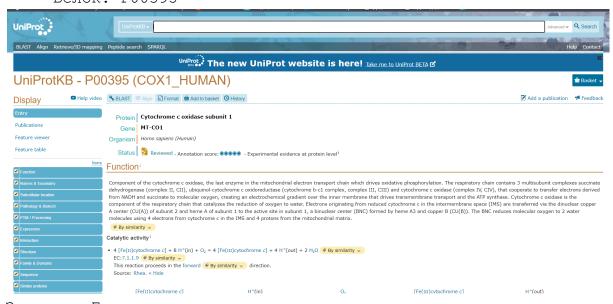
Отчет по ДЗ №3

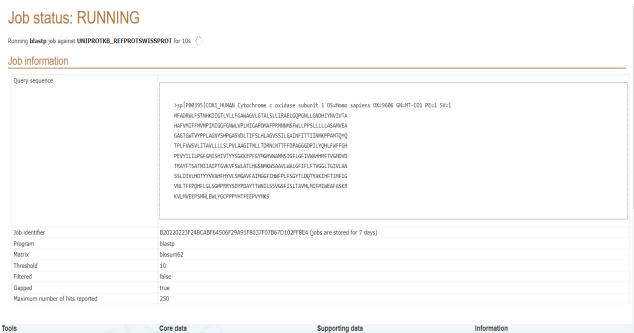
Дисциплина: Биоинформатика Исполнительница: Смолкина Ю.А

Группа: АДБМ2021

Исходные данные: Белок: P00395



Запускаю Бласт



Job information

Query sequence	
	>sp P00395 COX1_HUMAN Cytochrome c oxidase subunit 1 0S=Homo sapiens 0X=9606 GN=MT-C01 PE=1 SV=1 MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA HAFVMIFFMYMPINIGGFGNMLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMKPPAMTQYQ TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLFWFFGH PEVVILILPGFGMISHTVTTYYSGKKEPFGYMGMVWAMMSIGFLGFIYWAHHMFTVGMDVD TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIHFTIMFIG VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIWEAFASKR KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS
Date of job execution	2022-02-23
Job identifier	B20220223F248CABF64506F29A91F8037F07B67D102FF8E4 (jobs are stored for 7 days)
Running time	115.5 seconds
Scheduled time	92.6 seconds
Program	blastp (blastp BLASTP 2.9.0+)
Database	uniprotkb_refprotswissprot (Protein) generated for BLAST on Nov 17, 2021
Sequences	64,899,224 sequences consisting of 25,452,742,688 letters
Matrix	blosum62
Threshold	10
Filtered	false
Gapped	true
Maximum number of hits reported	250

Результаты:



То есть нашлось 250 совпадений

Лучшии находки : laя - P00395 c Identity = 100% 2ая находка Q9T9W1 c Identity = 98.8%

Для 1ого: (нашел сам себя, никто такое не запрещал)

Данные из Swiss-Prot

Protein

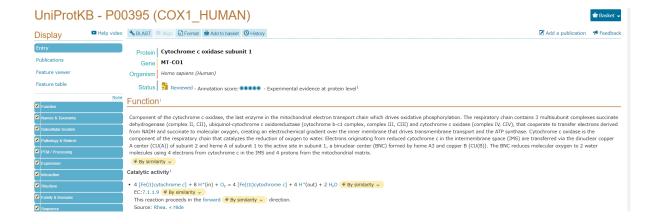
Cytochrome c oxidase subunit 1

Gene

MT-CO1

Organism

Homo sapiens (Human)



E-value: 0.0
Score: 2732
Ident.: 100.0%
Positives: 100.0%
Query Length: 513
Match Length: 513



Покрытие:

P00395 COX1_HUMAN	1	MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA	60
		MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA	
P00395 COX1_HUMAN	1	MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA	60
P00395 COX1_HUMAN	61	${\tt HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA}$	120
		HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA	
P00395 COX1_HUMAN	61	HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA	120
P00395 COX1_HUMAN	121	${\tt GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMKPPAMTQYQ}$	180
		GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMKPPAMTQYQ	
P00395 COX1_HUMAN	121	GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMKPPAMTQYQ	180
P00395 COX1_HUMAN	181	TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLFWFFGH	240
_		TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYOHLFWFFGH	
P00395 COX1_HUMAN	181	${\tt TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLFWFFGH}$	240
P00395 COX1_HUMAN	241	PEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGFIVWAHHMFTVGMDVD	300
		PEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGFIVWAHHMFTVGMDVD	
P00395 COX1_HUMAN	241	PEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGFIVWAHHMFTVGMDVD	300
P00395 COX1_HUMAN	301	TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN	360
		TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN	
P00395 COX1_HUMAN	301	TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN	360
P00395 COX1_HUMAN	361	SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIHFTIMFIG	420
		SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIHFTIMFIG	
P00395 COX1_HUMAN	361	SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIHFTIMFIG	420
P00395 COX1_HUMAN	421	VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIWEAFASKR	480
		VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIWEAFASKR	
P00395 COX1_HUMAN	421	VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIWEAFASKR	480
P00395 COX1_HUMAN	481	KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS	513
_		KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS	
P00395 COX1_HUMAN	481	KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS	513

Для 2ого:



Protein

Cytochrome c oxidase subunit 1

Gene

MT-CO1

Organism

Pan troglodytes (Chimpanzee)



E-value: 0.0Score: 2707Ident.: 98.8%

Positives: 99.2%Query Length: 513Match Length: 513

Покрытие:

P00395	COX1_HUMAN	1	MFADRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA MF DRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGOPGNLLGNDHIYNVIVTA	6
Q9T9W1	COX1_PANTR	1	MFTDRWLFSTNHKDIGTLYLLFGAWAGVLGTALSLLIRAELGQPGNLLGNDHIYNVIVTA	60
P00395	COX1_HUMAN	61	HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA	12
Q9T9W1	COX1 PANTR	61	HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA HAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPRMNNMSFWLLPPSLLLLLASAMVEA	12
P00395	COX1 HUMAN	121	GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSSILGAINFITTIINMKPPAMTQYQ	180
	_		GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAG+SSILGAINFITTIINMKPPAMTQYQ	
Q9T9W1	COX1_PANTR	121	${\tt GAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGISSILGAINFITTIINMKPPAMTQYQ}$	180
P00395	COX1_HUMAN	181	${\tt TPLFVWSVLITAVLILLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLFWFFGH}$	24
			TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLFWFFGH	
Q9T9W1	COX1_PANTR	181	TPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGGDPILYQHLFWFFGH	24
P00395	COX1_HUMAN	241	PEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGFIVWAHHMFTVGMDVD	30
			PEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGFIVWAHHMFTVGMDVD	
Q9T9W1	COX1_PANTR	241	PEVYILILPGFGMISHIVTYYSGKKEPFGYMGMVWAMMSIGFLGFIVWAHHMFTVGMDVD	300
P00395	COX1_HUMAN	301	${\tt TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN}$	36
			TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN	
Q9T9W1	COX1_PANTR	301	TRAYFTSATMIIAIPTGVKVFSWLATLHGSNMKWSAAVLWALGFIFLFTVGGLTGIVLAN	36
P00395	COX1_HUMAN	361	SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIHFTIMFIG	42
			SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKI F IMFIG	
Q9T9W1	COX1_PANTR	361	SSLDIVLHDTYYVVAHFHYVLSMGAVFAIMGGFIHWFPLFSGYTLDQTYAKIQFAIMFIG	420
P00395	COX1_HUMAN	421	VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWNILSSVGSFISLTAVMLMIFMIWEAFASKR	480
			VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWN+LSSVGSFISLTAVMLMIFMIWEAFASKR	
Q9T9W1	COX1_PANTR	421	VNLTFFPQHFLGLSGMPRRYSDYPDAYTTWNVLSSVGSFISLTAVMLMIFMIWEAFASKR	480
P00395	COX1_HUMAN	481	KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS	513
	-		KVLMVEEPS NLEWLYGCPPPYHTFEEPVYMKS	
09T9W1	COX1 PANTR	481	KVLMVEEPSANLEWLYGCPPPYHTFEEPVYMKS	51

Все находки имеют Е-значение = 0.0, сотых не видно. Я думаю, что в данном случае лимит стоит на максимальным числом находок, но 250 является максимально допустимым (там где я запускала- а точнее не изменяемым). *Запускала я из $\frac{\text{https://www.uniprot.org/blast/uniprot/B20220223F248CABF64506F2}}{9A91F8037F07B67D102FF8E4} (UNIPROT)$

^{*}По цифрам справа можно посчитать несовпадения-гепы

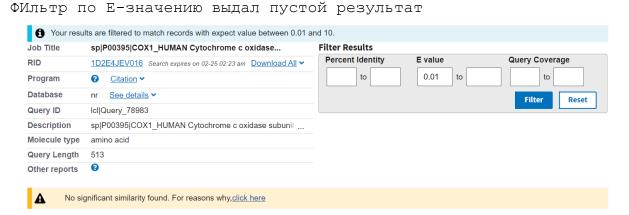
Поменяем параметры в Blastp:

General Parameters		
Max target sequences	♦ 1000 ➤ Select the maximum number of aligned sequences to display ?	
Short queries	Automatically adjust parameters for short input sequences ?	
Expect threshold	1 0	
Word size	6 ▼ ②	
Max matches in a query range	0	
Scoring Parameters		
Matrix	BLOSUM62 ✔ ②	
Gap Costs	Existence: 11 Extension: 1 🗸	
Compositional adjustments	Conditional compositional score matrix adjustment ✓	
Filters and Masking		
Filter	✓ Low complexity regions 🔞	
Mask		

Я оставила таким же threshold, и увеличила кол-во таргетных последовательностей (чтобы проверить предположение выше)

https://blast.ncbi.nlm.nih.gov/Blast.cgi

Действительно, нашлось еще больше последовательностей, но все тоже с E-значением = 0.0

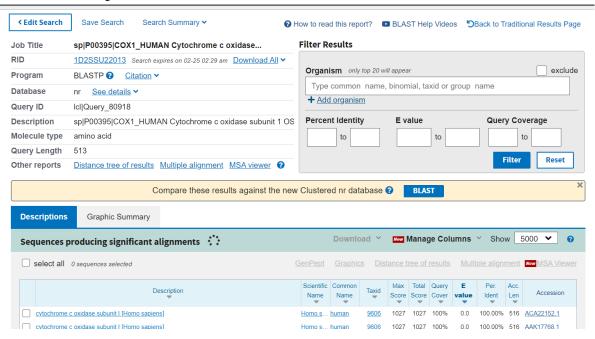


Поменяю параметры еще раз:

General Para	emeters
Max target sequences	♦ 5000 ♥ Select the maximum number of aligned sequences to display ?
Short queries	Automatically adjust parameters for short input sequences ?
Expect threshold	
	Expected number of chance matches in a random model. more You Tibe Expect value tutorial
Word size	6 v 3
Max matches in a query range	0
Scoring Para	nmeters
Matrix	BLOSUM62 ▼ 3
Gap Costs	Existence: 11 Extension: 1 🗸
Compositional adjustments	Conditional compositional score matrix adjustment ✓
Filters and M	lasking
Filter	♦ Low complexity regions
Mask	Mask for lookup table only Mask lower case letters

Таким образом я попытаюсь получить меньше результатов, но возможно с увеличенным E-значением

Мой компьютер завис:

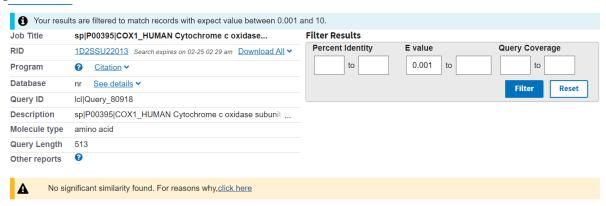


Теория не сработало, нашлось таки 5000 мэтчей

Sequences producing significant alignments

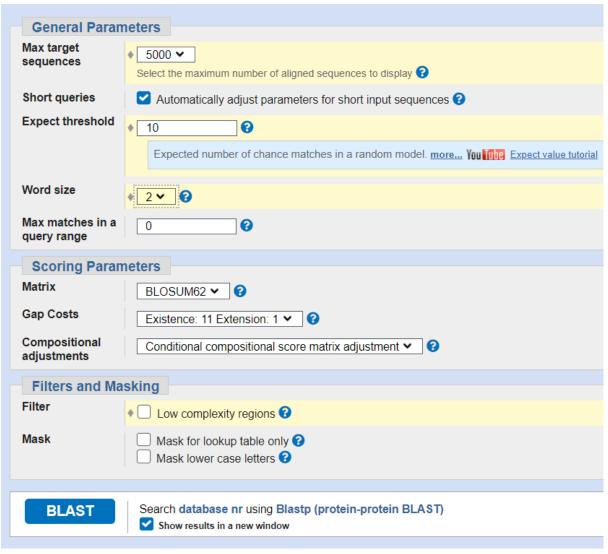
select all 5000 sequences selected

Фильтрация по E-значению снова не принесла желаемого результата

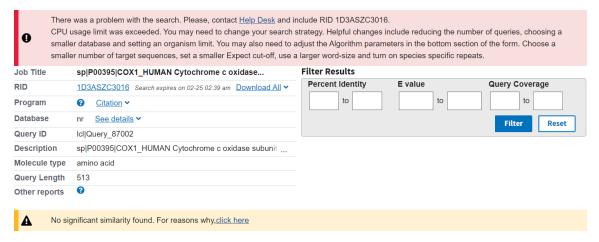


то есть все метчи имеют Е-значение = 0.0

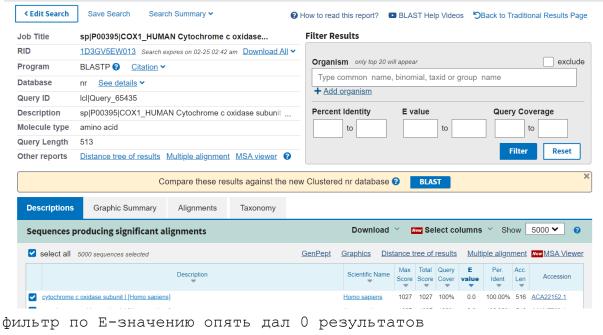
Теперь поменяю длину слова

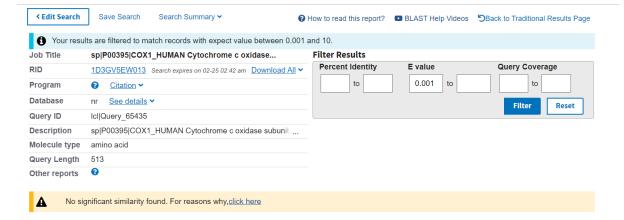


Я сломала Бласт

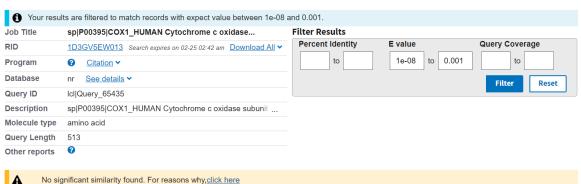


По итогу опять получили 5000 метчей

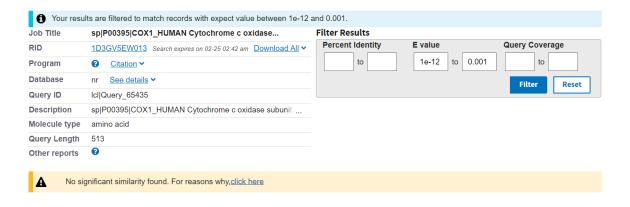




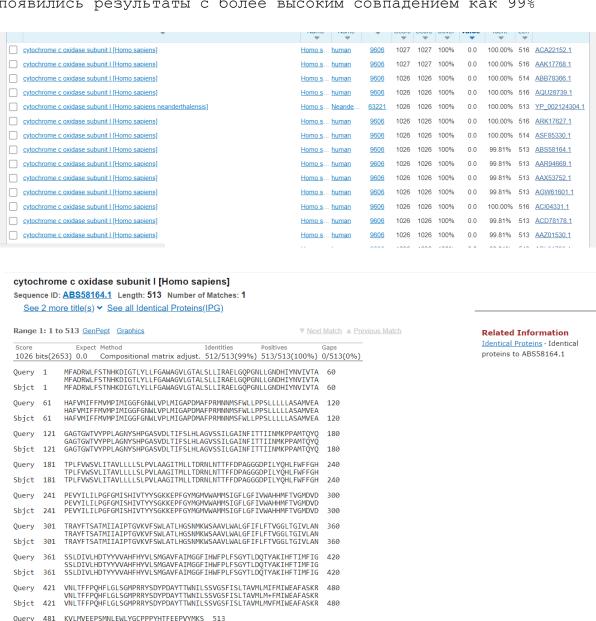
Е-значение от 0.0000001 до 0.001



No significant similarity found. For reasons why, click here



Тем ни менее из отличий с заданием N1 могу отметить, что появились результаты с более высоким совпадением как 99%



KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS KVLMVEEPSMNLEWLYGCPPPYHTFEEPVYMKS 513

Sbjct 481

 Score
 Expect
 Method
 Identities
 Positives
 Gaps

 1026 bits(2653)
 0.0
 Compositional matrix adjust.
 512/513(99%)
 513/513(100%)
 0/513(0%)

нет возможности выбрать е-значение != 0.0

Ссылки для проверки:

https://www.uniprot.org/blast/uniprot/B20220223F248CABF64506F2 9A91F8037F07B67D102FF8E4

Мой белок

https://www.uniprot.org/blast/uniprot/B20220223F248CABF64506F2 9A91F8037F07B67D102FF8E4?alignment=1

1ое совпадение

https://www.uniprot.org/uniprot/P00395

2ое совпадение

https://www.uniprot.org/uniprot/Q9T9W1