

Практикум 13

Ссылка на скрипты:

https://colab.research.google.com/drive/1iJR-Ot_EvF8BP_5zntj5-xl9DKzQkYNC?usp=sharing

Задание 1. Число старт-кодонов.

<i>Escherichia coli str. K-12 substr. MG1655:</i>	<i>Candidatus Gracilibacteria bacterium 28_42_T64:</i>	<i>Mycoplasma pneumoniae M29:</i>
ATT 4 TTC 1 ATG 3890 GTG 338 CTG 2 TTG 80	ACA 1 ATG 1129 GTG 41 TCA 1 TCT 1 TTG 23	ACT 1 ATA 4 TTA 3 ATC 1 ATT 8 CAA 2 TTG 53 CTC 2 CTG 1 GAA 1 GGA 1 GTG 60 GTT 1 TCT 1 AAA 1 ATG 629

Проанализировав полученные значения, можно заметить, что канонический кодон ATG превалирует, хотя GTG и TTG, отличающиеся от ATG на 1 нуклеотид (вероятно, вследствие единичной мутации), также достаточно часто встречаются. Также мы видим большое количество старт-кодонов, которые образовались в результате мутаций и отсутствия стабилизирующего отбора, и находятся в псевдогенах.

Также можно отметить разнообразие старт-кодонов у бактерии *Mycoplasma pneumoniae M29*, что может быть связано с ее паразитическим образом жизни.

Задание 2

В данных последовательностях у *E. coli* содержатся стоп-кодоны не в конце:

lcl|U00096.3_cds_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]

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]

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]

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]

Из четырех последовательностей с нетерминирующими стоп-кодонами первая является псевдогеном, в остальных трех стоп-кодон кодирует селеноцистеин.

Задание 3

<i>Escherichia coli str. K-12 substr. MG1655:</i>	<i>Candidatus Gracilibacteria bacterium 28_42_T64:</i>	<i>Mycoplasma pneumoniae M29:</i>
TGA 1246 TAA 2761 TAG 306	TGA 1 TAA 1000 TAG 188	TGA 0 TAA 533 TAG 221

У *Candidatus Gracilibacteria bacterium 28_42_T64* и *Mycoplasma pneumoniae M29* стоп-кодона TGA практически нет. У первой TGA кодирует глицин, у второй - триптофан.

Задание 4

<i>Escherichia coli str. K-12 substr. MG1655:</i>	<i>Candidatus Gracilibacteria bacterium 28_42_T64:</i>	<i>Mycoplasma pneumoniae M29:</i>
TTA 18505 TTG 18301 CTA 5203 CTG 71305 CTC 14952 CTT 14728	TTA 33582 TTG 26349 CTA 10064 CTG 75452 CTC 19443 CTT 22781	TTA 42828 TTG 32903 CTA 13569 CTG 78506 CTC 21636 CTT 27404