

# BLAST

## Задание 1

Найдите сходные последовательности в базе данных Refseq\_protein.

Я нашла последовательности, в которых находится мой белок с идентификатором NP\_950114.1

Для исходного запроса были выбраны критерии E-value max = 0.001 и 100000 находок

Первая последовательность: лучшая находка

Находка очень хорошая, о чем свидетельствуют показатели E-value и Ident.

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Rhodopseudomonas palustris]</a>	669	669	100%	0.0	100%	<a href="#">WP_011160312.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Rhodopseudomonas palustris]</a>	664	664	100%	0.0	99%	<a href="#">WP_012497885.1</a>
<input type="checkbox"/>	<a href="#">high-affinity phosphate transport protein ABC superfamily_peri_bind [Bradyrhizobium oligotrophicum]</a>	603	603	100%	0.0	87%	<a href="#">WP_015668928.1</a>
<input type="checkbox"/>	<a href="#">MULTISPECIES: phosphate ABC transporter substrate-binding protein [Bradyrhizobium]</a>	599	599	100%	0.0	86%	<a href="#">WP_044541082.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Alfipia birqiae]</a>	593	593	100%	0.0	85%	<a href="#">WP_019197628.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Rhodopseudomonas palustris]</a>	592	592	100%	0.0	90%	<a href="#">WP_011475321.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Bradyrhizobium sp. S23321]</a>	592	592	100%	0.0	86%	<a href="#">WP_015689159.1</a>

Переход из 70% в 69%: Зеленым последняя взятая находка, красным первая не взятая

видно, что e-value не коррелирует с Query cover

<input type="checkbox"/>	<a href="#">putative serine/threonine-protein kinase/receptor [Acanthamoeba polyphaga mimivirus]</a>	90.5	90.5	70%	2e-16	28%	<a href="#">YP_003987358.1</a>
<input type="checkbox"/>	<a href="#">periplasmic binding protein-like II [Coccomyxa subellipsoidea C-169]</a>	86.3	86.3	70%	1e-15	30%	<a href="#">XP_005645136.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein [Ruminococcus sp. JC304]</a>	72.0	72.0	70%	1e-10	25%	<a href="#">WP_044997403.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein [actinobacterium SCGC AAA023-D18]</a>	71.6	71.6	70%	1e-10	29%	<a href="#">WP_018227229.1</a>
<input type="checkbox"/>	<a href="#">serine/threonine protein kinase [Acanthamoeba castellanii str. Neff]</a>	70.1	139	70%	7e-10	30%	<a href="#">XP_004335056.1</a>
<input type="checkbox"/>	<a href="#">serine/threonine kinase [Acanthamoeba castellanii str. Neff]</a>	64.3	64.3	70%	5e-08	28%	<a href="#">XP_004335059.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Desulfovibrio salexigens]</a>	54.7	54.7	70%	3e-05	31%	<a href="#">WP_015853337.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Salmonella enterica]</a>	254	254	69%	8e-80	54%	<a href="#">WP_045713570.1</a>
<input type="checkbox"/>	<a href="#">MULTISPECIES: phosphate ABC transporter substrate-binding protein [Prochlorococcus]</a>	140	140	69%	4e-35	38%	<a href="#">WP_011125725.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Synechococcus sp. KORDI-52]</a>	136	136	69%	2e-33	39%	<a href="#">WP_038555396.1</a>
<input type="checkbox"/>	<a href="#">phosphate binding protein [Leptospirillum ferriphilum]</a>	137	137	69%	2e-33	35%	<a href="#">WP_014962148.1</a>
<input type="checkbox"/>	<a href="#">putative phosphate ABC transporter substrate-binding protein [Actinoplanes friuliensis]</a>	119	119	69%	3e-27	35%	<a href="#">WP_023362254.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Herbaspirillum sp. B501]</a>	118	118	69%	4e-27	33%	<a href="#">WP_034312399.1</a>

## Задание 2

Найдите сходные последовательности среди белков из какой-нибудь таксономической группы.

Была выбрана систематическая категория *Escherichia coli*

Организмы из *Escherichia coli* присутствуют в исходном результате запроса

<input type="checkbox"/>	<a href="#">ABC transporter substrate-binding protein [Pasteurella bettyae]</a>	51.6	51.6	51%	3e-04	30%	<a href="#">WP_005760590.1</a>
<input type="checkbox"/>	<a href="#">ABC transporter substrate-binding protein [Mannheimia sp. MG13]</a>	50.8	50.8	51%	5e-04	30%	<a href="#">WP_044470677.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Lactococcus garvieae]</a>	50.8	50.8	51%	5e-04	29%	<a href="#">WP_003134993.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	223	223	50%	4e-68	62%	<a href="#">WP_042106643.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Herbaspirillum sp. B39]</a>	213	213	50%	5e-64	65%	<a href="#">WP_034332664.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Burkholderia pseudomallei]</a>	180	180	50%	1e-51	51%	<a href="#">WP_038758802.1</a>
<input type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Kineosphaera limosa]</a>	115	115	50%	8e-26	44%	<a href="#">WP_040562822.1</a>

После этого провели поиск в категории *Escherichia coli*:

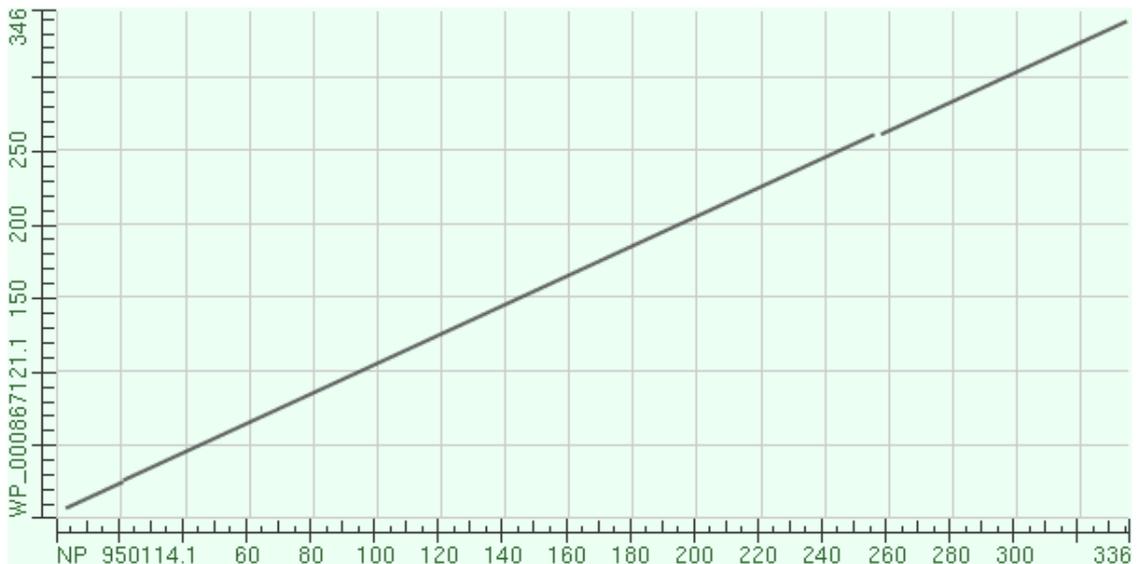
Видно, что E-value стал меньше, поскольку сама база стала меньше.

<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter, phosphate-binding protein PstS [Escherichia coli]</a>	365	365	95%	6e-123	56%	<a href="#">WP_001712948.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate-binding protein PstS [Escherichia coli]</a>	364	364	95%	7e-123	55%	<a href="#">WP_021514062.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	364	364	95%	7e-123	56%	<a href="#">WP_033545264.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	364	364	95%	8e-123	56%	<a href="#">WP_032257638.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	364	364	95%	8e-123	56%	<a href="#">WP_024231677.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate-binding protein PstS [Escherichia coli]</a>	364	364	95%	9e-123	56%	<a href="#">WP_000867155.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	364	364	95%	1e-122	56%	<a href="#">WP_032221449.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	364	364	95%	1e-122	56%	<a href="#">WP_000867137.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter, phosphate-binding protein PstS [Escherichia coli]</a>	363	363	95%	1e-122	56%	<a href="#">WP_001719149.1</a>
<input checked="" type="checkbox"/>	<a href="#">MULTISPECIES: phosphate ABC transporter substrate-binding protein [Enterobacteriaceae]</a>	363	363	95%	2e-122	56%	<a href="#">WP_039080358.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	363	363	95%	2e-122	56%	<a href="#">WP_032286772.1</a>
<input checked="" type="checkbox"/>	<a href="#">phosphate ABC transporter substrate-binding protein [Escherichia coli]</a>	363	363	95%	2e-122	56%	<a href="#">WP_033802673.1</a>

## Задание 3

Для одной из найденных последовательностей сохраните карту локального сходства.

Исходная последовательность NP\_950114.1 была выровнена с последовательностью WP\_000867121.1 в режиме парного выравнивания BLAST



У нас очень хорошие E-value и процент идентичности, поэтому у нас хорошая карта локального сходства.

#### Задание 4

Используйте BLAST для поиска в своей базе данных.

Была сделана база, состоящая из 40 белков в асцидиях (файл sequence.fasta)

Команды:

```
makeblastdb -dbtype prot -in sequence.fasta -out
my_align
blastp -db my_align -query my_first.fasta
```

```
> gi|319440115|dbj|BAJ61821.1| natural resistance-associated macrophage
protein [Ascidia sydneiensis samea]
Length=586

Score = 22.3 bits (46), Expect = 0.42, Method: Compositional matrix adjust.
Identities = 22/84 (26%), Positives = 35/84 (42%), Gaps = 8/84 (10%)

Query 127 LLAAGETGRAVNVKKGQFLAPWDMKNVADKVA STGNRRILLCERGT SFGYNTLVDFRGL 186
      +LAAG++      GQF+      M+  D  S  R  +L  R  +  G  L+  F  +
Sbjct 400 ILAAGQSSTMTGTYSGQFV---MEGFLDLRWSRFQR--VLLTR SIAIGPTFLLA AFSDI 453

Query 187 PTMAATGWPVVF DATHSVQQPGGL 210
      + TG  ++  S+Q P L
Sbjct 454 QHL--TGMNDLLNVLQSLQLPFAL 475
```

Видно, что плохой E-value, плохой Identities и Positives, и выравнивание не имеет смысла