

Additional materials to the article

Shavkunov K.S., Markelova N.Yu., Alikina O.V., Glazunova O.A., Panyukov V.V., Kolzhetsov N.P., Kiselev S.S., Ozoline O.N.. *Mathematical Biology and Bioinformatics*. 2024;19(2):453-471.
doi: [10.17537/2024.19.453](https://doi.org/10.17537/2024.19.453)

Supplementary Figures

a) Fragments of AspUVT- tRNA		Number of reads	
		correct	errors
GGAGCGGTAGTTTCAGTCGGTTAGAATACC		58	
GGAGCGGTAGTTTCAGTCGGTTAGAATACA			1
GGAGCGGTAGTTTCAGTCGGTTAGAATAC T			1
GGAGCGGTAGTTTCAGTCGGTTAGATATAT			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCA			33
GGAGCGGTAGTTTCAGTCGGTTAGAATACCC			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCT		47	
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTA			3
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTG		49	
GGAGCGGTAGTTTCAGTCGGTTAGAATACCT ACT			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGA			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGC			4
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTG C		603	
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGG			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGT			6
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTACA			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTACC			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTCCA			1
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGCA			59
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGCC		1009	
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGCG			4
GGAGCGGTAGTTTCAGTCGGTTAGAATACCTGCT			31
b) Products from <i>fimA-fimI</i> spacer		Number of reads	
		correct	errors
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCC		562	
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCG			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACTA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCCTCCA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCAA			3
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCCA			3
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			4
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA		145	
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			60
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			13
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			89
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA		739	
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			119
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			2
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			154
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			46747
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			923
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			64
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA		46441	
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			32
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			6
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			6
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			16
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			6
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			2
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			1
GGTTCAGGGACGTCATTACGGGCAGGGATGCCACCCAA			2
c) Fragments of AlaVUT- tRNA		Number of reads	
		correct	errors
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGC		1154	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGC			961
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGC			46
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCT		563	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCT			2
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCT			462
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTT		1351	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTT			1115
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTT			15
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTT			1
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTT		7539	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTT			5948
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTT			5
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTT			91
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTT			2
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTG		1943	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTG			1149
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTG			29
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTG		324	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTG			238
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTG			3
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTGCA		70	
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTGCA			64
GGGGCTATAGCTCAGCTGGGAGAGCGCCTGCTTTGCA			1
d) Fragments of AlaXW- tRNA		Number of reads	
		correct	errors
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC		454	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC			6
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC			1
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC		1858	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC			1725
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC			54
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGC			2
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCAT		330	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCAT			267
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCAT			7
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCAT			1
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATG		95	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATG			88
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATG			2
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGG		34	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGG			27
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGC		54	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGC			76
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGCA		116	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGCA			125
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGCA			2
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGCAT		33	
GGGGCTATAGCTCAGCTGGGAGAGCGCTTGCATGGCAT			17

Fig. S1. Examples of random (a) and non-random (b-d) presence of non-template extensions (marked in red) at the ends of correct oligonucleotides (bolded). a) AspUVT-tRNA fragments illustrate the median (13.8%) level of non-template additions. b-d) Non-template extensions reflecting the outbreaks in the percentage of misreads in Fig. 2a. b) 3'-terminal extensions in oligonucleotides transcribed from the *fimA-fimI* intergenic region, demonstrating the scale and frequency of adenylation, polyadenylation and polynucleotidylation of bacterial RNAs. c) and d) Examples of similar 5'-terminal extensions in oligonucleotides derived from the *alaVUT* and *alaXW* genes. The indicated numbers of correct (bolded) and erroneous oligonucleotides correspond to Eco_in_2 dataset. The number of modified oligonucleotides is indicated in the column highlighted in orange. To simplify their comparison with genomic sequences, detected oligonucleotides are printed in the DNA code (ACGT).

a) Fragments of SokC sRNA

GTTCAGC**ATATAGGAGGCCTCGGGTTGATGGTAAAAATACACTCGGGGCTTTTCTCTA**
 ATATAGGAGGCCTCGGGTTGATGGTAAAA**AT**ATCACT
 ATATAGGAGGCCTCGGGTTGATGGTAAAA**AT**ATCACTCGG
 ATATAGGAGGCCTCGGGTTGATGGTAAAA**AT**ATCACTCGGG
GTTC-GCATATAGGAGGCCTCGGGTTGATGGTAAAAATACACTCGGG
 ATATAGGAGGCCTCGGGTTGATGGTAAAA**AT**ATCACTCGGGGCTTTTCT
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGG-CTTTTCTC
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGGGCTTTTCT
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCA-TCGGGGCTTTTCTC
 ATATAGGAGGCCTCGGGTTGATGGTAAAA**AT**ATCACTCGGGGCTTTTCTC
 ATATAGGAGGCCTCGGGTTGATGGTAAAA**AT**ATCACTCGGGGCTTTTCT
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGG-CTTT-CTCT
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGG-CTTTTCTCT
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGGGCTTT-CTCT
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGGGCTTTTCTC
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGGGCTTTTCTC
 ATATAGGAGGCCTCGGGTTGATGGTAAAAATATCACTCGGG-CTTT-CTCTA

b) Fragments of RdlBC sRNAs

TGTCAGGTTTTACCTCTCAACGTGCGGGGGTTTTCTCT	2674
GTCAGGTTTTACCTCTCAACGTGCGGGG- TTTT C	1
GTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTC	2
GTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTC	8
GTCAGGTTTTACCTCTCAACGTGCGGGG- TTTT CT	9
GTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCT	247
TGTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCT	2
GTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCT	7
TGTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCT	2
GTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCT	46
GTCAGGTTTTACCTCTCAACGTGCGGGGG GGT TTTTCT	2
GTCAGGTTTTACCTCTCAA AC GTGCGGGGG GGT TTTTCT	1
GTCAGGTTTTACCTC --AAC GTGCGGGGGTTTTCTC	2
GTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCTC	2
GTCAGGTTTTACCTCTCAA AC GTGCGGGGG GT TTTTCTC	1
TGTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCTC	1
GTCAGGTTTTACCTCTCAACGTGCGGGGG GGT TTTTCTC	1
TGTCAGGTTTTACCTCTCAACGTGCGGGGG GT TTTTCTCT	1

Fig. S2. Examples of typical (a) and biased (b) distribution of InDels illustrated by fragments with average (2.4%) and excessive (13.58%) percentages of misreads in Eco_in_2 dataset. In both cases, changes were associated with homonucleotide tracts and only the same oligonucleotides were incorporated. However, in fragments derived from RdlBC sRNAs (b) two oligonucleotides were significantly overrepresented (247 and 46 copies), of which one has double insertions of G and T in the junction point between two tracts. In line with similar percentage of InDels in two Eco_in experiments (Fig. 2), the other set with internal oligonucleotides (Eco_in_1) also contained the same misreads in comparable numbers (130 and 18 reads, respectively). Genomic and anchor sequences are shown in bold and green bold, respectively. To simplify their comparison with oligonucleotides, the latter are printed in the DNA code (ACGT). InDels are marked in blue (a) and red (b). Numbers of correct and erroneous (highlighted in blue) oligonucleotides are indicated on the right.

CGGGCGTAGTTC AATGGTAGAACGAGAGCTTCCAAGCTCTATACGAG

CGGGCGTAGTTCAACGGTAGAACG
CGGGCGTAGTTCAATGGTAGGACG
CGGGCGTAGTTCAATGGCAGAACGAG
CGGGCGTAGTTCAATGGTAGAGCGGAGA
CGGGCGTAGTTCAATGGTGGAAACGAGA
CGGGCGTAGTTCAATGGTAGAACGGGAGC
CGGGCGTAGTTCAATGGTAAGACGAGAGCTT
CGGGCGTAGTTCAATGGTAGAACGAGAACTTCC
CGGGCGTAGTTCAATGGTAGAACGAGAGTTTCC
CGGGCGTAGTTCAATGGTAGAACGAGGGCTTCC
CGGGCGTAGTTCAATGGTAGAATGAGAGCTTCC
CGGGCGTAGTTCAATGGTAGAGCGGAGAGCTTCC
CGGGCGTAGTTCAATGATAGAACGAGAGCTTCCC
CGGGCGTAGTTCAATGGTAGAACAGAGCTTCCC
CGGGCGTAGTTCAATGGTAGAACGAGGGCTTCCC
CGGGCGTAGTTCAATGGTAGAATGAGAGCTTCCC
CGGGCGTAGTTCAATGGTAGAGCGAGAGCTTCCC
CGGGCGTAGTTCAATGGTGAACGAGAGCTTCCC
CGGGCGTAGTTCAATGGTAGAACGAGAGCTTCCCAAGC
CGGGCGTAGTTCAATGCTAGAACGAGAGCTTCCCAGACT
CGGGCGTAGTTCAATGGTAGAACGAGGGCTTCCCAGCT
CGGGCGTAGTTCAATGGTAGAACGAGAGCTTCCCAACTATACGAG

b) Fragments of TrpT- tRNA

Sequence	Count
CAAAACCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	5089
GTGGGAGTTCGAGTCTCCGCCCCAGCCA	36
TGTTGGGAGTTCGAGTCTCCGCCCCAGCCA	5
GTGGTGGGAGTTCGAGTCTCCGCCCCAGCCA	1
GGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	5
GGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	36
CGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	20
CCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	1
ACCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	5
AACCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	1
AAACCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	3
AAAAACCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	1
CAAAACCGGGTGTGGGAGTTCGAGTCTCCGCCCCAGCCA	2

c) Fragments of TrpT- tRNA

CAAAACCGGGT	GTTGGGAGTT	CGAGTCTCT	CCGCCCT	GC	5089
	GTTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	147
	TGTTGGGAGTT	CGAGTCTCT	CCGCCCT	GACCA	3
	TGTTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	36
	GTGTTGGGAGTT	CGAGTCTCT	CCGCCCT	GACCA	2
	GTGTTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	12
	GGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	4
	GGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	12
	GGGTGTTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	120
	CGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	GACCA	1
	CGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	1
	CGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	16
	CGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	40
	CGGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	3
	CCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	5
	ACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	2
	ACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	16
	AACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	1
	AACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	6
	AAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	6
	AAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	7
	AAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	11
	CAAAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	1
	CCAAAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	ACCA	1
	CTCCAAAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	2
	CTCCAAAACCGGGTGTGGGAGTT	CGAGTCTCT	CCGCCCT	CACCA	1

Fig. S3. Examples of random (**a**) and biased (**b**, **c**) distribution of mismatches (marked in red) in fragments with median (GlyU-tRNA in **a**) and excessive (TrpT-tRNA in **b**, **c**) percentages of misreads. (**b**) and (**c**) show different substitutions in the 3'-terminal CCCUGCCA motif. Genomic and anchor sequences are shown in bold and green bold, respectively. To simplify their comparison with oligonucleotides, the latter were printed in the DNA code (ACGT). Numbers of correct and erroneous oligonucleotides are indicated on the right (highlighted in light-green).

