

14 АПРЕЛЬ, 2022



Olympiad League

НАЦИОНАЛЬНАЯ ОЛИМПИАДА ПО БИОЛОГИИ НВО-4 2022

4-Этап Практика –Биоинформатика Задания

OLYMPIAD LEAGUE

Биоинформатика. Часть А

Ген SMA-4 *Caenorhabditis elegans* является членом семейства генов *dwarfin*, также называемого семейством MAD, которое играет роль в преобразовании передачи сигнала, опосредованной бета-фактором роста.

Используя аминокислотную последовательность SMA4 из *C.elegans*, найдите гомологи белка SMA-4 (SMA4_CAEEEL, Accession P45897) у человека (*Homo sapiens*), кошки (*Felis catus*) и плодовой мушки (*D.melanogaster*) с использованием UCSC Genome. браузерный инструмент - BLAT.

```
>sp|P45897|SMA4_CAEEEL Dwarfin sma-4 OS=Caenorhabditis elegans OX=6239
GN=sma-4 PE=2 SV=1
MFHFGMTSQPSTSNQMYDPLYGAEQIVQCNPMDYHQANILCGMQYFNNSHNRYPLLPQM
PPQFTNDHPYDFPNVPTISTLDEASSFNGFLIPSQPSSYNNNNISCVFTPTPTCTSSQASS
QPPPTPTVNPTPIPPNAGAVLTMTAMDSCQQISHVLQCYQQGGEDSDFVRKAIESLVKKLK
DKRIELDALITAVTSNGKQPTGCVTIQRSLDGRLQVAGRKGVPVHVYARIWRWPKVSKNE
LVKLVQCQTSSDHPDNICINPYHYERVVSNRITSADQSLHVENSPMKSEYLG DAGVIDSC
SDWPNTPPDNNFNGGFAPDQPQLVTPIISDIPIDLNQIYVPTPPQLLDNWCSIYYELDT
PIGETFKVSARDHGKVIYVDGGMDFHGENEGRCLGALSNVHRTEASEKARIHIGRGVELT
AHADGNISITSNCKIFVRSGLDYTHGSEYSSKAHRFTPNESFTVFDIRWAYMQMLRRS
RSSNEAVRAQAAAVAGYAPMSVMPAIMPDSGVDRMRDFCTIAISFVKAWGDVYQRKTIK
ETPCWIEVTLHRPLQILDQLLNSSQFGSS
```

По результатам BLAT ответьте на следующие вопросы.

1. На каких хромосомах человека, кошки и мухи расположены гомологи? (4 балла)
2. На основании оценок выравнивания BLAT белка SMA-4, какие организмы (человек, кошка или муха) являются ближайшими и самыми дальними родственниками *C.elegans*? (4 балла)
3. На какой хромосоме *C.elegans* находится исходный SMA-4? Какие начальное и конечное положения нуклеотидов? Каков размер SMA-4 гена в парах нуклеотидов (bp)? Экзоны и интроны в основном представлены в виде блоков и линий в браузере UCSC genome, соответственно. Проанализировав браузер, напишите, сколько интронов и экзонов имеет ген SMA-4? (8 баллов)

Редактирование генома — это группа технологий, которые дают ученым возможность изменять ДНК организма. Эти технологии позволяют добавлять, удалять или изменять генетический материал в определенных местах генома. Было разработано несколько подходов к редактированию генома. Хорошо известный из них называется CRISPR-Cas9, что является сокращением от сгруппированных регулярно расположенных коротких палиндромных повторов и CRISPR-ассоциированного белка 9. Давайте представим, что с помощью CRISPR-Cas9 вы добавили ген устойчивости к пуромицину (*puroR*) размером 603 п.н. -конец гена *GAPDH* человека. Линкер между *GAPDH* и *puroR* имеет длину 40 п.н. Теперь нужно проверить успешность вставки, выполнив простую ПЦР с использованием праймеров, нацеленных на *GAPDH* и регион, куда должен быть вставлен *puroR*.

Используйте инструмент UCSC In-silico PCR, чтобы ответить на следующие вопросы.

4. Какие праймеры подойдут для этой цели? (7 баллов)

- А) Прямой праймер-ТТГACTCCCTAGTGTCTGCT, обратный праймер-
CCTTCCCCATGGTGTCTGAG
- Б) Прямой праймер — GATGCCCCCATGTTCTGTCAT, обратный праймер —
CCATTCCCCAGCTCTCATACC
- В) Прямой праймер — CCATGTTCTGTCATGGGTGTGA, обратный праймер —
GAGCGTGTCCATAGGGTGCC.
- Д) Прямой праймер — GGGGAGCCAAAAGGGTCAT, обратный праймер —
GAGAAAGGTGGGAGCCTCAGT.

5. Каков размер продукта ПЦР, в котором вы использовали праймеры, выбранные вами в вопросе 4, и геном человека дикого типа? (3 балла)
6. Каков размер продукта ПЦР, если вы использовали праймеры, которые вы выбрали в вопросе 4, и геном, который вы отредактировали, вставив puoR, как указано выше? (4 балла)



Olympiad League

Биоинформатика. Часть Б

В этой части биоинформатики вы будете восстанавливать филогенетическое древо гена Sonic Hedgehog используя базу данных портала NCBI, алгоритм выравнивания BLAST, и метод UPGMA.

Sonic Hedgehog (SHH, с англ. «Сверхзвуковой ёжик») — семейство генов и соответствующих им белков, управляющих эмбриональным развитием нервной системы и скелетной системы организма, а также выполняющих ряд других физиологических функций. Белки, кодируемые этими генами, оказывают влияние на сегментацию тела дрозофилы. При выключении гена по всему телу мухи развиваются мелкие шипики. Гиперактивация гена вызывает развитие медуллобластомы мозга у детей и рака слизистой оболочки рта.

Задание 1. Собрать данные гена 'sonic hedgehog' у организмов:

- Человек
- Макака
- Собака
- Мышь
- Курица



Шаг 1 - Перейдите на главную страницу портала NCBI. Выберите категорию “HomoloGene” и введите название гена – “sonic hedgehog”. Затем нажмите кнопку “Search”.

Olympiad League

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

HomoloGene sonic hedgehog Search

NCBI Home

Resource List (A-Z)

- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

MANE is published in Nature! 06 Apr 2022

We are delighted to announce that three and a half years of hard work by the

NCBI hidden Markov models (HMM) release 8.0 now available! 31 Mar 2022

Release 8.0 of the NCBI Hidden Markov

BLAST+ 2.13.0 now available with SRA BLAST, ARM Linux executables, and database metadata 29 Mar 2022

[More...](#)

Шаг 2 – Перед вами хранилище гомологичных генов sonic hedgehog. Вам нужно скачать эту базу данных нажав на "Download"

NCBI Resources How To Sign in to NCBI

HomoloGene HomoloGene sonic hedgehog Search

Create alert Limits Advanced Help

Display Settings: HomoloGene Send to:

HomoloGene:30961. Gene conserved in Bilateria

Genes
Genes identified as putative homologs of one another during the construction of HomoloGene.

SHH, *H.sapiens*
sonic hedgehog
SHH, *P.troglodytes*
sonic hedgehog
SHH, *M.mulatta*
sonic hedgehog
SHH, *C.lupus*
sonic hedgehog
SHH, *B.taurus*
sonic hedgehog
Shh, *M.musculus*
sonic hedgehog
Shh, *R.norvegicus*
sonic hedgehog
SHH, *G.gallus*
sonic hedgehog
shh, *X.tropicalis*
sonic hedgehog
shha, *D.renio*
sonic hedgehog a
shhb, *D.renio*
sonic hedgehog b
hh, *D.melanogaster*
hedgehog
AgaP_AGAP001412, *A.gambiae*
AgaP_AGAP001412

Proteins
Proteins used in sequence comparisons and their conserved domain architectures.

NP_000184.1 462 aa
XP_001147185.2 465 aa
XP_001106515.2 465 aa
XP_850450.1 461 aa
XP_002687031.2 473 aa
NP_033196.1 437 aa
NP_058917.1 437 aa
NP_990152.1 425 aa
XP_002932544.2 370 aa
NP_571138.1 418 aa
NP_571274.2 416 aa
NP_001034065.1 471 aa
XP_321721.5 509 aa

Download , links

Protein Alignments
Protein multiple alignment, pairwise similarity scores and evolutionary distances.

Conserved Domains
Conserved Domains from CDD found in protein sequences by rpsblast searching.

Шаг 3 - Перед вами появится загрузчик. Выберите формат мРНК и гены интересующих видов. *H. sapiens*, *M. mulatta*, *C. lupus*, *M. musculus*, *G. gallus* – это человек, макака, собака, мышь и курица соответственно.

Затем нажмите “Download”. Вы скачаете текстовый документ с выбранными генами в формате FASTA.

HomoloGene Downloader

[Homologene:30961](#). Gene conserved in Bilateria

Download **mRNA** sequences (in FASTA format)

Include 0 bp upstream of gene

Include 0 bp downstream of gene

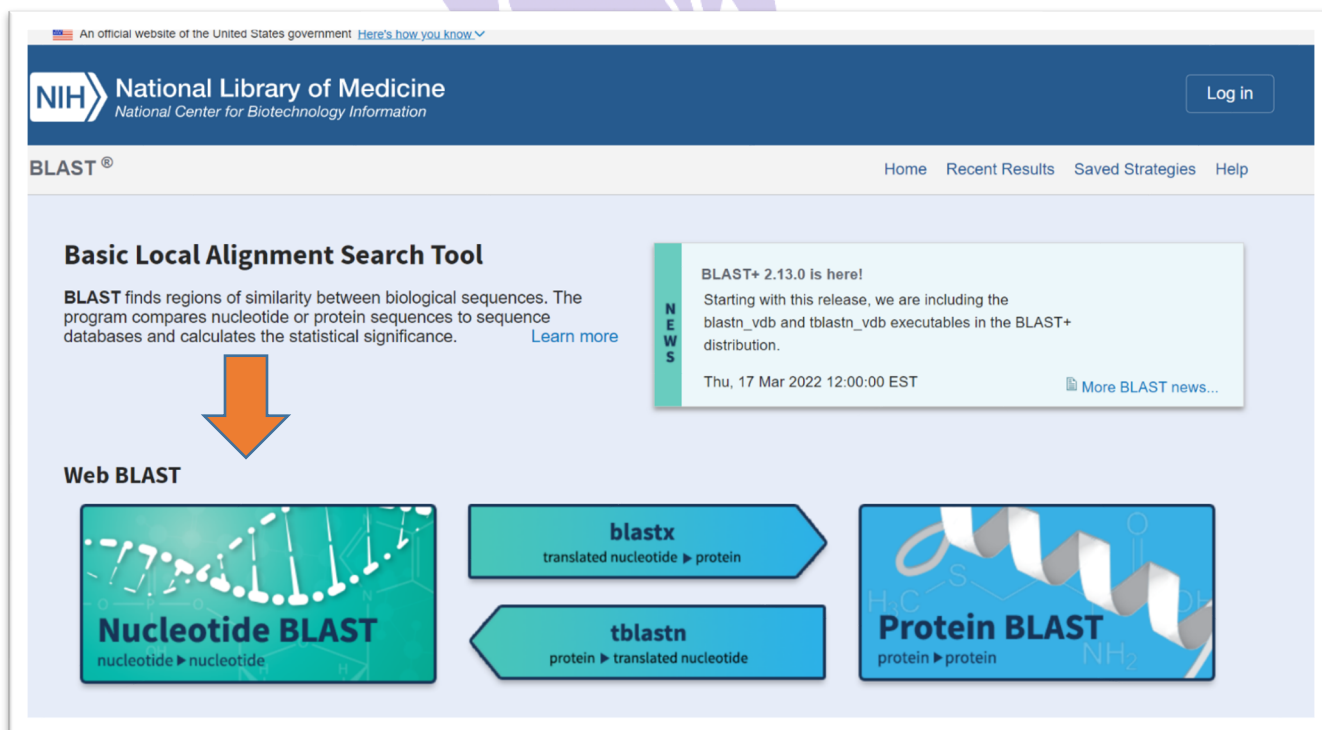
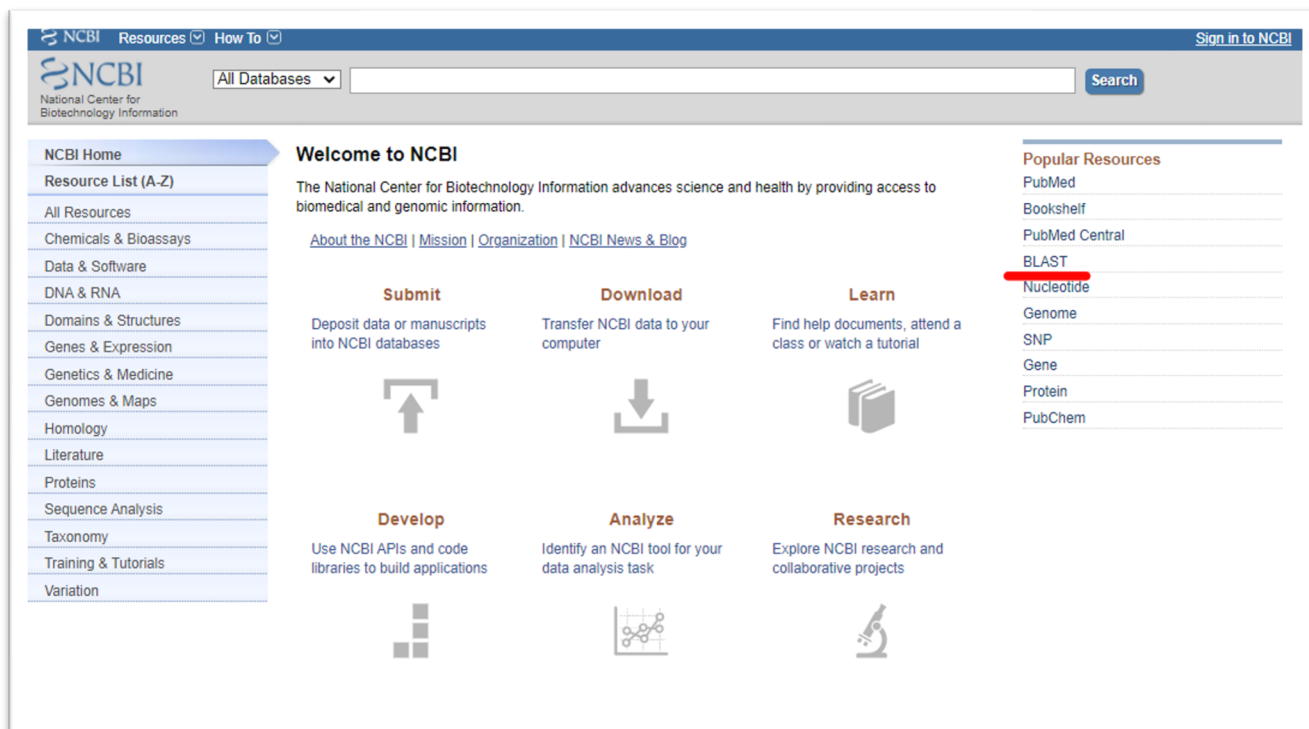
Select which sequences should be included

Select All Unselect All

	Species	Gene	mRNA	Protein
<input checked="" type="checkbox"/>	<i>H.sapiens</i>	SHH	NM_000193.2	NP_000184.1
<input type="checkbox"/>	<i>P.troglodytes</i>	SHH	XM_001147185.2	XP_001147185.2
<input checked="" type="checkbox"/>	<i>M.mulatta</i>	SHH	XM_001106515.2	XP_001106515.2
<input checked="" type="checkbox"/>	<i>C.lupus</i>	SHH	XM_845357.2	XP_850450.1
<input type="checkbox"/>	<i>B.taurus</i>	SHH	XM_002686985.2	XP_002687031.2
<input checked="" type="checkbox"/>	<i>M.musculus</i>	Shh	NM_009170.3	NP_033196.1
<input type="checkbox"/>	<i>R.norvegicus</i>	Shh	NM_017221.1	NP_058917.1
<input checked="" type="checkbox"/>	<i>G.gallus</i>	SHH	NM_204821.1	NP_990152.1
<input type="checkbox"/>	<i>D.erio</i>	shha	NM_131063.1	NP_571138.1
<input type="checkbox"/>	<i>D.erio</i>	shhb	NM_131199.2	NP_571274.2
<input type="checkbox"/>	<i>D.melanogaster</i>	hh	NM_001038976.1	NP_001034065.1
<input type="checkbox"/>	<i>A.gambiae</i>	AgaP_AGAP001412	XM_321721.5	XP_321721.5
<input type="checkbox"/>	<i>X.tropicalis</i>	shh	XM_002932498.2	XP_002932544.2

Задание 2. Восстановить филогенетическое древо при помощи BLAST.

Шаг 1 – На главной странице NCBI переходим на портал BLAST. Затем выбираем режим – “nucleotide to nucleotide”.



Шаг 2 – Нажмите на галочку “Align two or more sequences”. В “Query Sequence” введите цепь ДНК человека, ген которого будем сравнивать с остальными видами. В секцию “Subject Sequence” введите цепи оставшихся видов.



NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite

Align Sequences Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide subjects using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>Human
GCGAGGGCAGCCAGCGAGGGGAGAGCGAGCGGGCGAGCCGGAGCGA
GGAAGGGAAAGCGCAAGAGAGAGC
GCACACGCACACCCCGCGCGCACTCGCGCACGGACCCGCACG

Query subrange [?](#)

From

To

Or, upload file No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☒ Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>Macaca
CAGCCGGCGAGGGAGCGAGCGGGGCGAGCCGGAGCGAGGAAGG
GAGAGCGCAAGAGAGAGCGCACAC
GCACACACCCCGCGCGCACTCGCGCACGGACCCGCACGGGGACAG

Subject subrange [?](#)

From

To

Or, upload file No file chosen [?](#)

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)


Choose a BLAST algorithm [?](#)

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

+ Algorithm parameters

Убедившись, что данные введенны правильно, нажмите “BLAST” для выравнивание последовательностей.



National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST » blastn suite-2sequences » results for RID-5A99RECR114

Home
Recent Results
Saved Strategies
Help

Edit Search
Save Search
Search Summary

How to read this report?
BLAST Help Videos
Back to Traditional Results Page

Job Title
Human

RID
5A99RECR114
Search expires on 04-13 15:07 pm
Download All

Program
Blast 2 sequences
Citation

Query ID
lcl|Query_12107 (dna)

Query Descr
Human

Query Length
1576

Subject ID
lcl|Query_12109 and 4 more subject(s) (dna)

Subject Descr
See details

Subject
10143

Length

Filter Results

Percent Identity

to

E value

to

Query Coverage

to

Filter
Reset

Descriptions
Graphic Summary
Alignments

Sequences producing significant alignments

Download
Select columns
Show
100

☒ select all
4 sequences selected

Graphics
Distance tree of results
MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Macaca		2590	2590	99%	0.0	98.43%	1570	Query_12109
<input checked="" type="checkbox"/>	Dog		1889	1889	99%	0.0	88.25%	1941	Query_12110
<input checked="" type="checkbox"/>	Mouse		1164	1485	94%	0.0	88.30%	2727	Query_12111
<input checked="" type="checkbox"/>	Chicken		854	854	54%	0.0	84.73%	1535	Query_12112

BLAST®

U.S. National Library of Medicine | NCB | National Center for Biotechnology Information

Sign in to NCBI

Home Recent Results Saved Strategies Help

Blast Tree View

This tree was produced using BLAST pairwise alignments. [More...](#)

Result Tree

BLAST RID: **5A3BFFV114** Query ID: **klQuery_21431** Database: **n/a**

Tree method: Fast Minimum Evolution Max Seq Difference: 0.75 Sequence Label: Sequence Title (if avail)

Hide Inlined

Label color map: Query (yellow), From type material (grey)

blast names color map: urbinense (blue)

Find: all + TMT [icons] [icons] [icons] [icons] [icons]

Tools • Upload • [icons]

Success Nodes 9(0 selected) View port at (0.0) of 1942604 (0.002)

Шаг 5 – скачайте филогенетическое древо в формате PDF.

The screenshot shows the NCBI PhyML web interface. At the top, there is a blue header with "Sign in to NCBI". Below it, a navigation bar contains links for "Home", "Recent Results", "Saved Strategies", and "Help". The main content area has a light blue background with the text "Database n/a". A yellow tooltip message reads: "Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download". A "Tools" dropdown menu is open, showing options: "Download", "Layout", "Sort", "Zoom behavior", "Clear selection", "Clear subtree", "Clear rerooting", "Expand all", and "Edit labels". The "Download" option is selected, and a sub-menu is visible with options: "ASN text file", "ASN binary file", "Newick file", "NEXUS file", and "PDF file". The "PDF file" option is highlighted. On the right side, there is a "Label color map" section with a "Query" input field and a "names color map" section with a "Unknown" input field. The phylogenetic tree is displayed with labels "Dog", "Human", and "Macaca". A mouse cursor is pointing at the "Human" label.

Шаг 6 – загрузите файл в папку со своим именем. (4 балла)

Задание 3 – реконструировать филогенетическое древо через UPGMA

Шаг 1 – на предыдущей странице с результатом BLAST найдите иконку “Alignments”.

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite-2sequences » results for RID-5A99RECR114

Home Recent Results Saved Strategies Help

[← Edit Search](#) [Save Search](#) [Search Summary ▾](#) [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title Human

RID [5A99RECR114](#) Search expires on 04-13 15:07 pm [Download All ▾](#)

Program Blast 2 sequences [Citation ▾](#)

Query ID Icl|Query_12107 (dna)

Query Descr Human

Query Length 1576

Subject ID Icl|Query_12109 and 4 more subject(s) (dna)

Subject Descr [See details ▾](#)

Subject 10143

Length

Filter Results

Percent Identity to E value to Query Coverage to

[Filter](#) [Reset](#)

Alignments

Sequences producing significant alignments

Download ▾ Select columns ▾ Show 100 ▾

☒ select all 4 sequences selected

[Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

	Description ▾	Scientific Name ▾	Max Score ▾	Total Score ▾	Query Cover ▾	E value ▾	Per. Ident ▾	Acc. Len ▾	Accession
<input checked="" type="checkbox"/>	Macaca		2590	2590	99%	0.0	98.43%	1570	Query_12109
<input checked="" type="checkbox"/>	Dog		1889	1889	99%	0.0	88.25%	1941	Query_12110
<input checked="" type="checkbox"/>	Mouse		1164	1485	94%	0.0	88.30%	2727	Query_12111
<input checked="" type="checkbox"/>	Chicken		854	854	54%	0.0	84.73%	1535	Query_12112

Olympiad League

[illegible]

Шаг 3 – перейдите к “Query range 5: 241 to 300”. Если допустить, что каждый отличающийся нуклеотид это единица условной длины расхождения видов, то заполните таблицу для филогенетического дерева UPGMA подсчитав различия в нуклеотидах между каждой цепью. Заполните таблицу в листе ответов. (5 баллов)

	A - Human	B - Macaca	C - Dog	D - Mouse	E - Chicken
A - Human	0				
B - Macaca		0			
C - Dog			0		
D - Mouse				0	
E - Chicken					0

Шаг 4 – Используйте таблицу из шага 3 чтобы закончить вычисления расстояний между видами и заполнить оставшиеся таблицы. Заполните таблицы в листе ответов. (6 баллов)

Шаг 5 – используйте полученные данные чтобы реконструировать филогенетическое древо. Древо зарисуйте в листе ответов. (4 балла)



Olympiad League

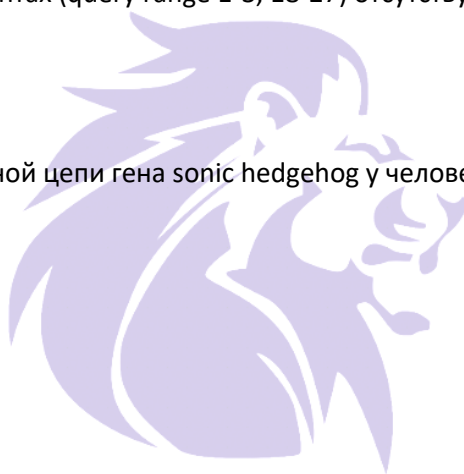
Задание 4. Письменные Вопросы.

4.1 Одинакова ли древо вычисленная алгоритмами BLAST и древо вами построенное с помощью UPGMA? Почему? (2 балла)

4.2 Если изготовить 60 нуклеотидный праймер специфический только к людям и макакам, то какой сегмент цепи (query range) вы бы использовали и почему? (2 балла)

4.3 Почему на некоторых сегментах (query range 1-3, 18-27) отсутствует днк цепь курицы? (2 балла)

4.4 Какова длина аминокислотной цепи гена sonic hedgehog у человека, макаки, собаки и курицы? (2 балла)



Olympiad League