

Практикум 4. Ортологи и паралоги

Задание 1. Составление списка гомологичных белков, включающих паралоги

Отберём скачанные из Uniprot полные протеомы бактерий отдела Pseudomonadota.

Мой выбор:

ACICJ.fasta HAEIN.fasta
AGRFC.fasta NEIMA.fasta
AROAE.fasta PARDP.fasta
BARHE.fasta PASMU.fasta

Запишем 8 белков в один файл:

```
cat A*.fasta BARHE.fasta HAEIN.fasta NEIMA.fasta PARDP.fasta PASMU.fasta >  
~/term4/practice4/prots.fasta
```

Создадим локальную базу данных, проведём по ней поиск blastp, установив пороговое значение E-Value 10e-4:

```
makeblastdb -in prots.fasta -dbtype prot
```

```
blastp -task blastp -query CLPX_ECOLI.fasta -db prots.fasta -out blastres.out -evalue  
0.0001
```

Выдача blast:

Sequences producing significant alignments:	Score (Bits)	E Value
sp Q5P160 CLPX_AROAE ATP-dependent Clp protease ATP-binding subunit	619	0.0
sp P57981 CLPX_PASMU ATP-dependent Clp protease ATP-binding subunit	612	0.0
sp Q8UFY5 CLPX_AGRFC ATP-dependent Clp protease ATP-binding subunit	596	0.0
sp Q6G3Z2 CLPX_BARHE ATP-dependent Clp protease ATP-binding subunit	588	0.0
sp P44838 CLPX_HAEIN ATP-dependent Clp protease ATP-binding subunit	587	0.0
sp A5FX05 CLPX_ACICJ ATP-dependent Clp protease ATP-binding subunit	583	0.0
sp A1B1H7 CLPX_PARDP ATP-dependent Clp protease ATP-binding subunit	580	0.0
sp Q9JTX8 CLPX_NEIMA ATP-dependent Clp protease ATP-binding subunit	557	0.0
sp A1B5T0 HSLU_PARDP ATP-dependent protease ATPase subunit HslU O...	103	3e-24
sp Q6G5G0 HSLU_BARHE ATP-dependent protease ATPase subunit HslU O...	97.1	4e-22
sp Q5P503 HSLU_AROAE ATP-dependent protease ATPase subunit HslU O...	93.6	8e-21
sp P57968 HSLU_PASMU ATP-dependent protease ATPase subunit HslU O...	93.2	1e-20
sp Q8UJ87 HSLU_AGRFC ATP-dependent protease ATPase subunit HslU O...	92.0	2e-20
tr A5FYD7 A5FYD7_ACICJ ATP-dependent protease ATPase subunit HslU...	92.0	2e-20
sp P43773 HSLU_HAEIN ATP-dependent protease ATPase subunit HslU O...	92.0	2e-20
tr A1B8N4 A1B8N4_PARDP ATP-dependent Clp protease, ATP-binding su...	50.1	1e-06
tr A0A0H3LXZ4 A0A0H3LXZ4_BARHE ATP-dependent zinc metalloprotease...	45.4	3e-05
tr Q7CT50 Q7CT50_AGRFC ATP-dependent zinc metalloprotease FtsH OS...	45.4	4e-05
tr A5FVF9 A5FVF9_ACICJ ATP-dependent zinc metalloprotease FtsH OS...	44.7	6e-05
tr A1AZV8 A1AZV8_PARDP ATP-dependent zinc metalloprotease FtsH OS...	44.3	7e-05
sp Q6G5R1 RUVB_BARHE Holliday junction branch migration complex s...	43.5	1e-04

Рис. 1. Находки blastp.

Задание 2. Реконструкция и визуализация

Из списка находок был составлен один файл, который подан на вход muscle.

Полученное множественное выравнивание конвертировано в .phy средствами BioPython, а затем подано на вход программе FastME (MT-Rev).

```
fastme -i res.phy -o protree.txt --protein=MtREV -b 100
```

Формула дерева:

```
((CLPX_AGRFC:0.089654,CLPX_BARHE:0.092496)82:0.058758,((((((HSLU_PASMU:0.046734,HSLU_HAEIN:0.0626  
26)100:0.159232,HSLU_AROAE:0.375102)82:0.139256,(A5FYD7_ACICJ:0.287852,((HSLU_BARHE:0.177959,HSLU_  
AGRFC:0.158872)100:0.163841,HSLU_PARDP:0.351467)50:0.072023)96:0.106156)93:0.638946,(((Q7CT50_AGRF  
C:0.226287,A0A0H3LXZ4_BARHE:0.182275)78:0.130323,A5FVF9_ACICJ:0.299305)77:0.199498,A1AZV8_PARDP:0.  
277581)100:2.330509,RUVB_BARHE:2.716402)38:0.304465,A1B8N4_PARDP:2.413242)71:0.226057)76:0.456599,(C  
LPX_HAEIN:0.180132,CLPX_PASMU:0.166857)90:0.092295)10:0.058702,CLPX_NEIMA:0.279369)24:0.028257,CLPX_  
_AROAE:0.220040)95:0.155743,CLPX_PARDP:0.133038)59:0.024741,CLPX_ACICJ:0.135951);
```

При помощи iTol было визуализировано дерево в формате Newick (с укоренением в среднюю точку).

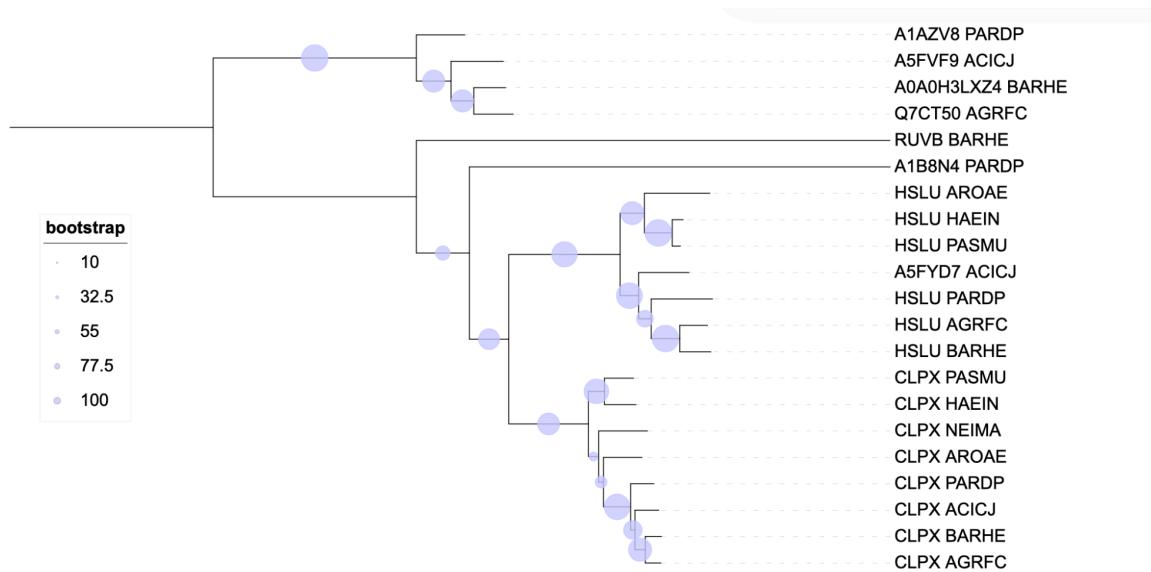


Рис. 2. Реконструкция дерева гомологов с использованием бутстреп-реплик.

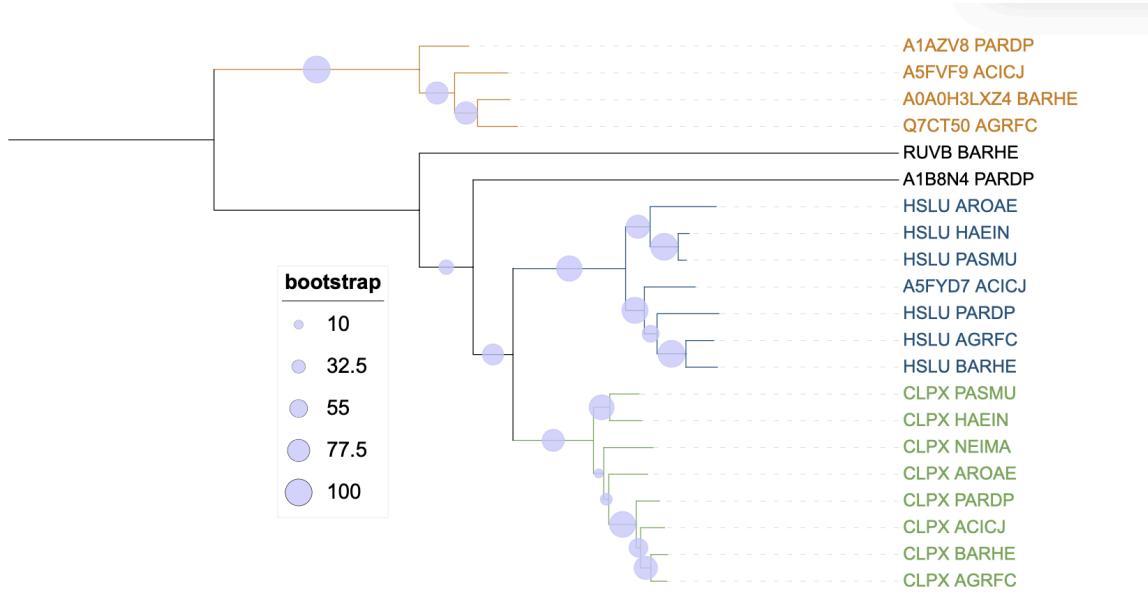


Рис. 3. Реконструкция дерева гомологов с раскраской по ортологическим группам.

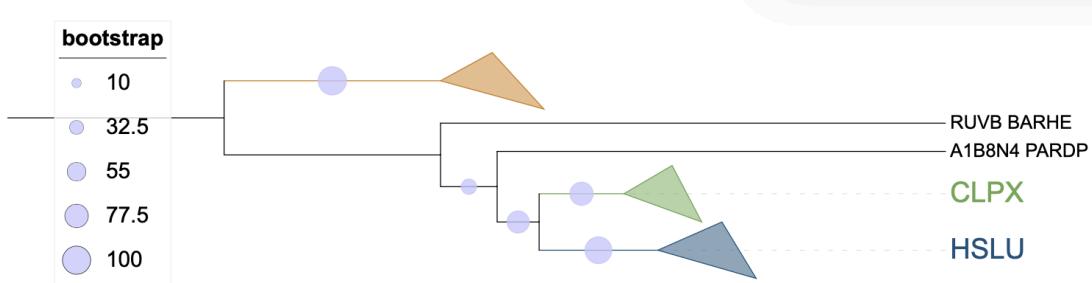


Рис. 4. Реконструкция дерева гомологов со схлопнутыми ортологическими группами.

Примеры пар ортологов: CLPX_PASMU и CLPX_HAEIN, CLPX_NEIMA и CLPX_BARHE, HSLU_PARDP и HSLU_AROAE

Примеры пар параполов: HSLU_BARHE и CLPX_BARHE, HSLU_PARDP и CLPX_PARDP, HSLU_AROAE и CLPX_AROAE

Можно выделить три ортологические группы по белку, встречающемуся чаще всего:
CLPX – туда входят CLPX всех 8 организмов

HSLU – HSLU 6 организмов (HSLU_AROAE, HSLU_HAEIN, HSLU_PASMU, HSLU_PARDP, HSLU_AGRFC, HSLU_BARHE) и A5FYD7_ACICJ

Третья группа – туда входят 4 разных белка 4-х разных бактерий

Теперь проанализируем филогению, опираясь на верную, представленную в задании.

По CLPX: неверно, PARDP и ACICJ переставлены местами, NEIMA должна быть дальше от PARDP, чем AROAE

По HSLU: всё верно

По третьей группе: неверно, PARDP и ACICJ переставлены местами