Практикум 13

Задание 1.

	ATG	ATT	GAG	CTG	GTG	TTG
Escherichia coli	3890	4	1	2	338	80
Candidatus Gracilibacteria bacterium	1129	2	0	0	41	23
Mycoplasma pneumoniae	629	19	0	2	64	55

Как можно видеть, наиболее распространённым старт-кодоном всё же является ATG, это связано со слабым отбором в сторону данного старт-кодона [1], так при этом условии эффективность инициации транскрипции намного выше, чем при использовании неканонических стартовых кодонов [2], но всё доля генов, использующих в качестве старт-кодона не ATG, достаточно велика, это можно связать с тем, что некоторые замены в старт кодонах компенсируются заменами в последовательности Шайна — Дальгарно, благодаря чему эффективность инициации транскрипции может вновь выходить на необходимый уровень. Также можно предположить, что в гене могут присутствовать несколько старт-кодонов (например, 2), один из которых может быть альтернативным и при разных условиях с одного гена могут синтезироваться белки с разными N-концевыми доменами [3].

Ссылка на код: [СтартКодоны1]

Задание 2.

В результате работы кода обнаружилось 4 таких полседрвательноти, где стоп-кодон находится не только в конце.

1) lcl|U00096.3_cds_b4587_250 [gene=insN] [locus_tag=b4587] [db_xref=ASAP:ABE-0285253,ECOCYC:G6130] [protein=IS911A regulator fragment] [pseudo=true] [location=join(270278..270540,271764..272190)] [gbkey=CDS]

2) lcl|U00096.3_cds_AAD13438.1_1459 [gene=fdnG] [locus_tag=b1474] [db_xref=UniProtKB/Swiss-Prot:P24183] [protein=formate dehydrogenase N subunit alpha] [transl_except=(pos:586..588,aa:Sec)] [protein_id=AAD13438.1] [location=1547401..1550448] [gbkey=CDS]

3) lcl|U00096.3_cds_AAD13456.1_3824 [gene=fdoG] [locus_tag=b3894] [db_xref=UniProtKB/Swiss-Prot:P32176] [protein=formate dehydrogenase O subunit alpha] [transl_except=(pos:586..588,aa:Sec)] [protein_id=AAD13456.1] [location=complement(4082772..4085822)] [gbkey=CDS]

4) lcl|U00096.3_cds_AAD13462.1_3997 [gene=fdhF] [locus_tag=b4079] [db_xref=UniProtKB/Swiss-Prot:P07658] [protein=formate dehydrogenase H] [transl_except=(pos:418..420,aa:Sec)] [protein_id=AAD13462.1] [location=complement(4297219..4299366)] [gbkey=CDS]

Первая последовательность является псевдогеном, в их полседовательности могут возникать самые различные мутации, поэтому неудивительно, что там возник стоп-кодон. Последующие 3 белка являются субъединицами формиатдегидрогеназы, в активном центре которой имеются остатки неканонической аминокислоты — селеноцистеина, сама аминокислота кодируется антикодоном TGA (который обычно является стоп-кодоном), но в данном случае благодаря определенным факторам транскрипции, при встрече TGA в последовательности мРНК к белку присоединяется селеноцистеин [4].

Ссылка на код: ∞ СтранныеСтопКодоны2

Задание 3.

	TGA	TAG	TAA
Escherichia coli	1246	306	2761
Candidatus Gracilibacteria bacterium	1	188	**1000**
Mycoplasma pneumoniae	0	221	533

У второй бактерии всего один кодон TGA читается как стоп-кодон, это можно объяснить тем, что у данной бактерии альтернативный генетический код, поэтому и стоп-кодоны другие, в данном случае TGA кодирует аминокислоту глицин [5].
У третьей бактерии вообще нет ни одного TGA в качестве стоп-кодона. Причины те же, что и у второй бактерии — альтернативный генетический код. Только на этот раз TGA кодирует аминокислоту триптофан (как видно из таблицы ниже) [6].

Таблица, в которой сравниваются генетические коды двух бактерий рода *Mycoplasma.*

Nucleic Acids Research, 1993, Vol. 21, No. 21 **4971**

Table 1. Codon Usage of *M.pneumoniae* and *M.capricolum*

	M.p.	M.c.		M.p.	M.c.		M.p.	M.c.		M.p.	M.c.
Phe(UUU)	86	269	Ser(UCU)	20	107	Tyr(UAU)	33	171	Cys(UCU)	7	41
Phe(UUC)	39	28	Ser(UCC)	67	1	Tyr(UAC)	52	24	Cys(UGC)	3	10
Leu(UUA)	68	433	Ser(UCA)	26	169	Stop(UAA)	–	–	Trp(UGA)	29	42
Leu(UUG)	68	22	Ser(UCG)	44	5	Stop(UAG)	–	–	Trp(UGG)	27	7
Leu(CUU)	30	36	Pro(CCU)	27	66	His(CAU)	14	72	Arg(CGU)	33	54
Leu(CUC)	79	0	Pro(CCC)	71	3	His(CAC)	30	22	Arg(CGC)	48	4
Leu(CUA)	20	62	Pro(CCA)	48	120	Gln(CAA)	106	256	Arg(CGA)	9	2
Leu(CUG)	34	1	Pro(CCG)	40	3	Gln(CAG)	70	10	Arg(CGG)	23	0
Ile(AUU)	74	480	Thr(ACU)	36	206	Asn(AAU)	82	440	Ser(AGU)	87	117
Ile(AUC)	39	56	Thr(ACC)	134	8	Asn(AAC)	132	78	Ser(AGC)	39	20
Ile(AUA)	8	111	Thr(ACA)	33	151	Lys(AAA)	84	730	Arg(AGA)	9	189
Met(AUG)	29	153	Thr(ACG)	59	1	Lys(AAG)	117	71	Arg(AGG)	14	3
Val(GUU)	52	284	Ala(GGU)	89	228	Asp(GAU)	107	294	Gly(GGU)	92	180
Val(GUC)	54	8	Ala(GCC)	91	3	Asp(GAC)	85	24	Gly(GGC)	65	5
Val(GUA)	31	157	Ala(GCA)	31	140	Glu(GAA)	78	373	Gly(GGA)	28	199
Val(GUG)	90	16	Ala(GCG)	72	4	Glu(GAG)	40	29	Gly(GGG)	78	16

The 3331 codons of *M.pneumoniae* are from the P1 operon (41) and the *deo*C gene (42). The 6814 codons of *M.capricolum* are from published data (6).

Ссылка на код: ∞ СчётСтопКодонов3

Задание 4.

	CTC	CTA	CTG	CTT	TTA	TTG	GC-состав(%)
Escherichia coli	14952	5203	71305	14728	18505	18301	50,6
Candidatus Gracilibacteria bacterium	4491	4861	4147	8053	15077	8048	28,9
Mycoplasma pneumoniae	2362	2715	2298	5771	7950	4487	40

Предположу, что различия между встречаемостями лейциновых кодонов в рамках одной бактерии связаны с тем, что у одной тРНК может иметься сразу

несколько генов в геноме, кодирующих её. А различия в этом плане между разными бактериями могут быть связаны с разным GC-составом,например, у *E. coli* с достаточно большим значением содержания GC-пар в геноме самым распространенным лейциновым кодоном является CTG, а у *Mycoplasma pneumoniae* со значением GC-состава 28,9% самым распространенным является уже TTA, где гуанина и цитозина нет вообще.
ссылка на код

Ссылка на код: ∞ Лейцин4

Список использованной литературы

[1] Научная статья про отбор старт-кодонов в пользу ATG.
Frida Belinky, Igor B. Rogozin & Eugene V. Koonin. **Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions, 2017,** *Scientific Reports*, https://www.nature.com/articles/s41598-017-12619-6 .

[2] Исследование влияния замены старт-кодона ATG на эффективность транскрипции гена GFP. Ariel Hecht, Jeff Glasgow, Paul R. Jaschke, Lukmaan A. Bawazer, Matthew S. Munson, Jennifer R. Cochran, Drew **Endy,corresponding author and Marc Salitcorresponding author. Measurements of translation initiation from all 64 codons in *E. coli*,** *2017, Nucleic Acids Res,* *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5397182/#SEC3title* .

[3] Статья про возможное участие альтернативных старт-кодонов в образовании разных белковых продуктов.
https://imperialbiosciencereview.wordpress.com/2020/09/18/alternative-start-codons-non-aug-translation/#:~:text=The%20use%20of%20alternative%20start,codons%20 have%20 distinct%20 biological%20 functions

[4] Научная статья про узнавание селеноцистеина в последовательности мРНК.
Jesse Donovan and Paul R. Copeland. **The Efficiency of Selenocysteine Incorporation Is Regulated by Translation Initiation Factors**, 2010, J Mol Biol, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3721751/

[5] Научная статья про альтернативный генетический код у *Candidatus Gracilibacteria bacterium.*

Anna Hanke, Emmo Hamann, Ritin Sharma, Jeanine S. Geelhoed, Theresa Hargesheimer, Beate Kraft,1 Volker Meyer, Sabine Lenk,1 Harald Osmers, Rong Wu,4 Kofi Makinwa, Robert L. Hettich, Jillian F. Banfield, Halina E. Tegetmeyer, and Marc Strous. **Recoding of the stop codon UGA to glycine by a BD1-5/SN-2 bacterium and niche partitioning between Alpha- and Gammaproteobacteria in a tidal sediment microbial community naturally selected in a laboratory chemostat**, 2014, Front Microbiol https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4032931/

[6] Научная статья про альтернативный генетический код у Mycoplasma pneumoniae.
PhilippeSimoneau, Cheng-ming Li, Steve Loechell, RainerWenzel, Richard Herrmann and Ping-chuan Hu. **Codonreadingscheme in Mycoplasmapneumoniae revealed bythe analysis ofthecompleteset oftRNA genes**, 1993, NucleicAcidsResearch, *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC311414/pdf/nar00070-0121.pdf*
.