

Задание 1

<i>E.coli</i>	<i>C.bacterium</i>	<i>M.pneumoniae</i>
ATG 3890 GTG 338 TTG 80 ATT 4 TTC 1 CTG 2	ATG 1129 GTG 41 TTG 23 TCA 1 ACA 1 TCT 1	GTG 60 ATG 627 TTG 49 ATT 7 CTG 2 TCC 2 CTC 3 ATA 3 TGA 1 CTA 1 ACT 1 TTC 1 AAA 1 ACA 1 TTA 1 GTT 1 TCT 1 CAA 1 CAC 1 ATC 1 GAA 1

ATG распознается формилметионин тРНК с антикодоном CAT, поэтому по механизму wobbl-взаимодействий распознаются и старт-кодны GTG, TTG, которых в геноме достаточно много. Удивляет количество старт-коднов у *M. pneumoniae*, вероятно, многие из последовательностей начинающихся с редких старт-коднов - псевдогены.

Задание 2

В последовательности я нашел четыре таких последовательности, первая из них - псевдоген, в остальных TGA кодирует селеноцистеин, трансляция не завершается, так как после кодона UGA на соответствующей мРНК содержится последовательность SECIS, связывающая особые трансляционные факторы, которые узнает тРНК^{Sec} и селеноцистеин встраивается в полипептидную цепь.

- 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]

Задание 3

<i>E.coli</i>	<i>C.bacterium</i>	<i>M.pneumoniae</i>
TGA 1246 TAA 2761 TAG 306 ATA 1 GAA 1	TAA 1000 TAG 188 TCT 2 TTA 1 AAA 1 CTT 1 ACA 1 TGA 1 GAA 1	ACT 1 TAG 220 TAA 526 ATA 1 CTA 1 GGG 5 GGC 1 AAT 1 CCC 1 CGG 1 GGT 3 GAT 1 TAT 1 TAC 1 AAA 1 TTA 1

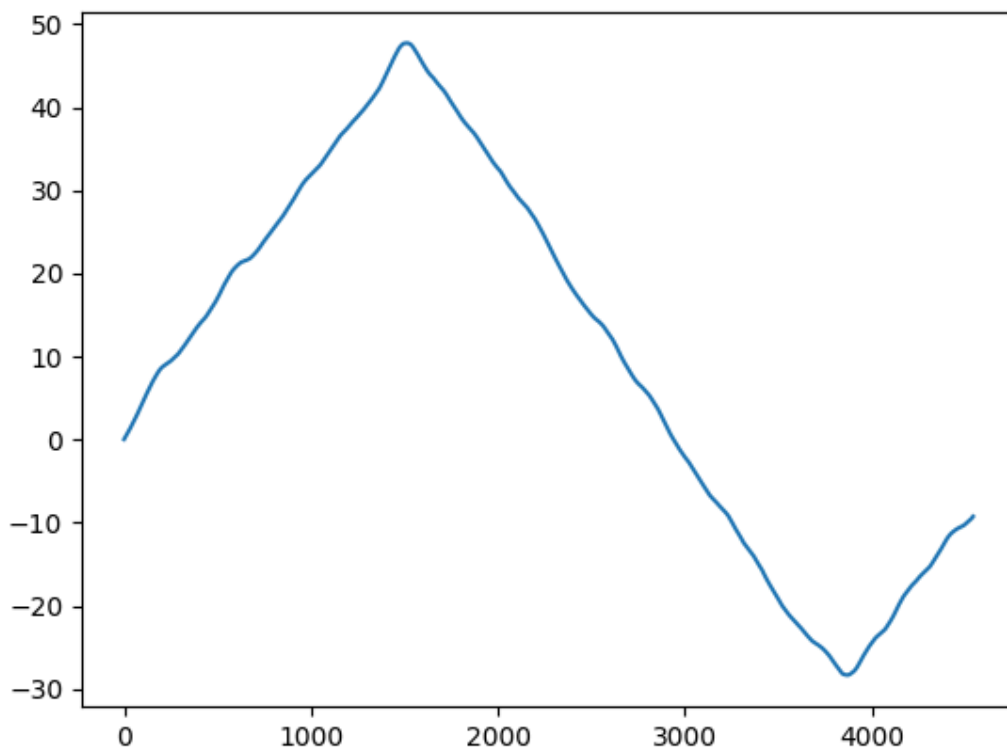
У *G. bacterium* TGA - кодон глицина, у *M. pneumoniae* - триптофана [1, 3]. Отсутствие стоп-кодона TGA в геноме этих двух бактерий, вероятно, связано с отсутствием у них RF2, как, например у *Mycoplasma capricolum* [3].

Задание 4.

<i>E.coli</i>	<i>C.bacterium</i>	<i>M.pneumoniae</i>
CTT 14728 CTC 14952 CTG 71305 CTA 5203 TTA 18505 TTG 18301	CTT 9332 CTC 3968 CTG 1714 CTA 3357 TTA 14766 TTG 3237	CTT 2782 CTC 3158 CTG 2470 CTA 2826 TTA 10294 TTG 5571

Явно видно, что в геноме каждой бактерии есть оптимальный кодон, встречающийся чаще всех, вероятно минорные кодоны - механизм тонкой регуляции трансляции. Предположу, что минорные кодоны встречаются между частями мРНК, кодирующими разные домены белка, чтобы успевал происходить здоровый фолдинг новосинтезированной полипептидной цепи. Насколько я знаю, в различных кладах бактерий выбраны различные оптимальные кодоны, этим и обусловлена разница в выборе оптимального кодона среди исследуемых бактерий.

Задание 5.



Максимум наблюдается в районе 1514000 нуклеотида, минимум в районе 3871000. Насколько я знаю, минимум кумулятивного GC-skew соответствует точке OriC, а максимум - точке ter [4].

Задание 6

<i>E.coli</i>	<i>C.bacterium</i>	<i>M.pneumoniae</i>
329 AAGGAG	181 TAAAAA	55 AATTAA
280 TAAGGA	179 ATAAAA	54 TTAAAA
255 CAGGAG	176 AAATAA	43 AAAGGA
255 AGGAGA	174 AATAAA	38 TTAAAA
224 AAAGGA	170 AAAAAA	38 ATAAAA

222 AAGGAA	149 TAATAA	37 TTAAAC
214 AGGAGT	137 TTTTAA	37 ATTTAA
204 GGAGAA	134 AAAATA	37 AATTTA
188 AGGAAA	134 AAAAAT	35 TTTTAA
177 ACAGGA	133 TAAATA	34 TTAATT
169 GAGGAA	126 AATAAT	34 TAATTA
149 AAAAGG	125 ATAATA	34 AATAAA
147 AGGAAT	122 TTAAAA	34 AAGGAG
138 CAAGGA	121 TTTTTA	32 AAATAA

AGGAG - известная последовательность Шайна-Дальгарно, связывающаяся с рРНК, предположу, что последовательность Шайна-Дальгарно у остальных двух бактерий иная(вероятно, это связано с GC-составом геномов этих бактерий).

Список литературы:

1. Inamine JM, Ho KC, Loechel S, Hu PC. Evidence that UGA is read as a tryptophan codon rather than as a stop codon by *Mycoplasma pneumoniae*, *Mycoplasma genitalium*, and *Mycoplasma gallisepticum*. *J Bacteriol.* 1990 Jan;172(1):504-6. doi: 10.1128/jb.172.1.504-506.1990. PMID: 2104612; PMCID: PMC208464.
2. Inagaki Y, Bessho Y, Osawa S. Lack of peptide-release activity responding to codon UGA in *Mycoplasma capricolum*. *Nucleic Acids Res.* 1993 Mar 25;21(6):1335-8. doi: 10.1093/nar/21.6.1335. PMID: 8464722; PMCID: PMC309316.
3. Hanke A, Hamann E, Sharma R, Geelhoed JS, Hargesheimer T, Kraft B, Meyer V, Lenk S, Osmer H, Wu R, Makinwa K, Hettich RL, Banfield JF, Tegetmeyer HE, Strous M. 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. *Front Microbiol.* 2014 May 16;5:231. doi: 10.3389/fmicb.2014.00231. PMID: 24904545; PMCID: PMC4032931.
4. Arakawa K, Tomita M. The GC skew index: a measure of genomic compositional asymmetry and the degree of replicational selection. *Evol Bioinform Online.* 2007 Sep 6;3:159-68. PMID: 19461976; PMCID: PMC2684130.