nature microbiology

SUPPLEMENTARY INFORMATION

VOLUME: 2 | ARTICLE NUMBER: 17003

In the format provided by the authors and unedited.

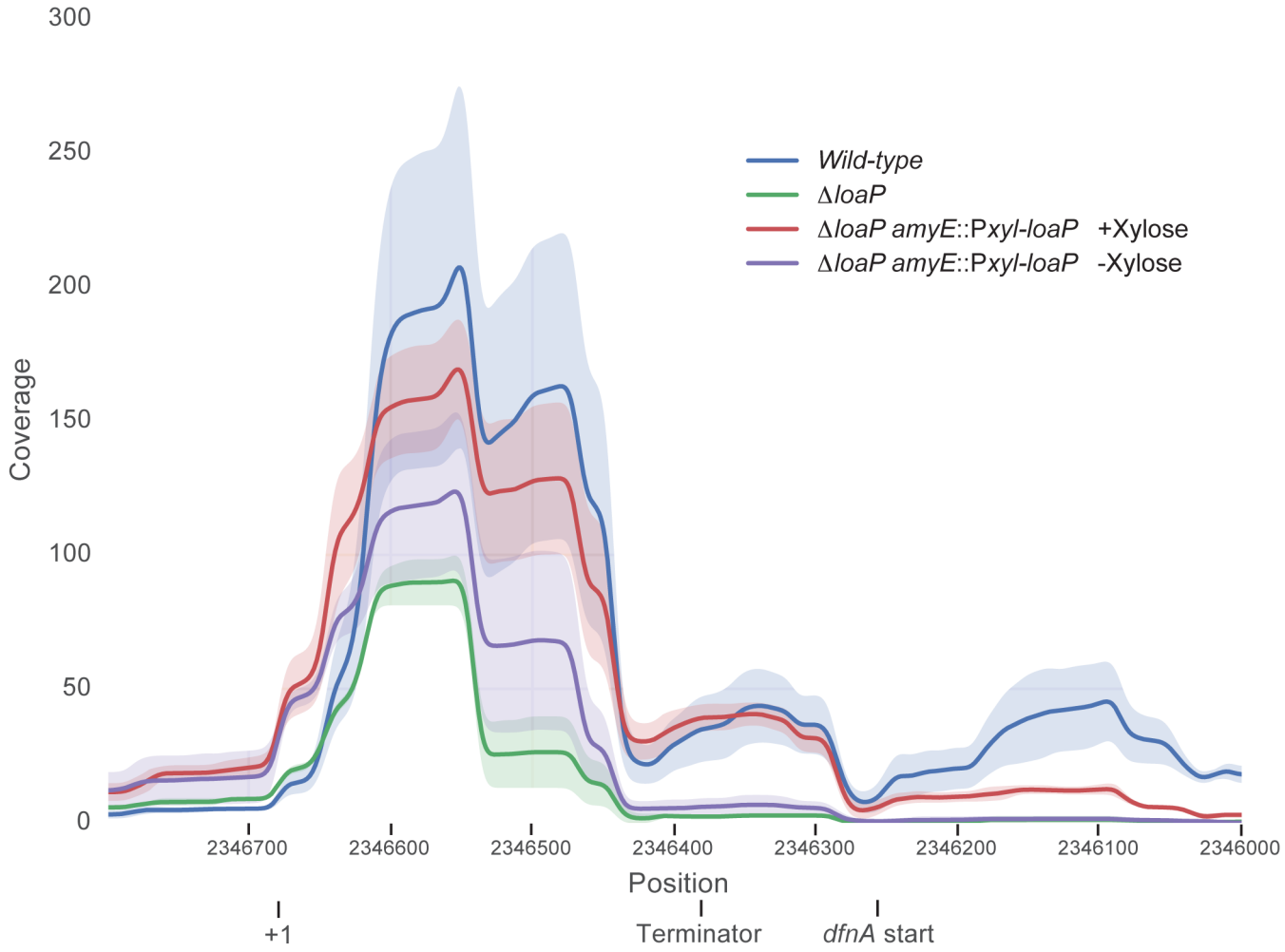
# LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens*

**Jonathan R. Goodson[1], Steven Klupt[1], Chengxi Zhang[2], Paul Straight[2,3], Wade C. Winkler[1,3]**

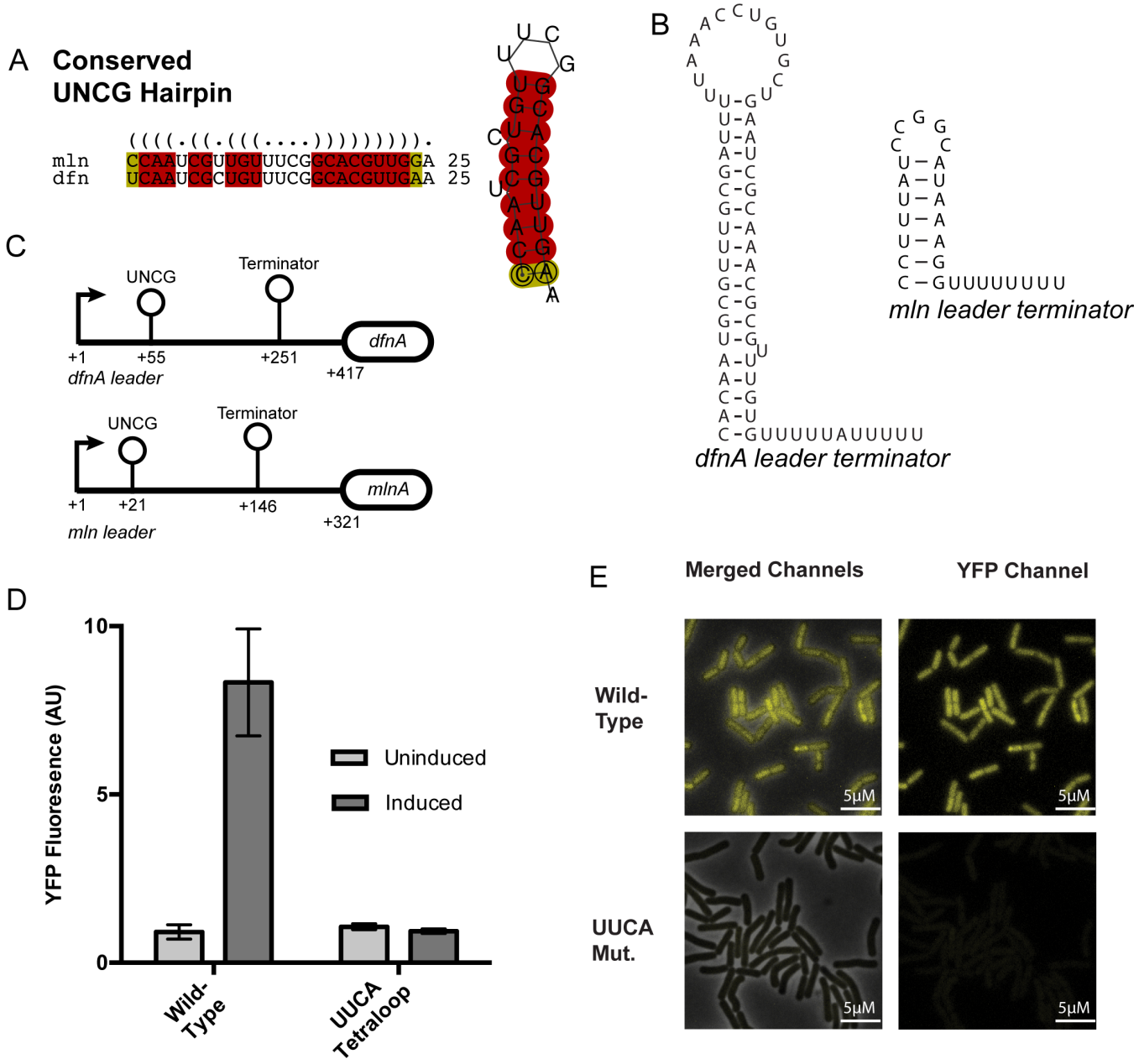
[1]The University of Maryland, Department of Cell Biology and Molecular Genetics, College Park, MD;

[2]Texas A&M University, Department of Biochemistry and Biophysics, College Station, TX

[3]Corresponding authors

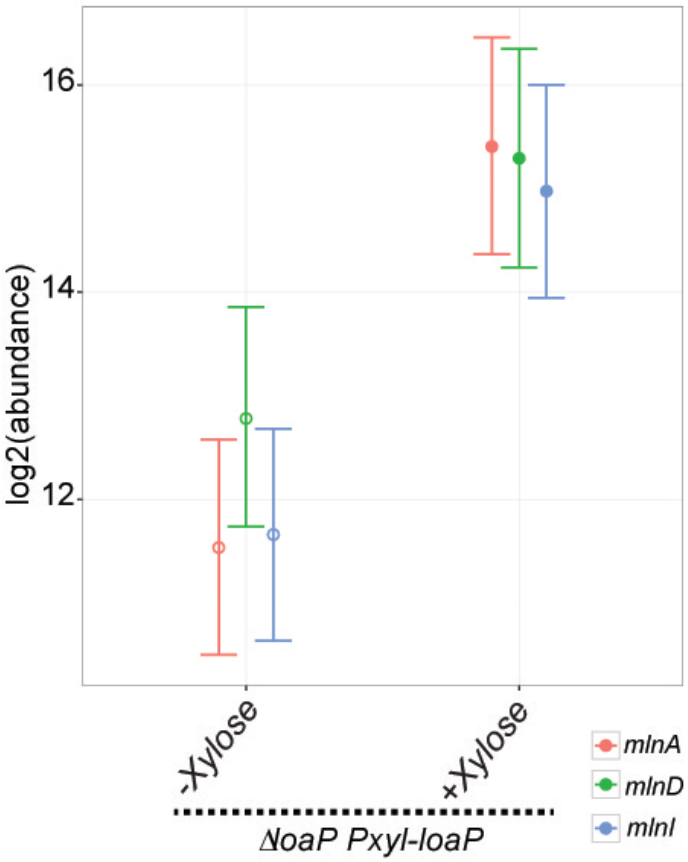



**Supplementary Figure 1.** Coverage of the *dfnA* leader region in RNA-seq data. RNA-seq coverage across the *dfnA* leader region normalized with DESeq2 normalization factors. Traces represent coverage data smoothed with Gaussian smoothing with a bandwidth of 5 nucleotides. Shading represents standard deviation from libraries from three independent cultures for each condition.

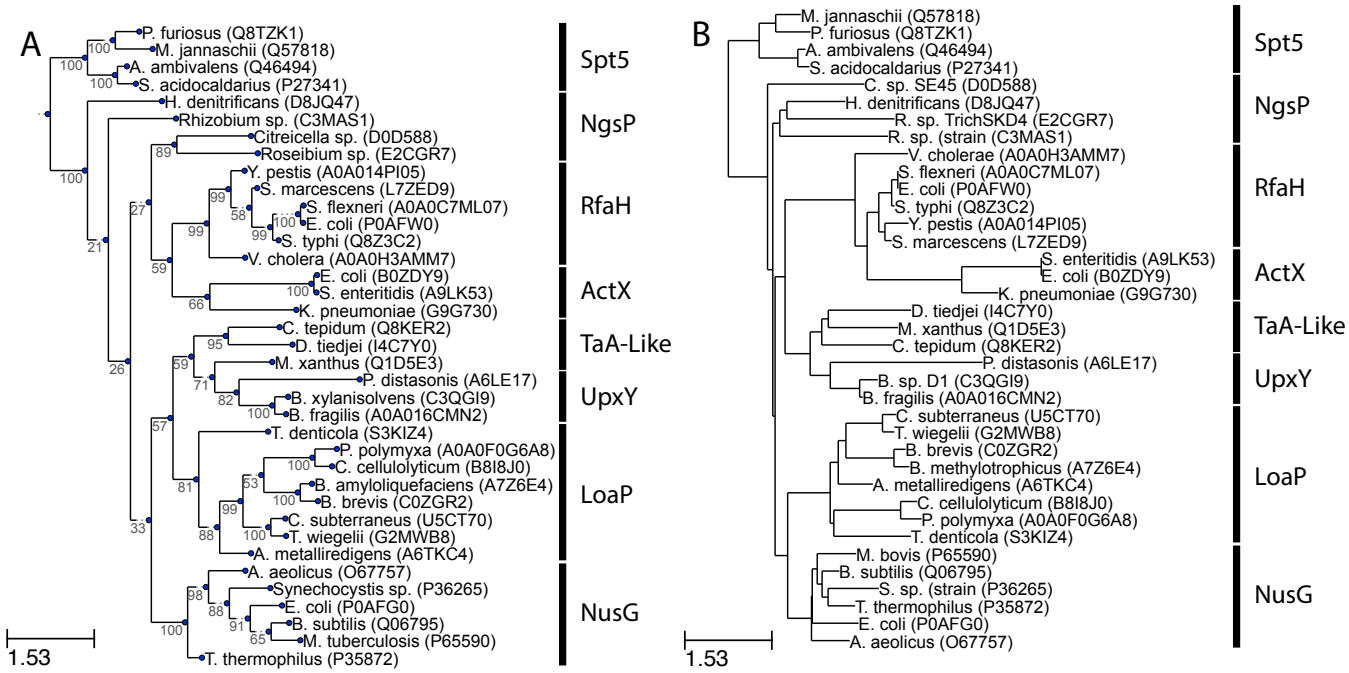



**Supplementary Figure 2.** Conserved RNA elements are present in the *dfn* and *mln* leader regions. (A) A highly similar RNA hairpin with a UUCG tetraloop is present in the upstream portion of both leader regions. (B) Predicted intrinsic terminator sequences are present near the center of both leader regions. (C) Both leader regions contain a similar UNCG tetraloop and intrinsic terminator in long 5' leader regions that do not contain any significant open reading frame. (D) Quantified fluorescence microscopy data of *B. subtilis* strains carrying a xylose-inducible *loaP* and a *dfnA* leader-*yfp* reporter cassette. The

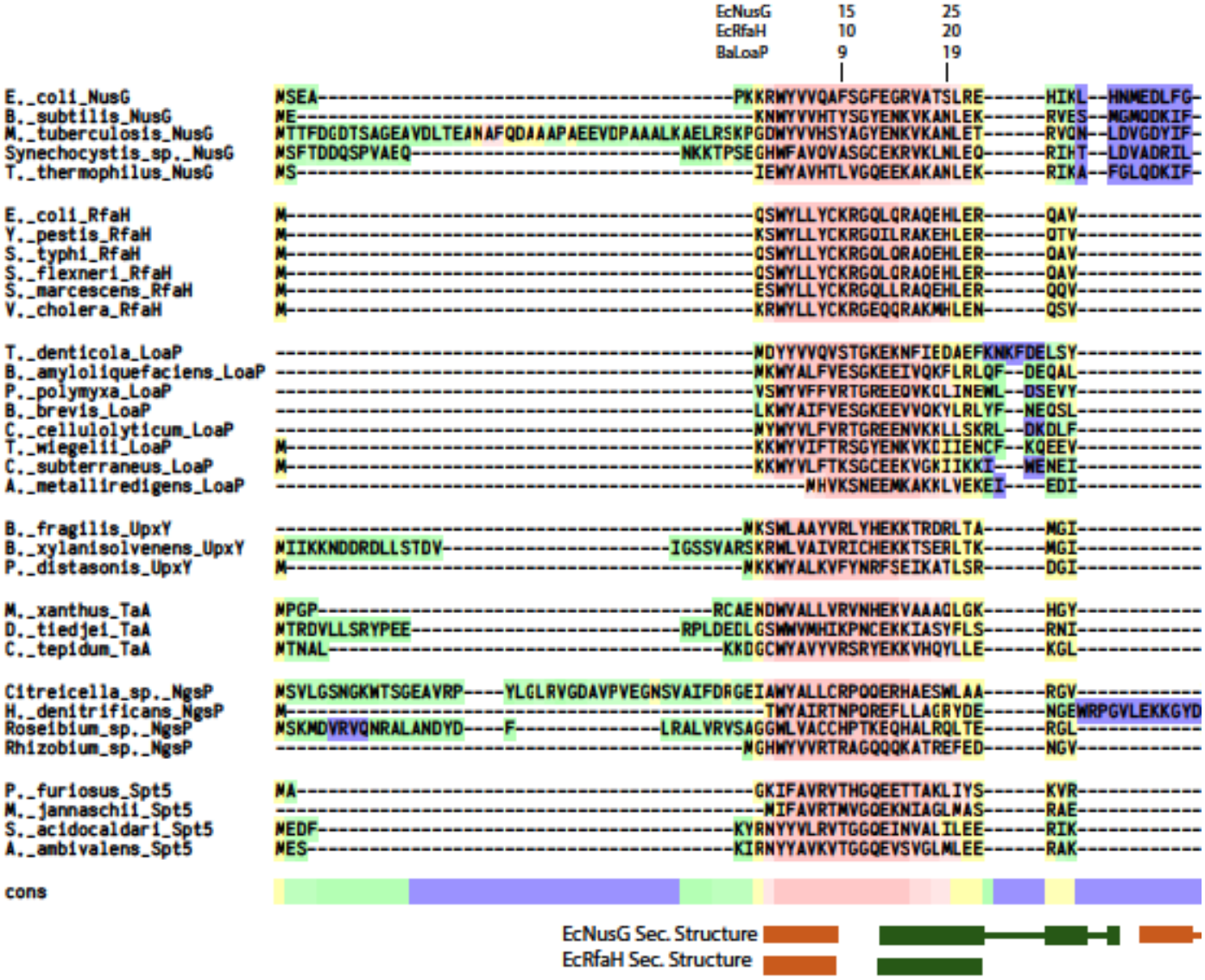
*yfp* reporter contained a constitutive promoter upstream of the *dfnA* leader region, which was transcriptionally fused to a downstream *yfp* gene. Reporter constructs were created with or without a mutation converted the UUCG tetraloop sequence to UUCA, a mutation that is predicted to abolish proper RNA hairpin formation. These data demonstrate that *B. amyloliquefaciens* LoaP antitermination can be recapitulated in the heterologous *B. subtilis* host. Data and error bars represent the mean fluorescence and 95% confidence interval (CI over means of each replicate) for all cells in three fields of view for each of three biological replicate cultures of each strain with and without induction. (E) Representative images of induced reporter strains quantified in (D) showing lack of reporter expression when the tetraloop sequence is mutated.

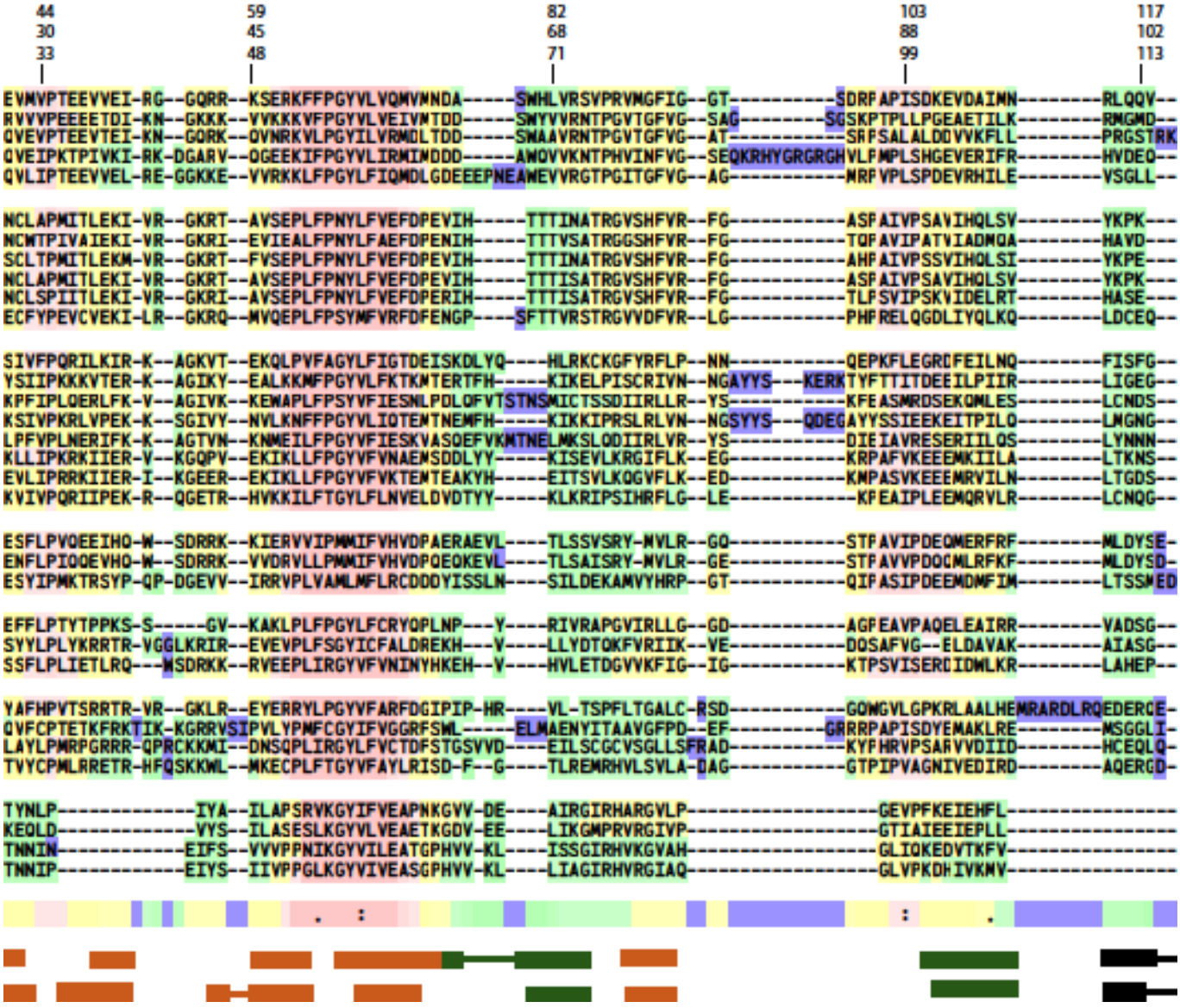


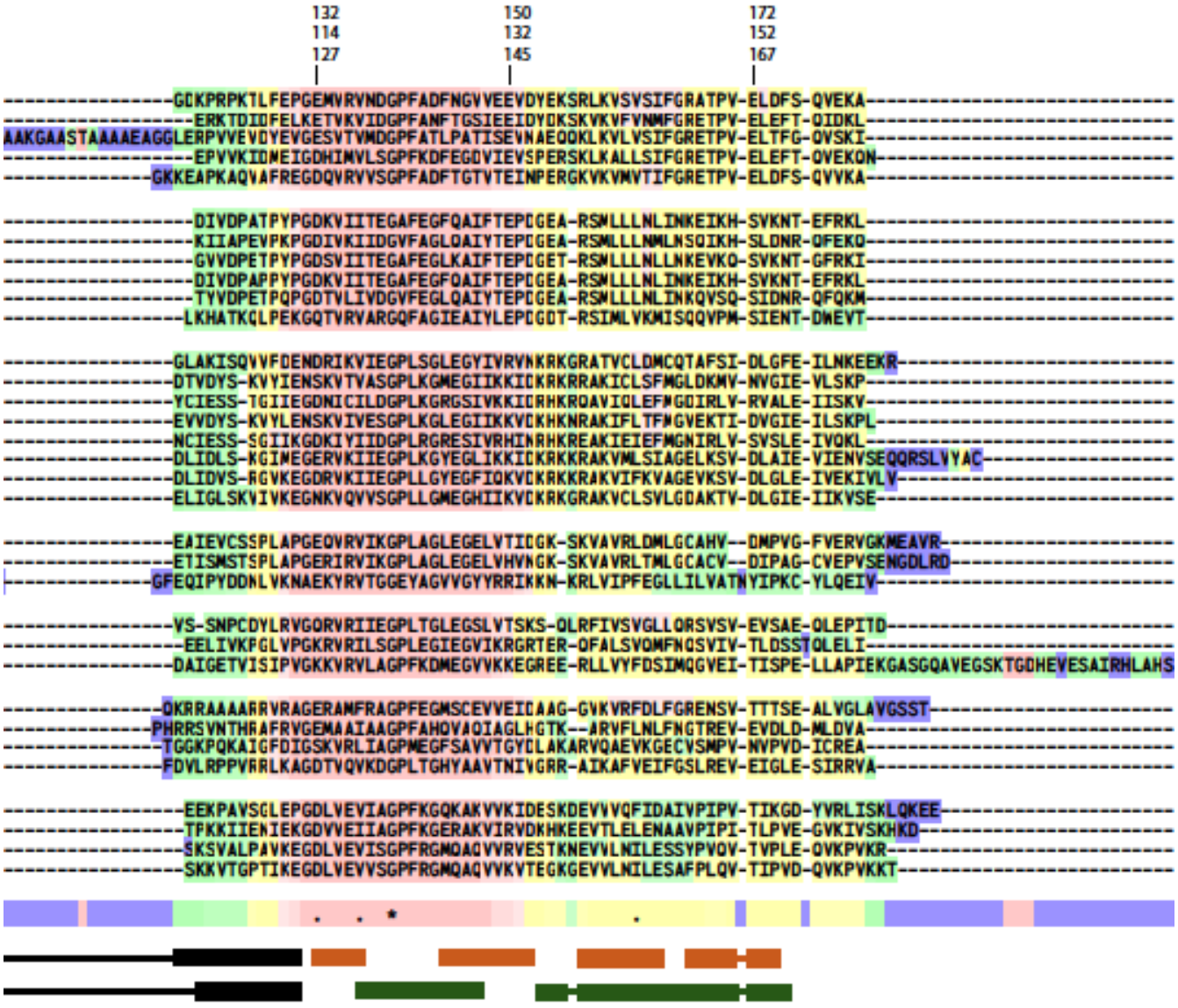

**Supplementary Figure 3.** RNA transcript levels for genes in the *mln* gene cluster are increased upon LoaP induction. Normalized transcript abundance at the beginning, middle, and end of the *mln* operon (*mlnA*, *mlnD*, and *mlnI*) as measured by qRT-PCR. Filled points represent samples with *loaP* expression and empty points represent samples with no or minimal *loaP* expression. Error bars represent Bayesian 95% highest posterior density estimates of mean expression. Data resulted from four biological replicate cultures.




**Supplementary Figure 4.** Large-scale phylogenetic analysis recapitulates targeted analysis of representative sequences. (A) Phylogenetic tree constructed from multiple sequence alignment of 37 representative protein sequences. (B) Subtree consisting of the same 37 protein sequences extracted from large-scale 1,207 sequence phylogenetic analysis. Both trees show very similar topology.

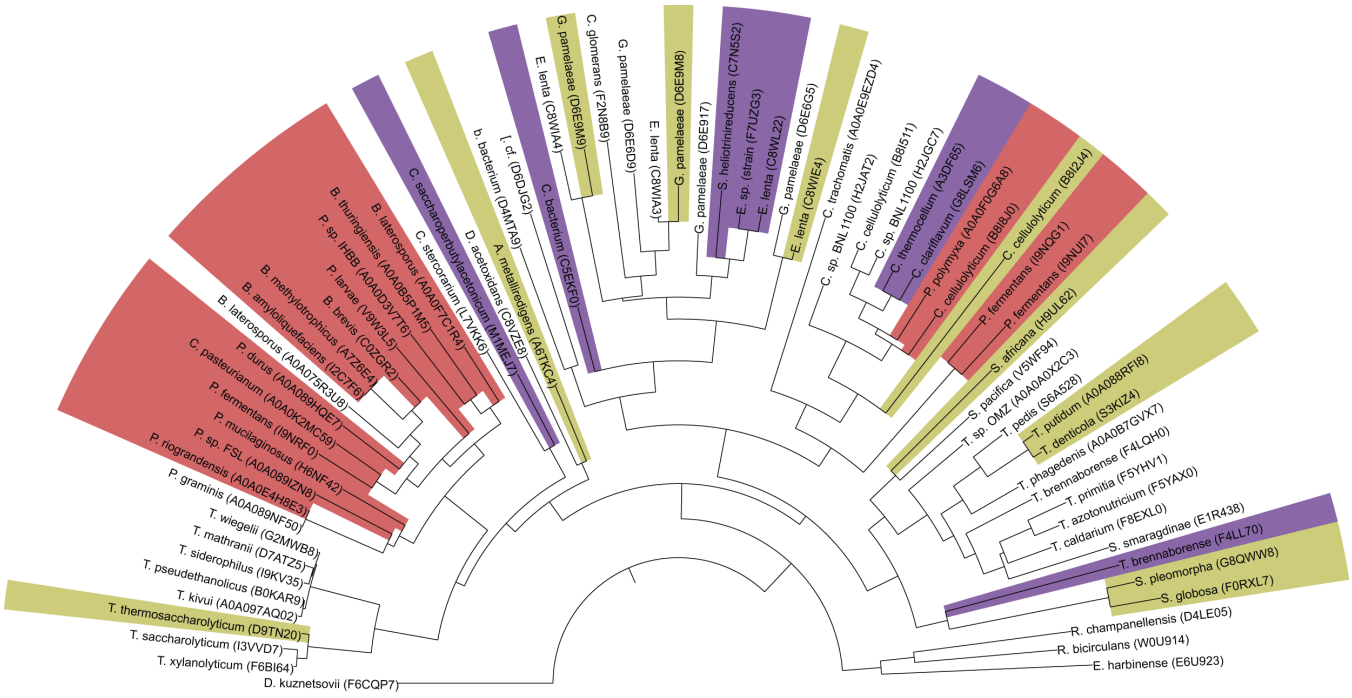

EcNusG 15 25
EcRfaH 10 20
BaLoaP 9 19
E._coli_NusG
B._subtilis_NusG
M._tuberculosis_NusG
Synechocystis_sp._NusG
T._thermophilus_NusG
E._coli_RfaH
Y._pestis_RfaH
S._typhi_RfaH
S._flexneri_RfaH
S._marcescens_RfaH
V._cholera_RfaH
T._denticola_LoaP
B._amyloliquefaciens_LoaP
P._polymyxa_LoaP
B._brevis_LoaP
C._cellulolyticum_LoaP
T._wiegelii_LoaP
C._subterraneus_LoaP
A._metalliredigens_LoaP
B._fragilis_UpxY
B._xylanisolvenens_UpxY
P._distasonis_UpxY
M._xanthus_TaA
D._tiedjei_TaA
C._tepidum_TaA
Citreicella_sp._NgsP
H._denitrificans_NgsP
Roseibium_sp._NgsP
Rhizobium_sp._NgsP
P._furiosus_Spt5
M._jannaschii_Spt5
S._acidocaldarii_Spt5
A._ambivalens_Spt5
cons
EcNusG Sec. Structure
EcRfaH Sec. Structure


44
30
33
59
45
48
82
68
71
103
88
99
117
102
113

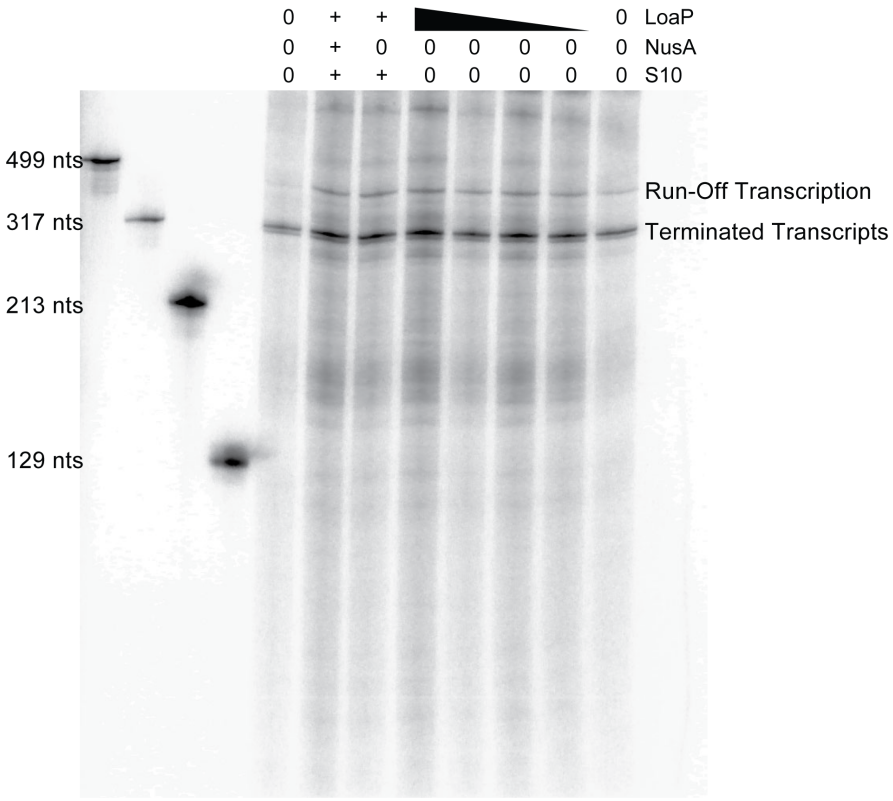



**Supplementary Figure 5.** Multiple sequence alignment of NusG family protein sequences reveals conserved differences between subtypes of specialized paralogs. Alignment contains 33 protein sequences containing representatives of each subtype of NusG paralog (excluding ActX) aligned by T-COFFEE using the "accurate" alignment method combining PSI-COFFEE and EXPRESSO alignment methods utilizing available crystal structure data. Residues are colored according to T-COFFEE consistency score representing pairwise reliability in alignment with all other sequences. Also shown are secondary structure diagrams representing the secondary structure from PDB structures of E. coli NusG

and RfaH (Orange represents beta-strands and green represents alpha-helices, black represents relatively unstructured inter-domain linker).




**Supplementary Figure 6.** LoaP represents a distinct group of NusG specialized paralogs and is commonly associated with large biosynthetic gene clusters. Subset of large-scale phylogenetic analysis showing LoaP homolog sequences. Background shading represents association of gene sequences with large gene clusters. Red shading denotes LoaP near PKS (polyketide synthase) or NRPS (nonribosomal peptide synthase) gene clusters. Purple shading denotes LoaP near polysaccharide gene clusters. Olive shading denotes LoaP near other types of antiSMASH gene clusters. Unlabeled sequences were not found nearby an antiSMASH predicted gene cluster, although some appear to be next to long stretches of coding sequences in one direction.




**Supplementary Figure 7.** Transcription of the *dfnA* leader sequence in vitro. The *B. amyloliquefaciens dfnA* leader sequence, including the endogenous promoter, was PCR amplified using a reverse oligonucleotide primer that would be predicted to result in either a 314-nucleotide transcript (for premature transcription termination at the putative intrinsic termination site) or a 420-nucleotide transcript (for run-off transcription). Reactions included 2 pmol PCR-generated DNA template, 20 mM Tris-HCl pH 8.0, 15 mM NaCl, 4 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT, 0.01% Triton X100, 1 mM NTPs, 1.5 pmol (5 μCi) α-$^{32}$P-UTP, and 160 nM $\sigma^A$-saturated *B. subtilis* RNAP. LoaP was added to final concentrations of 300 nM, 150 nM, 75 nM, and 3.7 nM. Purified hexahistidine-tagged *B. subtilis* S10 was added to 150 nM and purified hexahistidine-tagged *B. subtilis* NusA was added to 50 nM, as indicated. The reactions were incubated at 37 °C for 1 hour and products were resolved alongside *dfnA*

size markers by 6% urea-denaturing polyacrylamide gel electrophoresis. Gel image is representative of multiple transcription experiments.