Supplementary Information for

Patchy and widespread distribution of bacterial translation arrest peptides associated with the protein localization machinery

Keigo Fujiwara[1,2,*], Naoko Tsuji[1,2], Mayu Yoshida[1,2], Hiraku Takada[1,2] and Shinobu Chiba[1,2,*]

[1]Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-Ku, Kyoto 603-8555, Japan;

[2]Institute for Protein Dynamics, Kyoto Sangyo University, Japan

*Correspondence to:

Shinobu Chiba (schiba@cc.kyoto-su.ac.jp)

Keigo Fujiwara (kigfujiwara@cc.kyoto-su.ac.jp)

Content:

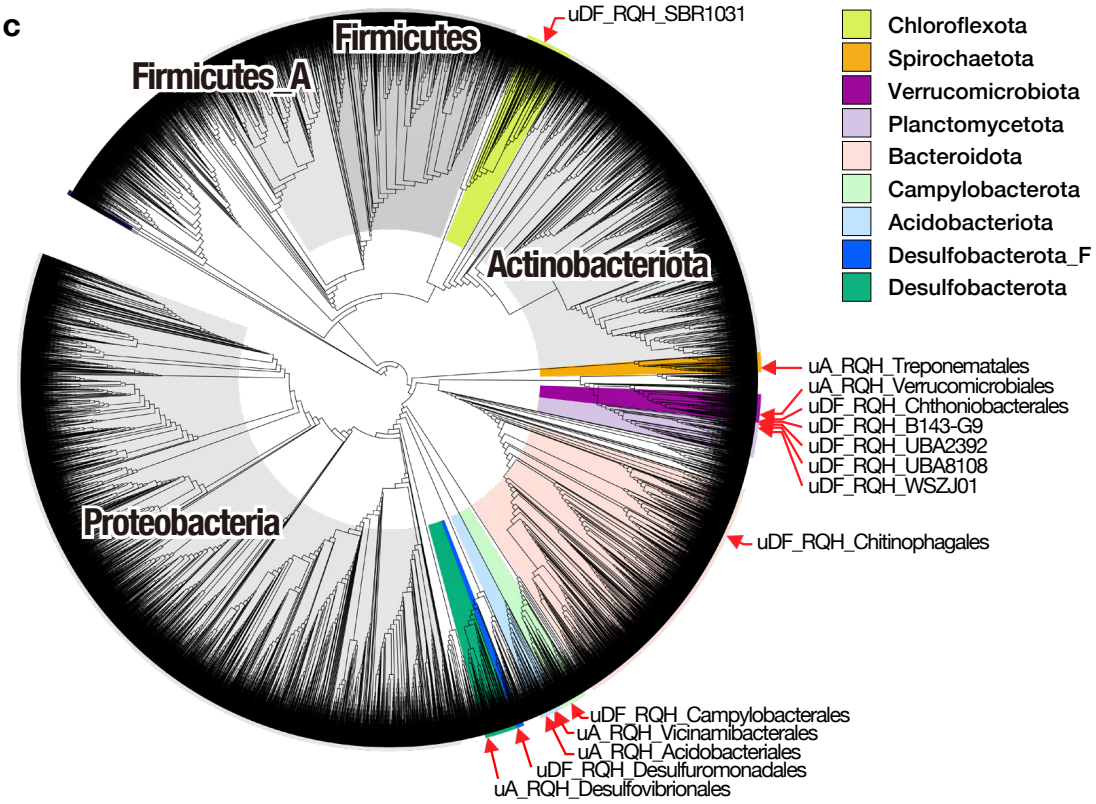
a

AP	phylum	order	mORF	n
SecM	Proteobacteria	-	*secA*	639
uA_LPPP	Planctomycetota	-	*secA*	50
uA_NSP-stop	Bacteroidota	-	*secA*	64
uA_RQH_Armatimonadales	Armatimonadota	Armatimonadales	*secA*	3
uA_RQH_Bacteroidales	Bacteroidota	Bacteroidales	*secA*	73
uA_RQH_Desulfurobacteriales	Aquificota	Desulfurobacteriales	*secA*	8
uA_RQH_Fimbriimonadales	Armatimonadota	Fimbriimonadales	*secA*	5
uA_RQH_Myxococcales	Myxococcota	Myxococcales	*secA*	6
uA_RQH_Opitutales	Verrucomicrobiota	Opitutales	*secA*	3
uA_RQH_Pseudomonadales	Proteobacteria	Pseudomonadales	*secA*	130
uA_RQH_UBA2386	Planctomycetota	UBA2386	*secA*	3
ApdA	Actinobacteriota	-	*secDF*	656
ApdP	Proteobacteria	-	*secDF*	370
VemP	Proteobacteria	-	*secDF*	342
uDF_DGMK-stop	Firmicutes_A	-	*secDF*	2
uDF_DGMK-stop	Firmicutes_B	-	*secDF*	2
uDF_NAP-stop	Bacteroidota	-	*secDF*	403
uDF_RQH_Burkholderiales	Proteobacteria	Burkholderiales	*secDF*	3
uDF_RQH_Caldicoprobacterales	Firmicutes_A	Caldicoprobacterales	*secDF*	5
uDF_RQH_Enterobacterales	Proteobacteria	Enterobacterales	*secDF*	44
uDF_RQH_Flavobacteriales	Bacteroidota	Flavobacteriales	*secDF*	272
uDF_RQH_Geobacterales	Desulfobacterota_F	Geobacterales	*secDF*	8
uDF_RQH_Lachnospirales	Firmicutes_A	Lachnospirales	*secDF*	4
uDF_RQH_Nitrosococcales	Proteobacteria	Nitrosococcales	*secDF*	6
uDF_RQH_Pseudomonadales	Proteobacteria	Pseudomonadales	*secDF*	52
uDF_RQH_Steroidobacterales	Proteobacteria	Steroidobacterales	*secDF*	3
uDF_RQH_UBA1845	Planctomycetota	UBA1845	*secDF*	6
uDF_RQH_UBA7662	Bacteroidota	UBA7662	*secDF*	3
ApcA	Actinobacteriota	-	*yidC*	1599
MifM	Firmicutes	-	*yidC*	592
uC_KYxIW	Firmicutes_A	-	*yidC*	27

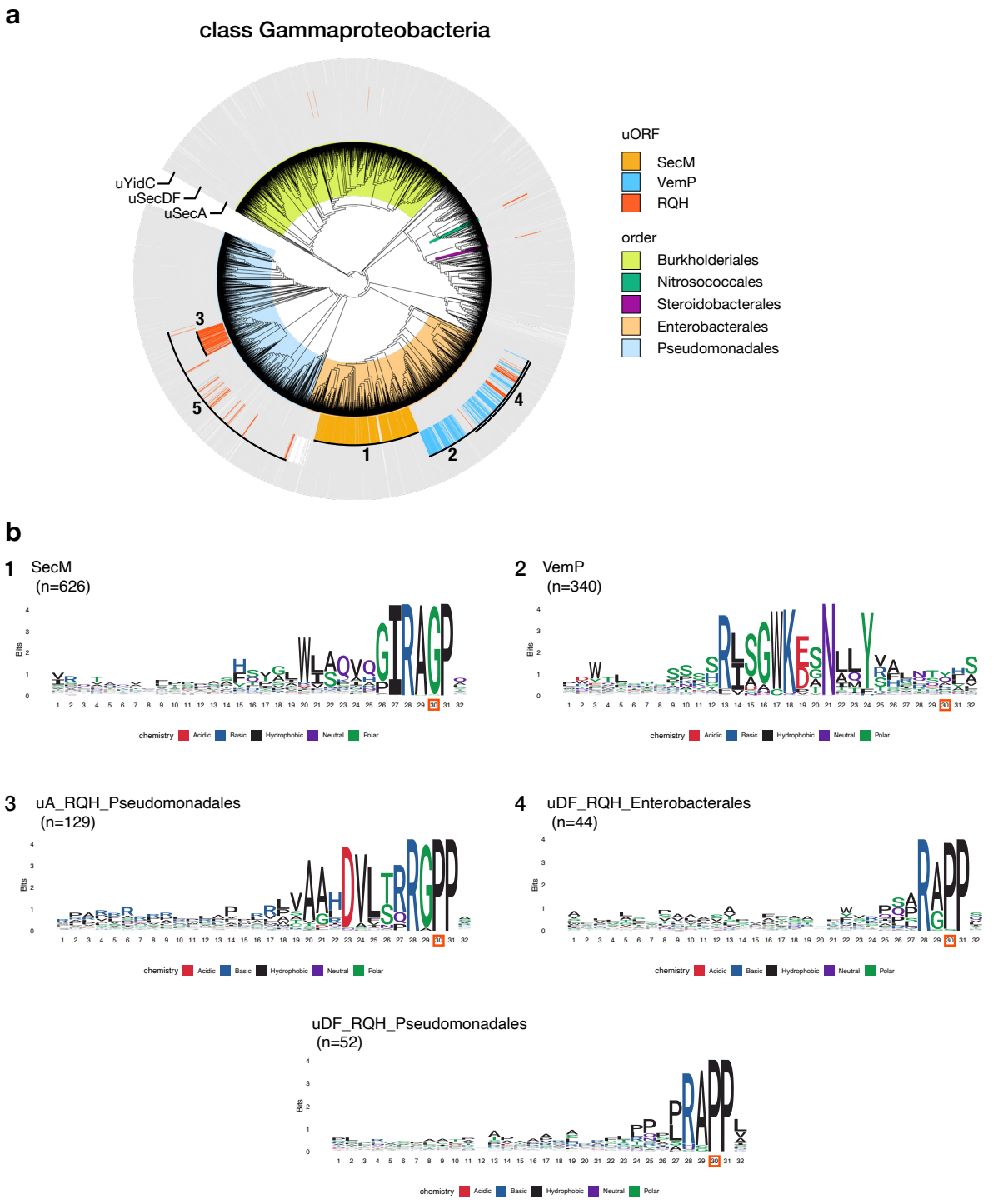
b

AP	phylum	order	mORF	n
uA_RQH_Acidobacteriales	Acidobacteriota	Acidobacteriales	*secA*	1
uA_RQH_Desulfovibrionales	Desulfobacterota	Desulfovibrionales	*secA*	1
uA_RQH_Treponematales	Spirochaetota	Treponematales	*secA*	1
uA_RQH_Verrucomicrobiales	Verrucomicrobiota	Verrucomicrobiales	*secA*	1
uA_RQH_Vicinamibacterales	Acidobacteriota	Vicinamibacterales	*secA*	2
uDF_RQH_B143-G9	Planctomycetota	B143-G9	*secDF*	1
uDF_RQH_Campylobacterales	Campylobacterota	Campylobacterales	*secDF*	1
uDF_RQH_Chitinophagales	Bacteroidota	Chitinophagales	*secDF*	2
uDF_RQH_Chthoniobacterales	Verrucomicrobiota	Chthoniobacterales	*secDF*	1
uDF_RQH_Desulfuromonadales	Desulfobacterota_F	Desulfuromonadales	*secDF*	1
uDF_RQH_SBR1031	Chloroflexota	SBR1031	*secDF*	1
uDF_RQH_UBA2392	Planctomycetota	UBA2392	*secDF*	2
uDF_RQH_UBA8108	Planctomycetota	UBA8108	*secDF*	2
uDF_RQH_WSZJ01	Planctomycetota	WSZJ01	*secDF*	2

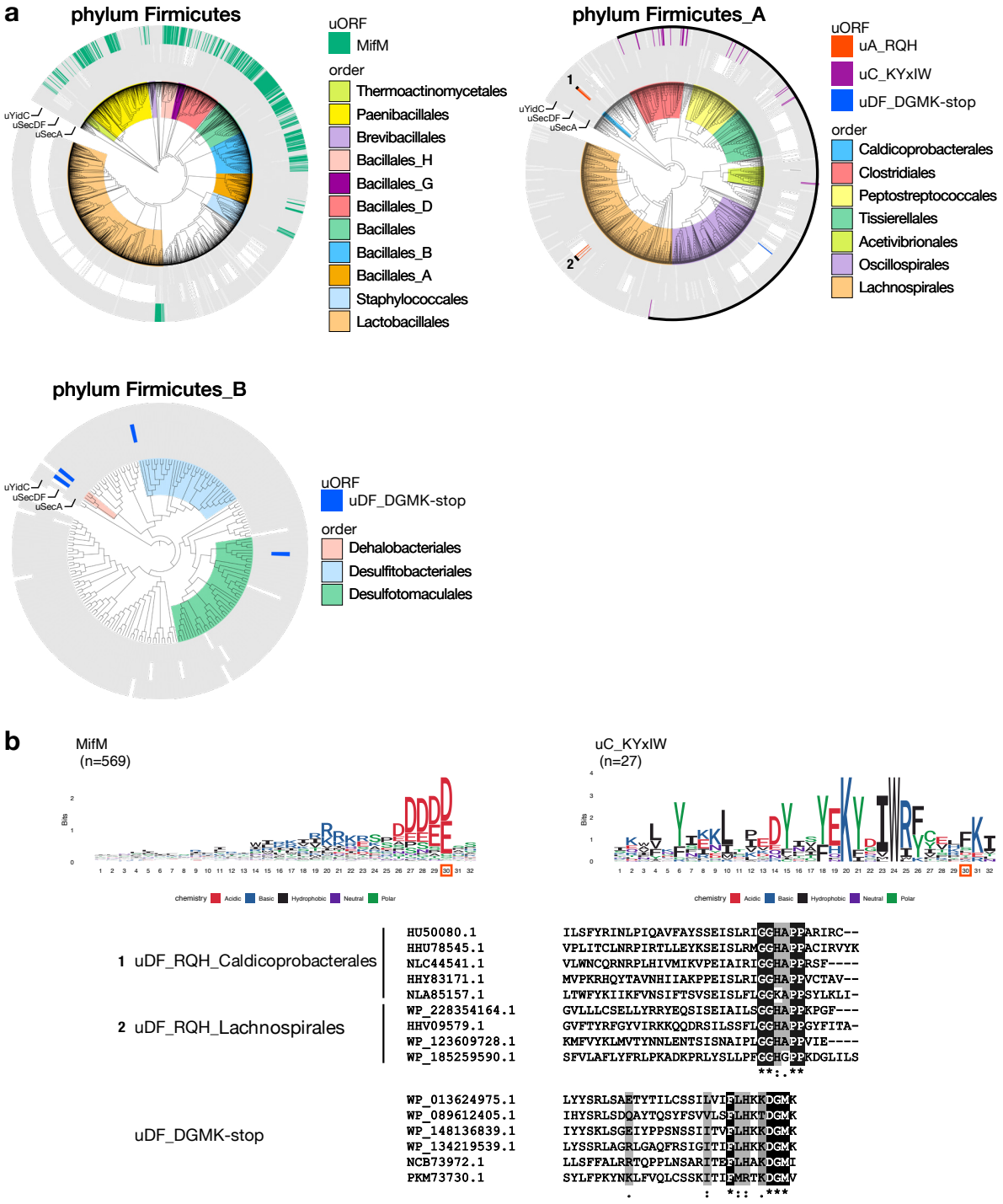
c



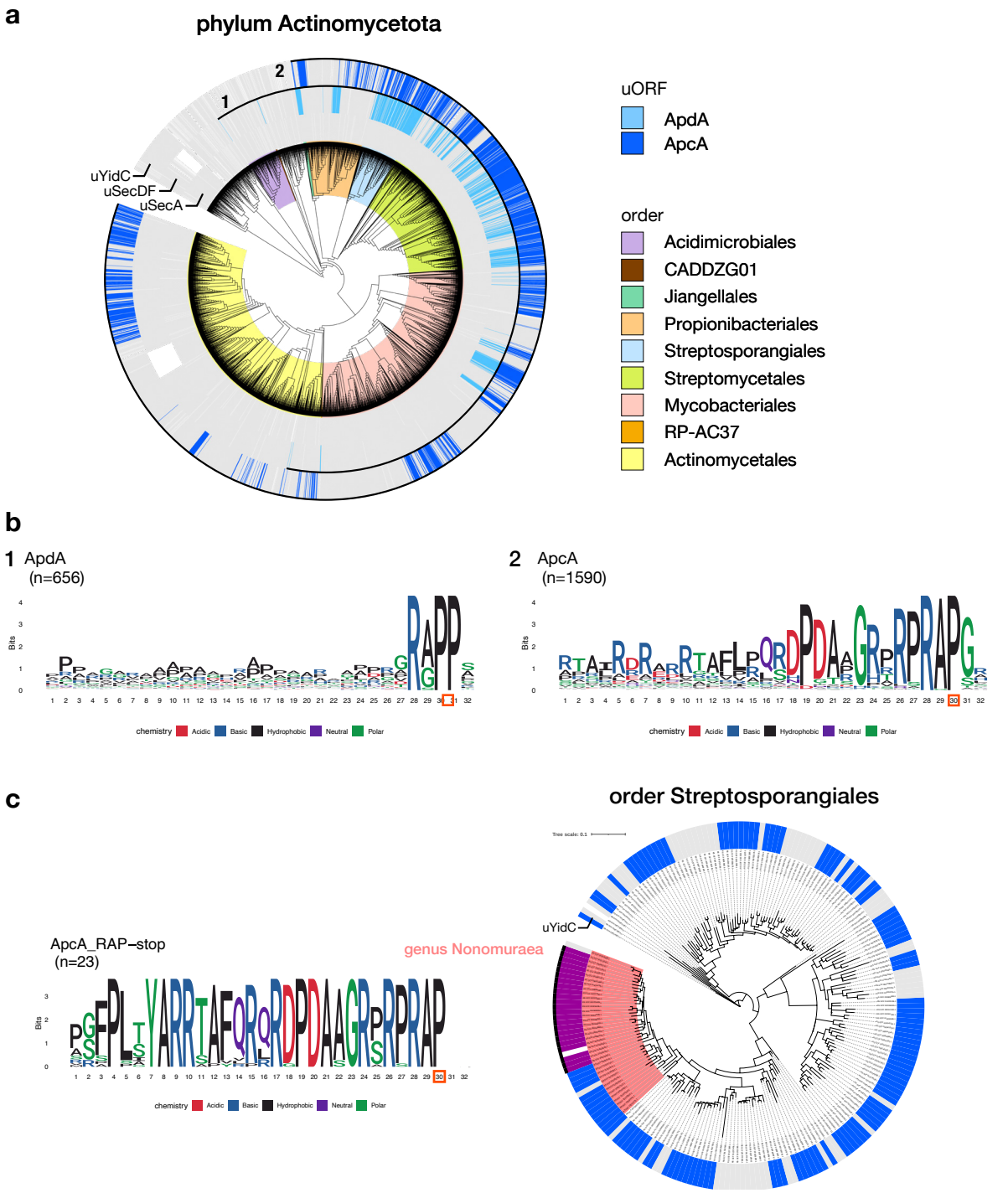
Supplementary Figure 1. General information of candidate monitoring substrates. **a** Names of previously and currently identified arrest peptides (AP), their phylogenetic distributions (phylum), and their downstream genes (mORF) are listed. 'n' represents the number of ORFs identified in this study. **b** Orphan uORFs with RAPP/RGPP motif. **c** A phylogenetic distributions of the orphan uORFs with RAPP/RGPP motif.



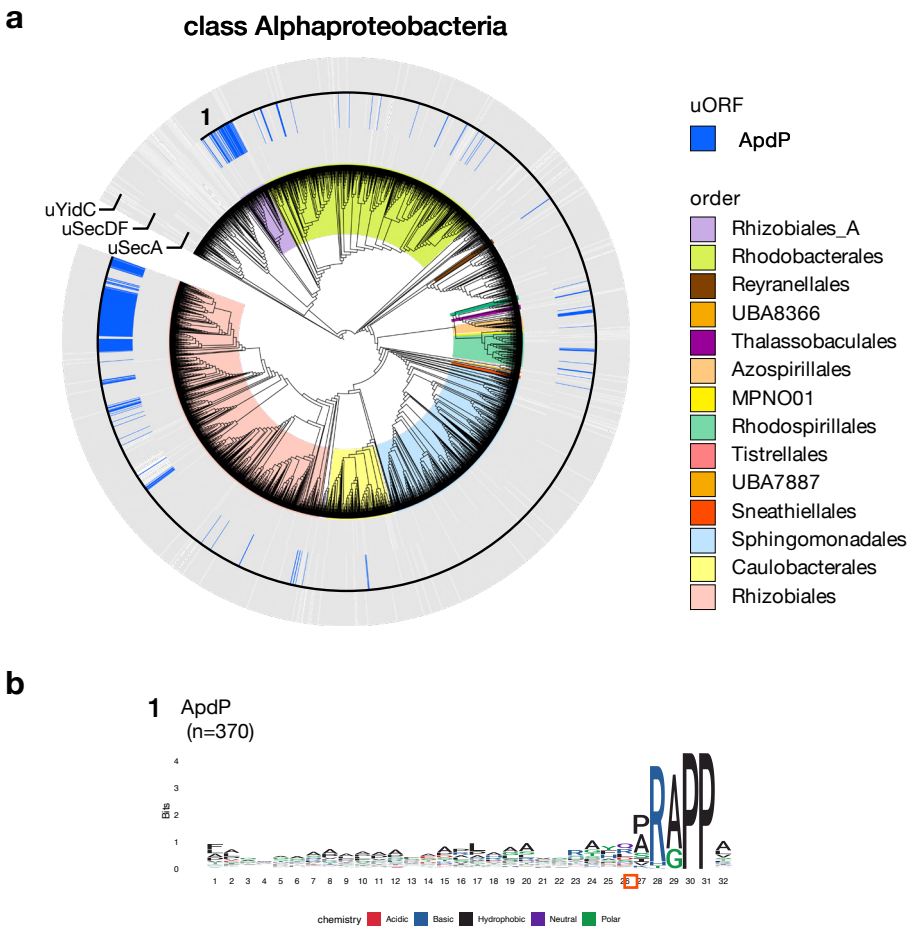
Supplementary Figure 2. Phylogenetic distribution of candidate monitoring substrates in Gammaproteobacteria. a Phylogenetic distribution of the candidate monitoring substrates in class Gammaproteobacteria. The bacterial genomes that encode SecA, SecD/F, or YidC homolog are indicated by grey strips around the bacterial phylogenetic tree. Genomes that possess genes for known monitoring substrates and putative arrest peptides upstream of *secA*, *secDF*, or *yidC* gene are represented by chromatic strips with names or numbering corresponding to those shown in **b**. The colors behind the tree indicate the bacterial orders listed on the right. **b** Sequence logos of known and candidate monitoring substrates. P-site residues at the 30th position (red square) with its N-terminal 29 residues, as well as the following two C-terminal residues, are used to plot sequence logos.



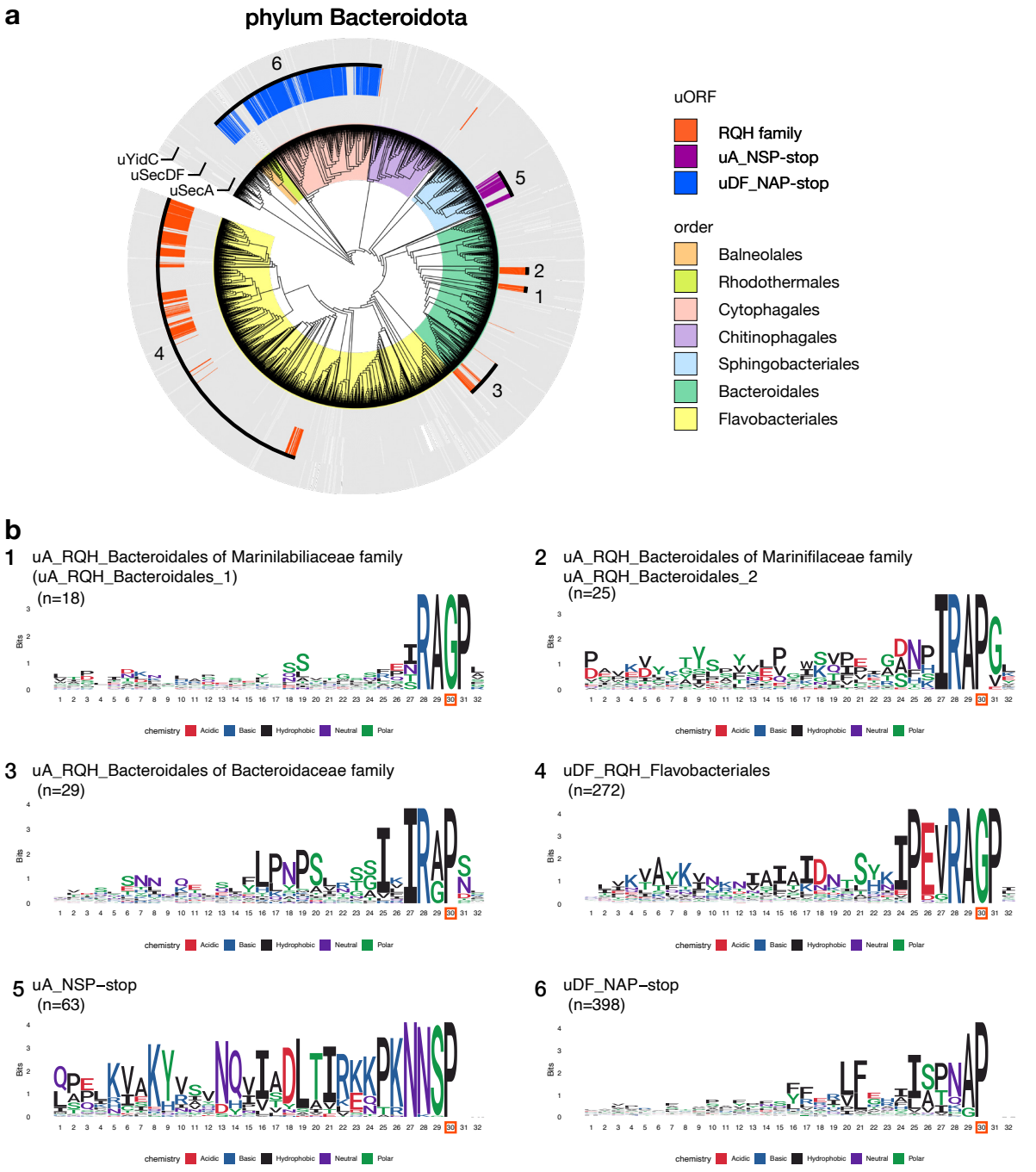
Supplementary Figure 3. Phylogenetic distribution of candidate monitoring substrates in phylum Firmicutes. a Phylogenetic distribution of the known and candidate monitoring substrates in the phylum Firmicutes, Firmicutes_A and Firmicutes_B. **b** Sequence logos and alignments of known and candidate monitoring substrates in the phylum Firmicutes, Firmicutes_A, and Firmicutes_B. The figures follow the format used in Supplementary Fig. 2.



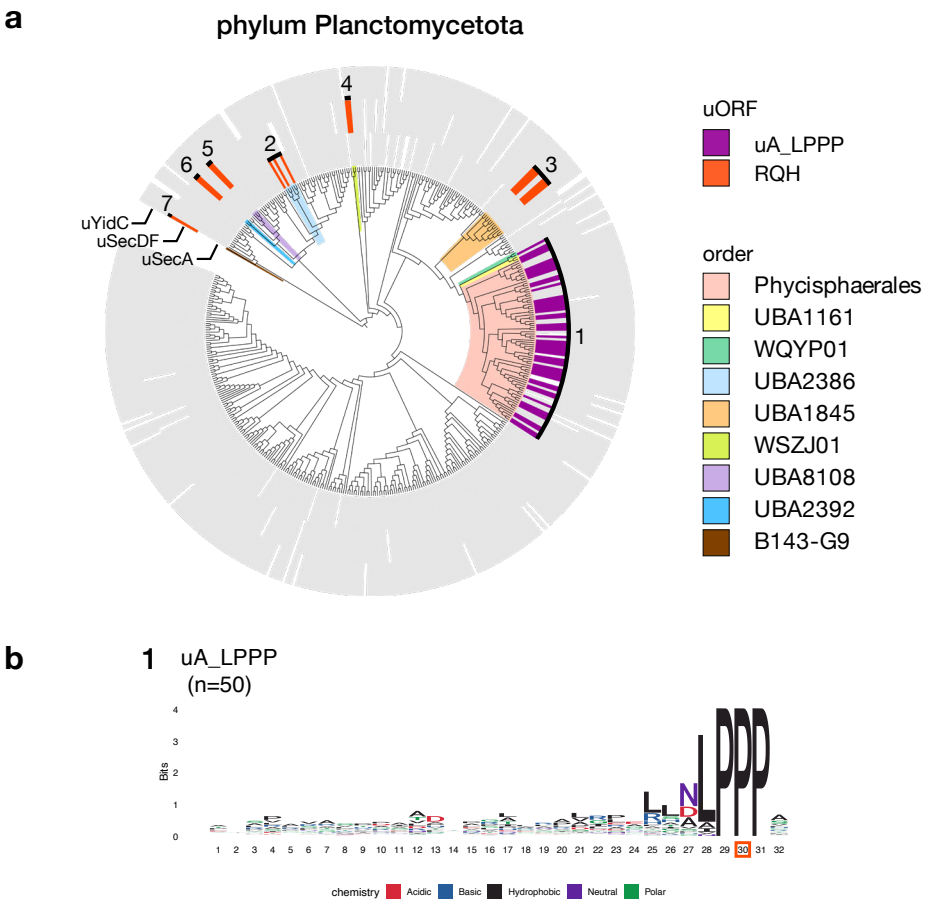
Supplementary Figure 4. Phylogenetic distribution of candidate monitoring substrates in phylum Actinomycetota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Actinomycetota. **b** Sequence logos of ApdA and ApcA. **c** Phylogenetic tree of order Streptosporangiales and a sequence logo of ApcA with RAP-stop found among genus Nonomuraea.



Supplementary Figure 5. Phylogenetic distribution of ApdP in class Alphaproteobacteria. **a,b** Phylogenetic distribution (**a)** and sequence logo (**b)** of ApdP. The figures follow the format used in Supplementary Fig. 2.



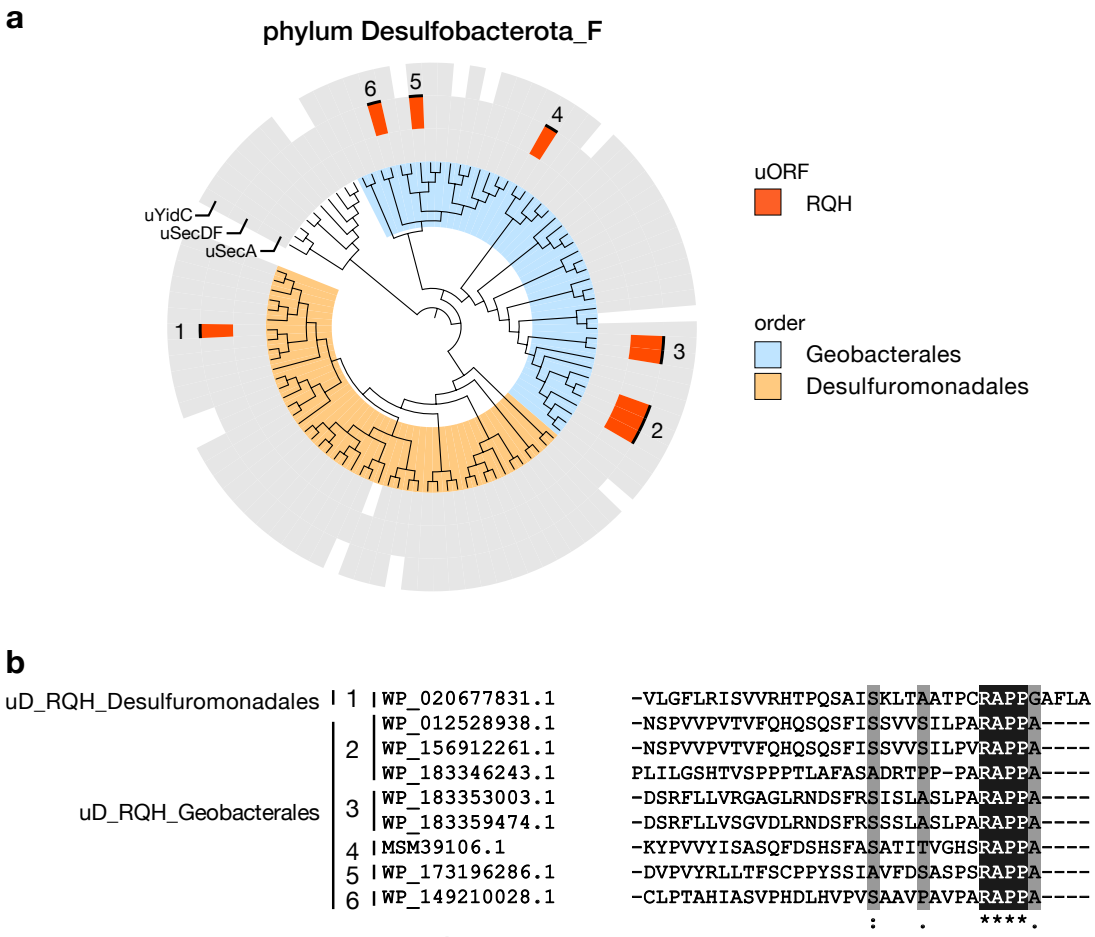
Supplementary Figure 6. Phylogenetic distribution of candidate monitoring substrates in phylum Bacteroidota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Bacteroidota. **b** Sequence logos of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.



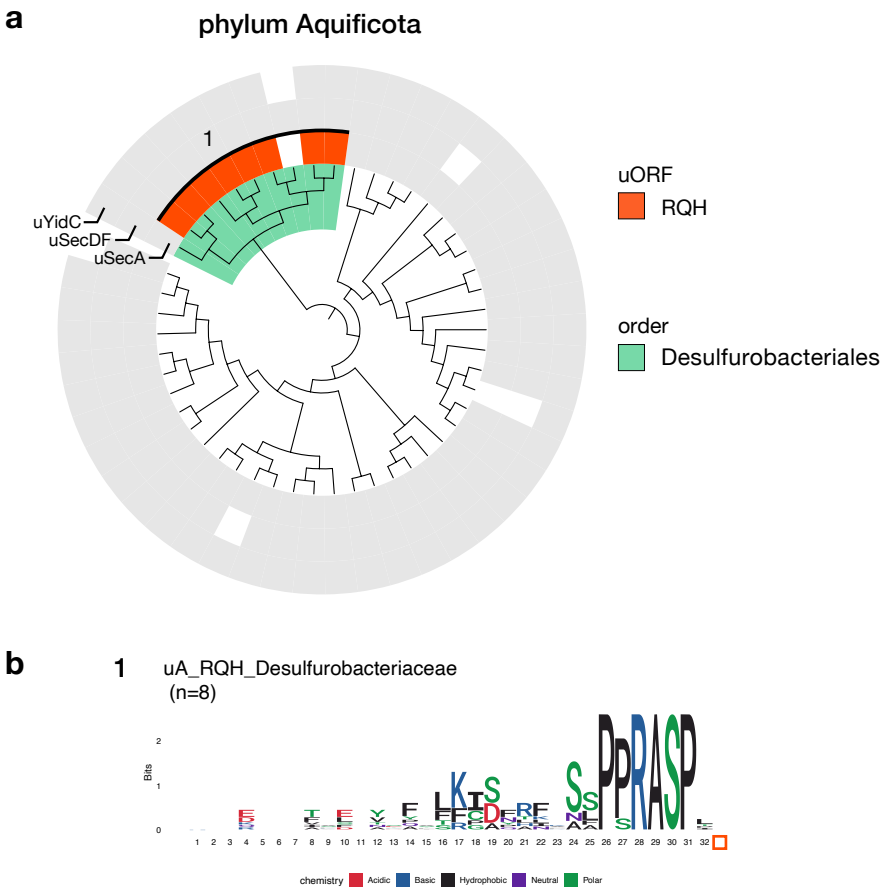

```
2  uA_RQH_UBA2386   MAD33348.1      VAVCFEVEPPRVAPRQARVDSWGLPPPRAPASARS
                    HHI78652.1      LFRSGFPAPRALGHARSGVRSFLLPAPRAPSS---
                    MAE77895.1      HCRSSDHAAVRHERFTPRRLRWVRPPARAPGC---

3  uD_RQH_UBA1845   NLU21220.1      -MAPHALESGAADRPVQATRAWLMAR-YPHAPPA-------
                    NLX24145.1      -ARCSRPKPADADQPSTAAAAWLVCR-YAHAPPLEH-----
                    NLX13139.1      -NGGRHFITVCRFPTGPTARYWLANR-YPHAPPGQV-----
                    RIK69839.1      -SLVKLERGNLSAGALVVAALHYPIYQ-AQAPPHAPALLQG
                    QDV91326.1      -THATLERGDSAATVTVVSALHYPIYQ-AQAPPPTQG----
                    TWT46075.1      -AMGAWRSHAAADARECFRTLLRSAIR-RTGPPPA------
4  uD_RQH_WSZJ01    KAF0245086.1    --LASTPGAGAPAALHPSFDAPAASRGLGRAPPAAL-----
                    NUN49545.1      -RFFLSAGRFESAPSTPPASRPGTLR-PGRAPPARS-----
5  uD_RQH_UBA8108   MBA20503.1      -AGELALHRPAVVRLAPPASQSWKHQ-PVRGPPVA------
                    HBO50953.1      -AGELALHRPADVRLAPPASQSWKHQ-PVRGPPAA------
6  uD_RQH_UBA2392   RIK65329.1      --TRPPPLAGPASFQDLVASQFCRAAQPVRAPPVALR----
                    NUQ33773.1      --LNMDPPAPWPSHERTEIGQAIAGYAGARAPPSLV-----
7  uD_RQH_B143-G9   RKY22358.1      -RGGGGLGLGLEMERVSPPSAESAPS-NPRAPPIL------
```

Supplementary Figure 7. Phylogenetic distribution of candidate monitoring substrates in phylum Planctomycetota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Planctomycetota. **b** A sequence logo and sequence alignments of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.

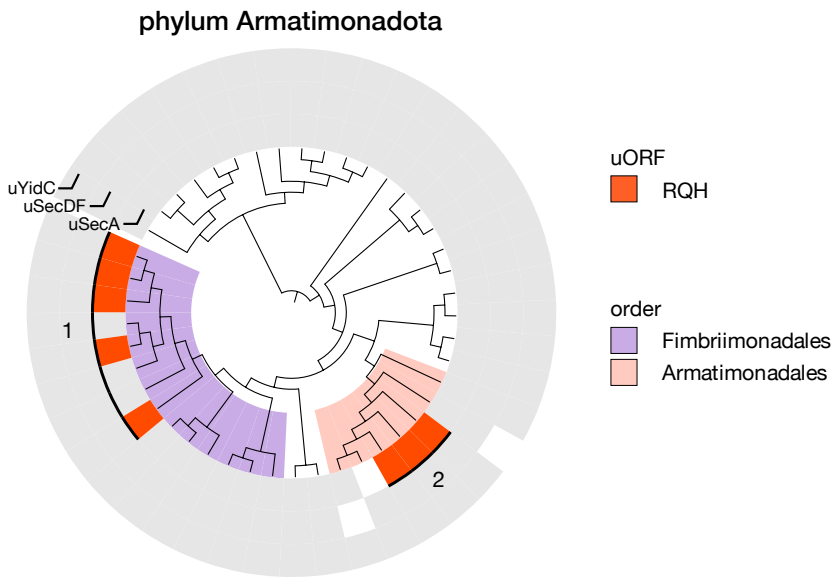


Supplementary Figure 8. Phylogenetic distribution of candidate monitoring substrates in phylum Desulfobacterota_F. a Phylogenetic distribution of the candidate monitoring substrates in phylum Desulfobacterota_F. **b** Sequence alignment of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.



Supplementary Figure 9. Phylogenetic distribution of candidate monitoring substrates in phylum Aquificota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Aquificota. **b** A sequence logo and alignment of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.

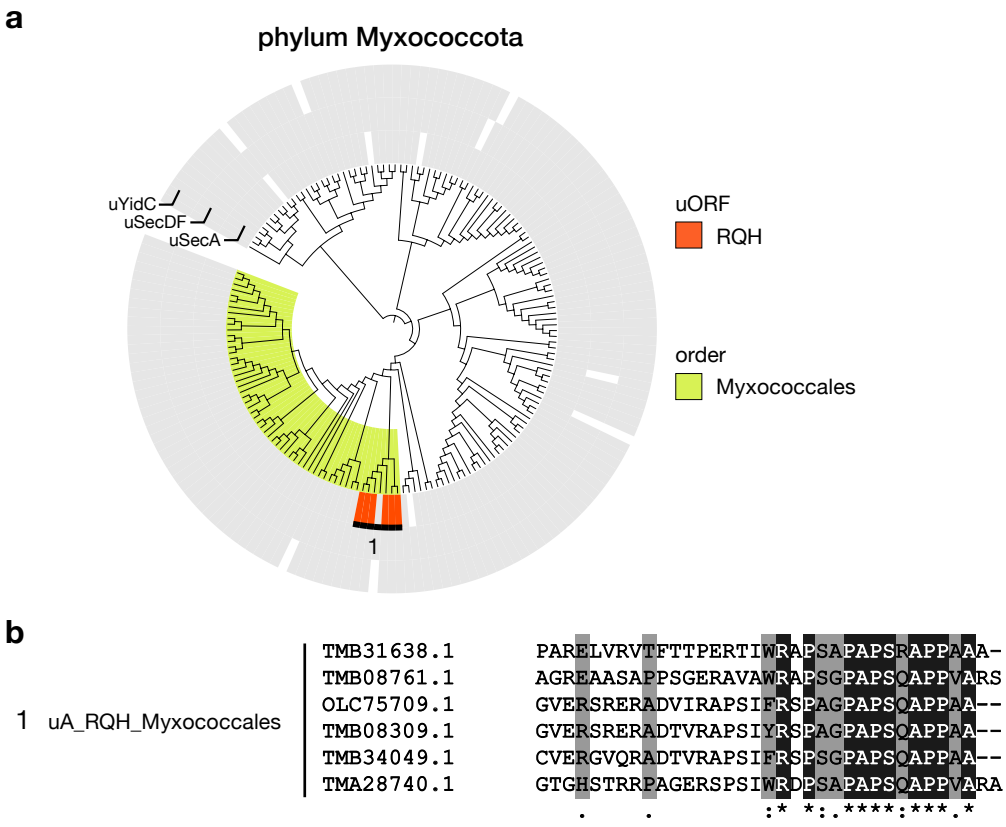
a



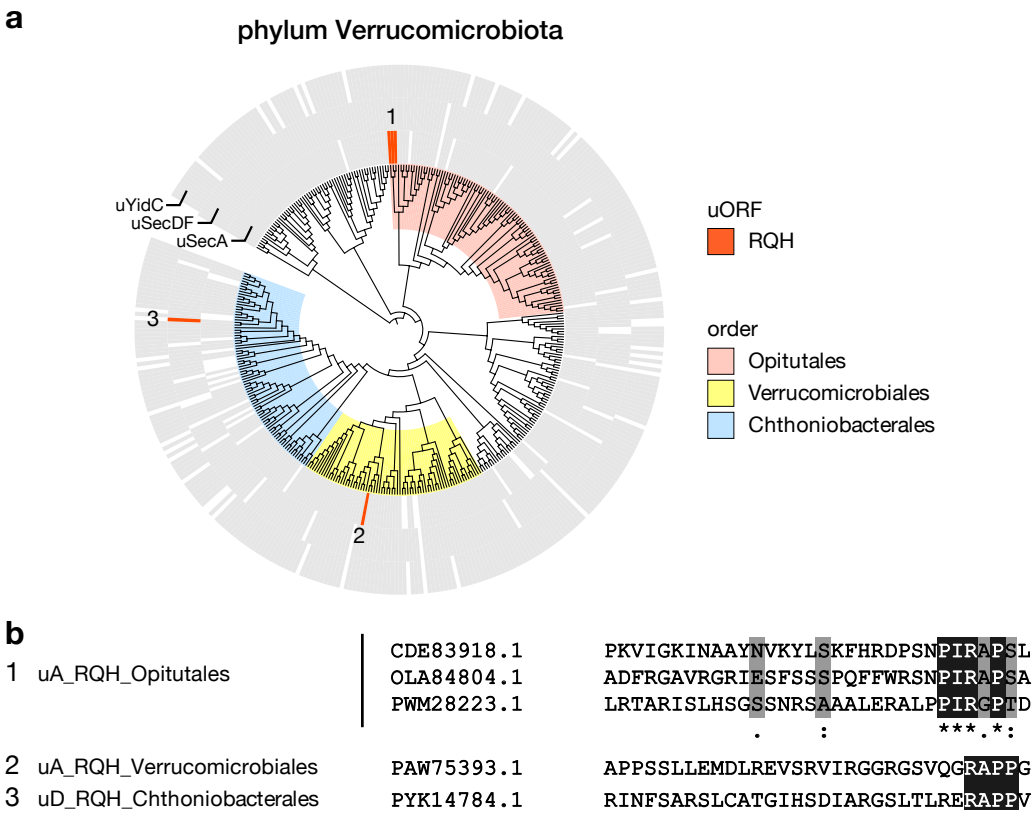
b

```
                                 OJU64000.1          --NLDDPPLPIEGRWIAALADGFAHSQRSRDGPEPLA
                                 KXK20322.1          KTQGTEGIRSEHLRQPPGPPSCR--SDLIRDGPLA--
1  uA_RQH_Fimbriimonadales       MBA4292110.1        -LVRSDSRSKANSIGFGS-TNVWPDSDRVRDGPLV--
                                 MBC8065182.1        --PDSDWNLNFYADEPSAVPEEILECQRSRDGPL---
                                 WP_144240978.1      --QQSEARPVISVGDSPTPPDGFEDCRRSRDGPLSIA
                                 MBC8103941.1        --PDDGPGLAGPTPAETLRSRFVVSPASPRAPPFA--
2  uA_RQH_Armatimonadales        WP_184199750.1      --PELCAARQLYPTLTTTLRSAERLSYGSRAPPHAA-
                                 MBC8142143.1        --LTACPSAPRSALRQAVVRPALPGRFSPRAPPV---
                                                                                 *  *
```

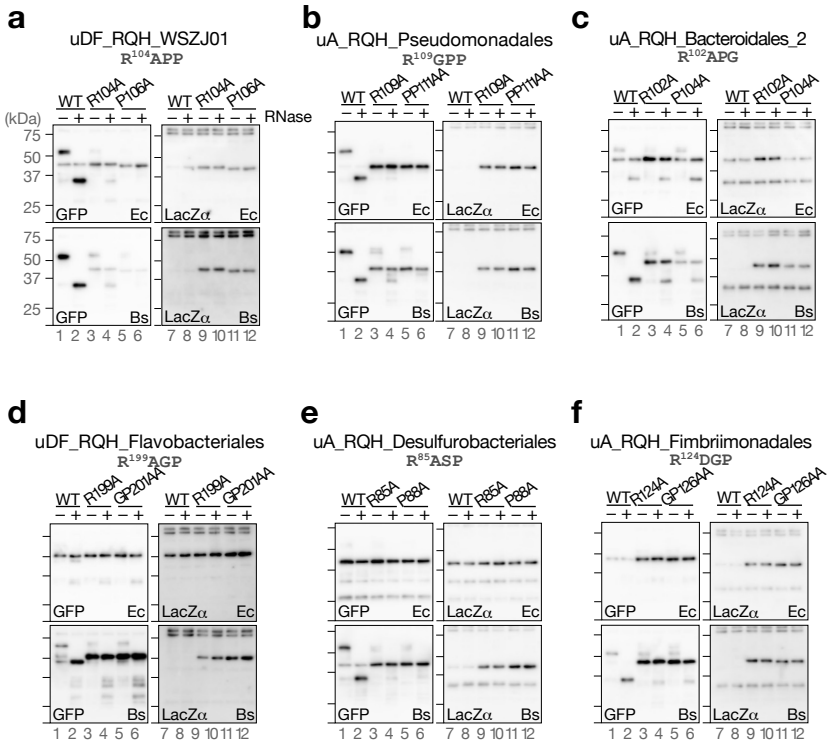
Supplementary Figure 10. Phylogenetic distribution of candidate monitoring substrates in phylum Armatimonadota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Armatimonadota. **b** A sequence alignment of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.



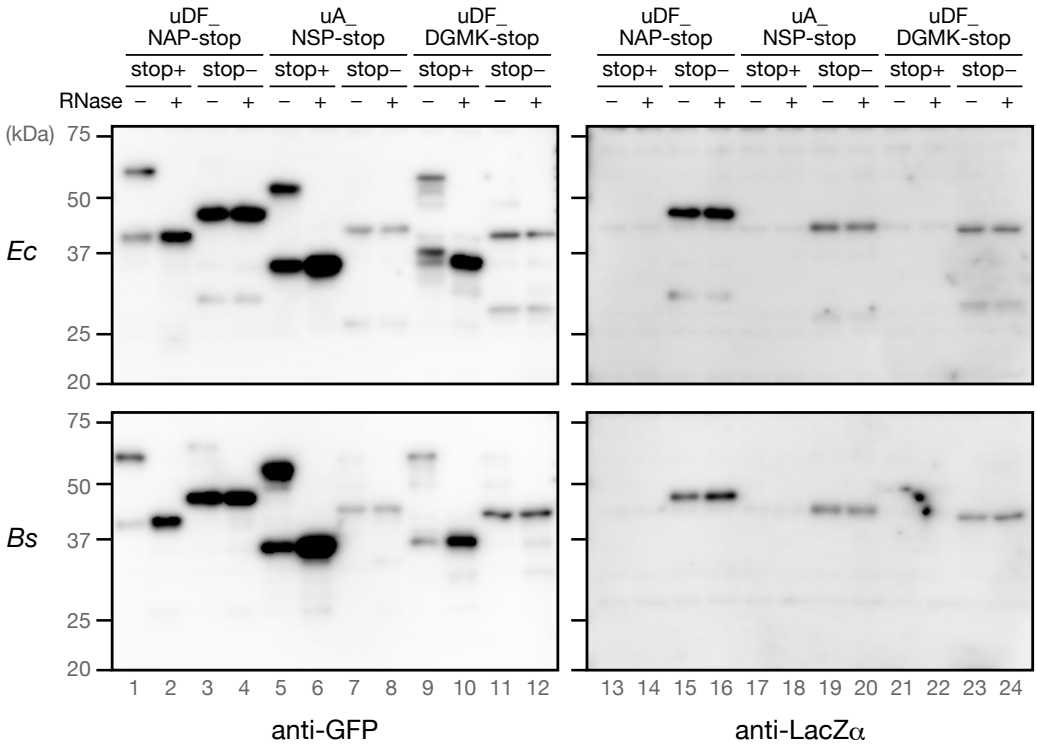
Supplementary Figure 11. Phylogenetic distribution of candidate monitoring substrates in phylum Myxococcota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Myxococcota. **b** A sequence alignment of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.



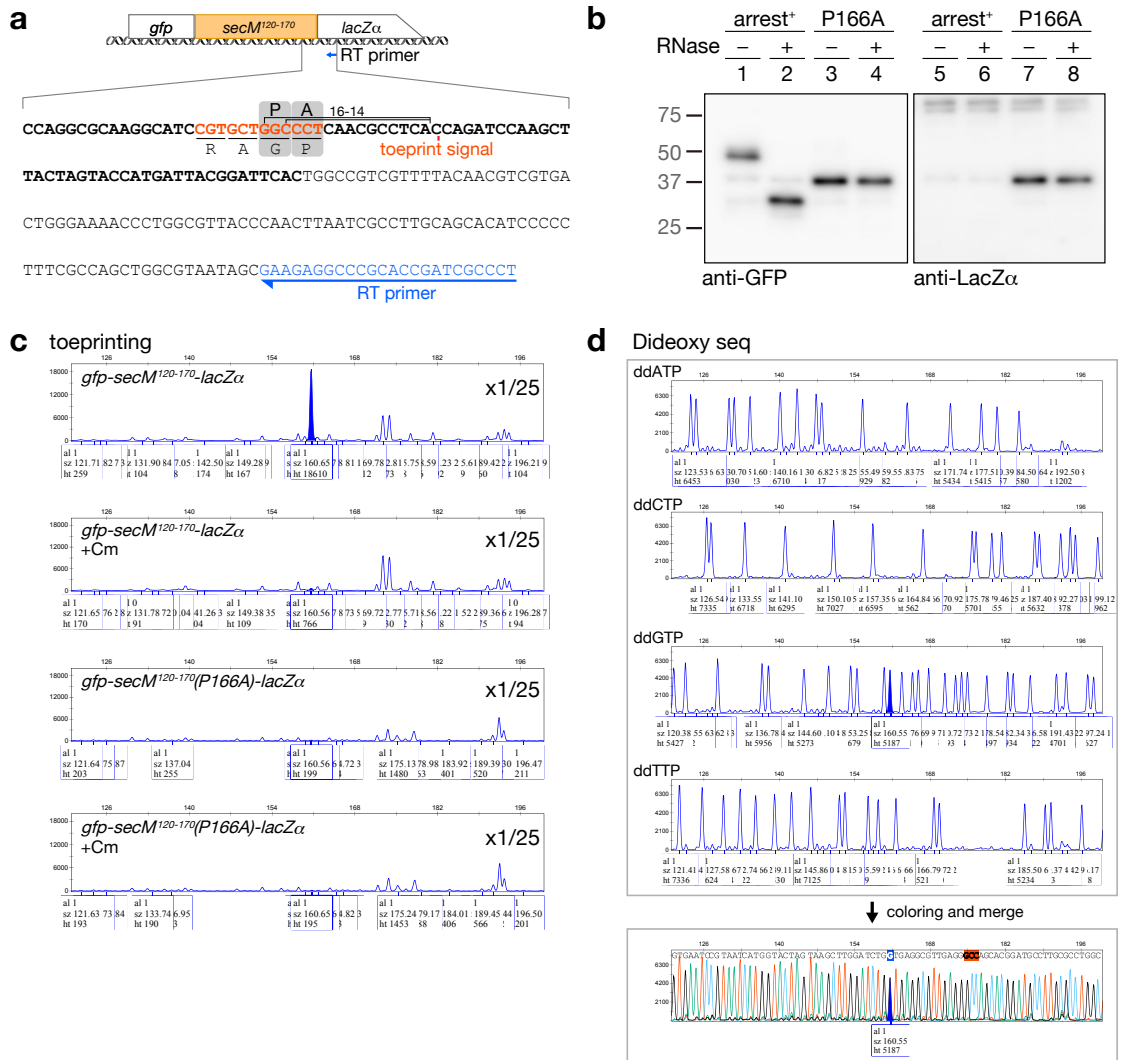
Supplementary Figure 12. Phylogenetic distribution of candidate monitoring substrates in phylum Verrucomicrobiota. a Phylogenetic distribution of the candidate monitoring substrates in phylum Verrucomicrobiota. **b** Sequence alignment of candidate monitoring substrates. The figures follow the format used in Supplementary Fig. 2.



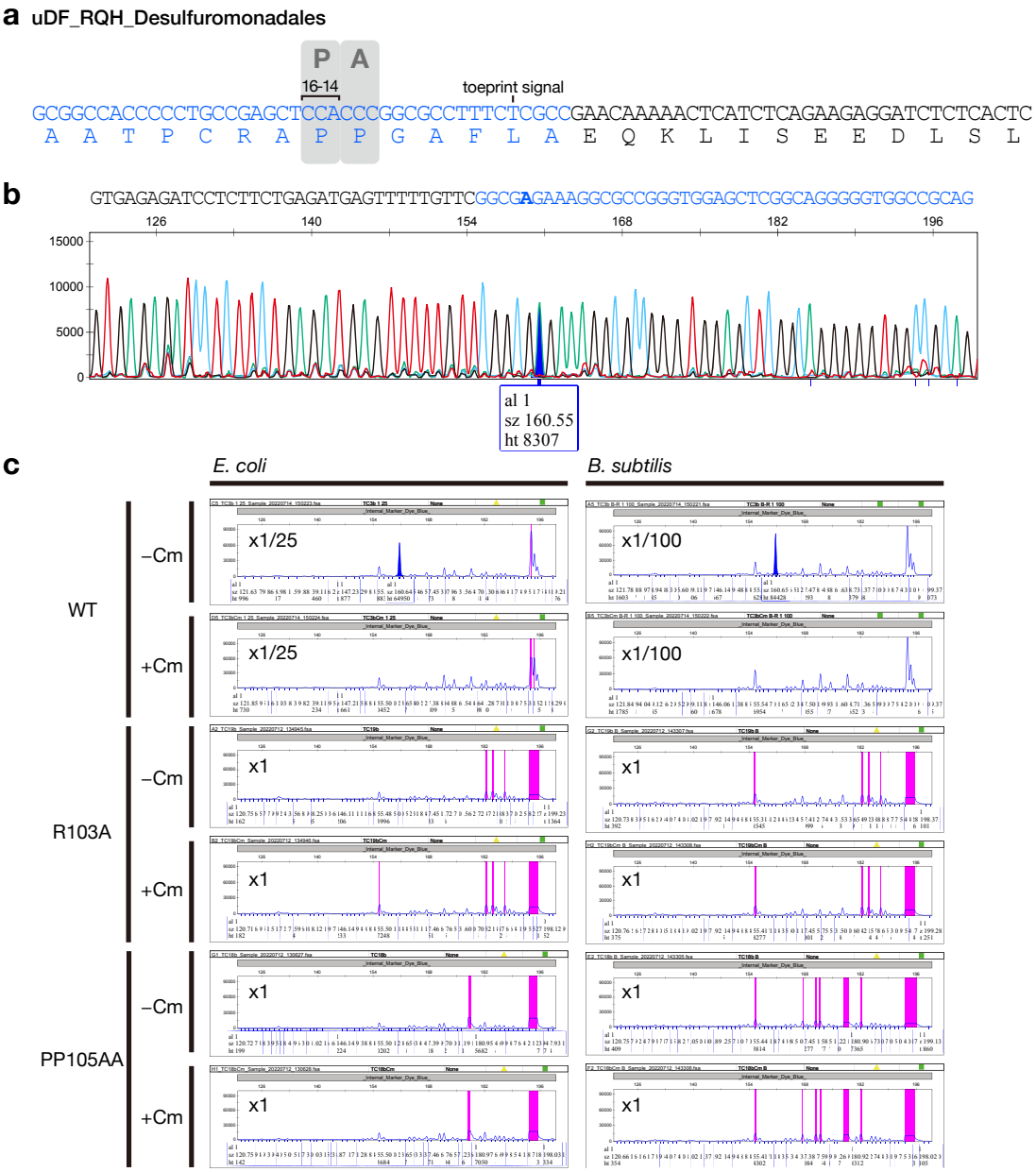
Supplementary Figure 13. In vitro analysis of translation arrest of candidate monitoring substrates. a-f Western blot analysis of the in vitro translation products. The reporter genes harboring wild-type (WT) or mutant derivatives of the putative arrest peptides indicated at the top of the figure were translated in the *E. coli* (Ec) and *B. subtilis* (Bs) PURE systems. The products were separated in neutral-pH gels and immunoblotted using anti-GFP (left) or anti-LacZα (right) antibodies. Before the separation, a portion of the samples were treated with RNase A (lanes indicated as +), to degrade the tRNA moiety. Molecular size standards are indicated as horizontal lines on the left of each membrane; from top to bottom: 75, 50, 37, and 25 kDa, respectively. Experiments were conducted twice independently, with similar results. Source data are provided as a Source Data file.



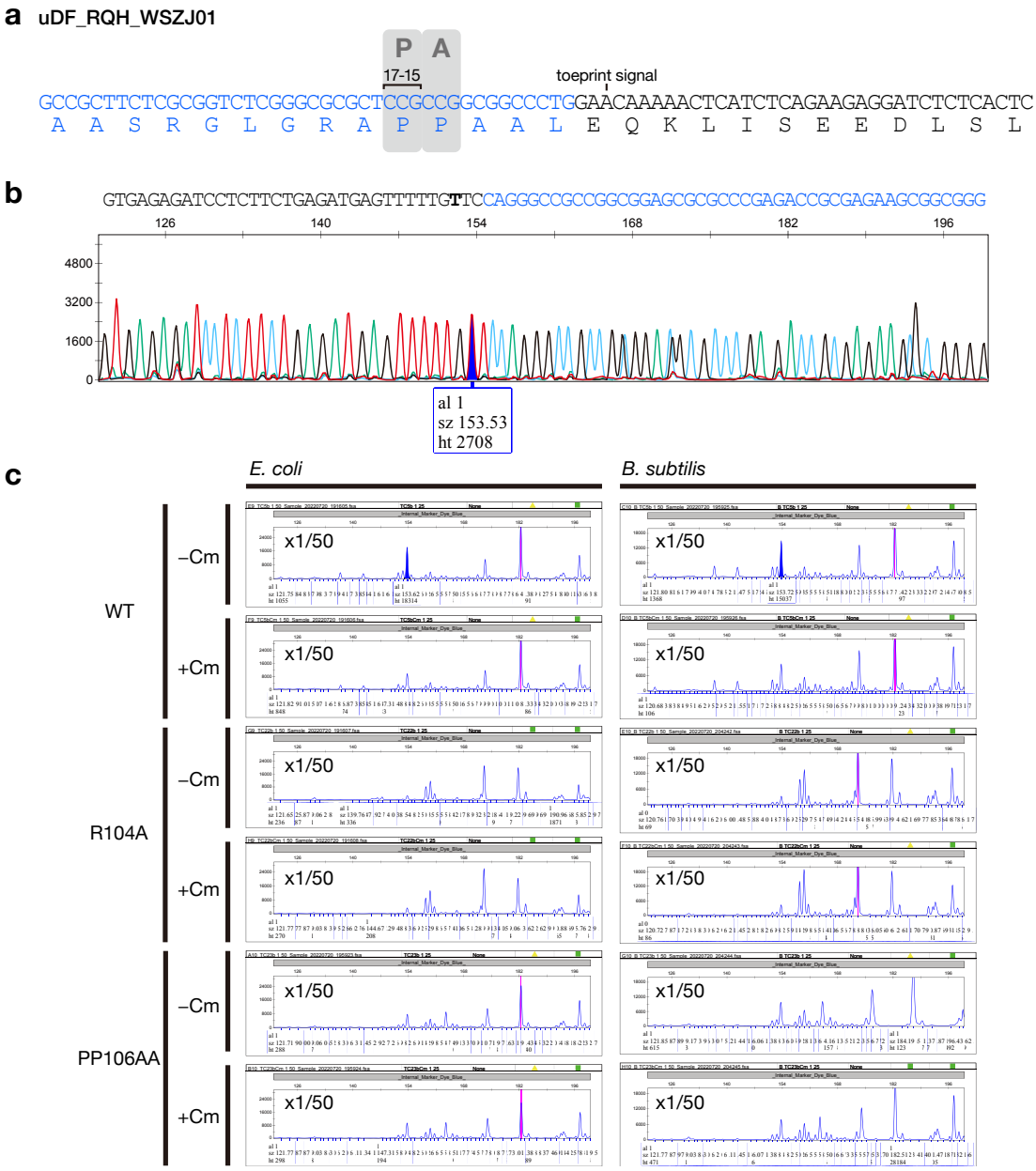
Supplementary Figure 14. The translation arrest of uDF_NAP-stop, uA_NSP-stop and uDF_DGMK-stop depends on the stop codon. The *gfp-ap-lacZa* reporters harboring uDF_NAP-stop, uA_NSP-stop and uDF_DGMK-stop C-terminal region with (stop+) or without (stop−) stop codon just after the NAP, NSP or DGMK motif were translated in *E. coli* (*Ec*, upper) or *B. subtilis* (*Bs*, lower) PURE. The translation products were immunoblotted using anti-GFP (left membranes) or anti-LacZα (right membranes). Before electrophoresis, portions of the sample were treated by RNase A (lanes indicated as +) to degrade tRNA moiety. Experiments were conducted twice independently, with similar results. Source data are provided as a Source Data file.



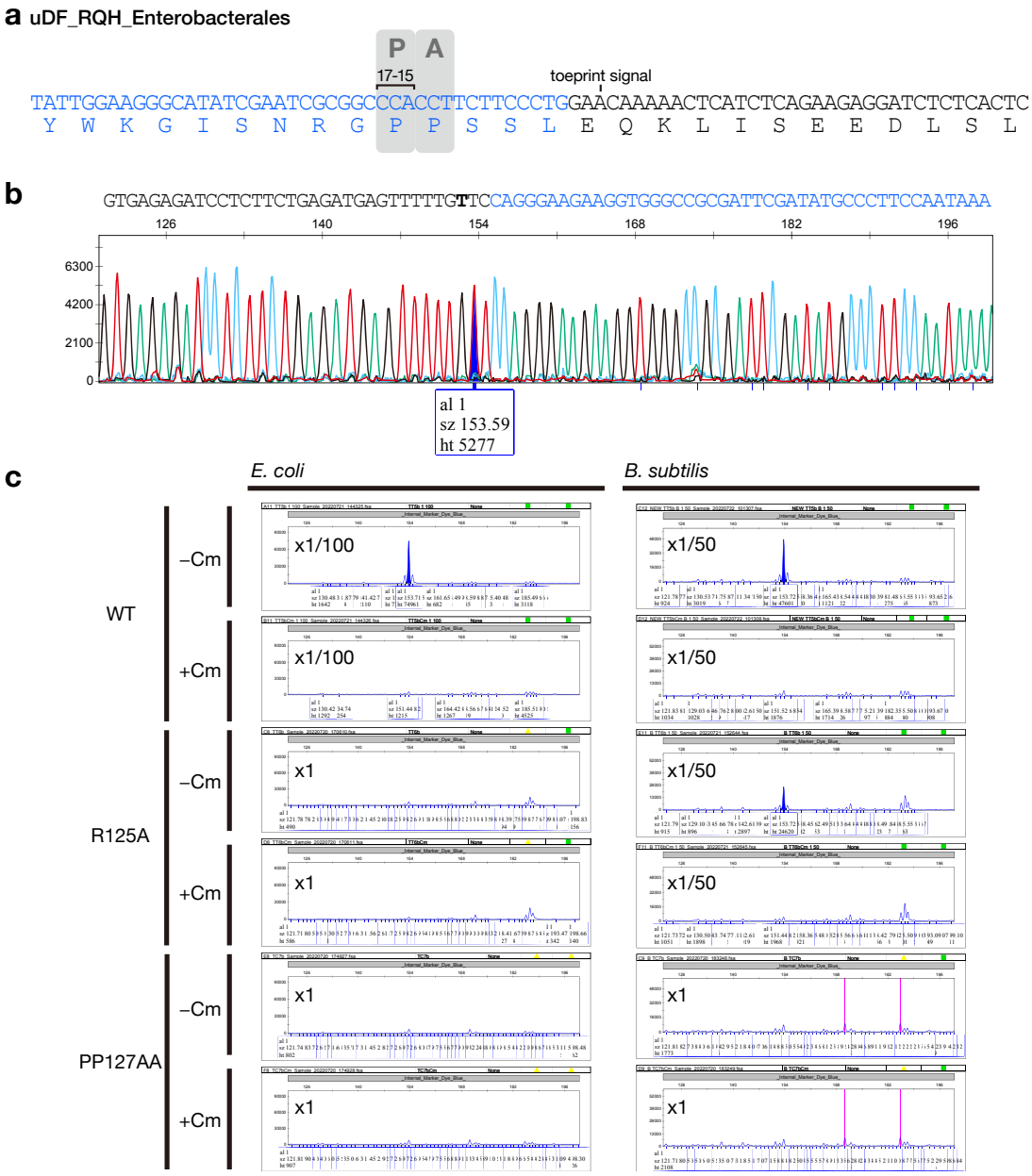
Supplementary Figure 15. Identification of the ribosome stalling site of *E. coli* SecM by toeprinting. a A schematic representation of the template gene for in vitro translation, in which the coding region for SecM residues 120-170 was sandwich-fused between *gfp* and *lacZα* (upper). The sequence between the coding region for SecM arrest site and the downstream primer annealing site for reverse transcription and dideoxy sequencing is shown at the bottom. The residue corresponding to the toeprint signal (results shown in **c** and **d**) and the estimated ribosome stalling site are shown. The gray boxes with 'P' and 'A' indicate P- and A-site codons, respectively. **b** Western blotting analysis of translation product of the reporter by *Ec* PURE system. Before electrophoresis, portions of the sample were treated by RNase A (lanes indicated as +) to degrade tRNA moiety. Experiments were conducted twice independently, with similar results. Source data are provided as a Source Data file. **c** Raw results of the fragment analysis of toeprint product using a capillary sequencer. The template gene shown in **a** was translated by *Ec* PURE with or without chloramphenicol, and subsequently subjected to reverse transcription using the reverse transcription (RT) primer shown in **a**. The size of the complementary DNA was then analyzed by the fragment analysis using a capillary DNA sequencer. The stalling-dependent toeprint signal specifically detected from the experiment using the wild-type but not from the arrest-defective P166A derivative of SecM in the absence of chloramphenicol is filled in blue. **d** Raw results of the fragment analysis of dideoxy sequencing products using the same RT primer. Each plot was colored (A; green, C; right blue, G; black, T; red) and then overlayed as shown at the bottom. The peaks whose fragment size are same as that of the stalling-dependent toeprint signal were filled in blue. The toeprint product generated by the SecM-stalled ribosome appeared as a single peak in our fragment analysis with the size indicating that the ribosome stalled with the P-site at the Gly_{165} codon and the A-site at the Pro_{166} codon (Fig. 3b).



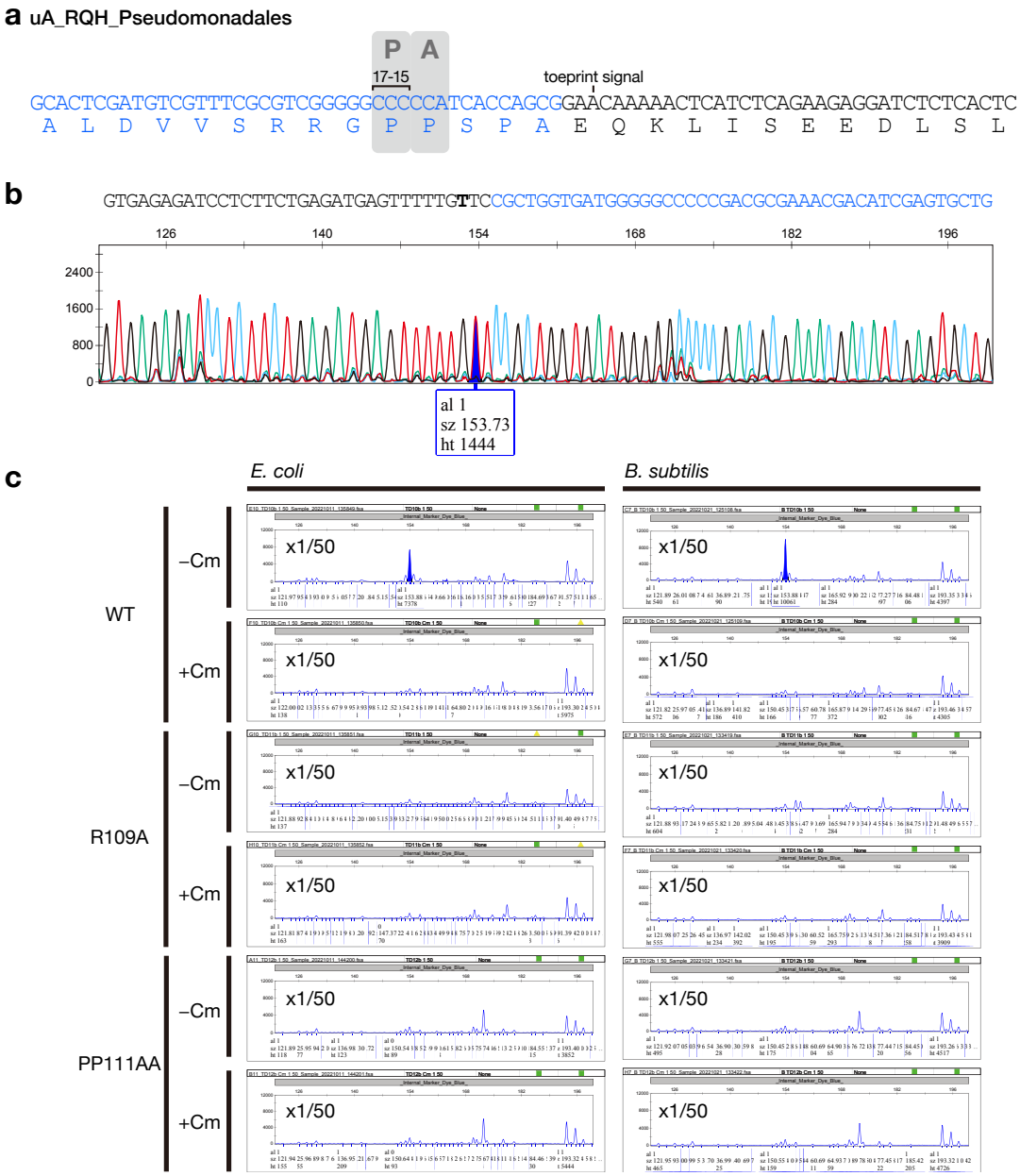
Supplementary Figure 16. Identification of the ribosome stalling site of uDF_RQH_Desulfuromonadales by toeprinting. a A schematic representation of the ribosome stalling site of uDF_RQH_Desulfuromonadales estimated by toeprinting. The nucleotide and amino acid sequences derived from the coding region of uDF_RQH_Desulfuromonadales are shown in blue characters. The estimated ribosome stalling site was shown by the P-site (P) and A-site (A) codons. **b** Overlayed results of dideoxy sequencing. The reverse complement sequence of the *uDF_RQH_Desulfuromonadales* region is presented in blue, and the peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. The numbers above the peak data indicate the estimated sizes (nucleotides) based on a molecular weight standard (500LIZ). **c** Raw results of the fragment analysis of toeprint products yielded by the reverse transcription after in vitro translation of wild-type (WT) and arrest-defective mutant derivatives (R103A, PP105AA) of uDF_RQH_Desulfuromonadales using *Ec* (left) and *Bs* (right) PURE. Each template DNA was subjected to translation in the presence (+Cm) or absence (–Cm) of chloramphenicol. When necessary, the toeprinting products were diluted with HiDi formamide prior to the capillary electrophoresis at the ratios indicated in each panel. The magenta lines represent saturated peaks.



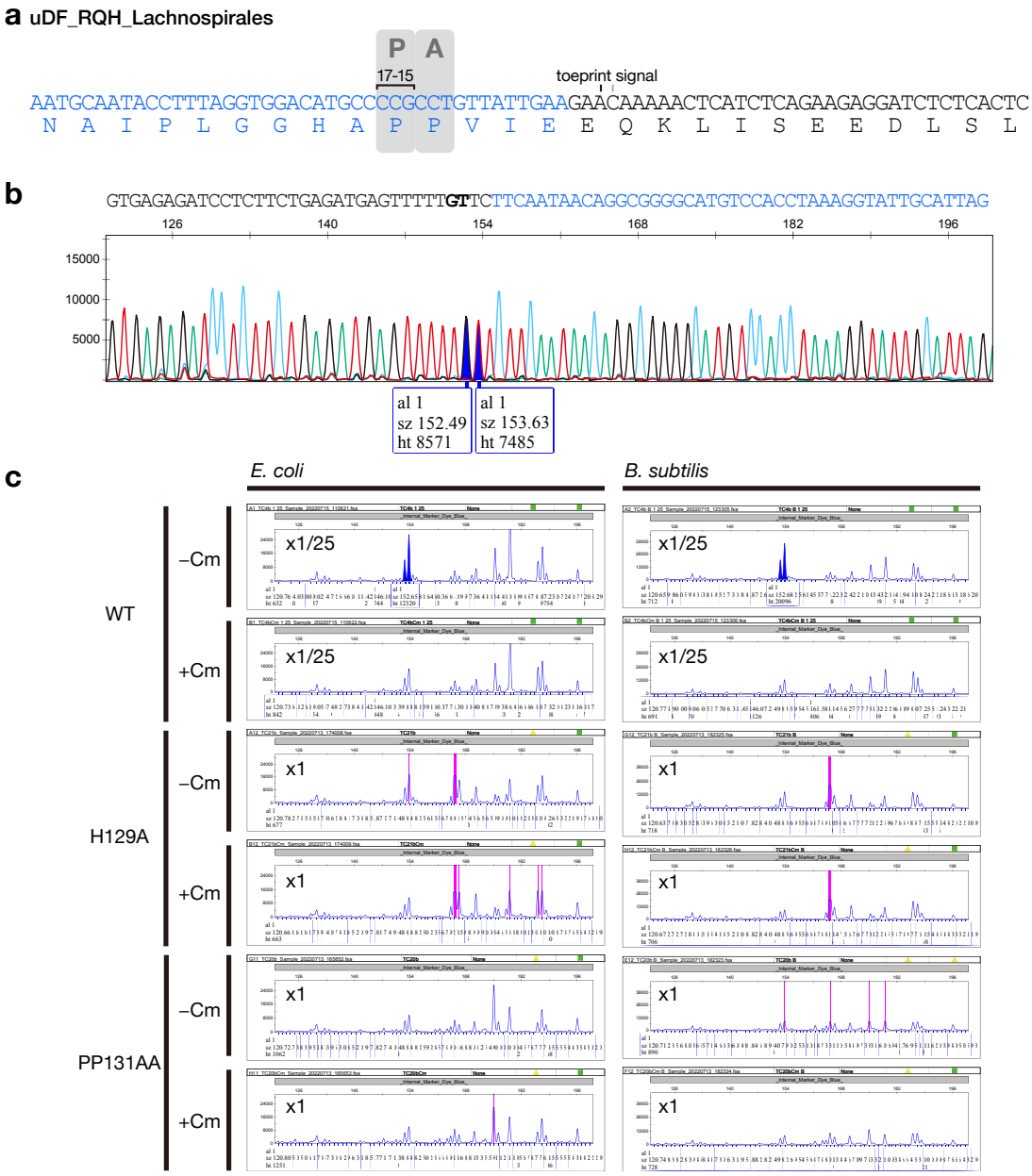
Supplementary Figure 17. Identification of the ribosome stalling site of uDF_RQH_WSZJ01 by toeprinting. **a** A schematic representation of the ribosome stalling site of uDF_RQH_WSZJ01 estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



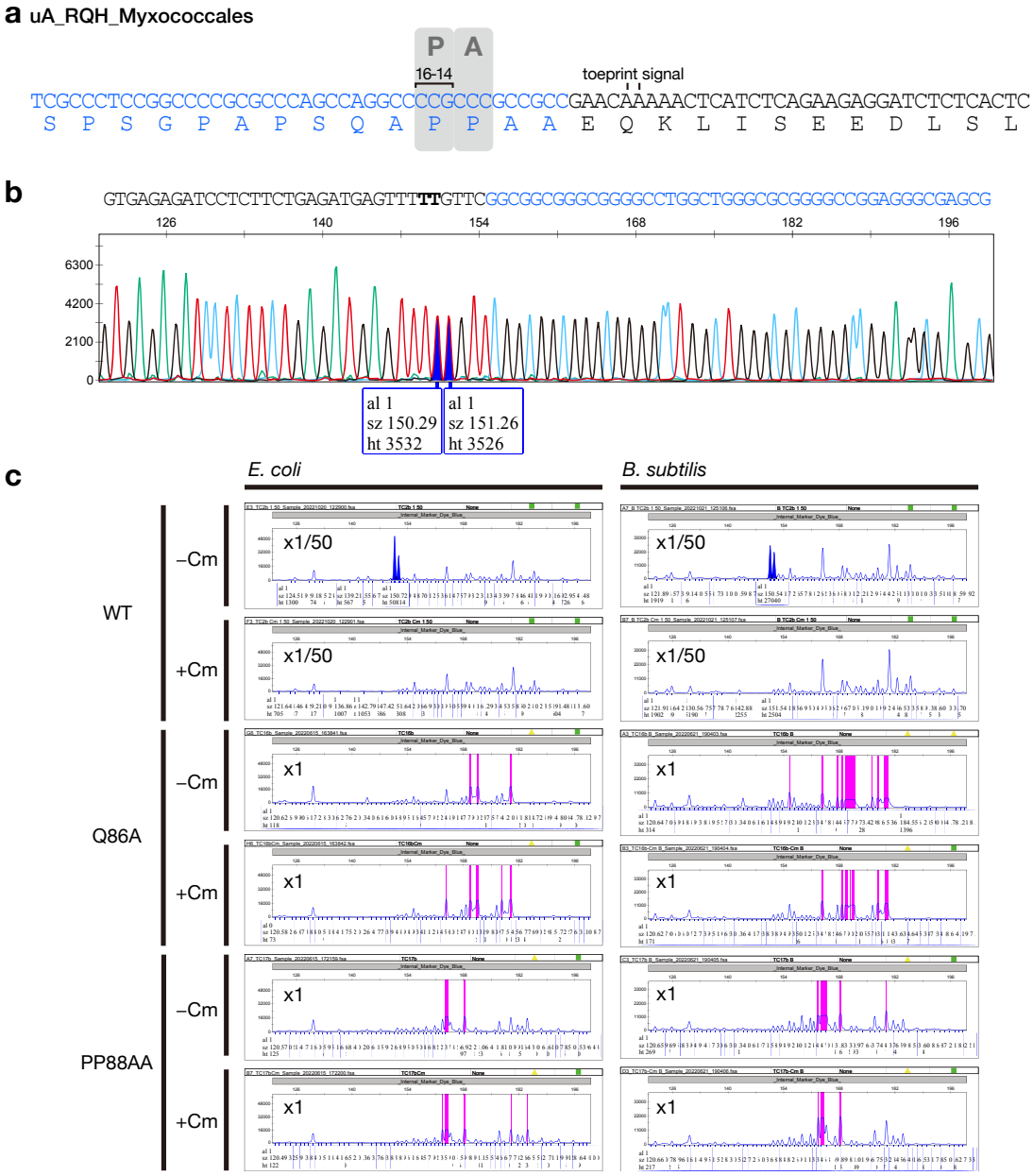
Supplementary Figure 18. Identification of the ribosome stalling site of uDF_RQH_Enterobacterales by toeprinting. a A schematic representation of the ribosome stalling site of uDF_RQH_Enterobacterales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



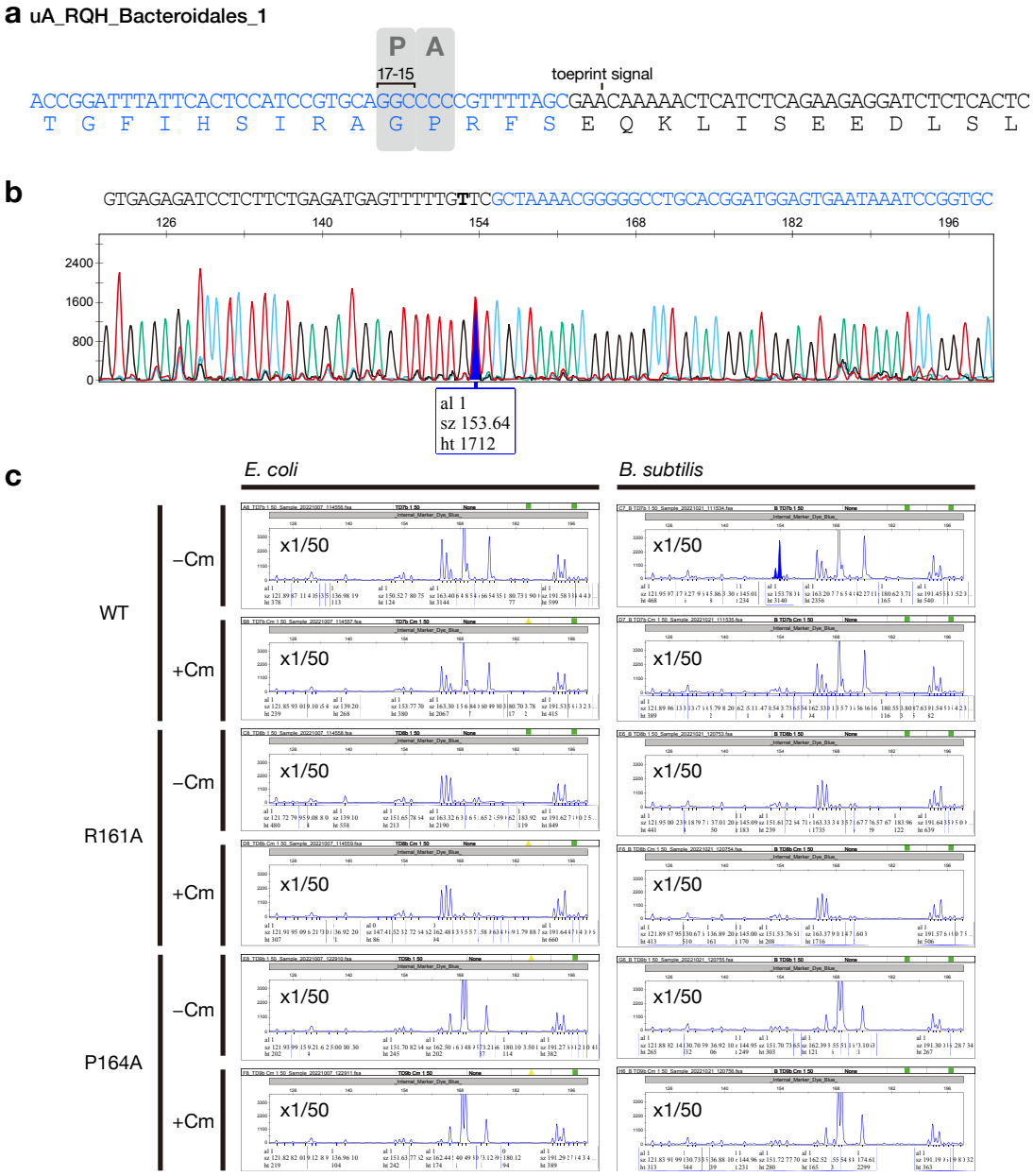
Supplementary Figure 19. Identification of the ribosome stalling site of uA_RQH_Pseudomonadales by toeprinting. a A schematic representation of the ribosome stalling site of uA_RQH_Pseudomonadales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



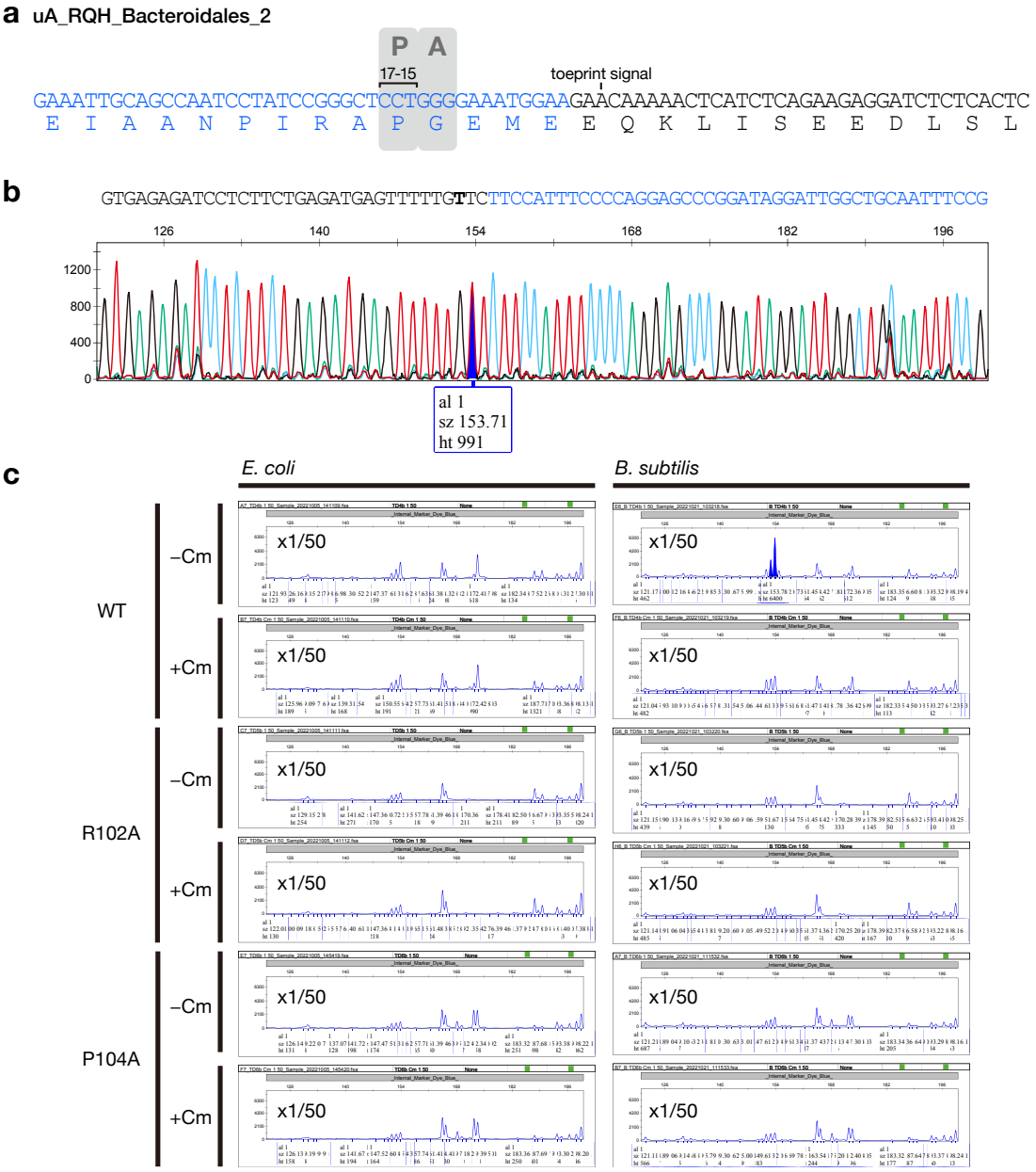
Supplementary Figure 20. Identification of the ribosome stalling site of uDF_RQH_Lachnospirales by toeprinting. a A schematic representation of the ribosome stalling site of uDF_RQH_Lachnospirales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabets, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



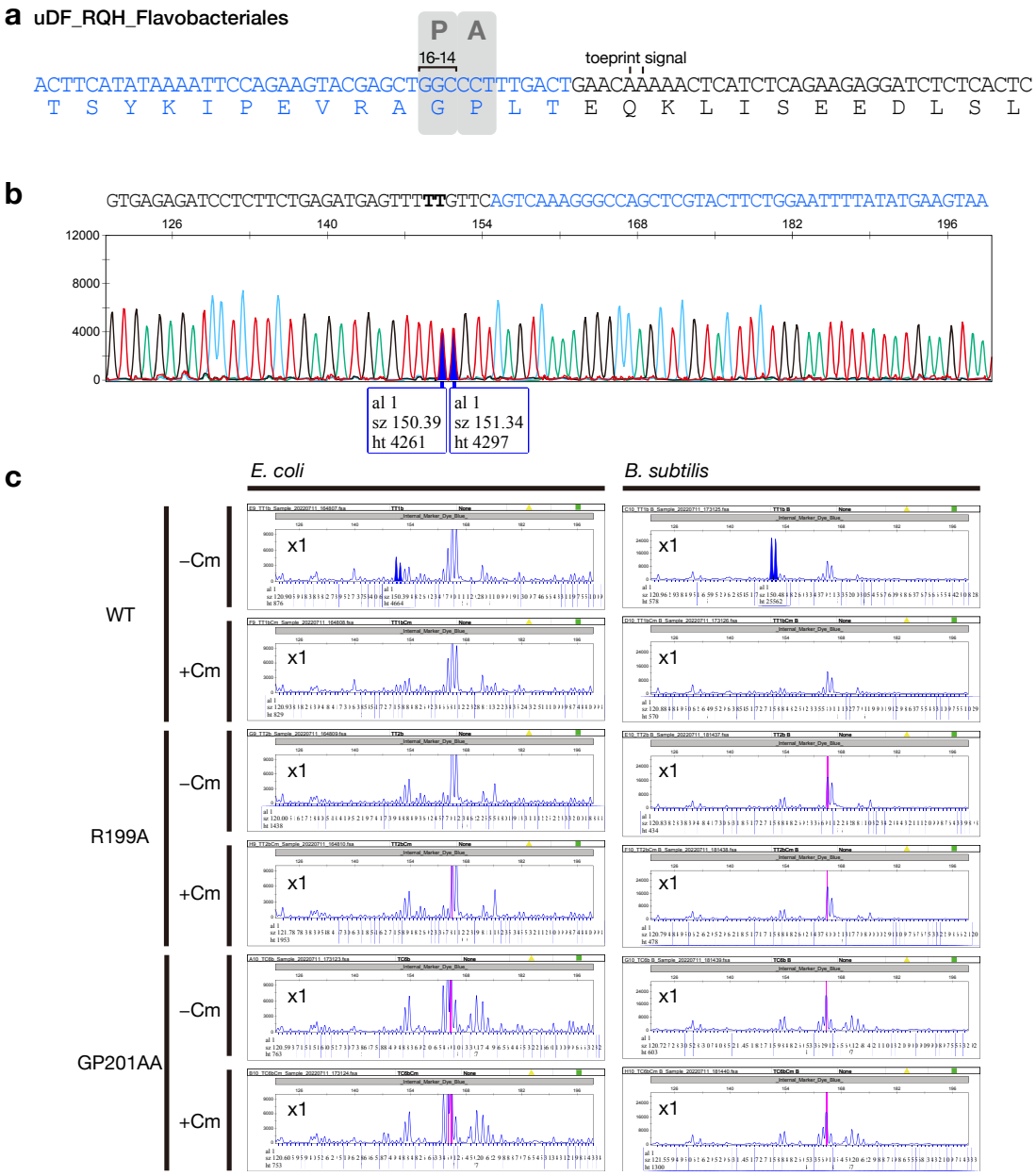
Supplementary Figure 21. Identification of the ribosome stalling site of uA_RQH_Myxococcales by toeprinting. a A schematic representation of the ribosome stalling site of uA_RQH_Myxococcales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



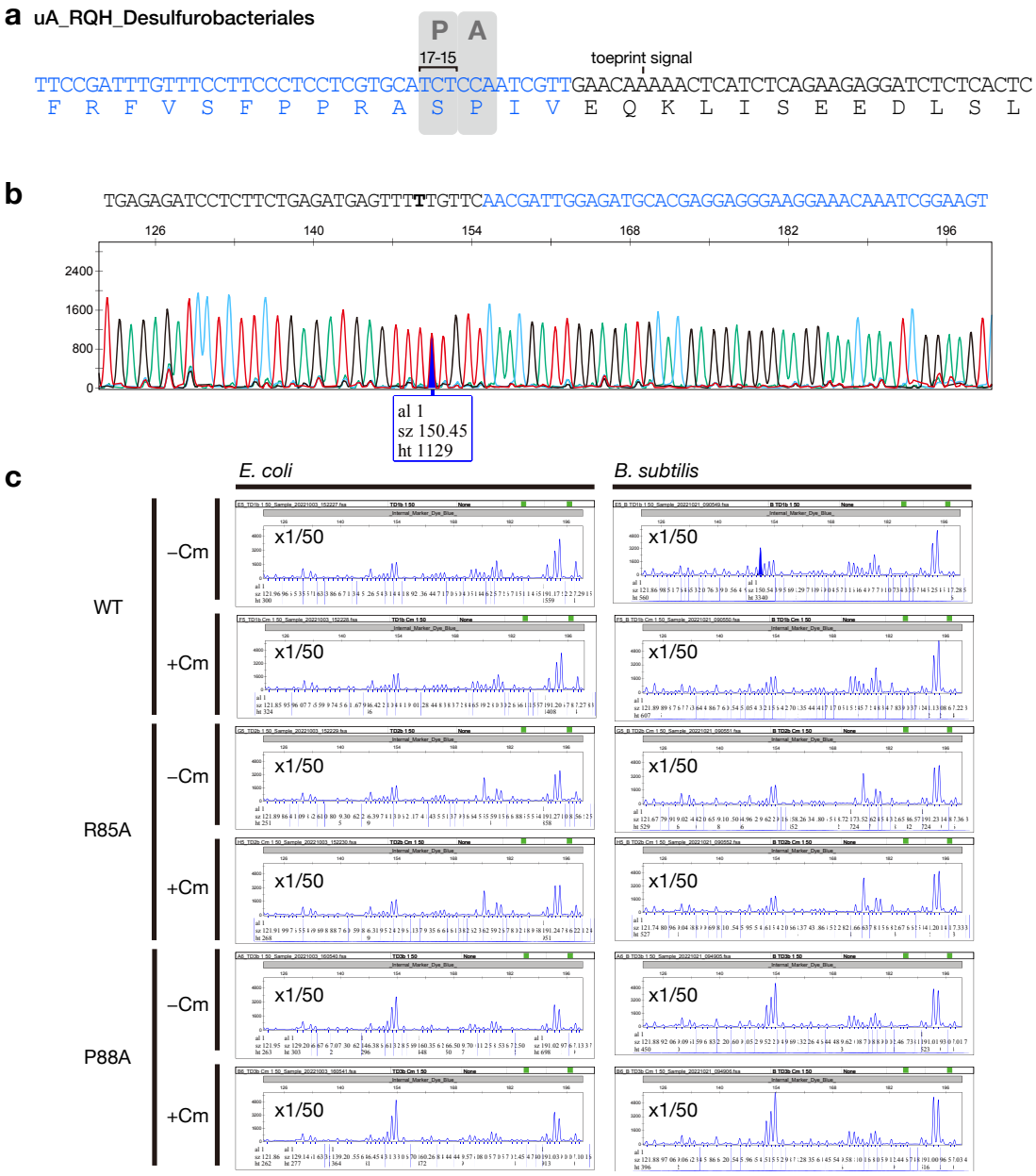
Supplementary Figure 22. Identification of the ribosome stalling site of uA_RQH_Bacteroidales_1 by toeprinting. a A schematic representation of the ribosome stalling site of uA_RQH_Bacteroidales_1 estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



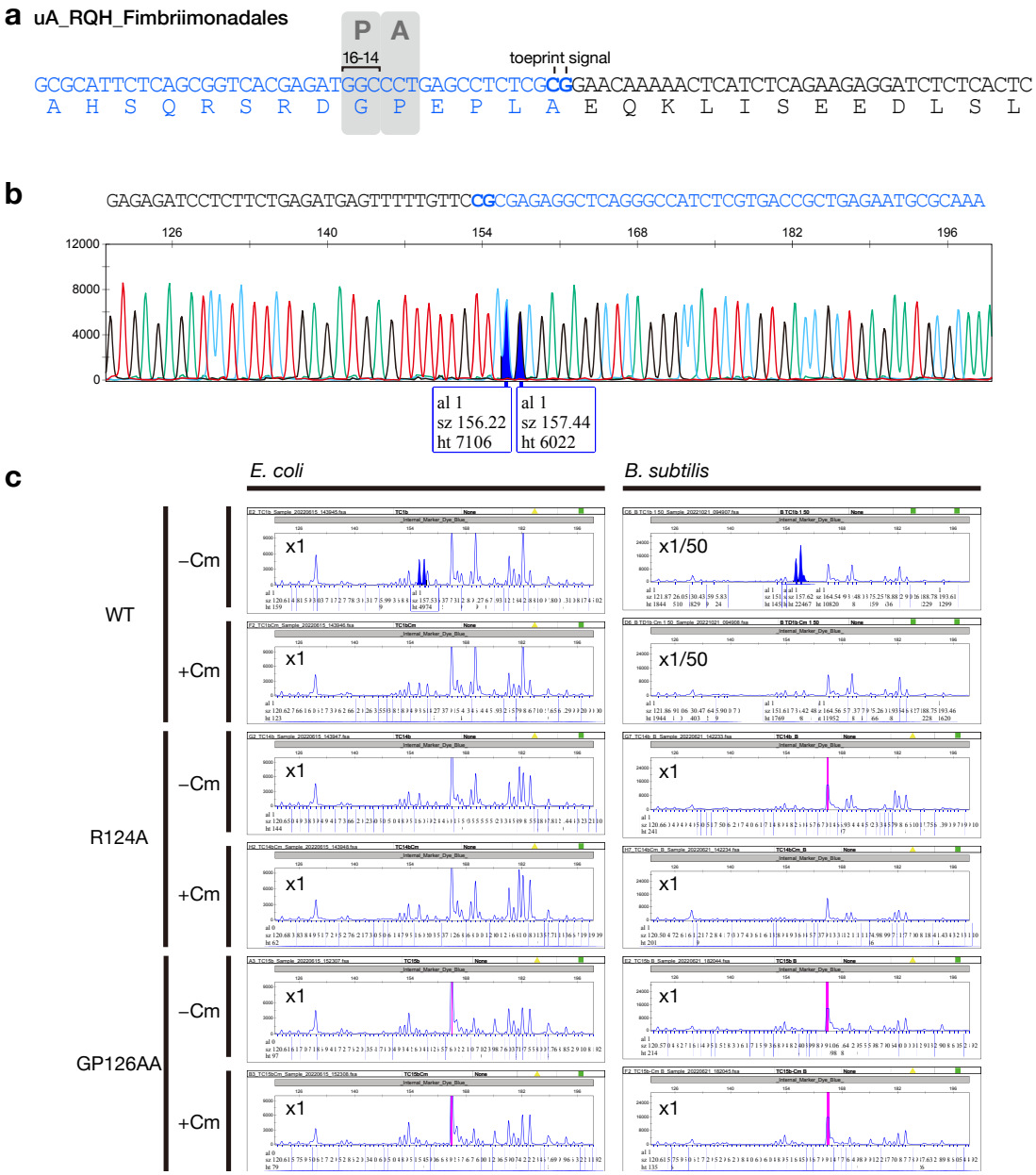
Supplementary Figure 23. Identification of the ribosome stalling site of uA_RQH_Bacteroidales_2 by toeprinting. a A schematic representation of the ribosome stalling site of uA_RQH_Bacteroidales_2 estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



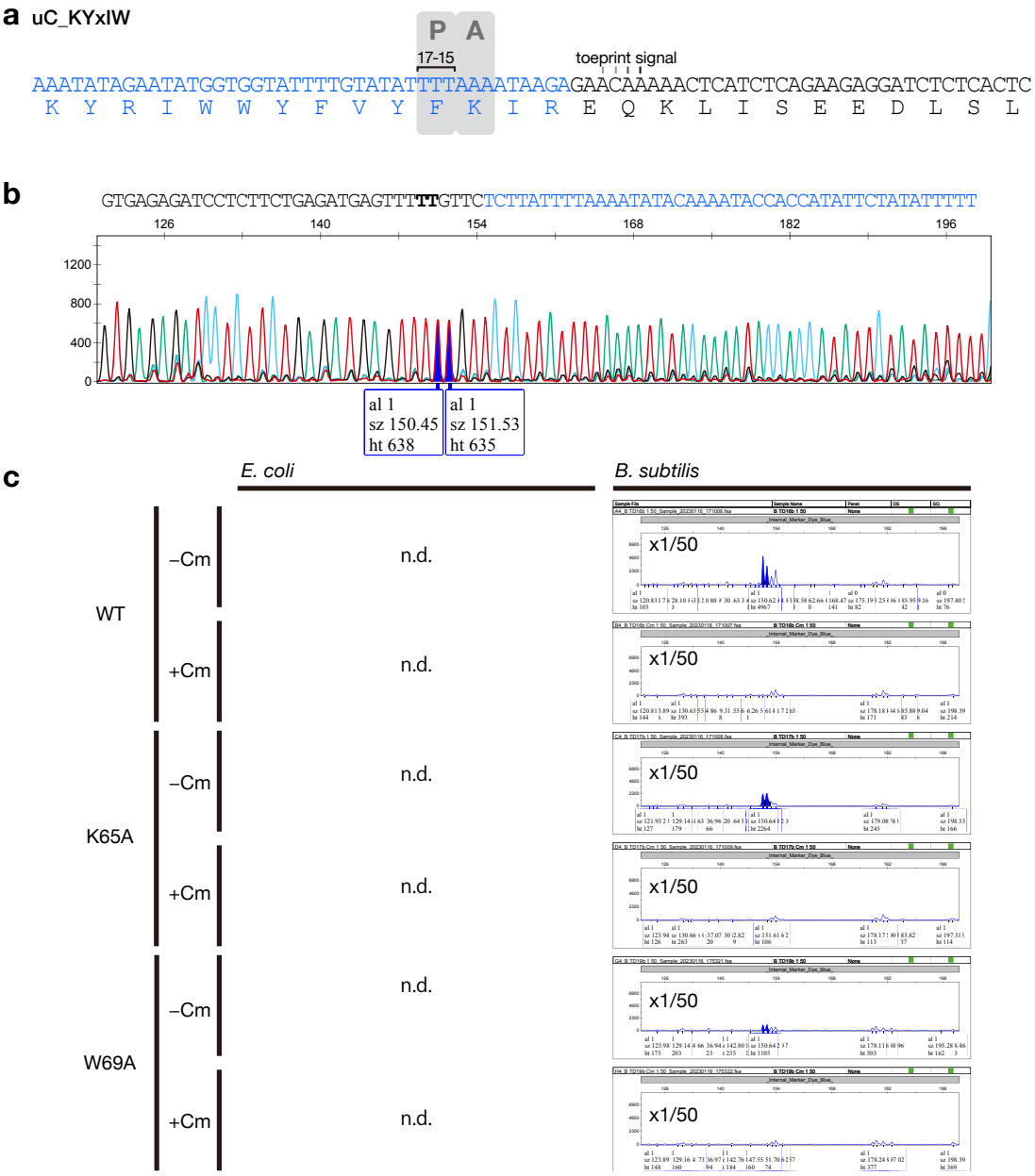
Supplementary Figure 24. Identification of the ribosome stalling site of uDF_RQH_Flavobacteriales by toeprinting. a A schematic representation of the ribosome stalling site of uDF_RQH_Flavobacteriales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



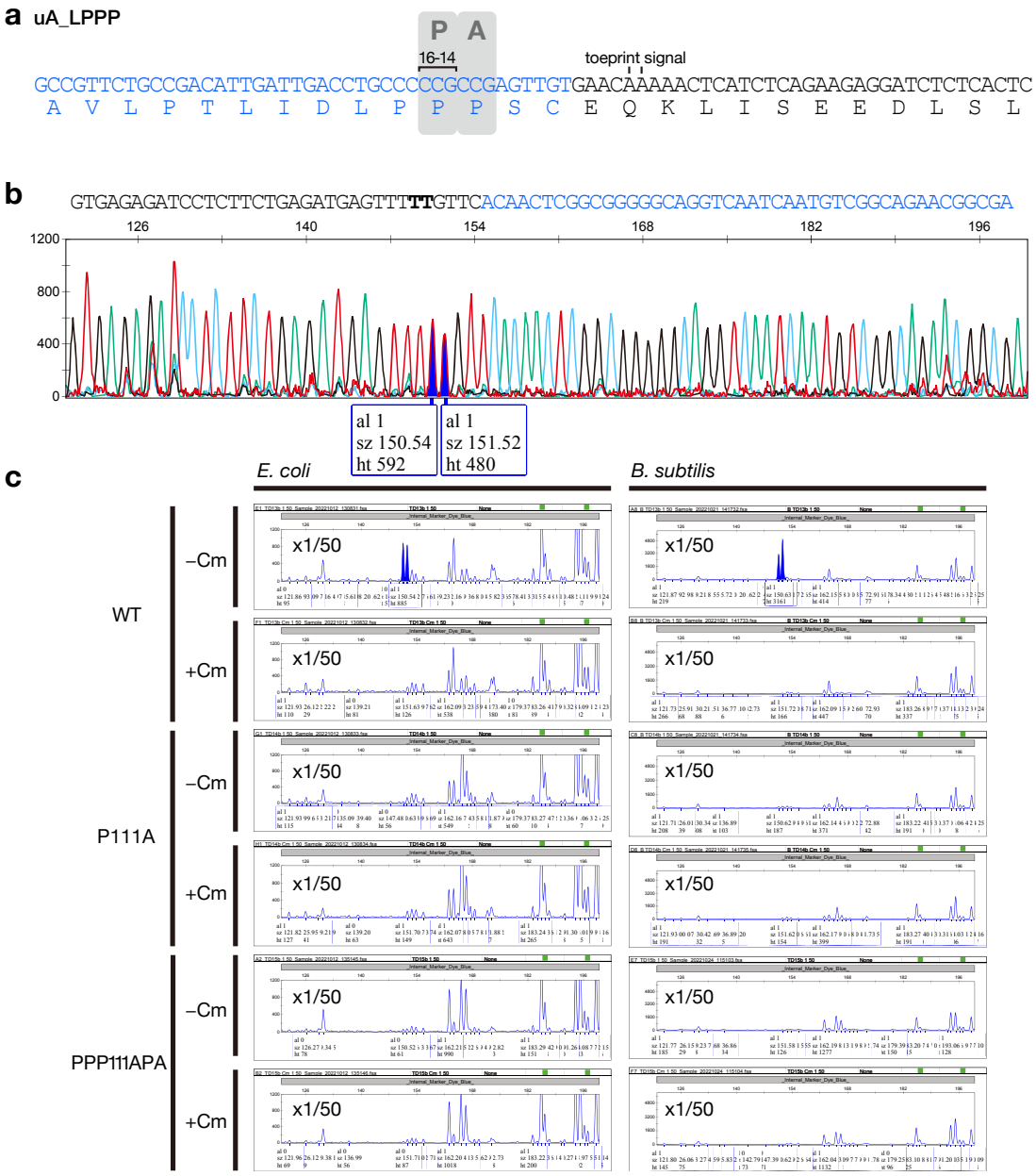
Supplementary Figure 25. Identification of the ribosome stalling site of uA_RQH_Desulfurobacteriales by toeprinting. a A schematic representation of the ribosome stalling site of uA_RQH_Desulfurobacteriales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



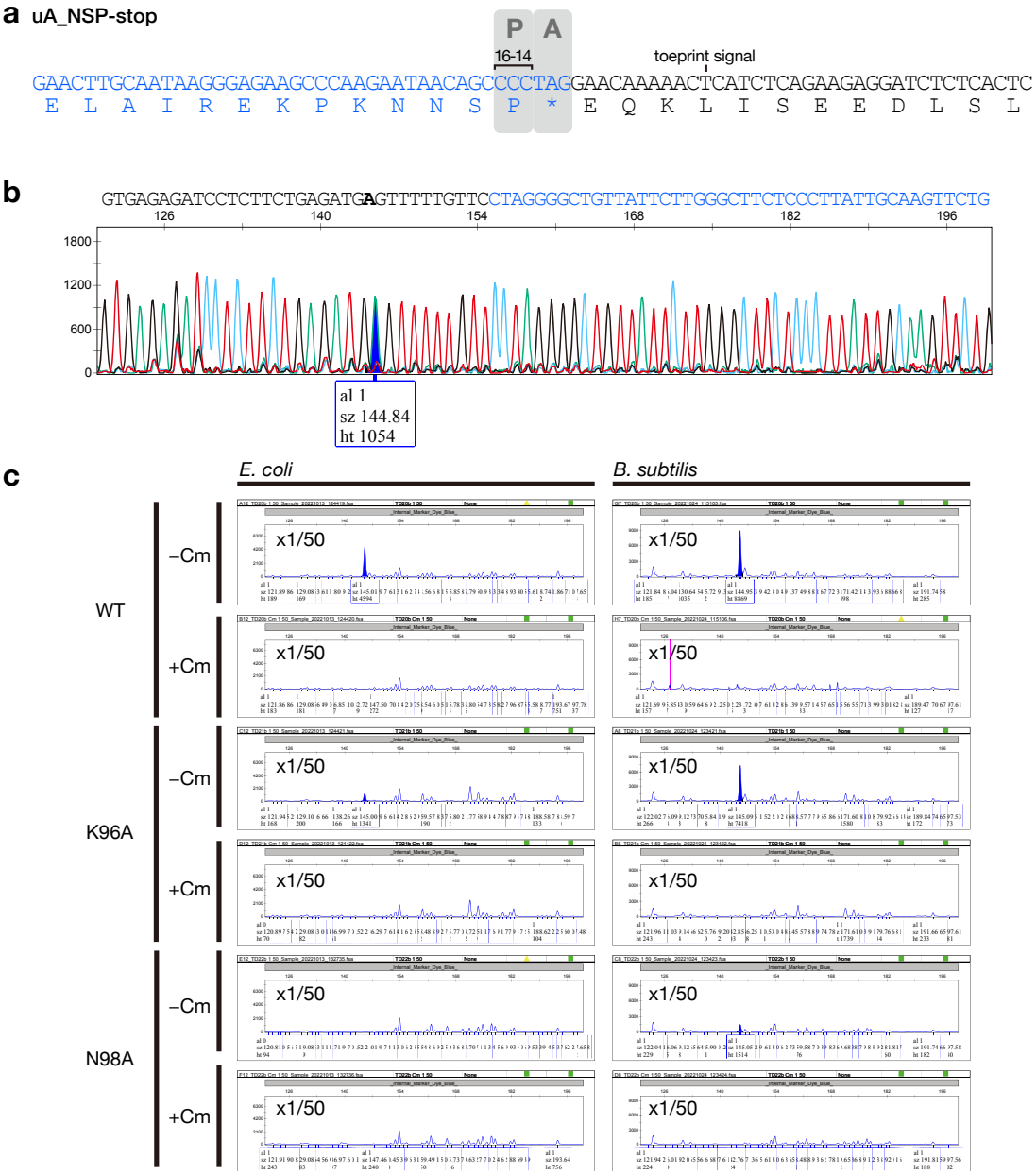
Supplementary Figure 26. Identification of the ribosome stalling site of uA_RQH_Fimbriimonadales by toeprinting. a A schematic representation of the ribosome stalling site of uA_RQH_Fimbriimonadales estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



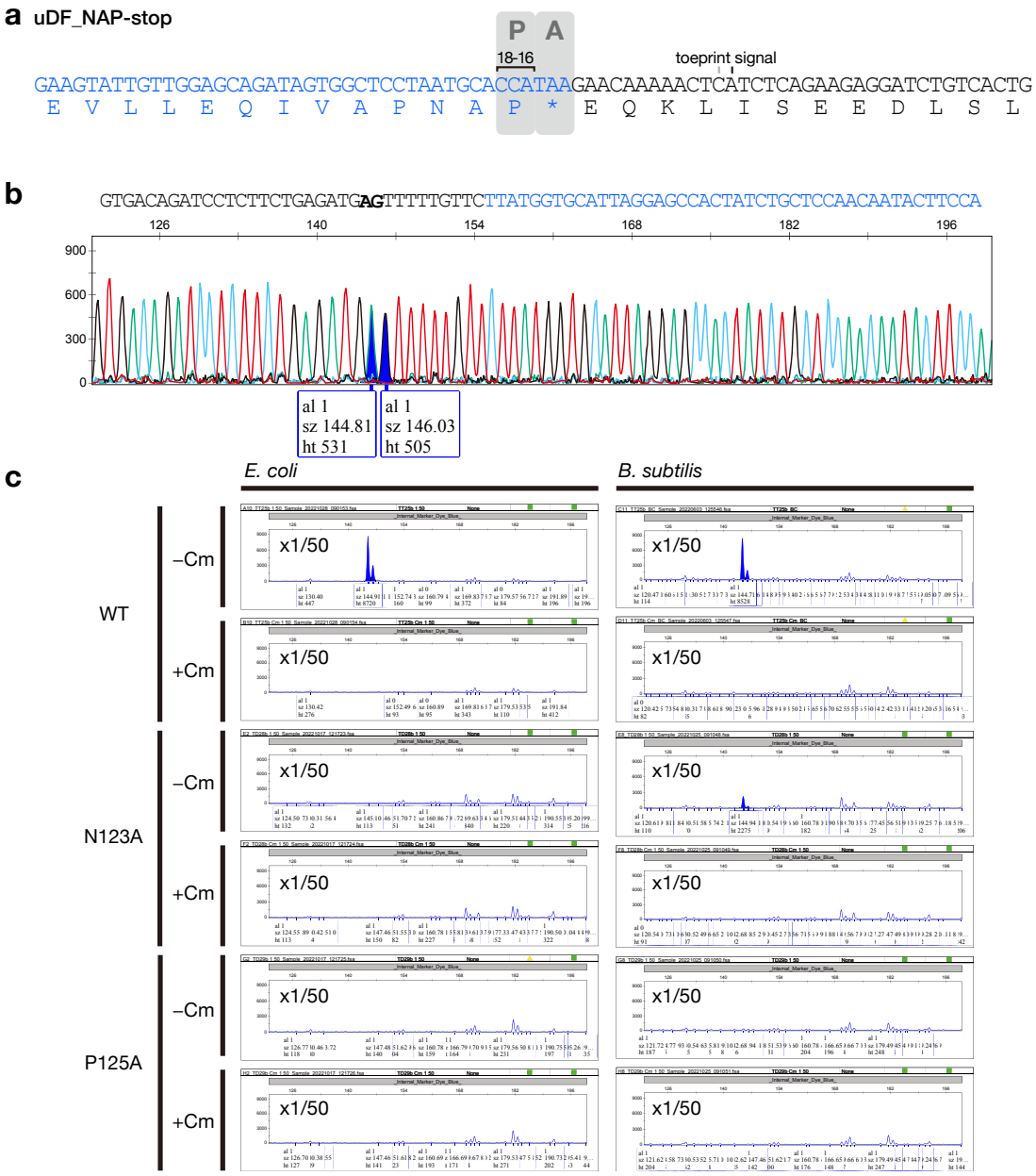
Supplementary Figure 27. Identification of the ribosome stalling site of uC_KYxIW by toeprinting. a A schematic representation of the ribosome stalling site of uC_KYxIW estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



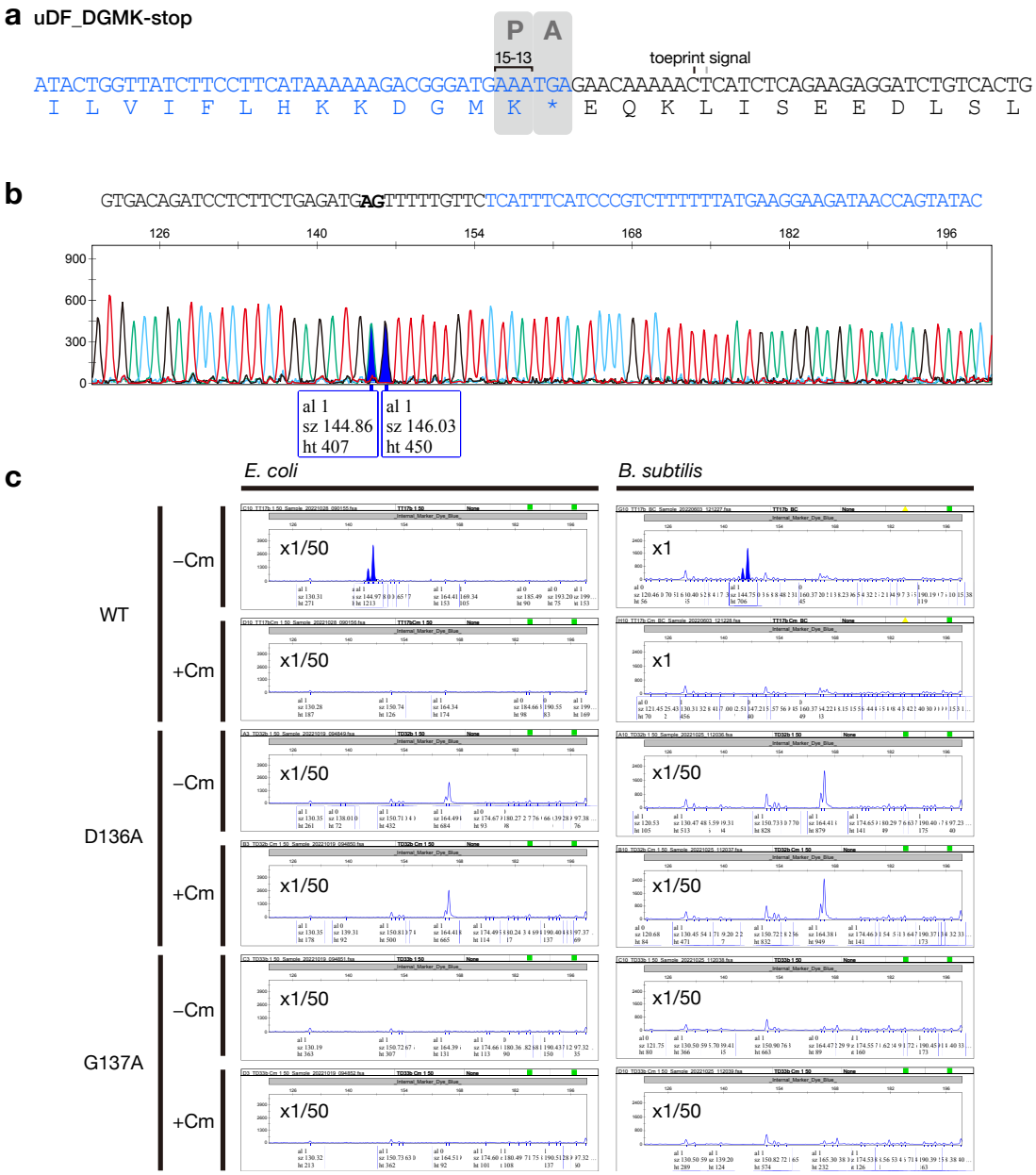
Supplementary Figure 28. Identification of the ribosome stalling site of uA_LPPP by toeprinting. a A schematic representation of the ribosome stalling site of uA_LPPP estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



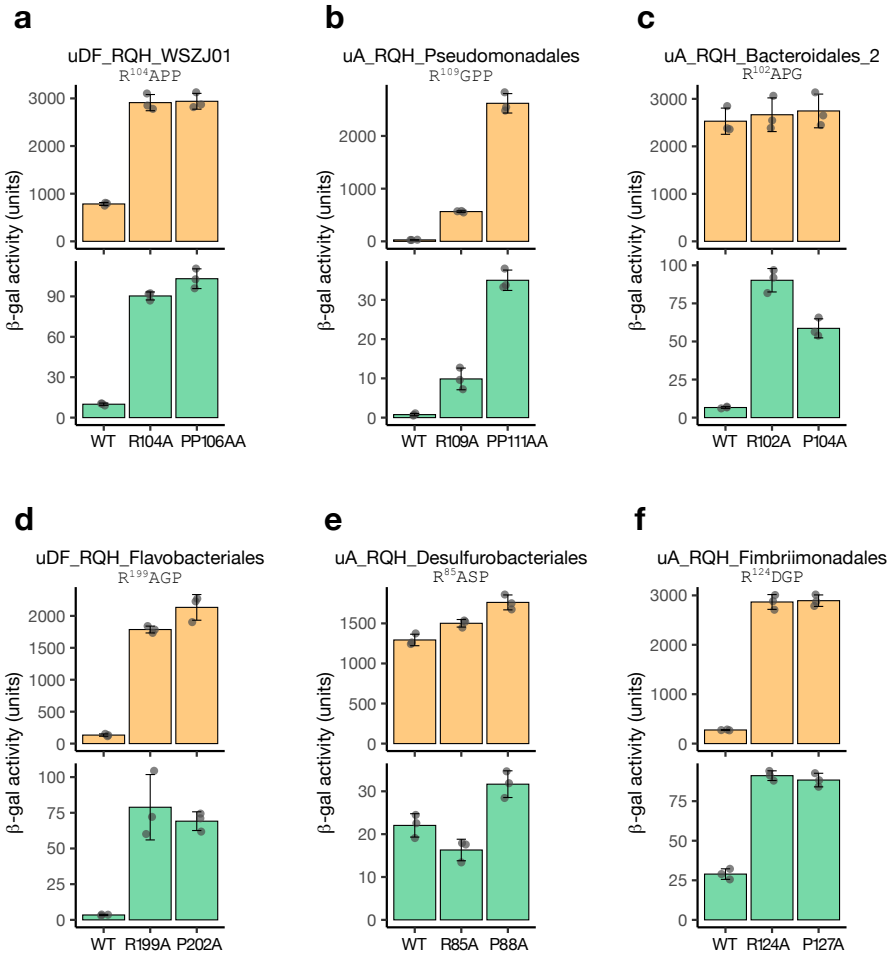
Supplementary Figure 29. Identification of the ribosome stalling site of uA_NSP-stop by toeprinting. a A schematic representation of the ribosome stalling site of uA_NSP-stop estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



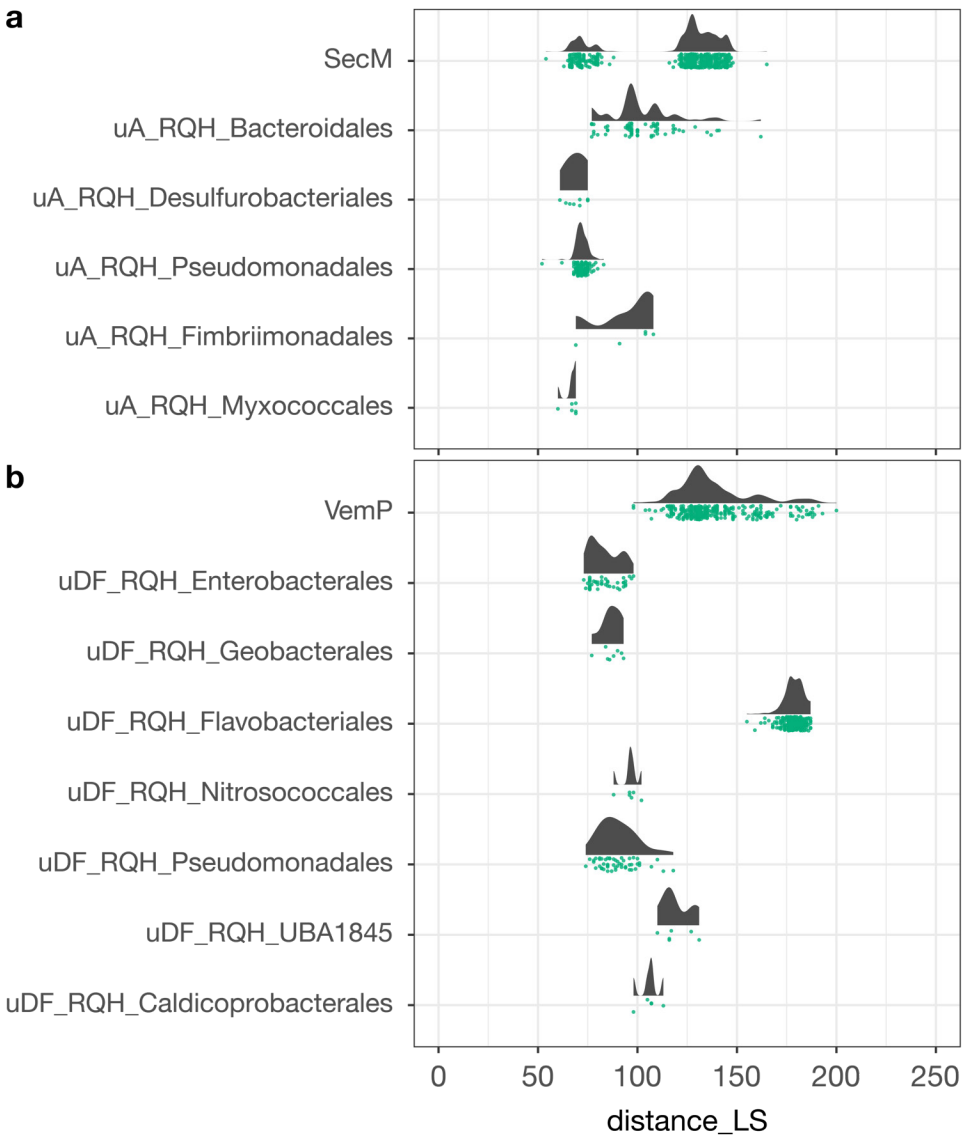
Supplementary Figure 30. Identification of the ribosome stalling site of uDF_NAP-stop by toeprinting. a A schematic representation of the ribosome stalling site of uDF_NAP-stop estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



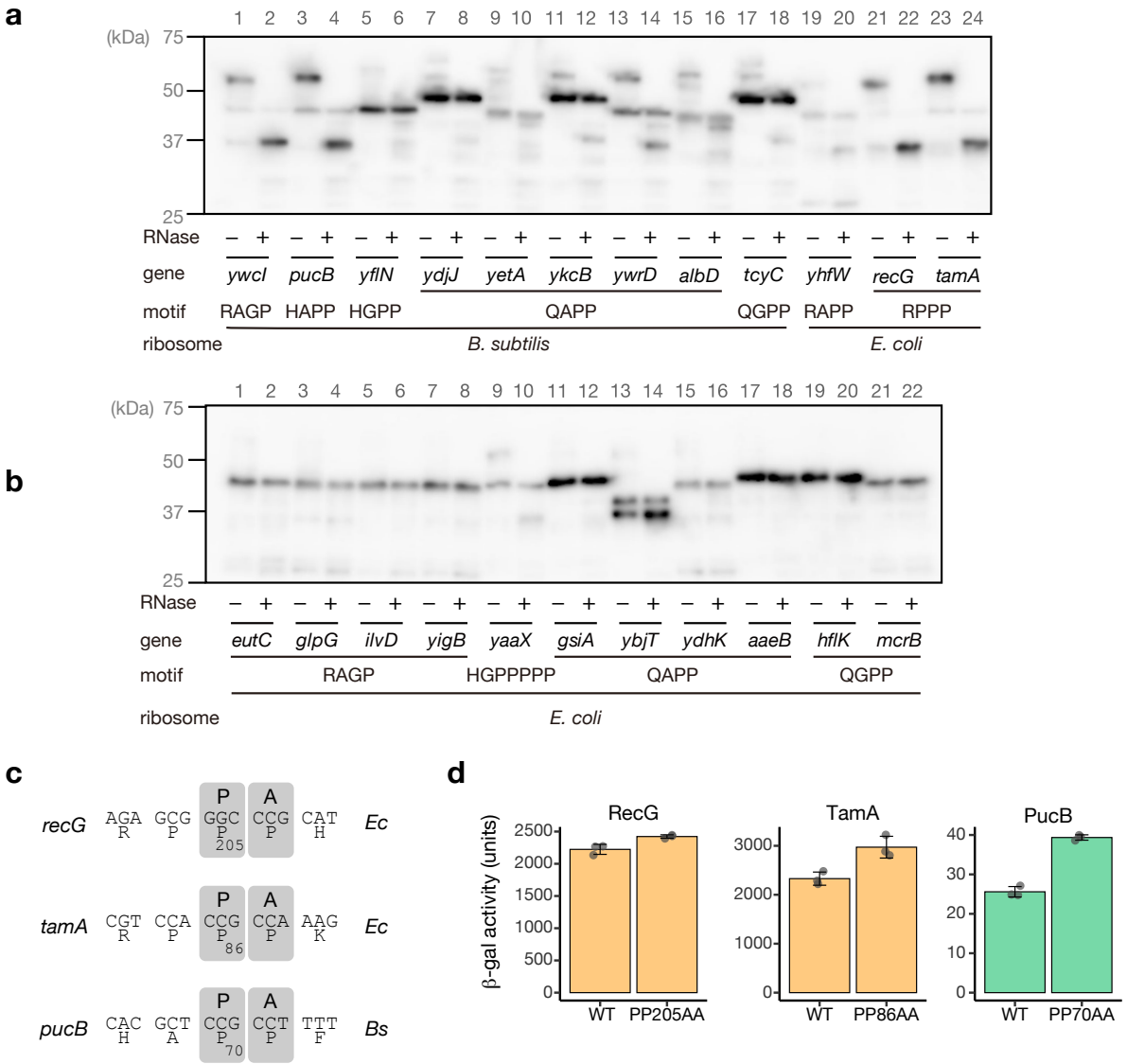
Supplementary Figure 31. Identification of the ribosome stalling site of uDF_DGMK-stop by toeprinting. **a** A schematic representation of the ribosome stalling site of uDF_DGMK-stop estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peaks corresponding to the stalling-specific toeprint signals and their nucleotides in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.



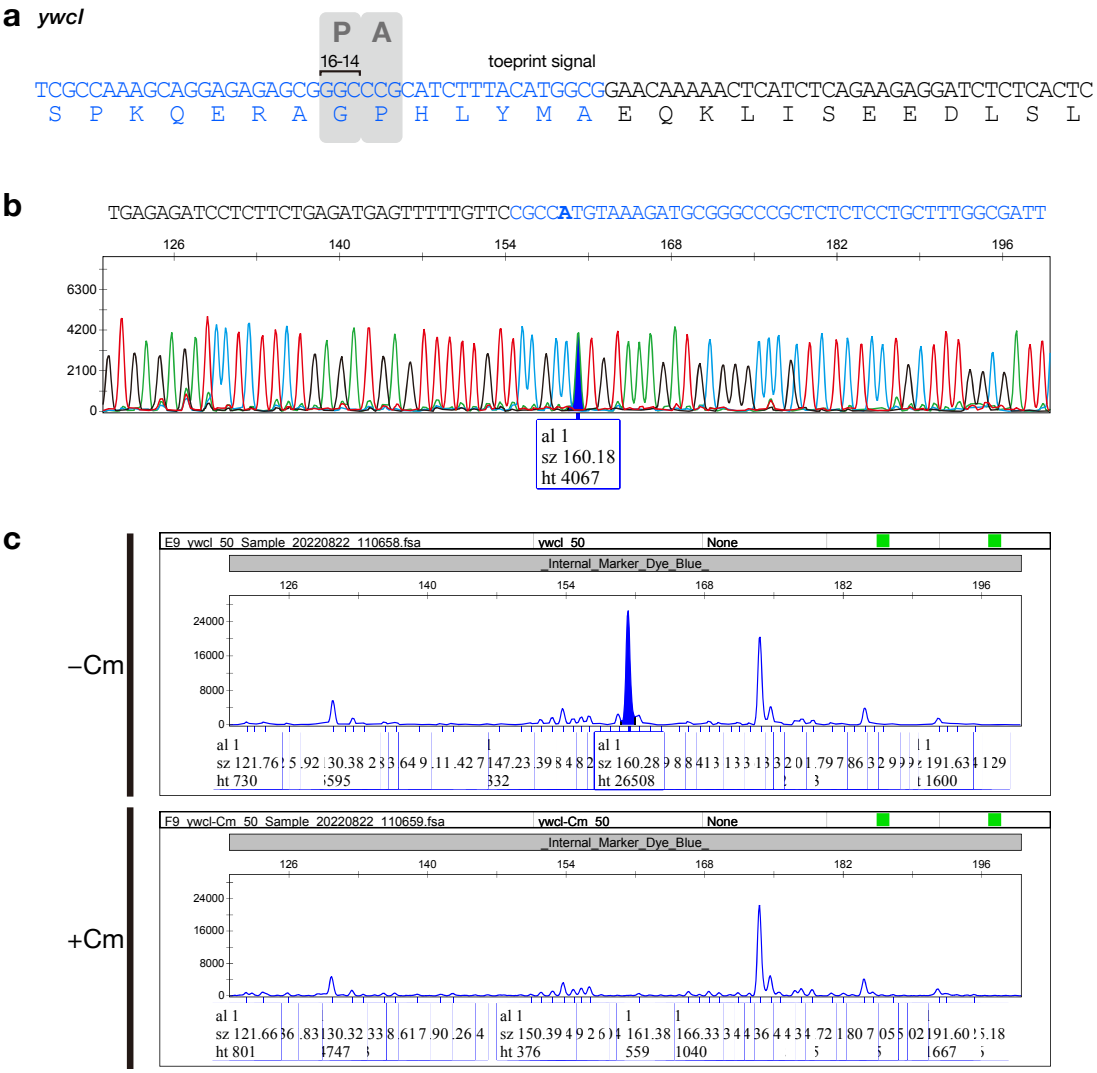
Supplementary Figure 32. In vivo translation arrest assay of candidate monitoring substrates. **a-f** the β-galactosidase activity of *E. coli* (orange bars) and *B. subtilis* cells (green bars), harboring wild-type (WT) or mutant derivatives of arrest peptide reporters. Error bars (mean ± standard deviations, n = 3 biologically independent cell cultures) and individual data points (dots) are indicated. Source data are provided as a Source Data file.



Supplementary Figure 33. Distances between the arrest sites and the localization signal of RQH family. a, b Raincloud plots of the distance_LS (see Fig. 6d) of RQH families found upstream of *secA* (a) and *secDF* (b). RQH families that contains at least 5 members were plotted. Source data are provided as a Source Data file.



Supplementary Figure 34. Investigation for translation arrest by amino acid sequences harboring RAPP-like motifs derived from *E. coli* and *B. subtilis* proteome. The genes tested are listed in Supplementary Data 5. **a, b** Western blotting of in vitro translation products. The *gfp-ap-lacZα* reporters were translated in *Ec* or *Bs* PURE. The translation products were then separated by neutral pH gels and immunoblotted using anti-GFP or anti-LacZα. Before the separation, portions of samples were treated by RNase A (lanes indicated as +) to degrade tRNA moiety. Experiments were conducted twice independently, with similar results. **c** Estimated ribosome stalling sites based on the results obtained by toeprinting analysis. **d** In vivo translation arrest assay. β-galactosidase activity of *E. coli* (orange bars) and *B. subtilis* cells (green bars), harboring wild-type (WT) or mutant derivatives of arrest peptide reporters. Error bars (mean ± standard deviations, n = 3 biologically independent cell cultures) and individual data points (dots) are indicated. Source data are provided as a Source Data file.



Supplementary Figure 35. Identification of the ribosome stalling site of *ywcI* by toeprinting. a A schematic representation of the ribosome stalling site of *ywcI* estimated by toeprinting. **b** Overlayed results of dideoxy sequencing. The peak corresponding to the stalling-specific toeprint signal and its nucleotide in the dideoxy sequencing data are shown by the filled blue peak and the bold alphabet, respectively. **c** Raw results of toeprinting analysis of wild-type and arrest-defective mutant derivatives. Detailed information is described in the legend of Supplementary Fig. 16.