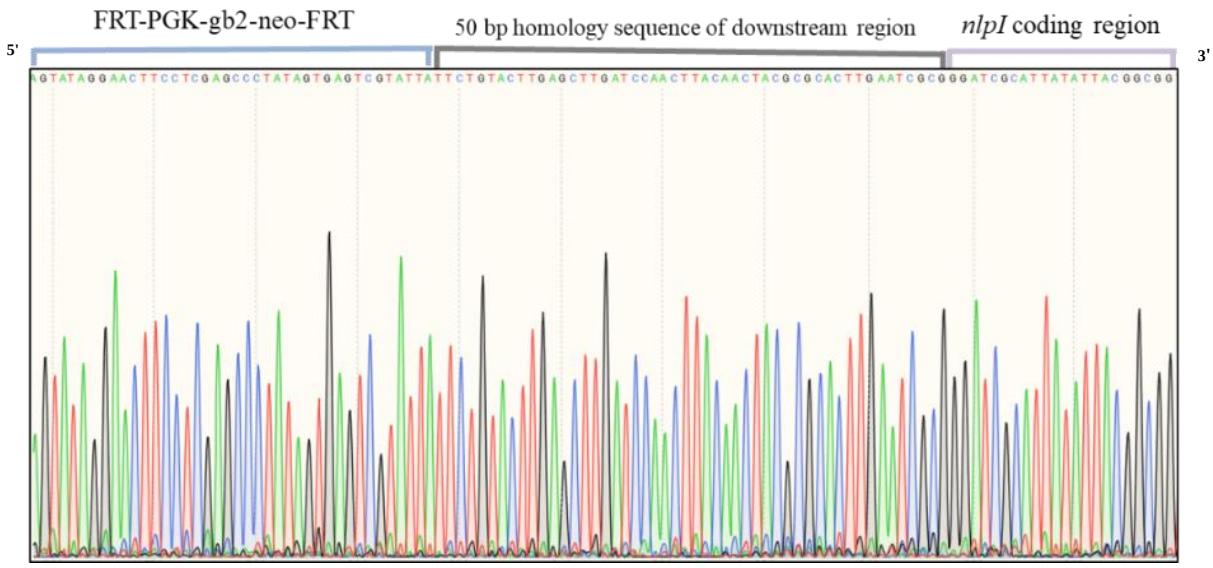


Fig. S1. Schematic representation of the original pTurboGFP–B vector. The red dotted box (inset) is indicating the sequences of the T5 promoter/lac operator element (PTS/LacO) of the original pTurboGFP–B vector along with its upstream and downstream regions. T5 promoter/lac operator element is indicated in black box. The transcription start site and the translation start codon of the GFP coding region are marked by black and green arrows respectively. The 93rd position is mentioned with the black arrow.

Figure S2: Sequence confirmation of the correct insertion of the FRT-PGK-gb2-neo-FRT fragment in the *E. coli* chromosome.

A.



B.

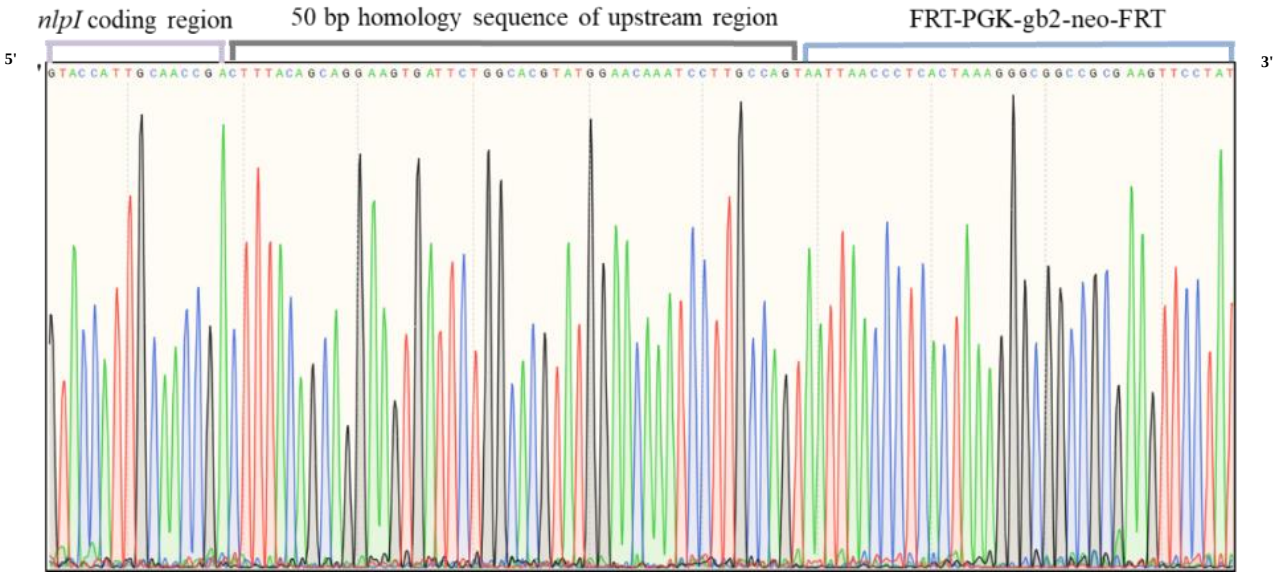


Fig. S2. Sequence confirmation of the correct insertion of the FRT-PGK-gb2-neo-FRT fragment in the *E. coli* chromosome by using (A) a forward primer (5'

TATCAGGACATAGCGTTGGCTACC 3'), complementary to a region located on the FRT-PGK-gb2-neo-FRT cassette to amplify its downstream region in the *E. coli* chromosome. (B) A reverse primer was used (5' CGAGACTAGTGAGACGTGCTAC 3') complementary to a region located on the FRT-PGK-gb2-neo-FRT cassette to amplify its upstream region in the *E. coli* chromosome.

Figure S3: Polyacrylamide gel image as loading control.

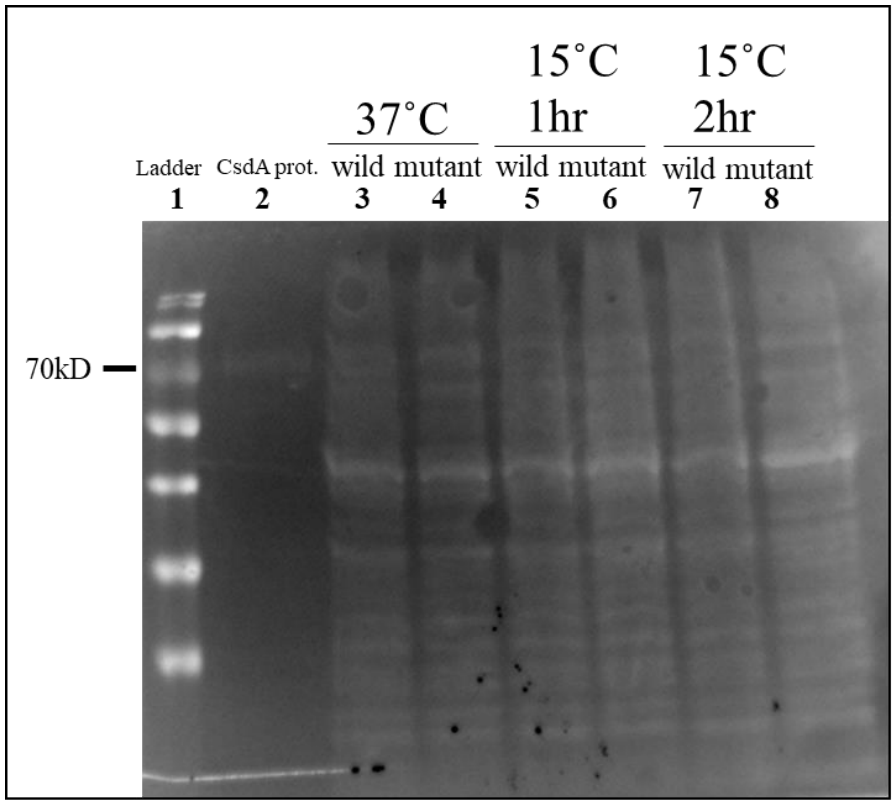


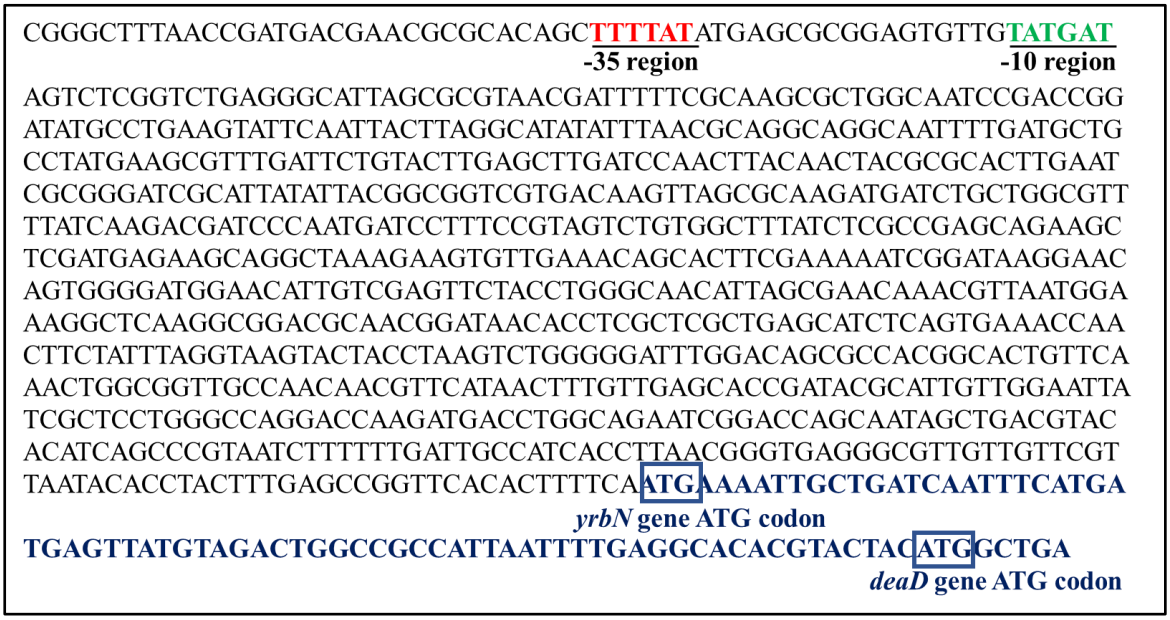
Fig. S3. Polyacrylamide gel image as loading control. Equal amount of the total cellular proteins was loaded. The gel was stained with Ponceau S stain.

Figure S4: *deaD* gene may transcribe along with *yrbN* gene and *nlpI* gene:

In continuation, our *in vivo* deletion experiment of the 200bp fragment region (Fig. 4A) from the *E. coli* chromosome revealed that the deletion could not completely abolish the expression of the *deaD* gene. Thus, we found *deaD* gene may also express along with the upstream *nlpI* gene. To check it, we have reverse-transcribed *deaD* cDNA from the total cellular RNA pool using the reverse primer 194 nucleotides downstream of the *deaD* start codon. Along with this reverse primer, we have amplified *deaD* cDNA with a forward primer which is 89 nucleotides upstream of the Δ100 region (905 nucleotides upstream of *deaD* gene coding region). We have found an approximately 1188bp long amplified product (Fig. S4C). Furthermore, we have reverse transcribed the *deaD* mRNA in the total cellular RNA pool using that same reverse primer 194 nucleotides downstream of the start codon of the *deaD* gene and ran the cDNAs on the denaturing 5% polyacrylamide gel containing 8M urea. We found an approximately 1000 nucleotide long cDNA product both at 37˚C and 15˚C. As the reverse primer is complementary to the region 194 nucleotides downstream of the *deaD* start codon, this cDNA product may represent the 800 nucleotides long 5'UTR (Fig. S4D). We also found two other cDNA products larger than 1000 nucleotides (Fig. S4D). One of these products (second largest) is present at both the 37˚C and 15˚C samples. As the start codon of the *nlpI* gene is 1064 nucleotides upstream of the start codon of the *deaD* coding region (promoter and the transcription start site of *nlpI* gene are still not known in *E. coli*), the approximate length of this transcript indicating that the *deaD* gene may also be transcribed along with the *nlpI* gene. The largest cDNA product is only present at the 15˚C sample (Fig. S4D). As the transcription start site of the *pnp* gene (immediately upstream of *nlpI* gene) is 3466 nucleotides upstream of the start codon of the

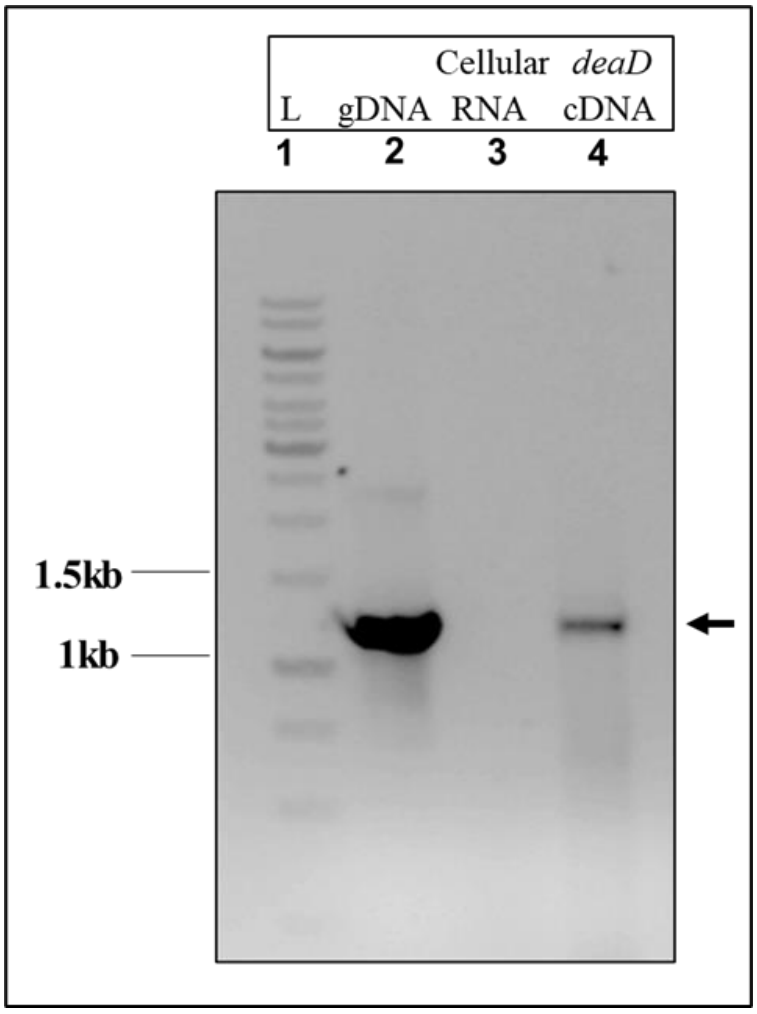
deaD coding region, the length of this cDNA product indicates that *deaD* mRNA may also be transcribed from the promoter of the *pnp* gene during the cold-shock condition.

A.



B.

C.



D.

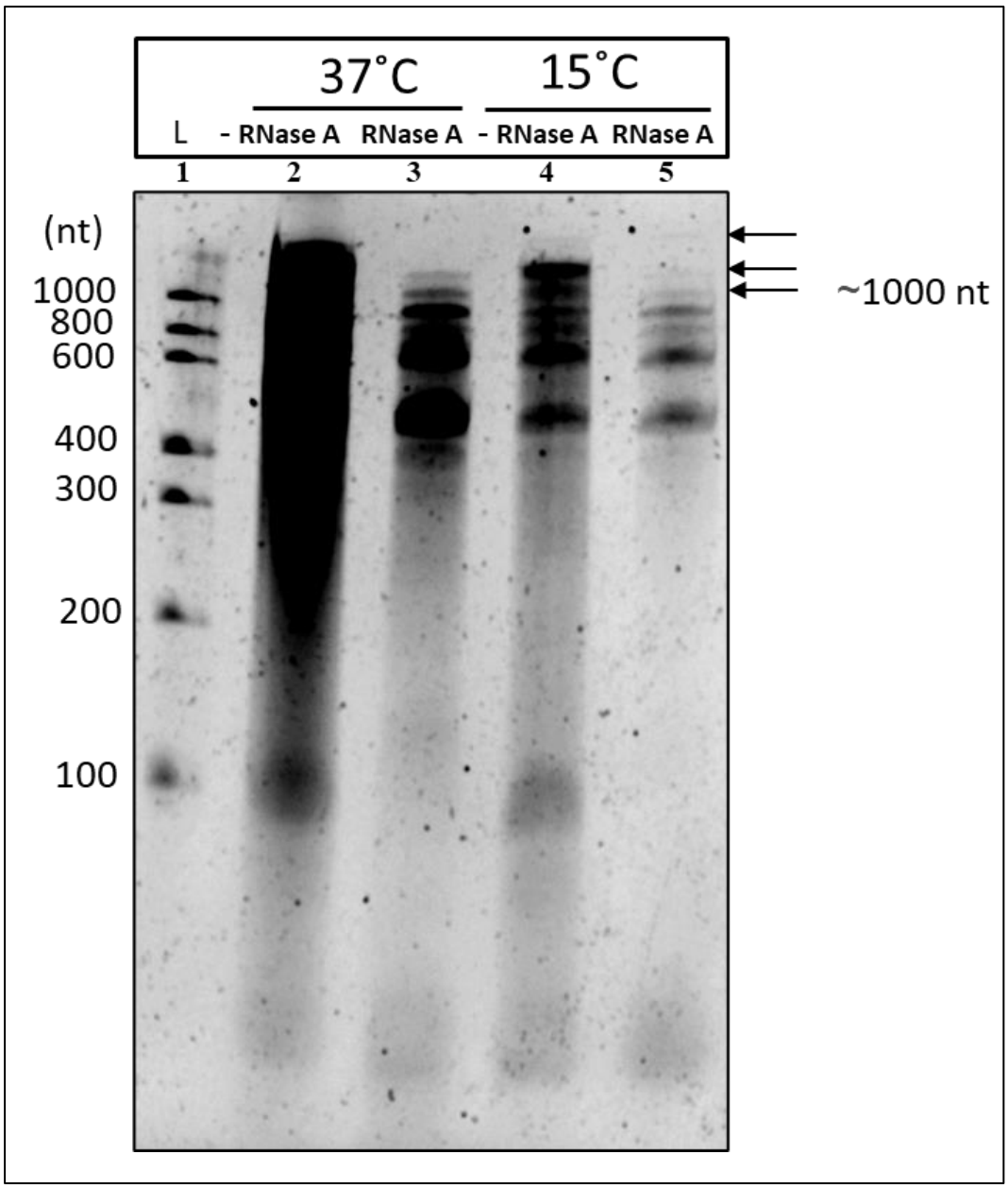


Fig. S4. Evidence of *deaD* gene as operon with *yrbN* gene and *nlpI* gene. (A) -35 region and -10 region are underlined. 'ATG' initiation codon of the *yrbN* gene and *deaD* gene are marked in boxes. The *yrbN* ORF is shown in blue and bold font. (B) Alignments of short ORFs of *yrbN* gene. The conserved sequences are indicated as '*. The 'ATG' initiation codon and

'TGA' stop codon of the *yrbN* gene are shown in red boxes. The 'ATG' initiation codon of *deaD* gene is shown in blue boxes. (C) PCR amplified products of cDNA of *deaD* mRNA. Lane 1 (L) is the DNA ladder. Lane 2 represents the amplified product from wild-type genomic DNA. Lane 3 represents the amplified product from the total cellular RNA pool (no reverse transcriptase added). Lane 4 represents the amplified product from *deaD* cDNA. The length of the amplified product is ~ 1188bp (shown by arrow). (D) *Urea-polyacrylamide gel image of the reverse transcribed product of the *deaD* mRNA. Lane 1 is the RNA ladder. Lane 2 and lane 4 are the cDNA samples of 37˚C and 15˚C respectively that are not treated with the RNase A enzyme. Lane 3 and lane 5 are the cDNA samples of 37˚C and 15˚C respectively are treated with the RNase A enzyme.

***Methods : Urea-polyacrylamide gel image of the reverse transcribed product of the *deaD* mRNA.**

5% denaturing polyacrylamide gel electrophoresis for *deaD* cDNA population study:

The *E. coli* cells were overnight grown at 37ºC. 1% inoculum from the overnight grown culture was inoculated in the fresh LB media and grown at 37ºC. When the culture reached the mid-log phase (O.D. ~ 0.5-0.6) half of the culture was pelleted down (i.e., 37ºC 0hr sample). The remaining culture was transferred to 15˚C and after 1hr of cold-shock, the cold-shock culture was pelleted down (i.e., 15ºC 1hr sample). The total cellular RNA was isolated using RNA isolation kit (Direct-zol™ RNA kit, Zymo Research) from the 37ºC 0hr sample and the cold-shock 15ºC 1hr sample. Any trace amount of contaminated genomic DNA was removed by DNase I enzyme treatment at 37˚C for 30 min. DNase I enzyme was inactivated at 65˚C for 10 min in presence of EDTA. For the cDNA synthesis purposes, the RNA samples were denatured at 80˚C for 5min and kept immediately on ice. *deaD* cDNA was synthesized with the reverse transcriptase enzyme (RevertAid First Strand cDNA Synthesis Kit (Thermo scientific)) at 42˚C

for 1hr using *deaD* mRNA specific reverse primer 5′AACAGAGGTAAAGAGAATGCTGCAGTTTTTCCGCTCCCC 3′ (complementary to the 194 nucleotides downstream of the *deaD* start codon). The specificity of this 39 nucleotides long reverse primer was checked in the BLAST web server. The reaction was terminated at 70˚C for 10 min. After the cDNA synthesis, the reverse transcribed product was treated with the RNase A enzyme at 37˚C for 30 min to remove the RNA samples. The amplified cDNA species were run on the denaturing 5% polyacrylamide gel containing 8M urea in 1X TAE buffer at 100V for 2hrs. Before the sample loading, all the cDNA samples were incubated at 80˚C for 5min and were chilled on ice immediately.