Supporting Information

7-deazaguanine derivatives in DNA: a complex modification system encoded on a widespread genomic island

Jennifer J. Thiaville, Stefanie M. Kellner , Yifeng Yuan, Geoffrey Hutinet, Patrick C. Thiaville, Watthanachai Jumpathong, Susovan Mohapatra, Celine Brochier-Armanet, Andrey V. Letarov, Roman Hillebrand, Chanchal K. Malik, Carmelo Rizzo, Peter C. Dedon, and Valérie de Crécy-Lagard

SI Material and Methods

Bioinformatics Analyses

Analysis of the gene clusters was performed using tools on the public SEED server (http://pubseed.theseed.org/SubsysEditor.cgi) (1, 2) and, unless otherwise stated, we restricted our analyses to genomes only present in the public SEED database. Genes physically clustered with the *tgtA5* gene were analyzed in all organisms containing a *tgtA5* gene using the the SEED subsystem coloring tool or the Seedviewer Compare regions tool. Presence of each *dpd* gene in the clusters was identified using the Seedviewer Compare regions tool, "Edit Empty Cells" tool on the SEED spreadsheet page, or "Find this gene in organism" on the SEEDviewer page for a gene of interest. The tools employ similarity based searches (BLAST or tBLASTn) to identify the closest matches to the gene/protein of interest. Once an open reading frame was identified, we assessed physical clustering with *tgtA5* and analyzed domain architecture using the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (9). Genes annotated as *dpd* genes were required to meet two criteria, in addition to sequence similarity: 1) they must be encoded in the neighborhood of a *tgtA5* gene; 2) the encoded protein must have similar predicted domains (See Fig. S5 for domains present in *S.* Montevideo Dpd proteins).

Taxonomic distribution of *tgtA5* was also visualized using the Interactive Tree of Life (iToL, itol.embl.de) (3). All information for NCBI TaxIDs and presence or absence of genes was exported from the SEED subsystem "dpd cluster". A text file containing the NCBI TaxIDs of representative organisms was uploaded into iToL and the tree was generated using the "Generate Tree" tool. The tree was added to the project on the user's account. Datasets for each gene containing the TaxID and 1 for presence or 0 for absence of the *tgtA5* were uploaded to the project, binary was selected for "Data Type", and the tree containing the data for presence or absence of the *tgtA5* was generated.

For other sequence analyses, the BLAST tools (4) and resources at NCBI (http://www.ncbi.nlm.nih.gov/) were routinely used. Sequence alignments were built using MUSCLE (5) and Multialin (6), and sequence logos were built from the alignments using WebLogo (7). Protein domain analysis was performed using the Pfam database tools (http://pfam.xfam.org/) (8), the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (9), HHpred (http://toolkit.tuebingen.mpg.de/hhpred) (10), and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/) (11). For detailed domain analysis of *S.* Montevideo sequences presented in Fig. S5 and Table S2, protein sequences for DpdA to DpdK were acquired from the SEED database and used as queries against the Pfam database through HHpred (10) with the default parameters selected. The top ranking hits, with a probability over 70%, were used for identification and are reported in Table S2. For regions of a protein with either no hits or low probability scores, the sequence of that region was resubmitted as a separate query.

The sequence alignment of select aTGT, bTGT, and TgtA5 proteins presented in Figure 2B was performed using MUSCLE (5). The protein sequences were retrieved from the SEED database using the following accession numbers: *S.* Montevideo, fig|745015.3.peg.1828; *Ferrimonas balearica*, fig|550540.3.peg.3256; *Sphingopyxis alaskensis*, fig|317655.17.peg.2735; *Comamonas testosteroni*, fig|1009852.3.peg.2809; *Kineococcus radiotolerans*, fig|266940.5.peg.4549; *E. coli*, fig|340185.4.peg.3772; *Zymomonas mobilis*, fig|264203.3.peg.859; *Shigella flexneri*, fig|766155.3.peg.537; *Bacillus subtilis*, fig|1148.72.peg.5059; *Aquifex aeolicus*, fig|224324.1.peg.910; *Pyrococcus horikoshii*, fig|70601.1.peg.1088; *Methanococcus aeolicus*, fig|419665.8.peg.915; *Thermoplasma volcanium*, fig|273116.1.peg.1497; *Picrophilus torridus*, fig|263820.1.peg.454; *Ferroplasma acidarmanus*, fig|97393.1.peg.1817.

Phylogenetic analysis of TgtA5/DpdA and DpdB

A survey of the 30578 proteomes deduced from complete and ongoing genome projects available at the NCBI allowed retrieval of 284 DpdA homologues. The corresponding protein sequences were aligned with MAFFT version 7 (option -L-INS-i) (12) The resulting alignments were trimmed with BMGE (BLOSUM30 option) (13). Maximum Likelihood (ML) trees were inferred with PhyML version 3.1 (14), with the Le and Gascuel model (LG) (15) and a gamma distribution to take into account evolutionary rate among site heterogeneity. A NNI + SPR strategy was used for topology search. The robustness of the ML tree was evaluated by the non-parametric bootstrap strategy implemented in PhyML (100 replicates of the original alignment). Bayesian trees were inferred with MrBayes v3.2.2 (16). MrBayes was run with four chains for 1 million generations and trees were sampled every 100 generations. To construct the consensus tree, the first 2000 trees were discarded as "burn-in".

For analysis of the DndB-like family of proteins, sequences of the DndB-like Dpd proteins clustering with DpdA were acquired from the SEED database through the "dpd cluster" subsystem and DndB proteins clustering with DndC and other phosphorothioate proteins were acquired from the SEED database through the "DNA phosphorothioation" subsystem. Sequences were submitted to phylogeny.fr (http://www.phylogeny.fr/advanced.cgi) (17) using default parameters in "Advanced" mode.

Bioinformatic identification of phage Tgt-like proteins

Bacteriophage Tgt-like proteins were identified by 3 iterations of psiBLAST, with standard parameters, using the phage 9g Tgt-like protein (18) sequence as query against the Virus database of NCBI. Each of the protein hits was verified by HHpred (10) using default parameters on the Bioinformatic toolkit website from the Max-Plank institute (http://toolkit.tuebingen.mpg.de/hhpred). The pdb70 HMM database was searched, and the multiple alignment methods were performed with HHblitz with three iterations. The sequences returned either the archaeal Tgt or the bacterial Tgt as predicted homologs. For each phage genome containing a *tgt*-like gene, the presence of Q biosynthesis pathway genes (*folE, queD, queE, queC, queF, queA, and queG*) was investigated by BLASTp and then confirmed by HHpred. All protein sequences verified by HHpred returned their respective predicted homolog with a probability of 100 %.

Strains, media, and growth conditions

All strains used in this study are listed in Table S3. *S.* Montevideo strains and *E. coli* C600 were grown in LB (Tryptone 10 g/L, yeast extract 5 g/L, NaCl 5 g/L) at 37 °C. *Kineococcus radiotolerans* was grown in TYG (Tryptone 5.0 g/L, yeast extract 5.0 g/L, Glucose 5.0 g/L) at 30 °C. Strains obtained from DSMZ were grown in conditions recommended by the company. *Comamonas testosteroni* DSM 50244 was grown in Nutrient Broth (Peptone 5.0 g/L, Beef extract 3.0 g/L) at 30 °C. *Ferrimonas balearica* DSM 9799 was grown in Difco Marine Broth 2216 (BD, Franklin Lakes, New Jersey) at 28 °C. *Meiothermus chliarophilus* DSM 9957 was grown in Castenholz medium (DSMZ Medium 86) at 50 °C. *Sphingopyxis alaskensis* DSM 13593 was grown in Tryptone Soya Broth (TSB, Oxoid) at 28 °C. For all solid media, 15 g/L agar was added.

Construction of mutations in *Salmonella* Montevideo genes

All deletion constructs were made using the linear recombination method described by Datsenko and Wanner (19). Deletion primers (Table S4) were designed with 50 bp of homology to the gene to be deleted. The P1 or P2 sites were added to the 3' end of each primer to amplify the chloramphenicol or kanamycin resistance cassettes from pKD3 and pKD4, respectively (19). For the 21-kb deletion of nearly the entire *dpd* cluster, the primers matched the 5' end of the *dpdC* gene and the 3' portion of the *dpdD* gene. PCR products were cleaned using the Zymo Clean and Concentrator-5 kit (Zymo Research), and 200 ng of PCR product was transformed via electroporation into freshly prepared electrocompetent *S.* Montevideo carrying the temperature-sensitive pKD46. Chloramphenicol or kanamycin resistant isolates were selected on LB-agar plates with 15 µg/mL of chloramphenicol or 40 µg/mL of kanamycin, respectively. All mutations were verified by PCR for loss of the target gene(s) and presence of the antibiotic resistance marker in the correct locations.

Plasmid Restriction Test

Restriction of the plasmid pUC19 was tested using a modified transformation efficiency method (20, 21). The pUC19 plasmid was propagated through *S.* Montevideo WT or YYF3022 (Δ*dpdB-dpdC*::kan) and isolated from each strain using QIAprep Spin Miniprep Kit (Qiagen). The presence or absence of the modification on pUC19 was confirmed as described in DNA analysis. 10 ng of each plasmid (modified or unmodified) was transformed by electroporation into freshly prepared electrocompetent *S.* Montevideo WT and YYF3022. Transformants were recovered in 1 mL LB for 1 hour at room temperature and dilutions were plated onto LB agar plates with or without ampicillin (100 µg/mL). Transformation efficiency was determined by calculating the number of Amp^R transformants per 10^6 viable CFU and dividing by the ng of DNA transformed.

Bacterial DNA isolation

A 500 mL culture of each strain was grown in the proper conditions described above. Cells were harvested by centrifugation at 10,000 x g for 20 min. Pellets were washed once in 40 mL TEN buffer (100 mM Tris, pH 8.0; 100 nM NaCl; 10mM EDTA, pH 8.0) and centrifuged again. Pellets were suspended in 20 mL TEN buffer with 600 µg/ml lysozyme, 1 % SDS, 6 mAU/ml Proteinase K (Qiagen), and incubated at 56 °C for 1 h. Following incubation, an equal volume of Tris-buffered phenol, pH 7 was added to each sample and incubated with shaking at room temperature overnight. The samples were centrifuged at 4,000 x g at room temperature for 10 min and the aqueous phase was transferred to the new tubes. The extractions were repeated with

the mixture of 25:24:1 phenol:chloroform:isoamyl alcohol followed by a chloroform extraction. After the final extraction, the aqueous phase was treated with 200 µg/mL RNAse A (Qiagen) for 30 min at 37 °C. The RNAse was removed by another round of phenol:chloroform extraction. The DNA was precipitated with 0.1 volume sodium acetate and 1 volume isopropanol and spooled with a glass rod. After washing briefly in 70% ethanol, the DNA was suspended in sterile deionized water.

Phage 9g DNA isolation

50 mL of LB medium was inoculated with 1 mL of an overnight culture of *E. coli* C600 and incubated at 37 °C with shaking at 250 rpm until an OD_{600} 0.6 was reached. The culture was inoculated with 10^9 PFU of bacteriophage 9g and incubated in the same conditions for an additional 5 h. The lysate was cleared by centrifugation at 10,000 x g for 10 min and the phage titer was determined (expected yield, approximately 10^{11} PFU/mL). The lysate was treated with DNAse (0.01 mg/mL) and RNAse (0.05 mg/mL) for 1 h at 37 °C. The phage was pelleted by ultracentrifugation in JS-28.35 rotor (Beckmann-Coulter, USA) at 70,000 x g for 1 h through 1 cm of a 20% glycerol pad prepared on SM buffer without gelatin (100 mM NaCl, 1 mM of $MgCl_2$, 50 mM of TrisHCl, pH 7.5). The supernatant was discarded and the phage was resuspended in 1 mL of SM buffer. Phage DNA was extracted with an equal volume Tris-HCl-saturated phenol, pH 8 in 1.5 ml Eppendorf tubes with gentle rotation for 1-2 min. The tubes were centrifuged for 3 min at 10,000 x g, the aqueous phase was transferred to the new tubes containing tips with cut ends, and extracted with 0.4 volumes of TE-saturated chloroform according to the same procedure. The DNA was precipitated by addition of 2.5 vol of 96% ethanol, pelleted at 13,000 x g at 4 °C at 20 min, washed with 70% ethanol and dried in a Vacuum Concentrator System (Labconco) at 40 °C for 30 min. The DNA pellet was suspended in 100 µl of sterile deionized water.

Synthesis of 7-deaza-2'-deoxyguanosine standards

Mass spectrometric analysis of standards. High resolution mass spectrometric (HRMS) measurements were performed on a Waters Synapt HDMS QTOF mass spectrometer with the following parameters: Capillary voltage 2.5 kV, sampling cone 20 V, extraction cone 3 V, source temperature 80 °C, desolvation temperature 40 °C, trap collision energy 4.0 eV, and transfer collision energy 8.0 eV.

2-Amino-7-(2-deoxy-β-D-*erythro*-pentofuranosyl)-4,7-dihydro-4-oxo-3*H*-pyrrolo[2,3-*d*]pyrimidine-5-carbonitrile (dPreQ$_0$): $dPreQ_0$ was prepared from 7-iodo-7-deaza-2'-deoxyguanosine (MyChem LLC, San Diego, CA) as previously described by Ramzaeva, *et al*. (22) and purified by HPLC using a Phenomenex Luna- 5µ C18(2) 100 Å column (250 mm × 4.6 mm, 5 µM particle size) equilibrated with 99% solvent A (0.1 M ammonium formate) and 1% solvent B (acetonitrile) and eluted with the following solvent gradient: 0 to 15 min, linear gradient from 99% to 90% A; 90% A for 5 min; 20 to 25 min, linear gradient to 80% A; 25 to 28 min, linear gradient to 20% A; 20% A for 2 min; 30 to 33 min, linear gradient to 99%. Under these conditions, $dPreQ_0$ eluted at t_r 19 min. HRMS: *m/z* calculated for $C_{12}H_{14}N_5O_4$ $[M + H]^+$, 292.1046; observed, 292.1038.

2-Amino-7-(2-deoxy-β-D-*erythro*-pentofuranosyl)-4,7-dihydro-4-oxo-1*H*-pyrrolo[2,3-*d*]pyrimidine-5-carboxylic acid (dCDG): dCDG was prepared from $dPreQ_0$ as previously described by Ramzaeva *et al.* (22) and purified by HPLC using a Phenomenex Luna- 5µ C18(2) 100 Å column (250 mm × 4.6 mm, 5 µM particle size) equilibrated with 99% solvent A (0.1 M

ammonium formate) and 1% solvent B (acetonitrile) and eluted with the following solvent gradient: 17 to 19 min linear gradient from 99% to 80% A; 80% A for 1 min; 20 to 22 min, linear gradient to 99% A. Under these conditions, dCDG eluted at 10 min. HRMS: *m/z* calculated for $C_{12}H_{15}N_4O_6$ $[M + H]^+$, 311.0992; observed, 311.1005.

2-Amino-7-(2-deoxy-β-D-*erythro*-pentofuranosyl)-4,7-dihydro-4-oxo-1*H*-pyrrolo[2,3-*d*]pyrimidine-5-carboxamide (dADG): dADG was prepared from $dPreQ_0$ as previously described by Ramzaeva *et al.* **(22)** and purified by HPLC using a Phenomenex Luna- 5μ C18(2) 100 Å column (250 mm × 4.6 mm, 5 μM particle size) equilibrated with 99% solvent A (0.1 M ammonium formate) and 1% solvent B (acetonitrile) and eluted with the following solvent gradient: 0 to 15 min, linear gradient from 99% to 90% A; 90% A for 5 min; 20 to 25 min, linear gradient to 80% A; 25 to 28 min, linear gradient to 20% A; 20% A for 2 min; 30 to 33 min, linear gradient to 99% A. Under these conditions, dADG eluted at 15 min. HRMS: *m/z* calculated for $C_{12}H_{16}N_5O_5$ $[M + H]^+$, 310.1151; observed, 310.1265.

2-Amino-7-(2-deoxy-β-D-*erythro*-pentofuranosyl)-4,7-dihydro-4-oxo-3*H*-pyrrolo[2,3-*d*]pyrimidine-5-carboximidamide (dG^+): dG^+ was prepared from $dPreQ_0$ as previously described by Brückl *et al.* (23). As noted by these authors, dG^+ reverts to $dPreQ_0$ upon purification, so dG^+ was therefore not isolated. HRMS: *m/z* calculated for $C_{12}H_{17}N_6O_4$ $[M + H]^+$, 309.1311; observed, 309.1251.

2-Amino-5-(aminomethyl)-7-(2-deoxy-β-D-*erythro*-pentofuranosyl)-3,7-dihydro-4*H*-Pyrrolo[2,3-*d*]pyrimidin-4-one ($dPreQ_1$): $dPreQ_1$ was prepared as previously described by Wang and Gold (24) and purified by HPLC using a a Phenomenex Luna-5μ C18(2) 100 Å column (250 mm × 4.6 mm, 5 μM particle size) equilibrated with 99% solvent A (0.1 M ammonium formate) and 1% solvent B (acetonitrile) and eluted with the following solvent gradient: 0 to 15 min, linear gradient from 99% to 90% A; 90% A for 5 min; 20 to 25 min, linear gradient to 80% A; 25 to 28 min, linear gradient to 20% A; 20% A for 2 min; 30 to 33 min, linear gradient to 99% A. Under these conditions, $dPreQ_1$ eluted at 10.8 min. LC-MS/MS: *m/z* calculated for $C_{12}H_{18}N_5O_4$ $[M + H]^+$, 296.1359; observed, 296.1.

DNA analysis

DNA (100 μg) was hydrolyzed in 10 mM Tris-HCl (pH 7.9) with 1 mM $MgCl_2$ with Benzonase (20U), DNase I (4U), calf intestine phosphatase (17U) and phosphodiesterase (0.2U) for 16 h at ambient temperature. Following passage through a 10 kDa filter to remove proteins, the filtrate was lyophilized and resuspended to a final concentration of 2 μg/μL (based on initial DNA quantity); a final concentration of 0.2 μg/μL was used for dADG quantification. Standards for dADG, dCDG, dG^+, $dPreQ_0$, and $dPreQ_1$were synthesized as described above.

Modified 2'-deoxynucleosides were initially detected by liquid chromatography-coupled triple quadrupole mass spectrometry (LC-MS/MS). Aliquots of hydrolyzed DNA (20 μg) were injected onto a Dionex Acclaimed Polar Advantage II C18 column (2.1 x 250 mm, 3 μm particle size) equilibrated with 95.5% solvent A (0.1% v/v formic acid in water) and 0.5% solvent B (acetonitrile) at a flow rate of 0.25 mL/min and eluted with the following solvent gradient: 22% B for 15 min, 0.1 min ramp to 100% B for 10 min, 1 in ramp to 0.5% B for 10 min. The HPLC column was coupled to an Agilent 6430 triple quadruple mass spectrometer (Agilent, Santa Clara, CA) with electrospray ionization in positive ion mode and the following parameters: spray voltage 3.8 kV, capillary temperature 325 °C, nebulizer gas 30 psi, fragmentor voltage 76 V, and collision energy 4 V. 2'-Deoxynucleosides were detected using selected reaction monitoring (SRM) for the following transitions involving loss of 2-deoxyribose: *m/z* 326→210 for $dCPH_4$,

311→195 for dCDG, 292→176 for $dPreQ_0$, 296→180 for $dPreQ_1$, 394→278 for dQ and 309→193 for dG^+.

Subsequent structural corroboration of $dPreQ_0$ and dADG was achieved by liquid chromatography-coupled quadrupole time-of-flight mass spectrometry (LC-QTOF). Using HPLC conditions and retention times defined earlier, hydrolyzed bacterial DNA was resolved by HPLC and fractions containing $dPreQ_0$ and dADG were collected, pooled (3-4 HPLC runs), dried, and redissolved in water. These samples were analyzed by LC-QTOF using the HPLC system noted earlier coupled to an Agilent 6510 QTOF mass spectrometer with an electrospray ionization source operated in positive ion mode with the following parameters: drying gas temperature, 325 °C; drying gas flow, 8 L/min; nebulizer, 25 psi; capillary voltage, 3500 V; fragmentor voltage of 90-140 V or 90 V; collision energy 5-15 V.

Quantification of $dPreQ_1$, $dPreQ_0$, dADG, dCDG, and dG^+ was achieved by LC-MS/MS using external calibration curves. A stock solution of standards was prepared with the following concentrations: 1 mM dC, dG, dT and dA; 10 nM each $dPreQ_1$, $dPreQ_0$ and dCDG; and 100 nM dADG; the concentration of $dPreQ_1$, $dPreQ_0$, dCDG, and dADG were quantified on the basis of the $PreQ_0$ extinction coefficient ε_{264} = 10,090 $M^{-1}cm^{-1}$ (25). Nine calibration solutions were then prepared with the lowest concentrations of 2 μM each dC, dG, dT and dA; 20 pM each $dPreQ_1$, $dPreQ_0$ and dCDG; and 200 pM dADG. Calibration samples (10 μL) were interspersed with DNA hydrolysate samples and injected on an Agilent 1290 series HPLC equipped with a Kinetex EVO C18 column (150 mm x 2.1 mm, 2.6μm particle size, 100 Å pore size; Phenomenex) and a diode array detector (DAD). The column was heated to 40 °C and eluted at 0.5 mL/min with 100% solvent A (5 mM ammonium acetate, pH 5.3) for 2 min followed by a linear gradient to 10% solvent B (acetonitrile) over 8 min, 60% solvent B at 9 min and held for 1 min. The column eluent was monitored by UV absorbance at 260 nm to determine the retention times of canonical 2'-deoxynucleosides (dC, 1.2 min; dG, 3.8 min; dT, 4.2 min; and dA, 5.5 min). The HPLC column was coupled to an Agilent 6430 triple quadruple mass spectrometer with an electrospray ion source in positive ion mode and parameters as follows: gas temperature 300 °C, gas flow 12 L/min, nebulizer pressure 40 psi, and capillary voltage 3800 V. 2'-Deoxynucleosides were identified and quantified in multiple reaction monitoring mode (MRM) using previously optimized conditions. Using the transitions noted earlier, the modified 2'-deoxy-7-deazaguanosines were detected at the following retention times: $dPreQ_1$, 3.2 min; dG^+, 6.0 min; dADG, 6.4 min; $dPreQ_0$, 6.7 min; and dCDG, 7.1 min. The limit of detection was determined to 1-2 fmol for the five modified 2'-deoxynucleosides, which corresponds to ~10 modifications per 10^6 nucleotides in 2 μg of injected DNA digest.

tRNA Extraction

For bulk tRNA extractions, *S.* Montevideo cultures were grown in 200 mL LB at 37 °C, with shaking at 250 rpm overnight and *K. radiotolerans* cultures were grown in 1 L TYG at 30 °C, with shaking at 250 rpm overnight. Cells were harvested by centrifugation at 8,000 x g for 20 min at 4 °C. Pellets washed once in 0.1 volume of 95% ethanol. Pellets were centrifuged again and suspended in 50 mM sodium acetate, pH 5.5 at 3 mL/g of cell weight. One volume of acidic-buffered phenol (pH 5.5) was added and incubated at 70 °C for 30 min. The samples were centrifuged at 4,000 x g at room temperature for 10 min. A mixture of 25:24:1 phenol:chloroform:isoamyl alcohol was added to the aqueous phase, and the samples were centrifuged again. After a second chloroform extraction, sodium chloride was added to a final concentration of 1 M and 0.2 volume of isopropanol was added to the aqueous phase. The

samples were centrifuged again, and 0.6 volume of isopropanol was added to the supernatant and incubated overnight at -20 °C. The samples were centrifuged again, and the pellets were washed with 80% ethanol and dried in the Vacuum Concentrator Systems (Labconco) at 40 °C for 30 min. The pellets were suspended in 100 µL TE buffer (100 mM Tris, pH 8.0; 10mM EDTA, pH 8.0).

Detection of queuosine in bulk tRNAs

Detection of the presence of Q in tRNA of *S.* Montevideo was adapted from a protocol developed by Igloi and Kossel (26) and recently used by Zaborske *et al.* (27). For each sample, 10 µg of total tRNA was deacylated by incubation in 100 mM Tris-HCl (pH 9) for 30 min at 37 °C. Deacylated tRNAs were precipitated using ammonium acetate, isopropanol, and linear polyacrylamide as a carrier. The pellet obtained was washed with 70% ethanol and dried in a Vacuum Concentrator System (Labconco) at 40 °C for 10 min. For each lane, 1 µg of deacylated tRNAs were resuspended in RNA Loading Dye (NEB) and loaded onto a denaturing 8M urea, 8% polyacrylamide gel containing 0.5% 3-(Acrylamido)phenylboronic acid (Sigma-Aldrich). The migration was performed at 4 °C in 1X TAE. Migrated tRNAs were transferred onto a Biodyne B precut Nylon membrane (Thermo Scientific) using a wet transfer apparatus in 1X TAE at 150 mA 4 °C for 90 min. After the transfer, the membrane was baked in an oven for 30 min at 80 °C and then UV irradiated in a UV Crosslinker (Fisher FB-UVXL-1000) at a preset UV energy dosage of 120 mJ/cm^2. $tRNA^{ASP}$ was detected with the North2South Chemiluminescent Hybridization and Detection Kit (Thermo) using the specific biotinylated primer (5' biotin-CCCTCGGTGACAGGCAGG 3'). The blot was exposed to X-Ray film (Thermo Scientific, CL-X Posure Film) for 5 seconds. The film was developed using a Film Processor (Konica QX-60A).

Detection of Q and $PreQ_0$ on the tRNA of *K. radiotolerans* was performed as described previously (28). Briefly, samples of each purified bulk tRNA preparation were enzymatically hydrolyzed as described by P. Crain (29) and liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) as described by Pomerantz and McCloskey (30).

Table S1. *tgtA5*/*dpdA* containing organisms

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Kineococcus radiotolerans SRS30216	266940	A6WGA1	1	2	1	1	1	1	1	1	1		1	1	1	1			1							
Kribbella flavida DSM 17836	479435	D2PYU7	1	2	1	1	1	1	1	1	1		1	1	1											
Nocardioidaceae bacterium Broad-1	408672	E9US40	1	2	1	1	1	1	1	1	1	1	1	1	1											
Streptomyces griseoflavus Tu4000	467200	UPI0001B50039	1	2	1	1	1	1	1	1	1	1	1	1	1	1			1							
Gordonia paraffinivorans NBRC 108238	1223543	M3V7F5	1	2	1	1	1	2	1	1	1		1	1	1				1					1	1	
Nocardia sp. 348MFTsu5.1	1172185	UPI000370E8C7	1	2	1	1	1	1	1	1	1	1	1		1				1					1	1	
Salinispora arenicola CNR107	1169167	UPI0003821849	1	2	1	1	1	1	1	1	1	1	1	1	1				1							
Salinispora arenicola DSM 45545	999546	UPI00036525C0	1	2	1	1	1	1	1	1	1	1	1	1	1				1							
Kribbella catacumbae DSM 19601	1122138	n.d.	1	2	1	1	1	1	1	1	1		1	1	1				3							
Amycolatopsis nigrescens CSC17Ta-90	1068980	n.d.	1	2	1	1	1	1	1	1	1	1	1	1	1				3		1	1	1			
Streptomyces chartreusis NRRL 12338	1079986	n.d.	1	2	1	1	1	1	1	1	1	1	1	1	1	1			2							
Streptomyces sp. ATexAB-D23	1157635	UPI0003813600	1	2	1	1	1	1	1	1	1	1	1	1	1				1						1	
Streptomyces sp. FXJ7.023	579932	UPI00055C92B7	1	2	1	1	1	1	1	1	1	1	1	1	1	1			1							
Meiothermus chliarophilus DSM9957	926560	UPI0004218CCF	1*				1	1	1	1	1		1	1	1	1		1	1					1	1	

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Desulfurobacterium thermolithotrophum DSM11699	868864	F0S2D6	1															2	1			1	1	1		1
Spirosoma spitsbergense DSM 19989	1123278	UPI00037258A1	1*									1	1	1	1		1	7	1	1		1		1	1	
Cyanothece sp. CCY0110	391612	A3IZM8	1*																3		1	1	1	1		1
Geitlerinema sp. PCC 7407	1173025	K9SAS3	1*			1		1	1	1	1		1	1	1				1		1	1	1	1		1
Microcoleus sp. PCC 7113	1173027	K9WRC9	1*	1						1	1		1				1	1	2			1	1	1		1
Oscillatoria sp. PCC 10802	1173028	UPI000345CE7C	1							1							1		1		1	1	1	1		1
Paenibacillus polymyxa	1406	G0VWT7	1*											1	1	1	1	3	1	1	1	1	1	1	1	1
Paenibacillus polymyxa SC2	886882	E3EGG6	1*											1	1	1	1	3	1	1	1	1	1	1	1	1
Paenibacillus vortex V453	715225	UPI0001F05DD6	1*			1	1	1	1	1	1		1	1	1		1	1	1		1	1	1	1		1
Blastopirellula marina DSM 3645	314230	A3ZSB0	1	2	1	1	2	1	1	1	1			1	1					1	2	1	1	1		1
Bradyrhizobium sp. S23321	335659	I0FZ26	1	2	1			1	1	1								1	2		1	1	1	1		
Methylobacterium extorquens AM1	272630	C5B4Q4	1																1		1	1	1	1		1
Sinorhizobium fredii USDA 257	1185652	I3WZI7	1	2	1	1	1	1	1	1	1		1	1	1				1		1	1	1	1	1	1
Sphingobium chlorophenolicum L-1	690566	F6EUE7	1	2	2								1		1				1			1	1	1	2	1
Sphingopyxis alaskensis RB2256	317655	Q1GPS0	1	2	1	1	1	1	1	1	1		1	1	1				1			1	1	1	2	1
Burkholderia sp. CCGE1003	640512	E1T3X2	1	2	1								1		1				1	1	1	1	1	1		1
Comamonas testosteroni	1009852	H1RRG1	1	2	1	1	1	1	1	1	1		1	1	1				1		1	1	1	1	1	1

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
ATCC 11996																										
Herbaspirillum sp. JC206	1095769	UPI000315EA36	1*										1	1			1	6	1	1	1	1	1	1	1	1
Aeromonas hydrophila ML09-119	1288394	UPI0002069B5B	1	1	1	1	1	1	1	1			1	1	1				1			1	1	1	1	1
Aeromonas veronii B565	998088	UPI0002069B5B	1	1	1	1	1	1	1	1			1		1				1			1	1	1	1	
Colwellia psychrerythraea 34H	167879	Q47ZY4	1	1	1	1	1	1	1	1	1	1	1	1	1				2	1		1	1	1		1
Ferrimonas balearica DSM 9799	550540	E1SVY3	1	1	1	1	1	1	1	1	1		1	1	1				1	1	1	1	1	1	1	1
Moritella dasanensis ArB 0140	1201293	UPI0002EB310B, UPI0002F37F12	2	2	1	1		1	1	1	1	1	1	1	1				1			1	1	1	1	2
Pseudoalteromonas spongiae UST010723-006	1117319	U1L931	1	1	1	1		1	1	1	1		1		1				2			1	1	1		1
Brenneria sp. EniD312	598467	G7LP70	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Cedecea davisae DSM 4568	566551	UPI0003A830BE	1	1	1	1	1	1	1	1									1	1	1	1	1	1	1	1
Enterobacter cloacae SCF1	701347	E3GCE4	1	1	1	1	1	1	1	1	1		1		1	1			1		1	1	1	1	1	1
Escherichia coli 3.2608	869679	I2TKQ1_ECOLX	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli 4.0967	869687	I2WPW9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli 541-1	752786	I4TKI2	1	1	1	1	1	1		1	1		1						1		1	1	1	1	1	1
Escherichia coli 93.0624	869680	I2U2E1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli DEC11A	868186	H5GLG9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster															Queosine genes not in cluster						
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Escherichia coli DEC11C	868188	H5HGC4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli DEC11D	868189	H5HY49	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli DEC11E	868190	H5ICF0	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli DEC14A	868201	H5N2Q5	1	1	1	1	1	1		1	1		1						1		1	1	1	1	1	1
Escherichia coli HM46	1308982	n.d., WP_000040318.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli KTE136	1182719	L4I0H9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli KTE14	1169333	S0UY80	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli KTE176	1169401	L5EC58	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli KTE227	1181773	L4T3N4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli KTE98	1182698	S1H135	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli NCCP15657	1169863	n.d., WP_000040321.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli E22	340185	B3I543	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli O103:H2 str. 12009	585395	C8U0N1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli O103:H2 str. CVM9450	1165941	I4NG76	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli O45:H2 str. 03-EN-705	1078033	n.d., KEU_RS0119955	1	1	1	1		1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0299483.1	1125648	N3LTF5	1	1	1	1	1	1	1	1			1						1		1	1	1	1	1	1

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster															Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-ike*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Escherichia coli P0299483.2	1125649	N3N4S5	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0299483.3	1125650	N3M0Z4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.10	1116079	N3S1A9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.2	1116073	N1SX55	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.3	1116074	N3QSH3	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.4	1116075	N3R2K6	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.6	1116076	N3RVF6	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.7	1125643	N3NYW1	1	1	1	1	1	1	1	1			1						1		1	1	1	1	1	1
Escherichia coli P0302293.8	1116077	N3SFA0	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302293.9	1116078	N3T1M3	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.1	1116153	M9FY36	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.10	1116158	N3FVR4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.11	1116159	N3G3K7	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.12	1116160	N4PX35	1	1	1	1	1	1	1	1			1						1		1	1	1	1	1	1
Escherichia coli P0302308.13	1116161	N4R2E0	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.14	1116162	N4PYS3	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.2	1116154	N3GRS0	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.3	1116155	N3GD14	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Escherichia coli P0302308.4	1116156	N3IIZ2	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli P0302308.5	1116157	N3HPT9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Escherichia coli TA124	656435	H1FHW2	1	1		1	1	1		1	1		1						1		1	1	1	1	1	1
Escherichia coli TA271	656443	F4UWR1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Salmonella enterica serovar Montevideo	745015	E7V8J4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	749951	E7Y6L7	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871591	E8DZP3	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871587	n.d., SEEM0077_RS0123085	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	859199	A0A023N8U6	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	858317	n.d., WP_001542917.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871588	E8D877	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	745020	E7WSI3	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871589	E8DB61	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	745017	E7VPX7	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	882865	E8FQ96	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Salmonella enterica serovar Montevideo	766762	E7Z4F9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	749949	E7XRG5	1	1	1	1	1	1	2	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	882866	n.d., EGA47221.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	882869	E8GVX4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	789332	n.d., EFY62506.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871593	n.d., EGA25998.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	858318	n.d., EFZ82503.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	745018	n.d., EFY18721.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	882863	n.d., EGA31819.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	882867	n.d., EGA49200.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	745021	n.d., EFY34580.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	882864	n.d., EGA35350.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871590	E8DXF6	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	871592	E8ED74	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo	749950	n.d., EFY47421.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar	745019	n.d.,	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster						
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Montevideo		EFY24387.1																								
Salmonella enterica serovar Montevideo str. 19N	763919	n.d., EFY56108.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 366867	858314	n.d., EFY72248.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 413180	858315	n.d., EFY76699.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 414877	858313	n.d., EFY68858.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 42N	763921	G9V749	1	1	1	1	1	1		1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 4441 H	766761	n.d., EHL61016.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 556152	871586	E8CAZ9	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 609460	859198	n.d., EFZ86583.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. 80959-06	763922	H0L381	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. CT_02035278	984240	n.d., EHN33861.1	1	1	1	1	1	1	2	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. CT_02035318	984241	n.d., EHN21388.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. CT_02035320	984242	n.d., EHN31461.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. CT_02035321	882860	H0MIU4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	

Organism	TaxID	UniProt Entry of DpdA[a]	Genes in *dpd* cluster																Queosine genes not in cluster							
			dpdA	*dpdB*	*dpdC*	*dpdD*	*dpdE*	*dpdF*	*dpdH*	*dpdJ*	*dpdK*	*folE*	*queD*	*queE*	*queC*	*yhhQ*	*tgt-like*	Other DBP[b]	*folE1*	*folE2*	*queD*	*queE*	*queC*	*tgt*	*yhhQ*	*queF*
Salmonella enterica serovar Montevideo str. CT_02035327	882862	H0MXB4	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. IA_2010008286	882868	n.d., EHP21752.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Montevideo str. SARB31	749952	n.d., EHL39173.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	
Salmonella enterica serovar Tennessee str. CDC07-0191	573395	n.d., ZP_04654487.1	1	1	1	1	1	1	1	1	1		1						1		1	1	1	1	1	1
Vibrio cyclitrophicus ZF99	1136170	UPI0002FEF6E0, UPI00031DBF90	2	2	1	1	1	1	1	1		1	1	1	1				3		1	1	1	1	1	1
Stigmatella aurantiaca DW4/3-1	378806	E3FUF0	1	1	1	1				1	1								1		1	1	2	3	1	1

Number indicates number of genes present. * indicates presence of long version of *dpdA*.

[a] UniProt entry given when available. n.d. indicates entry was not found in UniProt. When available, the RefSeq entry or locus tag is given.

[b] Other DBP: Other predicted DNA binding and processing proteins.

Table S2. HHpred predictions for *S.* Montevideo DpdA-DpdK proteins.

Query		HHpred result								
Protein	Residues	PFAM	Description	Probability	E-value	P-value	Score	Aligned	Query	Template
DpdA	total (417)	PF01702	TGT: Queuine tRNA-ribosyltransferase	100	3.00E-39	1.80E-43	303.6	210	113-415	1-215 (223)
DpdB	total (376)	PF14072	DndB: DNA-sulfur modification-associated protein	100	8.90E-52	5.50E-56	403.1	316	6-367	1-343 (346)
DpdC	total (319)	PF03883	DUF328: Protein of unknown function (DUF328)	98.7	1.00E-07	6.30E-12	85.6	150	34-232	53-211 (233)
	230-319	PF08069	Ribosomal_S13_N: Ribosomal S13/S15 N-terminal domain	80.6	1.4	8.90E-05	26	31	39-71	29-60 (60)
DpdD	total (724)	PF10087	DUF2325: Uncharacterized protein conserved in bacteria (DUF2325)	96.1	0.0095	5.90E-07	50.3	89	632-720	1-93 (94)
	1-600	no result								
DpdE	total (927)	PF12137	RapA_C: RNA polymerase recycling family C-terminal	100	4.20E-33	2.60E-37	297.8	316	564-916	1-322 (360)
		PF00176	SNF2_N: SNF2 family N-terminal domain	99.9	2.60E-27	1.60E-31	257.8	225	78-328	1-272 (307)
		PF00271	Helicase_C: Helicase conserved C-terminal domain	99.5	1.30E-14	8.30E-19	129.5	94	453-556	4-105 (105)
DpdF	total (817)	PF00270	DEAD: DEAD/DEAH box helicase	99.8	2.30E-20	1.40E-24	182.2	170	139-324	1-175 (175)
		PF00271	Helicase_C: Helicase conserved C-terminal domain	99.7	1.30E-17	8.30E-22	147.7	105	359-469	1-105 (105)
	500-817	PF16124	RecQ_Zn_bind: RecQ zinc-binding	89.6	0.35	2.20E-05	33.4	45	117-164	20-64 (64)
DpdG	total (300)	PF09821	AAA_assoc_C: C-terminal AAA-associated domain	94.3	0.13	7.80E-06	40.3	98	53-178	12-116 (119)
	200-300	PF00538	Linker_histone: linker histone H1 and H5 family	79.1	0.98	6.10E-05	26	53	9-67	8-60 (73)
DpdH	total (1049)	PF10923	DUF2791: P-loop Domain of unknown function (DUF2791)	97.7	0.00091	5.60E-08	74.2	329	68-524	53-407 (414)
	550-1049	PF08397	IMD: IRSp53/MIM homology domain	74	9.6	0.00059	33.8	76	424-499	10-85 (219)
	550-950	PF14276	DUF4363: Domain of unknown function (DUF4363)	70.8	19	0.0012	26.5	89	279-385	27-116 (120)
	550-800	no result								

Query		HHpred result								
Protein	Residues	PFAM	Description	Probability	E-value	P-value	Score	Aligned	Query	Template
DpdI	total (245)	PF03999	MAP65_ASE1: Microtubule associated protein (MAP65/ASE1 family	89.9	4.8	0.00029	37.1	182	3-223	65-253 (598)
DpdJ	total (1506)	PF00270	DEAD: DEAD/DEAH box helicase	99.8	2.40E-18	1.50E-22	182.1	171	159-430	1-175 (175)
		PF00271	Helicase_C: Helicase conserved C-terminal domain	99.4	6.10E-13	3.80E-17	127.9	105	480-636	1-105 (105)
	700-1506	PF09369	DUF1998: Domain of unknown function (DUF1998); DEAD/DEAH-box family of helicases	92.1	0.12	7.20E-06	42.5	64	477-552	1-64 (80)
	700-800	no result								
DpdK	total (182)	PF13091	PLDc_2: Phospholipase D-like domain	99.9	1.80E-24	1.10E-28	152.4	122	49-172	2-134 (134)

Table S3. Strains used in this study

Strain name	Genotype / Relevant Characteristics	Reference / Source
Salmonella enterica subspecies *enterica* serovar Montevideo ATCC BAA-710	Wild type	(31)/Max Tepliski
YYF3020	*S.* Montevideo Δ*tgt*::kan; kan^R	This study
YYF3022	*S.* Montevideo Δ*dpdC-dpdD*::kan; 20-kb deletion of gene cluster; kan^R	This study
Kineococcus radiotolerans	Wild type	(32)/John Battista
Comamonas testosteroni ATCC 11996	Wild type	(33)/DSMZ
Ferrimonas balearica DSM 9799	Wild type	(34)/ DSMZ
Meiothermus chliarophilus DSM 9957	Wild type	(35)/DSMZ
Sphingopyxis alaskensis RB2256	Wild type	(36)/DSMZ
Escherichia coli C600	F- *tonA*21 *thi*-1 *thr*-1 *leuB*6 *lacY*1 *glnV*44 *rfbC*1 *fhuA*1 λ-	(37)

Table S4. Oligonucleotides used in this study

Name	Sequence 5'-3'
cl2in-F	CGCAACGTGGTCAACTTCT
cl2in-R	GCCGCTATCACGAAAGACA
tgtA5in-F	CGAGCATCGCGTTCGTCAGC
tgtA5in-R	CGATCGCCCAAAGTACGTTCG
cl1in-F	ATGAGTGAGTTCCGTGTCCC
cl1in-R	GTTGCGAGGCTTCTTATCCC
SM1830-F	CTGATTATCCGTTCGGGCGAC
SM1830-R	CAGCATGTCGCGGGTACACAG
SM1832-F	GTGGCAATTACCAACAACGCTG
SM1832-R	TCATTTTCTTGCCCCCCCTGG
cl3in-F	TGGGTTCCGCTGACACTACTC
cl3in-R	GCGTTTCGCTGAAGGTTGAT
Tn5-F	GGATGGAAGCCGGTCTTGTCGATCAGGAT
SM1839-R	CTTCAAACCTACGCTGGATAGGC
tgtA5out-up	GAATATCGTGCTCACTAGACTG
SM-tgt-out-F	GTAACGGTTGCAGCGGGAAC
SM-tgt-out-R	GAAATCACCAGGTCACGATG
SM-tgt-in-F	CGCGGTCGTCTGGTGTTTG
SM-tgt-in-R	CGAAGCTCTCTAATTTACCCTC
SMclus2-deletion-F	GTGAGCCAAATTAATCTGATAACAACATGTACAAATG GTAAGGTGTAGGCTGGAGCTGCTTCG
SMclus3-deletion-R	TCAATCCAGCGCACTGACGAAGGCTGCGATCATACTC GATGGTCCATATGAATATCCTCCTTAG
SM-tgt_deletion-5	ATGAAATTTGAGTTAGATACCACCGATGGTCGCGCGC GTCGTGTAGGCTGGAGCTGCTTC
SM-tgt_deletion-3	TTAATCAACGTTCAAAGGTGGAACGGGTCGCCCCTGA CGTGTCCATATGAATATCCTCCTTAG

Table S5. Distributions *tgt* and $preQ_0$/$preQ_1$ synthesis genes in phage genomes

NCBI accession	Name	Host	*folE*	queD	*queE*	*queC*	*gat-queC*	*queF*	*tgt*	*parB*
KC821616.1	phiSM	*Cellulophaga baltica*	gp57	gp49	gp48	gp50		gp54	gp52	gp51
HQ268735	Dp-1	*Streptococcus pneumoniae*	gp5	gp3	gp4	gp2		gp1	gp7	
NC_022087.1	AnnaL29	*Mycobacterium smegamatis*	gp6	gp4	gp5	gp3			gp2	gp1
JN699004.1	Ares	*Mycobacterium avium*	gp6	gp4	gp5	gp3			gp2	gp1
JN698991	Hedgerow	*Mycobacterium avium*	gp6	gp4	gp5	gp3			gp2	gp1
KM101117.1	LizLemon	*Mycobacterium avium*	gp6	gp4	gp5	gp3			gp2	gp1
DQ398048	Qyrzula	*Mycobacterium avium*	gp6	gp4	gp5	gp3			gp2	gp1
AY129334	rosebush	*Mycobacterium avium*	gp6	gp4	gp5	gp3			gp2	gp1
NC_019515.1	BCD7	*Bacillus cereus*	gp20	gp14	gp15	gp17			gp21	
NC_009447.1	BcepGomr	*Burkholderia*	gp33	gp34	gp37	gp35			gp31	
JX483876.1	RHEph04	*Rhizobium etli*	gp53	gp54	gp56	gp55			gp50	
KM236243	Seurat	*Escherichia coli* ETEC	gp35	gp36	gp43	gp39			gp33	
NC_026610.1	VpKK5	*Vibrio parahaemolyticus*	gp69	gp68	gp66	gp67			gp70	
NC_024146.1	9g	*Escherichia coli* K-12	gp4	gp5	gp8		gp6		gp3	
KP719134.1	JenK1	*Escherichia coli* K-12	gp32	gp33	gp36		gp34		gp31	
KP719132.1	JenP1	*Escherichia coli* K-12	gp34	gp35	gp38		gp36		gp32	
JQ807243.1	eHP-23	*unknown*							gp130	
NC_019768.1	HK106	*Escherichia coli*							gp29	
NC_023724.1	Larva	*Mycobacterium smegmatis*							gp86	
JX042579.1	macncheese	*Mycobacterium smegmatis*							gp42	
NC_026598.1	Milly	*Mycobacterium smegmatis*							gp42	
NC_008197.1	Orion	*Mycobacterium smegmatis*							gp20	
NC_021558.1	PG1	*Paenibacillus glucanolyticus*							gp49	
NC_015938.1	phage 7-11	*Salmonella* Newport							gp32	
KJ743987.1	phiLM21	*Sinorhizobium* sp. LM21							gp25	
NC_021303.1	SDcharge11	*Mycobacterium smegmatis*							gp20	
JX509734.1	SfI	*Shigella flexneri*							gp34	

NCBI accession	Name	Host	*folE*	queD	*queE*	*queC*	*gat-queC*	*queF*	*tgt*	*parB*
NC_004087.1	SIRV1	*Sulfolobus islandicus*							gp6	
NC_004086.1	SIRV2	*Sulfolobus islandicus*							gp13	
NC_019545.1	SPN3UB	*Salmonella* Typhimurium							gp54	
NC_025375.1	SRV	*Stygiolobus* species							gp14	
NC_006268.1	STSV1	*Sulfolobus tengchongensis*							gp18	
NC_020077.1	STSV2	*Sulfolobus tengchongensis*							gp15	
NC_023563.1	Suffolk	*Mycobacterium smegmatis*							gp20	
KJ194579.1	Swish	*Mycobacterium smegmatis*							gp20	
NC_023498.1	Validus	*Mycobacterium smegmatis*							gp43	
AJ748296.1	rudivirus 1 variant XX	*Sulfolobus islandicus*							gp6	
NC_024147.1	ZoeJ	*Mycobacterium smegmatis*							gp42	
KC821625.1	phi13:1	*Cellulophaga baltica*	gp32	gp30	gp31					

References

1. Overbeek R, *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res Symp Series* 33(17):5691-5702.
2. Overbeek R, *et al.* (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). (Translated from eng) *Nucleic Acids Res* 42(Database issue):D206-214 (in eng).
3. Letunic I & Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. (Translated from eng) *Nucleic Acids Res* 39(Web Server issue):W475-478 (in eng).
4. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *J Mol Microbiol* 215(3):403-410.
5. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
6. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881-10890.
7. Crooks GE, Hon G, Chandonia J-M, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-1190.
8. Finn RD, *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res* 38(suppl_1):D211-222.
9. Marchler-Bauer A, *et al.* (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39(suppl 1):D225-D229.
10. Soding J, Biegert A, & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue):W244-248.
11. Kelley LA, Mezulis S, Yates CM, Wass MN, & Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845-858.
12. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
13. Criscuolo A & Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210.
14. Guindon S, *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307-321.
15. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307-1320.
16. Ronquist F, *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Sys Biol* 61(3):539-542.
17. Dereeper A, *et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36(Web Server issue):W465-469.
18. Kulikov EE, *et al.* (2014) Genomic sequencing and biological characteristics of a novel *Escherichia coli* bacteriophage 9g, a putative representative of a new Siphoviridae genus. *Viruses* 6(12):5077-5092.
19. Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia* coli K-12 using PCR products. *Proc Natl Acad Sci U S A* 97(12):6640-6645.

20. Kasarjian JK, Iida M, & Ryu J (2003) New restriction enzymes discovered from *Escherichia coli* clinical strains using a plasmid transformation method. *Nucleic Acids Res* 31(5):e22.
21. Xu T, Yao F, Zhou X, Deng Z, & You D (2010) A novel host-specific restriction system associated with DNA backbone S-modification in Salmonella. *Nucleic Acids Res* 38(20):7133-7141.
22. Ramzaeva N, Becher G, & Seela F (1998) Facile synthesis of 2'-deoxynucleoside analogs of preQ. *Synthesis* 1998(09):1327-1330.
23. Bruckl T, Klepper F, Gutsmiedl K, & Carell T (2007) A short and efficient synthesis of the tRNA nucleosides $PreQ_0$ and archaeosine. *Org Biomol Chem* 5(23):3821-3825.
24. Wang RW & Gold B (2009) A facile synthetic approach to 7-deazaguanine nucleosides via a Boc protection strategy. *Org Lett* 11(11):2465-2468.
25. Levic J & Micura R (2014) Syntheses of (15)N-labeled pre-queuosine nucleobase derivatives. *Beilstein J Org Chem* 10:1914-1918.
26. Igloi GL & Kossel H (1985) Affinity electrophoresis for monitoring terminal phosphorylation and the presence of queuosine in RNA. Application of polyacrylamide containing a covalently bound boronic acid. (Translated from eng) *Nucleic Acids Res* 13(19):6881-6898 (in eng).
27. Zaborske JM*, et al.* (2014) A nutrient-driven tRNA modification alters translational fidelity and genome-wide protein coding across an animal genus. *PLoS Biol* 12(12):e1002015.
28. Phillips G*, et al.* (2008) Biosynthesis of 7-deazaguanosine-modified tRNA nucleosides: a new role for GTP cyclohydrolase I. *J Bacteriol* 190(24):7876-7884.
29. Crain PF (1990) Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol* 193:782-790.
30. Pomerantz SC & McCloskey JA (1990) Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. *Methods Enzymol* 193:796-824.
31. Zhuang RY, Beuchat LR, & Angulo FJ (1995) Fate of *Salmonella montevideo* on and in raw tomatoes as affected by temperature and treatment with chlorine. *Appl Environ Microbiol* 61(6):2127-2131.
32. Phillips RW*, et al.* (2002) *Kineococcus radiotolerans* sp. nov., a radiation-resistant, gram-positive bacterium. (Translated from eng) *Int J Syst Evol Microbiol* 52(Pt 3):933-938 (in eng).
33. Tamaoka J, Ha D-M, & Komagata K (1987) Reclassification of *Pseudomonas acidovorans* den Dooren de Jong 1926 and *Pseudomonas testosteron*i Marcus and Talalay 1956 as *Comamonas acidovorans* comb. nov. and *Comamonas testosteroni comb*. nov., with an emended description of the genus Comamonas. *Int J Syst Evol Microbiol* 37(1):52-59.
34. Rosselló-Mora R, Ludwig W, Kämpfer P, Amann R, & Schleifer KH (1995) *Ferrimonas balearica* gen. nov., sp. nov., a new marine facultive Fe(III)-reducing bacterium. *Syst Appl Microbiol* 18(2):196-202.
35. Tenreiro S, Nobre MF, & da Costa MS (1995) *Thermus silvanus* sp. nov. and *Thermus chliarophilus* sp. nov., two new species related to *Thermus ruber* but with lower growth temperatures. *Int J Syst Bacteriol* 45(4):633-639.
36. Vancanneyt M*, et al.* (2001) *Sphingomonas alaskensis* sp. nov., a dominant bacterium from a marine oligotrophic environment. *Int J Syst Evol Microbiol* 51(Pt 1):73-79.

37. Appleyard RK (1954) Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12. *Genetics* 39(4):440-452.

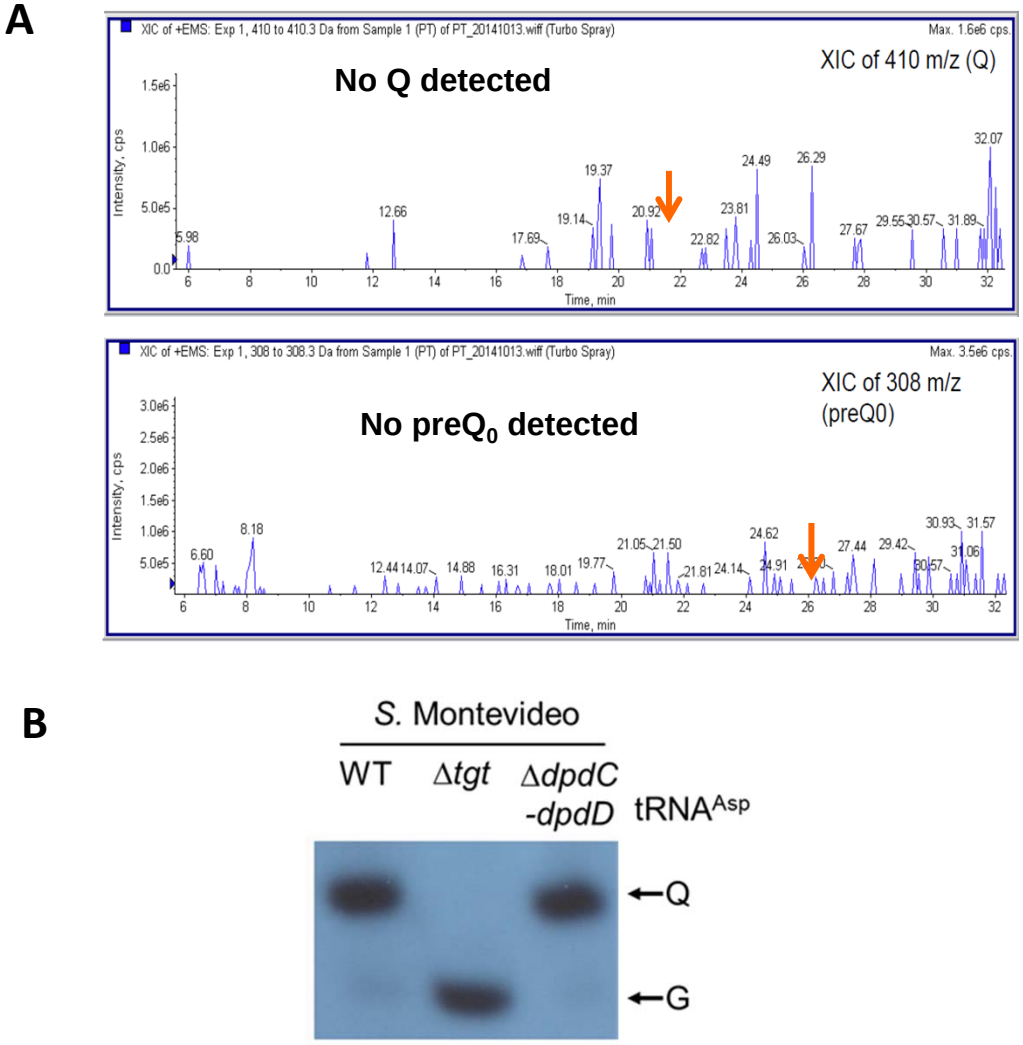

Figure S1. TgtA5 does not insert Q or $PreQ_0$ in tRNA. (A) LC-MS/MS analysis of bulk tRNA from *K. radiotolerans* shows the absence of queuosine (top) and lack of $preQ_0$ (bottom). Arrows indicate absence of peak. (B) Detection of queuosine (Q) on $tRNA^{Asp}$ of *S.* Montevideo wild-type (WT), YYF3020 (Δ*tgt*) and YYF3022 (Δ*dpdC-dpdD*). Bulk tRNA was resolved on an 8 M urea, 8% polyacrylamide gel containing 0.5% 3-(acrylamido)phenylboronic acid and transferred to a nylon membrane. The blotted tRNA was probed with a biotynylated primer to $tRNA^{Asp}$, and detected with the North2South Chemiluminescent Hybridization and Detection Kit. tRNA modified with Q migrates slower than unmodified tRNA.

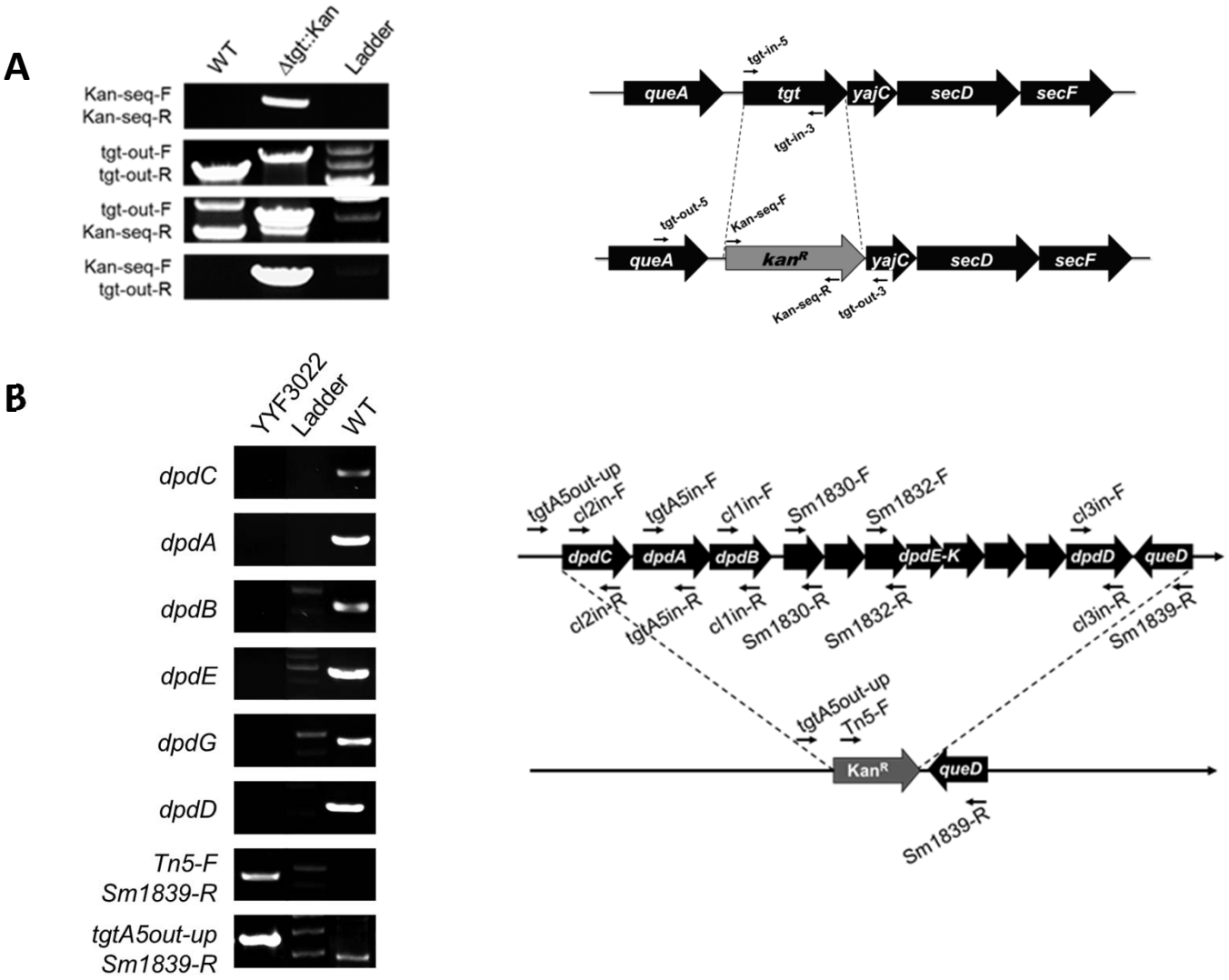

Figure S2. Confirmation of mutant strains of *S.* Montevideo. (A) (Right) PCR confirmation of YYF3020 (*S.* Montevideo Δ*tgt*::kan). Top row: Kan-seq-F and Kan-seq-R primers confirm the presence of kanamycin cassette in the mutant, but absence in the wild-type (WT). Second row: tgt-out-F and tgt-out-R primers give a larger PCR product for mutant than WT. Third row: kan-seq-F and tgt-out-R primers show insertion of kanamycin cassette at the correct location. (Left) *tgt* gene cluster of *S.* Montevideo and construct of Δ*tgt*::kan. Small arrows above or below indicate primers used for PCR confirmation (Right).
(B) PCR confirmation of YYF3022 (*S.* Montevideo Δ*dpdC-dpdD*::kan) (Left). Absence of *dpdC* was confirmed with cl2in-F and cl2in-R primers, Absence of *dpdA* was confirmed with tgtA5in-F and tgtA5in-R primers. Absence of *dpdB* was confirmed with cl1in-F and cl1-in-R primers. Absence of *dpdE* was confirmed with SM1830-F and SM1830-R primers. Absence of *dpdG* was confirmed with SM1832-R and SM1832-R primers. Replacement the 11 genes of the dpd cluster with kanamycin resistance cassette was confirmed using Tn5-F and SM1839-R primers and tgtA5-out-up and SM1839-R primers. (Right) Positions of the primers.

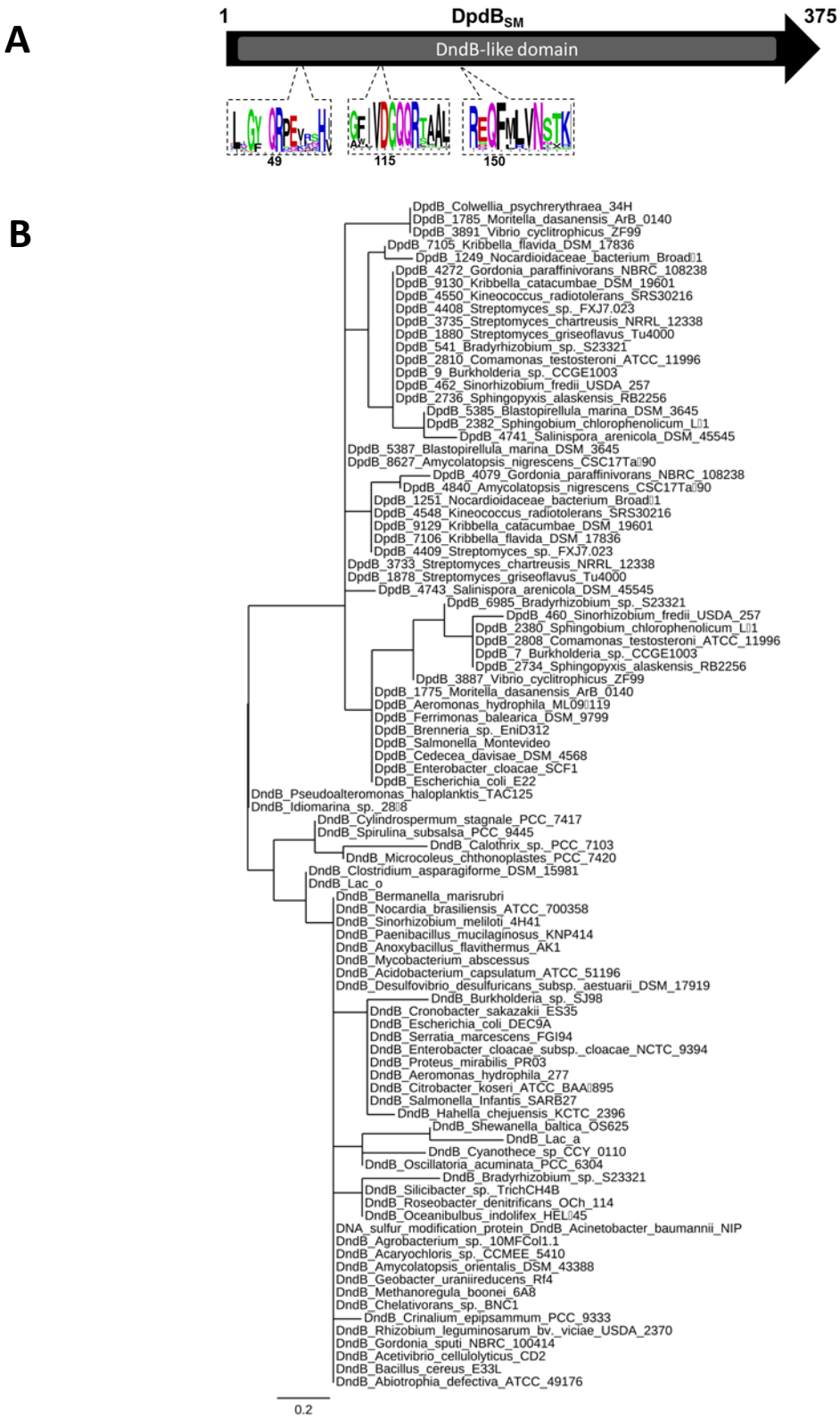

Figure S3. DndB-like proteins cluster with *tgtA5*. (A) Conserved sequence motifs of DndB are conserved in DpdB. Shown here is a schematic of *S.* Montevideo DpdB (Uniprot ID: A0A023N865). Grey bar depicts "DndB-like" domain identified by PFAM. Sequence logo beneath represents alignment of all DpdB sequences retrieved from the PubSEED database. Conserved residues shown are also highly conserved in the DndB proteins (QR, FXXXN, and DGQQ/HR). Numbers beneath the logo correspond to the *S.* Montevideo DpdB sequence. (B) Phylogenetic tree of representative DpdB and DndB proteins constructed using phylogeny.fr.

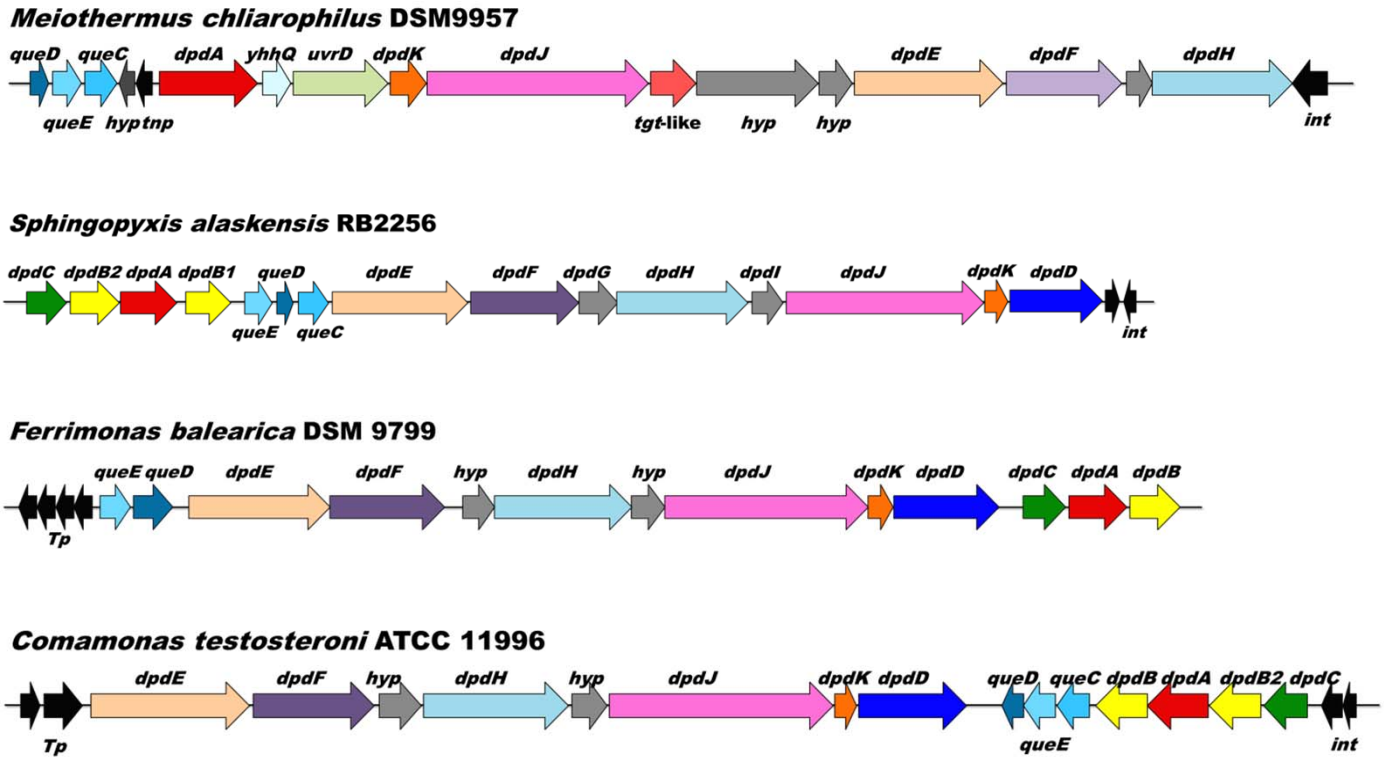

Figure S4. Gene cluster patterns for *tgtA5*/*dpdA* of diverse bacteria confirmed to have ADG on DNA. *Meiothermus chliarophilus* (GenBank Accession: NZ_KE387023; TgtA5 UniParc ID: UPI0004218CCF), *Sphingopyxis alaskensis* RB2256 (GenBank Accession: NC_008048.1; TgtA5 UniProt ID Q1GPS0), *Ferrimonas balearica* DSM9799 (GenBank accession: NC_014541.1; TgtA5 UniProt ID E1SVY3), *Comamonas testosteroni* ATCC 11996 (GenBank Accession: NZ_AHIL00000000.1; TgtA5 UniProt ID H1RRG1). All gene data was acquired from the SEED database. Labeling and colors follow Figure 1B. Red arrows represent *tgtA5*/d*pdA* genes, black arrows represent genes indicative of mobile elements, grey arrows represent genes of unknown function.

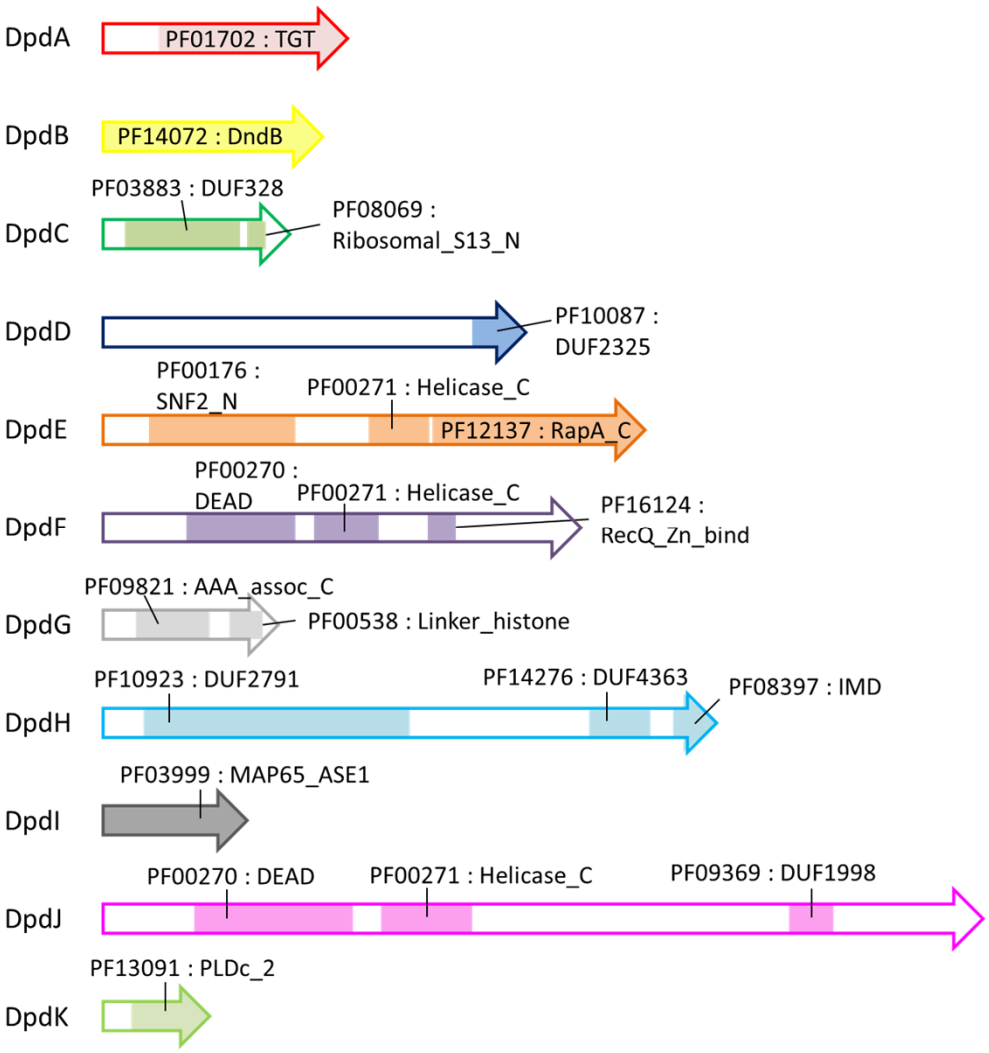

Figure S5. Domain predictions of proteins encoded in *S.* Montevideo *tgtA5*/*dpdA* gene cluster. Domains were predicted using HHpred running a query against the PFAM database.

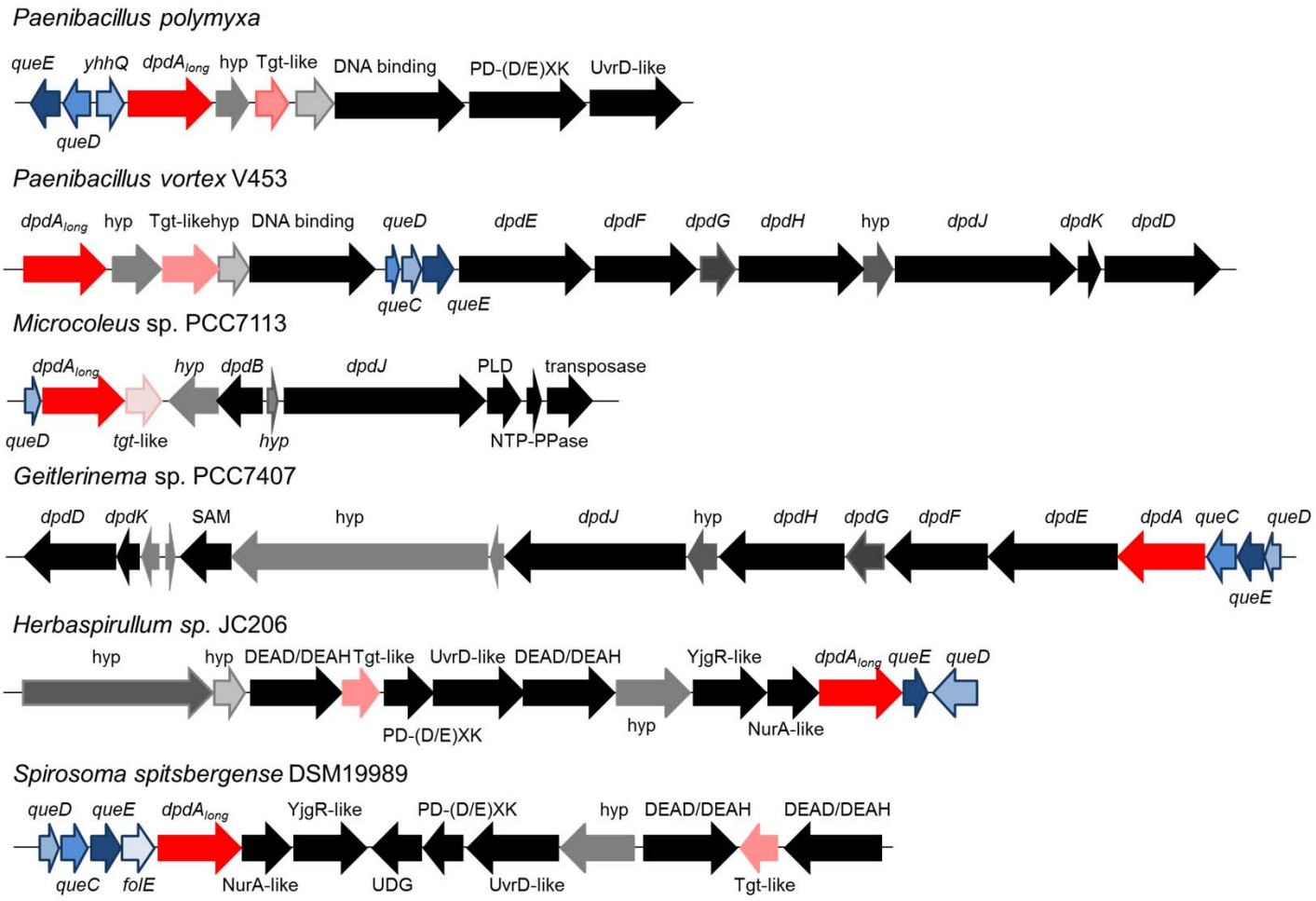

Figure S6. Gene cluster patterns for the long versions of *tgtA5*/*dpdA*. *Paenibacillus polymyxa* (DpdA UniProt ID: G0VWT7); *Paenibacillus vortex* V453 (DpdA UniProt ID: UPI0001F05DD6); *Micrococleus* sp. PCC7113 (DpdA UniProt ID: K9WRC9); *Geitlerinema* sp. PCC7407 (DpdA UniPRot ID: K9SAS3); *Herbaspirullum* sp. JC206 (DpdA UniParc ID: UPI000315EA36), and *Spirosoma spitsbergense* DSM19989 (DpdA UniParc ID: UPI00037258A1).

All gene data was acquired from the SEED database. Red arrows represent *tgtA5*/d*pdA* genes, blue arrows represent preQ$_0$ synthesis (*queE*, *queC*, *queD*) and transporter genes (*yhhQ*), grey arrows represent genes of unknown function (hyp), black arrows represent genes with predicted domains or functions. SAM: radical SAM superfamily domain; NTP-PPase: nucleotide triphosphate pyrophosphohydrolase; DEAD/DEAH: DEAD/DEAH box helicase; UvrD-like: ATP-dependent DNA helicase UvrD / PcrA; PD-(D/E)XK: PD-(D/E)XK nuclease superfamily / ATP-dependent helicase/deoxyribonuclease subunit B; NurA-like: NurA nuclease superfamily; YjgR-like: YjgR-type ATPase; UDG: Uracil-DNA glycosylase enzyme family.

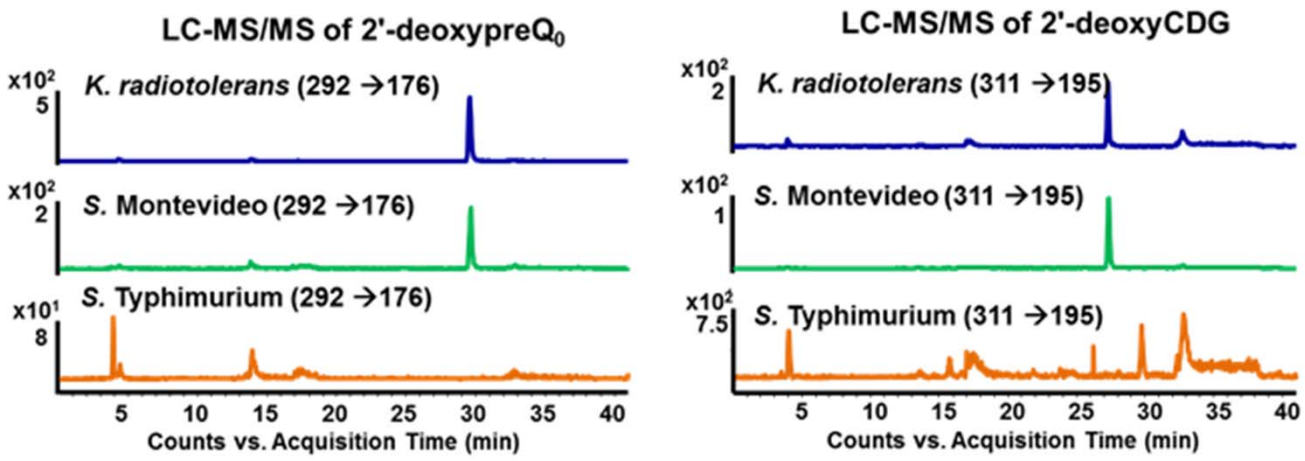

Figure S7. Detection of 2'-deoxy-7-deazaguanosine derivatives in *S.* Montevideo, *K. radiotolerans*, but not *S.* Typhimurium LT2. Left panel: MRM chromatogram of putative 2'-deoxypreQ_0 (mass transition: 292→176). Right Panel: MRM chromatogram of predicted 2'-deoxyCDG (mass transition: 311→195) that was later revealed to be the M+1 signal for 2'-deoxyADG (mass transition: 310→ 194) in DNA of *K. radiotolerans, S.* Montevideo, and *S.* Typhimurium. The signals for the two Q-precursors were reproducibly detected in the DNA of *K. radiotolerans* and *S.* Montevideo, but not in *S.* Typhimurium. High mass accuracy mass spectra for these 2-deoxynucleosides are shown in Figure S6.

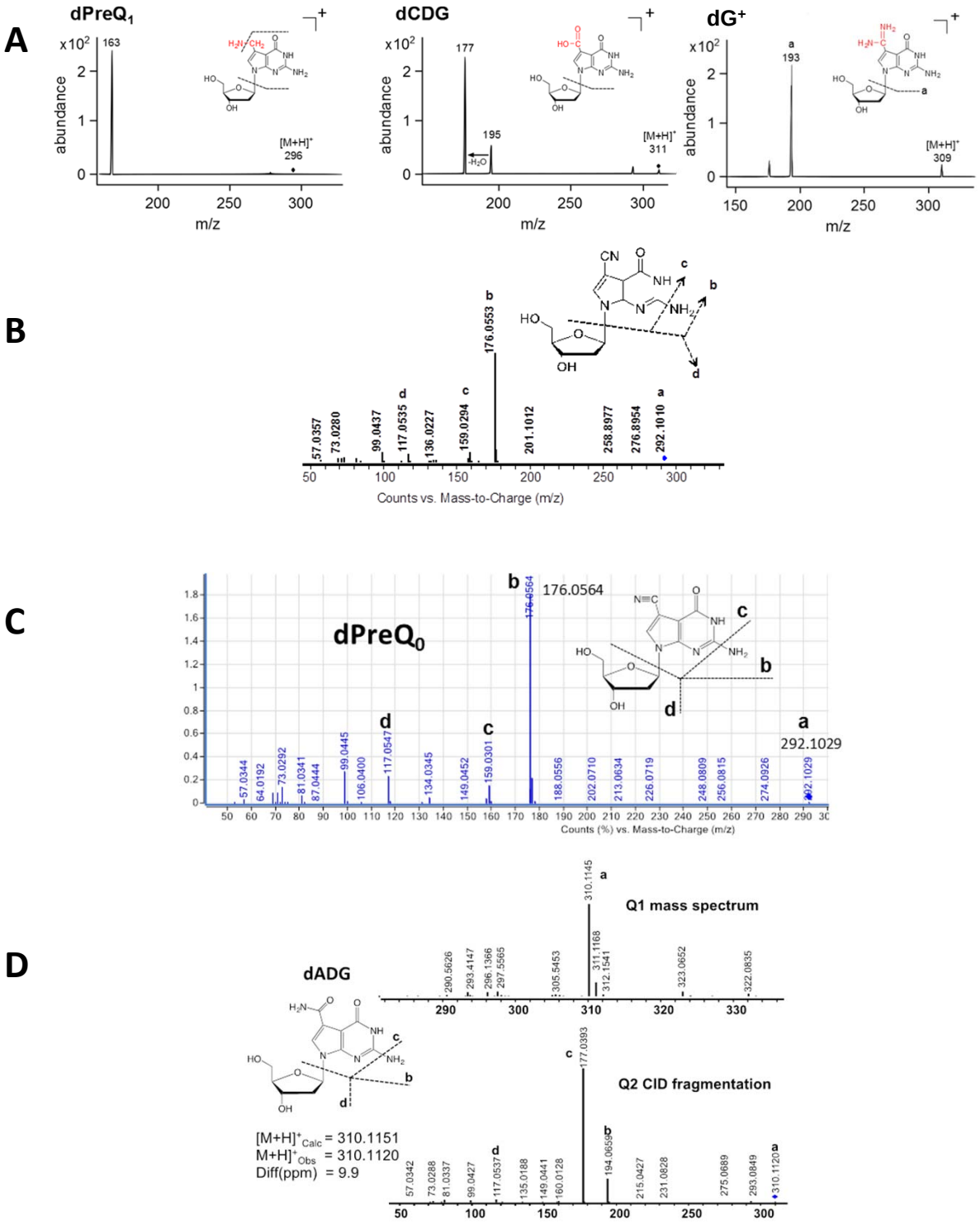

Figure S8. Mass spectrometric characterization of Q-like 2'-deoxynucleoside structures. (**A**) Collision-induced dissociation (CID) mass spectra from LC-MS/MS analysis of synthetic $dPreQ_1$, dCDG, and dG^+. The diamonds at 296 and 311 denote the parent molecular ion for $dPreQ_1$ and dCDG, respectively. CID fragmentation mass transitions used for subsequent LC-MS/MS detection and quantification are indicated with dashed lines. (**B**) High mass-accuracy CID mass spectra from LC-QTOF analysis of $dPreQ_0$ isolated from bacterial DNA. (**C**) High mass-accuracy CID mass spectra from LC-QTOF analysis of synthetic $dPreQ_0$ standard. (**D**) High mass-accuracy mass spectra from LC-QTOF analysis of dADG isolated from bacterial DNA. Upper: Q1 mass spectrum of dADG. Lower: Q2 CID fragmentation pattern. The blue diamond at 310.1120 denotes the parent molecular ion for dADG.

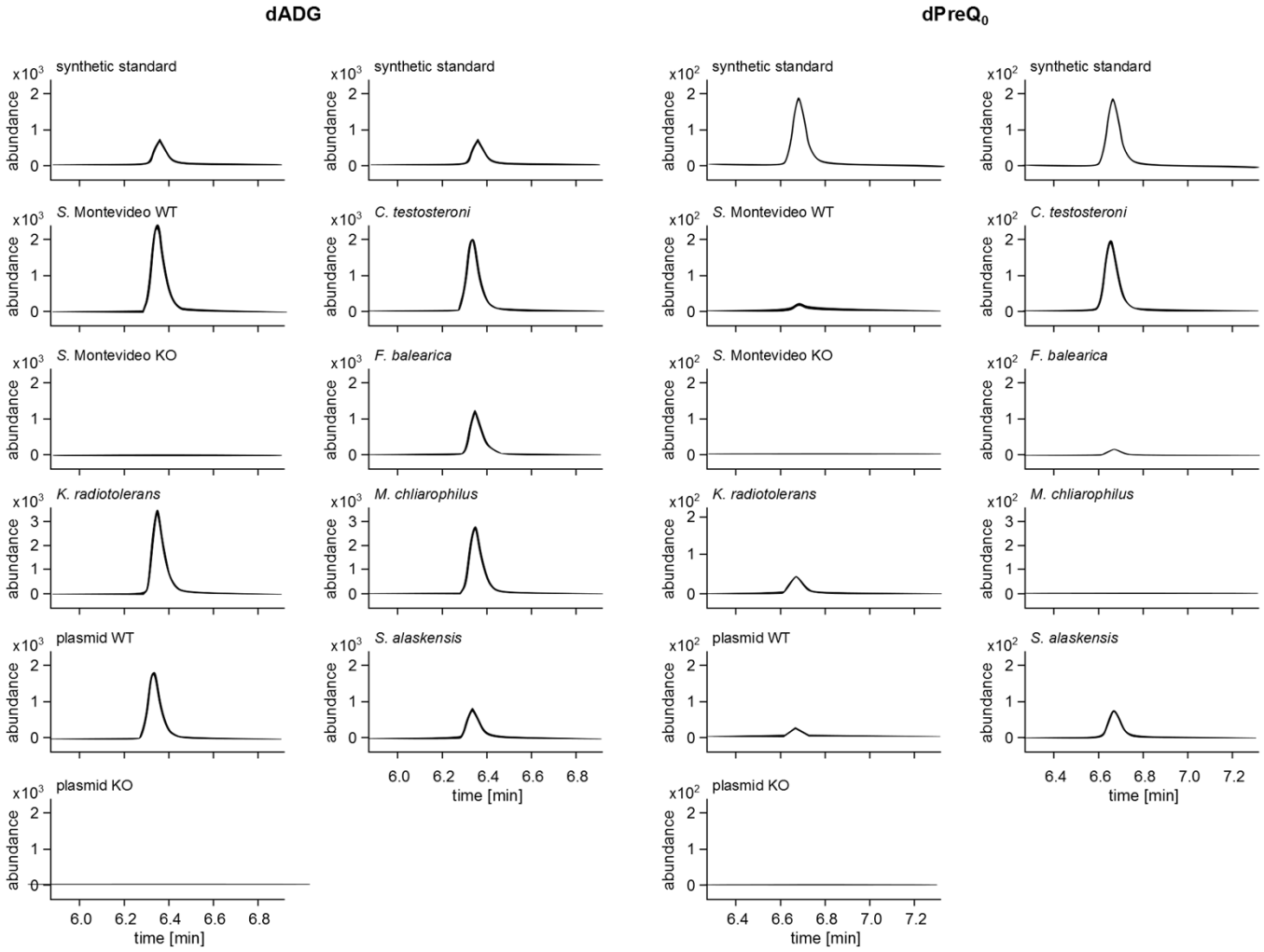

Figure S9. Extracted ion chromatograms for dADG and dPreQ$_0$ detection in various bacterial strains possessing or lacking *dpd* genes. The left two columns show the peaks detected using the mass transition for dADG (310 →194). The two columns on the right show the peaks detected for the corresponding dPreQ$_0$ mass transition (292→176). In both cases, the upper panels show the peaks from the corresponding synthetic standards as retention time reference.

Figure S10. Maximum likelihood (A) and Bayesian (B) trees of TgtA5 homologs found in 30,578 proteomes deduced from complete and ongoing genome projects (284 sequences, 362 positions). Numbers at branch represent bootstrap values (maximum likelihood tree, A) or posterior probabilities (Bayesian tree, B). For clarity bootstrap values < 50% and posterior probabilities < 0.5 are not shown. The scale bars represent the average number of substitutions per site. Colors correspond to taxonomic groups (see trees).


B
0.1

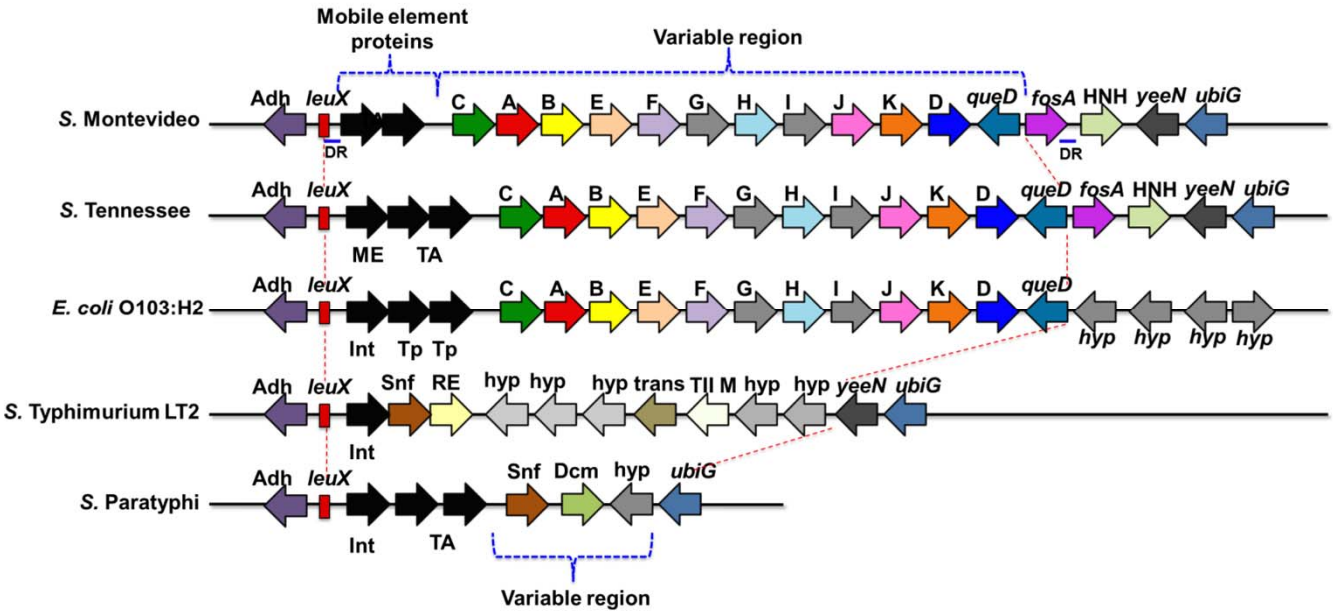

Figure S11. Examples of genomic islands inserted at the tRNA-leu gene (*leuX*; red rectangle) hotspot in *Salmonella* and *E. coli*. The red dotted lines delineate the insertions. Black arrows represent mobile element proteins (i.e., integrases, phage-derived genes, transposases, etc). The colors indicate homologs. *S.* Montevideo, *S.* Tennessee, and *E. coli* O103:H2 all have the ADG modification cluster. The gene lettering corresponds to the *dpd* genes in Fig. 1B. *S.* Typhimurium LT2 and *S.* Paratyphi represent variations in genomic islands inserted at this site. DR, 19-bp direct repeats; Adh, alcohol dehydrogenase; ME, mobile element; TA, toxin-antitoxin genes *ccdA-ccdB*; Int, integrase; TII M, Type II restriction methytransferase; *fosA*, fosfomycin resistance gene; HNH, HNH nuclease; YeeN, putative transcriptional regulator; *ubiG*, 3-methylubiquinone 3-methyltransferase; Dcm, DNA cytosine methyltransferase; Snf, Superfamily I-type helicase.
Uniprot/UniParc entry ID of Adh proteins: S. Montevideo (A0A023N8U4); *S.* Tennessee (UPI0001AF5B80); *E. coli* O103:H2 (C8U0M7); *S.* Typhimurium LT2 (Q8ZK20); *S.* Paratyphi B (UPI00000CD606).

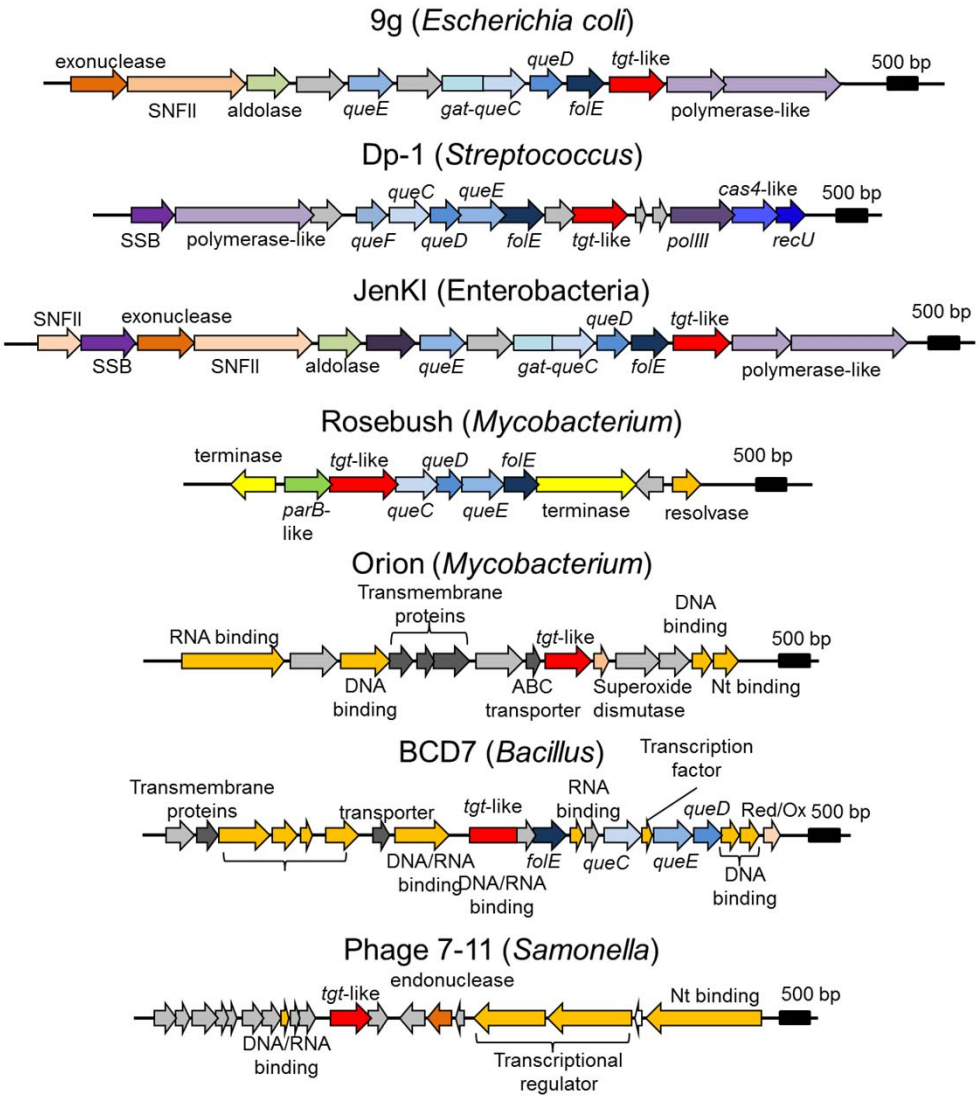

Figure S12. Representation of the genomic context of the *tgt*-like genes for 7 representative phages. 9g (NC_024146), Dp-1 (HQ268735), JenKI (KP719134), Rosebush (AY129334), Orion (NC_008197), BCD7 (NC_019515) and Page 7-11 (NC_015938). Each gene is colored according to predicted function. Red, *tgt*-like, dark blue to light blue, the $preQ_0$/$preQ_1$ biosynthetic pathway; green, *parB*; yellow and orange, predicted to process/bind DNA, RNA, or nucleotides; purple, DNA replication; dark grey, membrane/transporter proteins; light grey, genes of unknown function.

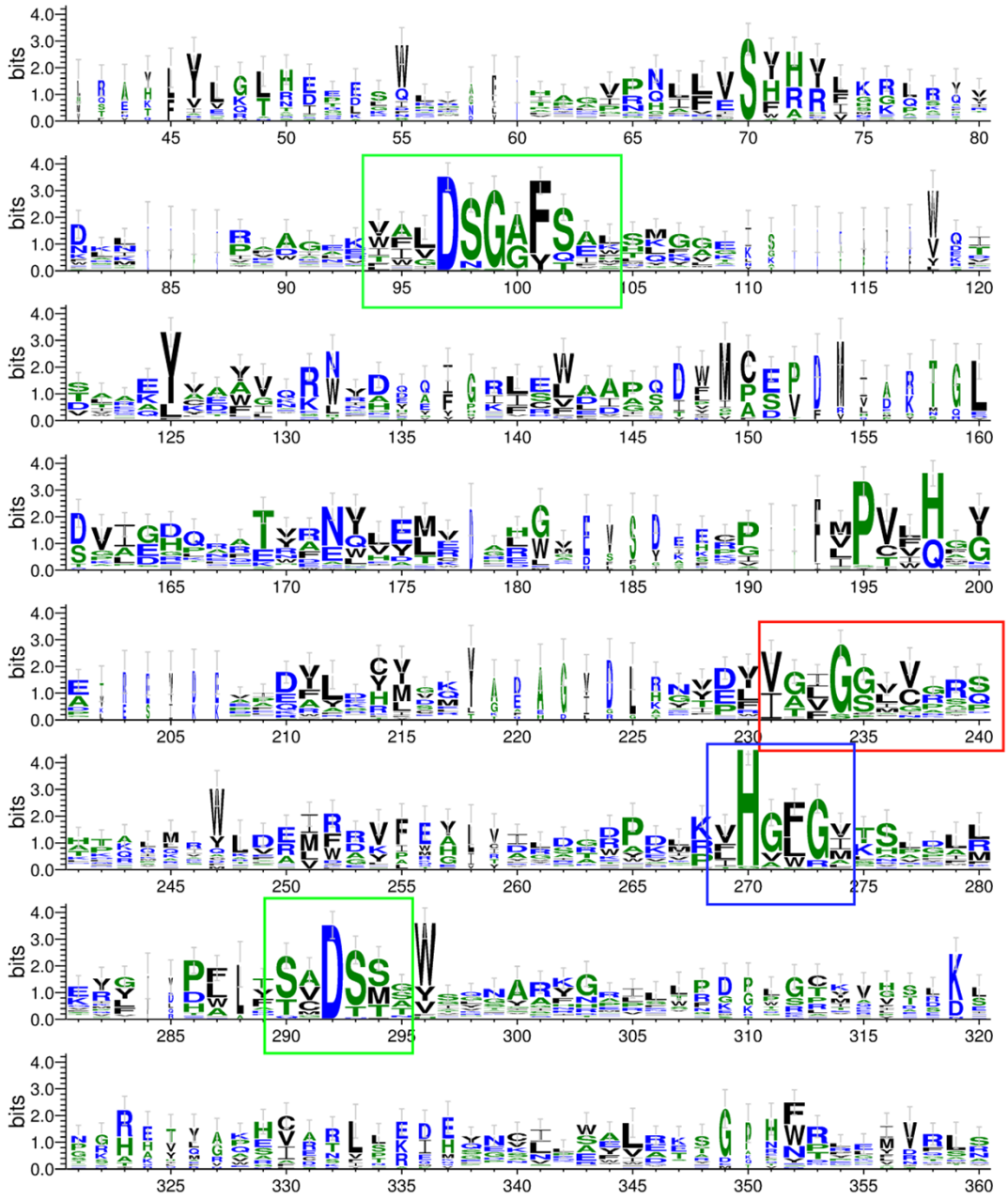

Figure S13. Sequence logo representing the alignment of the phage Tgt-like proteins. The putative catalytic domains are shown in the green boxes, the predicted substrate binding site is shown in the red box, and the conserved histidine is shown in the blue box.