Молекулярная биология

Некоторые параметры молекул ДНК и белка:

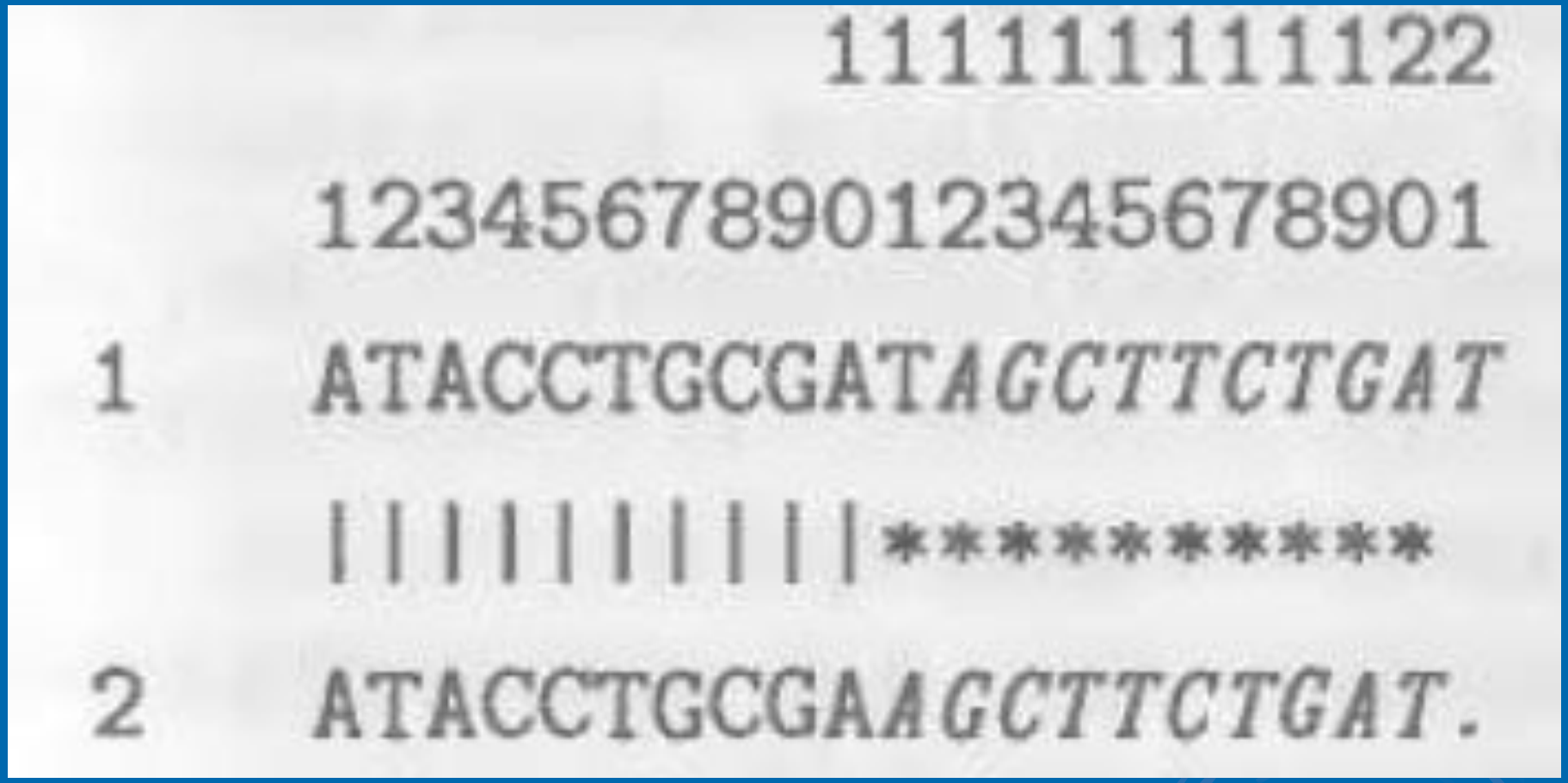
- Один шаг это полный виток спирали ДНК-поворот на 360^0
- Один шаг составляют 10 пар нуклеотидов
- Длина одного шага - 3,4 нм
- Расстояние между двумя нуклеотидами - 0,34 нм
- Молекулярная масса одного нуклеотида - 345 г/моль
- Молекулярная масса одной аминокислоты – 100 г/мол
- В молекуле ДНК: А+Г=Т+Ц (Правило Чаргаффа
- Комплементарность нуклеотидов: А=Т; Г=Ц
- Цепи ДНК удерживаются водородными связями, которые образуются между комплементарными азотистыми основаниями: аденин с тимином соединяются 2 водородными связями, а гуанин с цитозином тремя.
- В среднем один белок содержит 400 аминокислот

Выравнивание генетических последовательностей

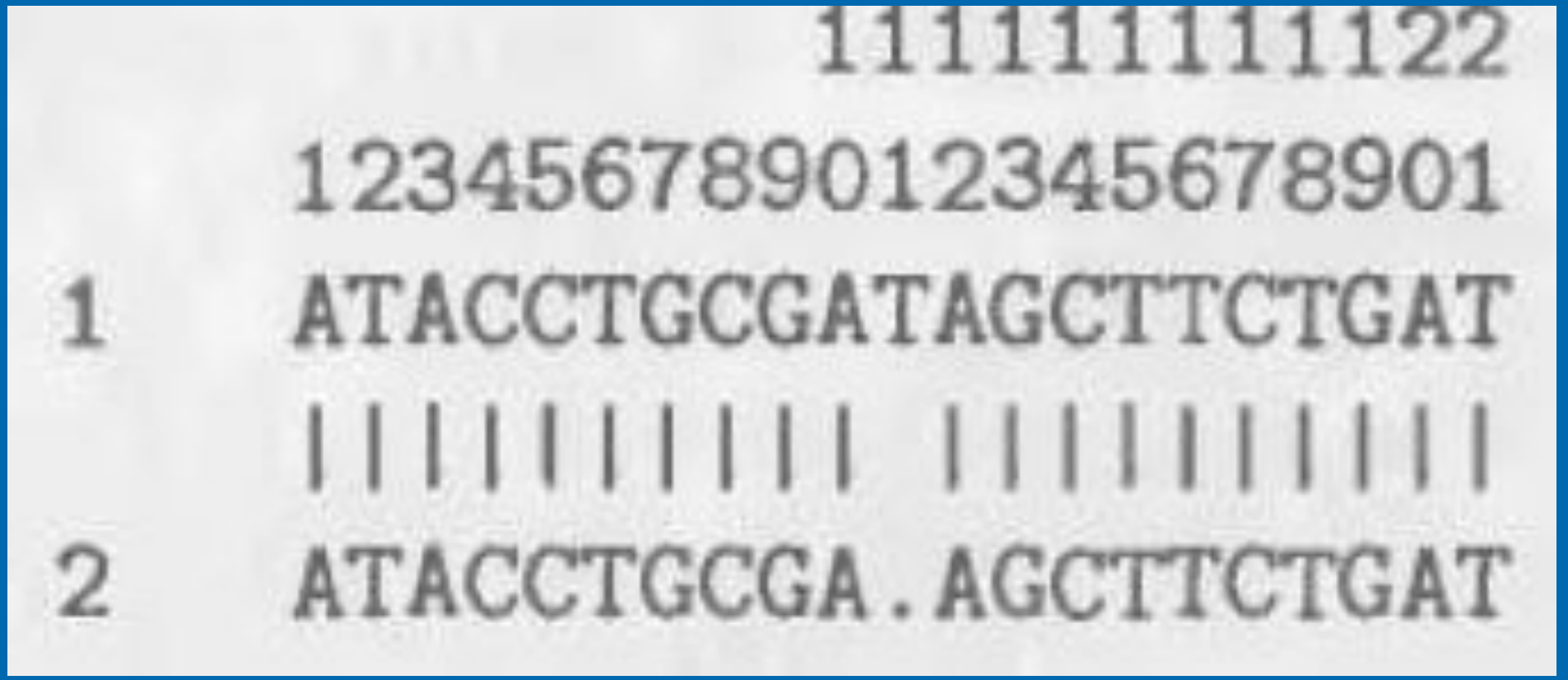
- В эволюции генетических последовательностей происходят как замены, так и вставки и делеции. Первым этапом филогенетического анализа является идентификация вставок и делеций, имевших место в эволюционной истории анализируемой группы последовательностей. Эту процедуру называют выравниванием (to align, alignment) последовательностей.

- Выравнивание последовательностей направлено на выявление гомологичных (имеющих общее эволюционное происхождение) позиций анализируемых последовательностей, установление наиболее вероятного, т.е. требующего наименьшего числа эволюционных событий, сценария эволюции анализируемой группы.

Выравнивание генетических последовательностей



Выравнивание генетических последовательностей



Выравнивание генетических последовательностей

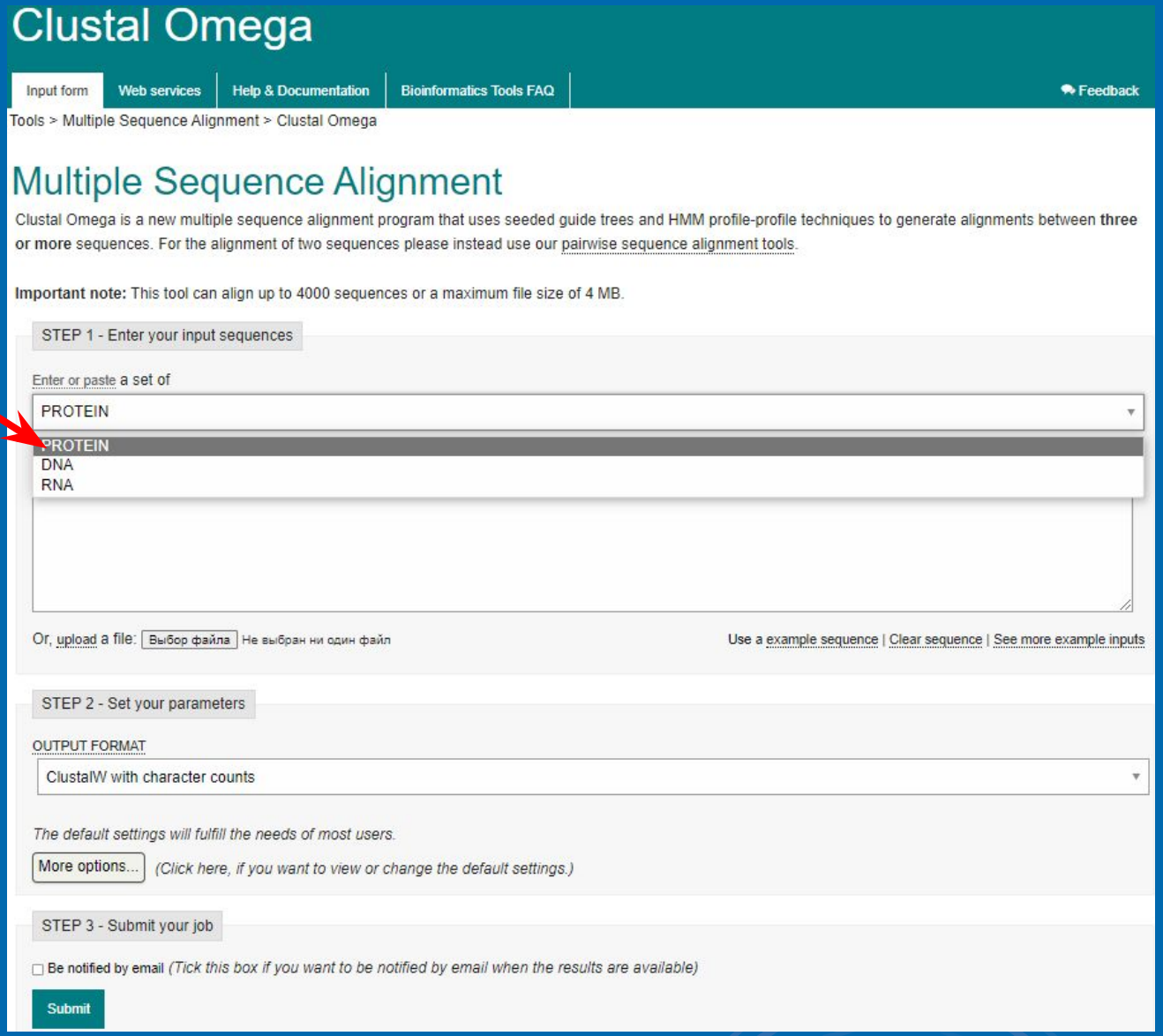
- **Clustal -- это одна из самых широко используемых компьютерных программ для множественного выравнивания нуклеотидных и аминокислотных последовательностей (multiple sequence alignment).**

- **Переходим по ссылке:**

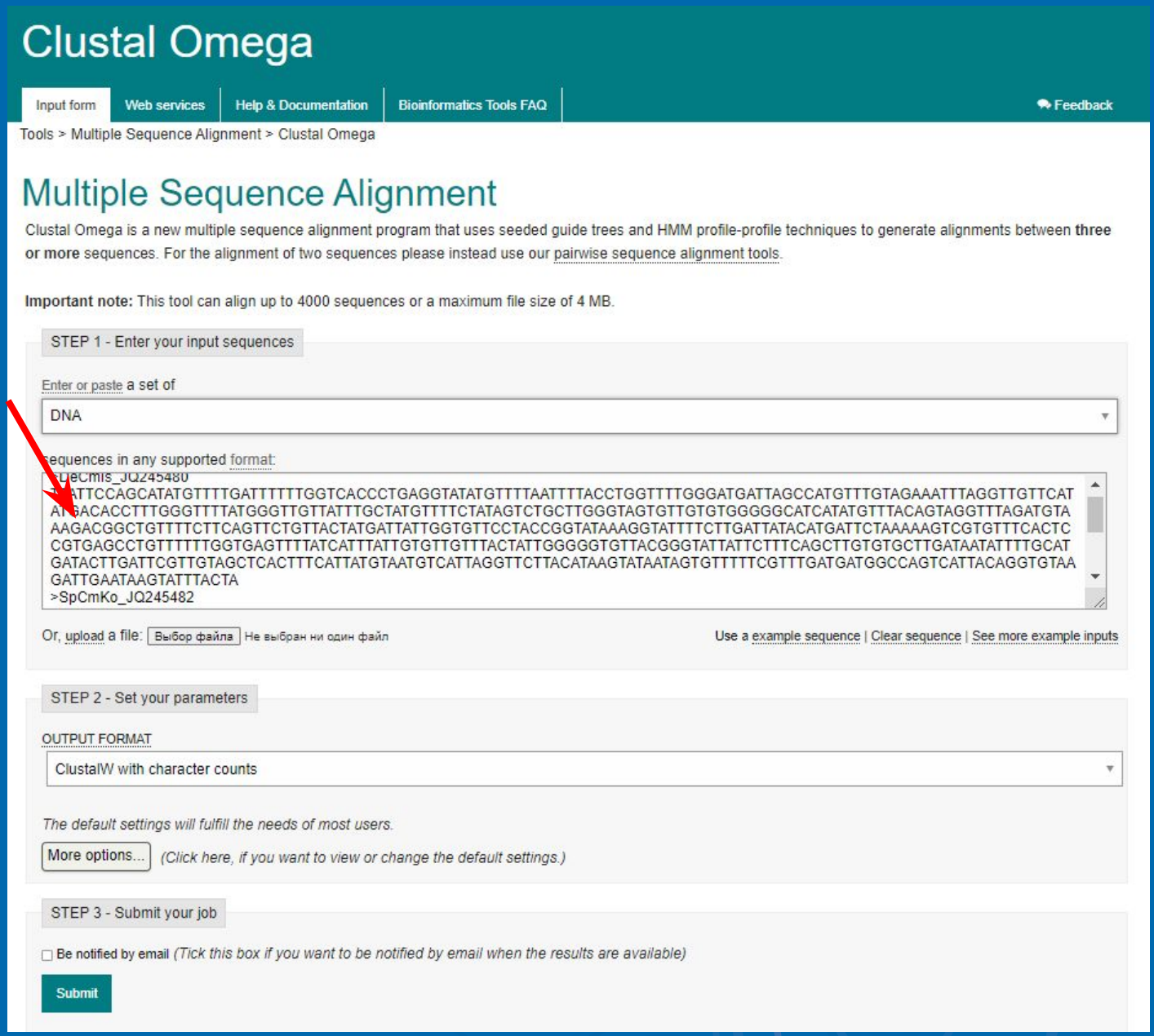
- **https://www.ebi.ac.uk/Tools/msa/clustalo/**

1. Выбираем нужный тип данных (белок, ДНК или РНК)

В нашем случае - DNA



2. Вставляем последовательности
Открываем папку Выравнивания.
Открываем файл B-11.fasta программой UltraEdit-32.
Копируем оттуда 2 или более понравившиеся последовательности



Clustal Omega

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ | Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

DNA

sequences in any supported format:

```
>DeCmIs_JQ245480
TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCTGGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCAT
ATGACACCTTTGGGTTTTATGGGTTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTTGTGTGGGGGCATCATATGTTTACAGTAGGTTTAGATGTA
AAGACGGCTGTTTTCTTCAGTTCTGTTACTATGATTATTGGTGTTCCTACCGGTATAAAGGTATTTTCTTGATTATACATGATTCTAAAAAGTCGTGTTTCACTC
CGTGAGCCTGTTTTTTGGTGAGTTTTATCATTTATTGTGTTGTTTACTATTGGGGGTGTTACGGGTATTATTCTTTCAGCTTGTGTGCTTGATAATATTTTGCAT
GATACTTGATTCGTTGTAGCTCACTTTCATTATGTAATGTCATTAGGTTCTTACATAAGTATAATAGTGTTTTTCGTTTGATGATGGCCAGTCATTACAGGTGTAA
GATTGAATAAGTATTTACTA
>SpCmKo_JQ245482
```

Or, upload a file: Выбор файла Не выбран ни один файл

Use a example sequence | Clear sequence | See more example inputs

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

More options... *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

☐ Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Submit

3. Нажимаем Submi

4. Получаем результат

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ

```
DeCmCh1.2-Hap1                  GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
SpTaKo_JQ245484                 GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeClUK_KC812045                 GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
SpTaKo_JQ245483                 GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo11-Hap18                  GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeOkCL_AB623150                 GGTTTTGGAATGATTAGTCATGTTTGTAGAAACTTAGGTTGTTCATATGATACCTTTGGG  120
SpCmKo_JQ245482                 GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeOkCL_AB623149                 GGTTTTGGAATGATTAGTCATGTTTGTAGAAACTTAGGTTGTTCATATGATACCTTTGGG  120
KlUaKa_AB375661                 GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACTTTTGGG  120
NiUaKm_JQ245471                 GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACTTTTGGG  120
DiClbKa1-HapB                   GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACTTTTGGG  120
DiClbKa11-HapH                  GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACTTTTGGG  120
DiClbKa8.9-HapF                 GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACTTTTGGG  120
DiClbKa16-HapL                  GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACTTTTGGG  120
DiClbKa10-HapG                  GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACTTTTGGG  120
SpCmCh12-HapA                   GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACTTTTGGG  120
Dipl.grandis_AB425840           GGGTTTGGAATGATTAGACATGTTTGTAGTAACTTAGGTTGTTCATATGATACTTTTGGA  120
Dipl.balaenopterae_AB822370     GGGTTTGGAATGATTAGACATGTTTGTAGTAACTTAGGTTGTTCATATGATACTTTTGGA  120
DeSlAz_AB374223                 GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo13-Hap20                  GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo7.8-Hap15                 GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo10-Hap17                  GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
SpCmCh9-Hap12                   GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeCmSe10.11-Hap7                GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeCmCh5-Hap4                    GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeHsNE_KC812048                 GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo15-Hap22                  GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeBlHo14-Hap21                  GGTTTTGGGATGATTAGCCATGCTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo18-Hap25                  GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
DeTnHo16-Hap23                  GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTCATATGACACCTTTGGG  120
                                ** ***** ******** **** ******:** **.** ** **:***** ** ** **.

DiOtUS_KY552872                 TTTTATGGGTTGTTATTTGCCATGTTTTCTATTGTTTGTCTAGGTAGCGTTGTATGGGGG  180
DiClbKa5.6-HapD                 TTTTATGGGTTGTTGTTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGGGGA  180
DiCsTy_JQ245472                 TTTTATGGGTTGTTGTTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGGGGA  180
SpOmKi_JQ245477                 TTTTATGGGTTGTTGTTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGGGGA  180
DiHpJP_AB979518                 TTTTATGGGTTGTTGTTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGGGGA  180
DiClbKa12-HapI                  TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGTGTTGTATGAGGG  180
DiClbKa14.15-HapK               TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGAGGG  180
DiClbKa7-HapE                   TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGAGGA  180
DiClbKa13-HapJ                  TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGAGGA  180
DiClbKa2.3.4-HapC               TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGAGGA  180
DiCsTy_JQ245474                 TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGAGGA  180
DiSaUK_FM209182                 TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTTTGTTTAGGTAGCGTTGTATGAGGA  180
SpHsJP_AB488497                 TTTTACGGGTTGTTATTTGCTATGTTTTCTATAGTTTGCTTAGGTAGTGTTGTGTGGGGG  180
UrUaUS_AB605762                 TTTTATGGGTTATTATTTGCTATGTTTTCTATAGTCTGTTTGGGTAGTGTTGTGTGGGGG  180
UrUaUS_AB605763                 TTTTATGGGTTATTATTTGCTATGTTTTCTATAGTCTGTTTGGGTAGTGTTGTGTGGGGG  180
NiOkJP_AB521677                 TTTTACGGATTGTTATTTGCTATGTTTTCTATAGTTTGCTTAGGTAGTGTTGTGTGGGGG  180
DeTnHo4.5.6-Hap14               TTTTATGGGTTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTTGTGTGGGGG  180
DeTnHo2.3-Hap13                 TTTTATGGGTTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTTGTGTGGGGG  180
DeHsCH_AM412738                 TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTCTGCTTGGGTAGTGTTGTGTGGGGG  180
DeHsCZ_KC812047                 TTTTATGGGTTGTTATTTGCCATGTTTTCTATAGTCTGCTTGGGTAGTGTTGTGTGGGGG  180
DeCmSe14-Hap10                  TTTTATGGGTTGTTATTTGCTATGTTTTCTATAGTTTGCTTGGGTAGTGTTGTGTGGGGG  180
DeCmIs_JQ245480                 TTTTATGGGTTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTTGTGTGGGGG  180
```

5. Интерпретация

- **Звездочка (*) – различия по данной позиции (нуклеотид или аминокислота) отсутствуют между разными последовательностями**
- **Пробел ( ), точка (.) , двоеточие (:) – по данной позиции имеются различия между разными последовательностями**

Чтение выходных данных выравнивания нескольких последовательностей

Символ	Определение	Имея в виду
*	звездочка	позиции, которые имеют единственный и полностью консервативный остаток
:	двоеточие	сохранение между группами сильно схожих свойств с результатом более 0,5 по матрице PAM 250
.	период	сохранение между группами слабо сходных свойств с оценкой меньше или равной 0,5 по матрице PAM 250

- **Одни и те же символы показаны как для выравнивания ДНК / РНК, так и для выравнивания белков, поэтому, хотя символы * (звездочка) полезны для обоих, другие согласованные символы следует игнорировать при выравнивании ДНК / РНК.**

6. Здесь же можно посмотреть предварительное филогенетическое дерево

Clustal Omega

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-I20220228-023416-0897-41339244-p1m

Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details

Download Alignment File | Show Colors

```
CLUSTAL O(1.2.4) multiple sequence alignment


DiOtUS_KY552872                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiClbKa5.6-HapD                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAAGTGTATGTCTTAATTTTACCG   60
DiCsTy_JQ245472                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAAGTGTATGTCTTAATTTTACCG   60
SpOmKi_JQ245477                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAAGTGTATGTCTTAATTTTACCG   60
DiHpJP_AB979518                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAAGTGTATGTCTTAATTTTACCG   60
DiClbKa12-HapI                  TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiClbKa14.15-HapK               TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiClbKa7-HapE                   TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiClbKa13-HapJ                  TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiClbKa2.3.4-HapC               TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiCsTy_JQ245474                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
DiSaUK_FM209182                 TTATTTCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCG   60
SpHsJP_AB488497                 TTATTTCAGCATATGTTTTGATTTTTTGGCCATCCCGAAGTATATGTTTTAATTTTACCT   60
UrUaUS_AB605762                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCC   60
UrUaUS_AB605763                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTGTATGTCTTAATTTTACCC   60
NiOkJP_AB521677                 TTATTTCAGCATATGTTTTGATTTTTTGGACACCCCGAAGTATATGTTTTAATTTTACCT   60
DeTnHo4.5.6-Hap14               TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCCGAGGTATATGTTTTAATTTTACCT   60
DeTnHo2.3-Hap13                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCCGAGGTATATGTTTTAATTTTACCT   60
DeHsCH_AM412738                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATCTTACCT   60
DeHsCZ_KC812047                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATCTTACCT   60
DeCmSe14-Hap10                  TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeCmIs_JQ245480                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeCmBa_KC812046                 TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeCmSe13-Hap9                   TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeTnHo12-Hap19                  TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeCmSe12-Hap8                   TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeCmCh3-Hap2                    TTATTCCAGCATATGTTTTGATTTTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
DeTnHo17-Hap24                  TTATTCCAGCATATGTTTTGATTCTTTGGTCACCCTGAGGTATATGTTTTAATTTTACCT   60
```

=clustalo-I20220228-023416-0897-41339244-p1m&analysis=phylotree

Input form | Web services | Help & Documentation | Bioinformatics Tools FAQ

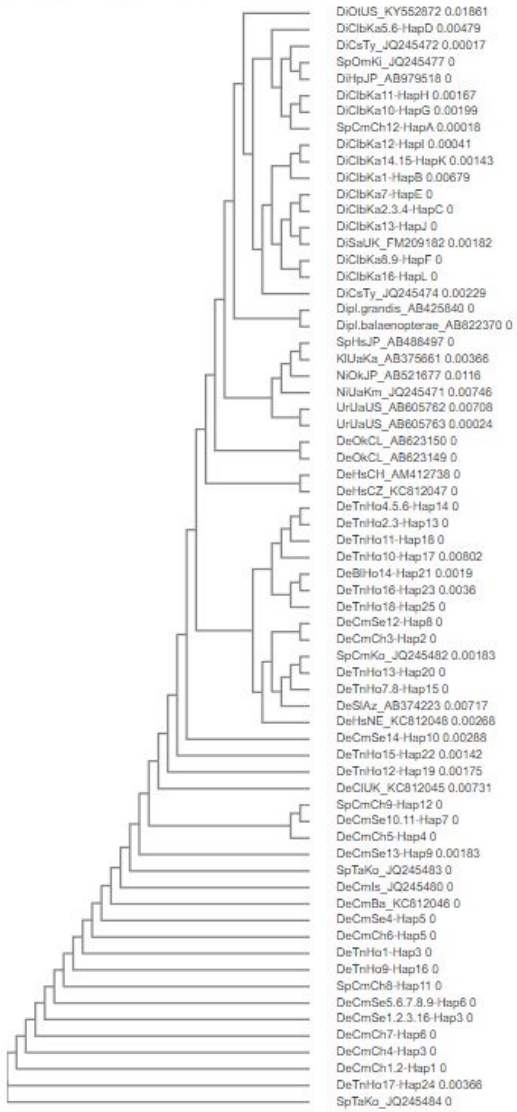
Alignments | Result Summary | Guide Tree | Phylogenetic Tree | Results Viewers | Submission Details

Download Phylogenetic Tree Data

Phylogenetic Tree

...ance corrections.

Branch length: ◉ Cladogram ○ Real



BLAST

- **BLAST (англ. Basic Local Alignment Search Tool — средство поиска основного локального выравнивания) — семейство компьютерных программ, служащих для поиска гомологов белков или нуклеиновых кислот, для которых известна первичная структура (последовательность) или её фрагмент.**
- **Используя BLAST, исследователь может сравнить имеющуюся у него последовательность с последовательностями из базы данных и найти последовательности предполагаемых гомологов.**
- **Является важнейшим инструментом для молекулярных биологов, биоинформатиков и систематиков.**

Классификация программ серии BLAST
Нуклеотидные

- предназначены для сравнения изучаемой нуклеотидной последовательности с базой данных секвенированных нуклеиновых кислот и их участков:

- blastn — медленное сравнение с целью поиска всех сходных последовательностей и др.
- megablast — быстрое сравнение с целью поиска высоко сходных последовательностей,
- dmegablast — быстрое сравнение с целью поиска дивергировавших последовательностей, обладающих незначительным сходством,

Классификация программ серии BLAST

Белковые

- предназначены для сравнения изучаемой аминокислотной последовательности белка с имеющейся базой данных белков и их участков.

- blastp — медленное сравнение с целью поиска всех сходных последовательностей,
- cdart — сравнение с целью поиска гомологичных белков по доменной архитектуре,
- rpsblast — сравнение с базой данных консервативных доменов,
- psi-blast — сравнение с целью поиска последовательностей, обладающих незначительным сходством,
- phi-blast — поиск белков, содержащих определённый пользователем паттерн и др.

Классификация программ серии BLAST

Транслирующие

- способны транслировать нуклеотидные последовательности в аминокислотные:
- blastx — переводит изучаемую нуклеотидную последовательность в кодируемые аминокислоты, а затем сравнивает её с имеющейся базой данных аминокислотных последовательностей белков,
- tblastn — изучаемая аминокислотная последовательность сравнивается с транслированными последовательностями базы данных секвенированных нуклеиновых кислот,
- tblastx — переводит изучаемую нуклеотидную последовательность в аминокислотную, а затем сравнивает её с транслированными последовательностями базы данных секвенированных нуклеиновых кислот.

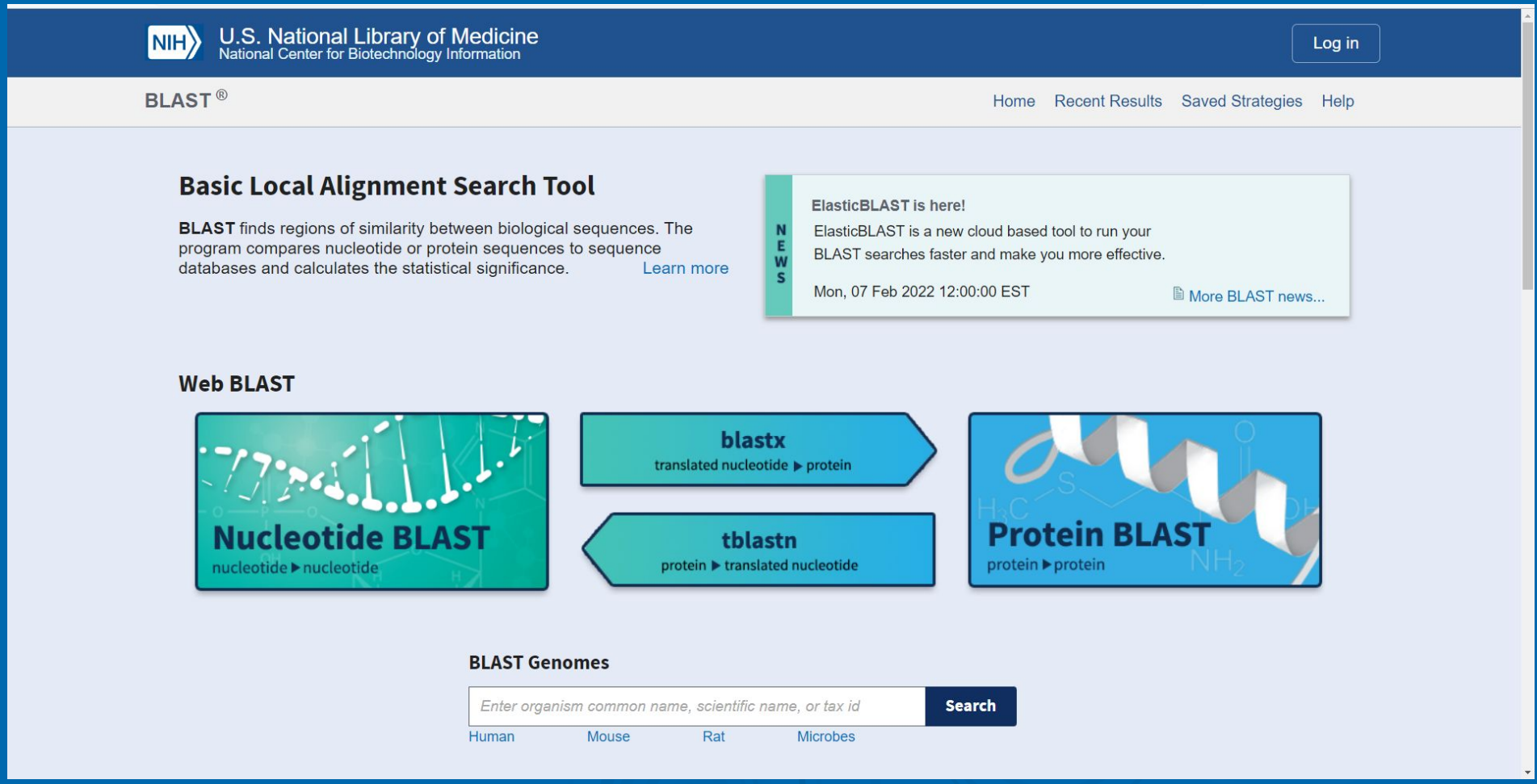
Классификация программ серии BLAST

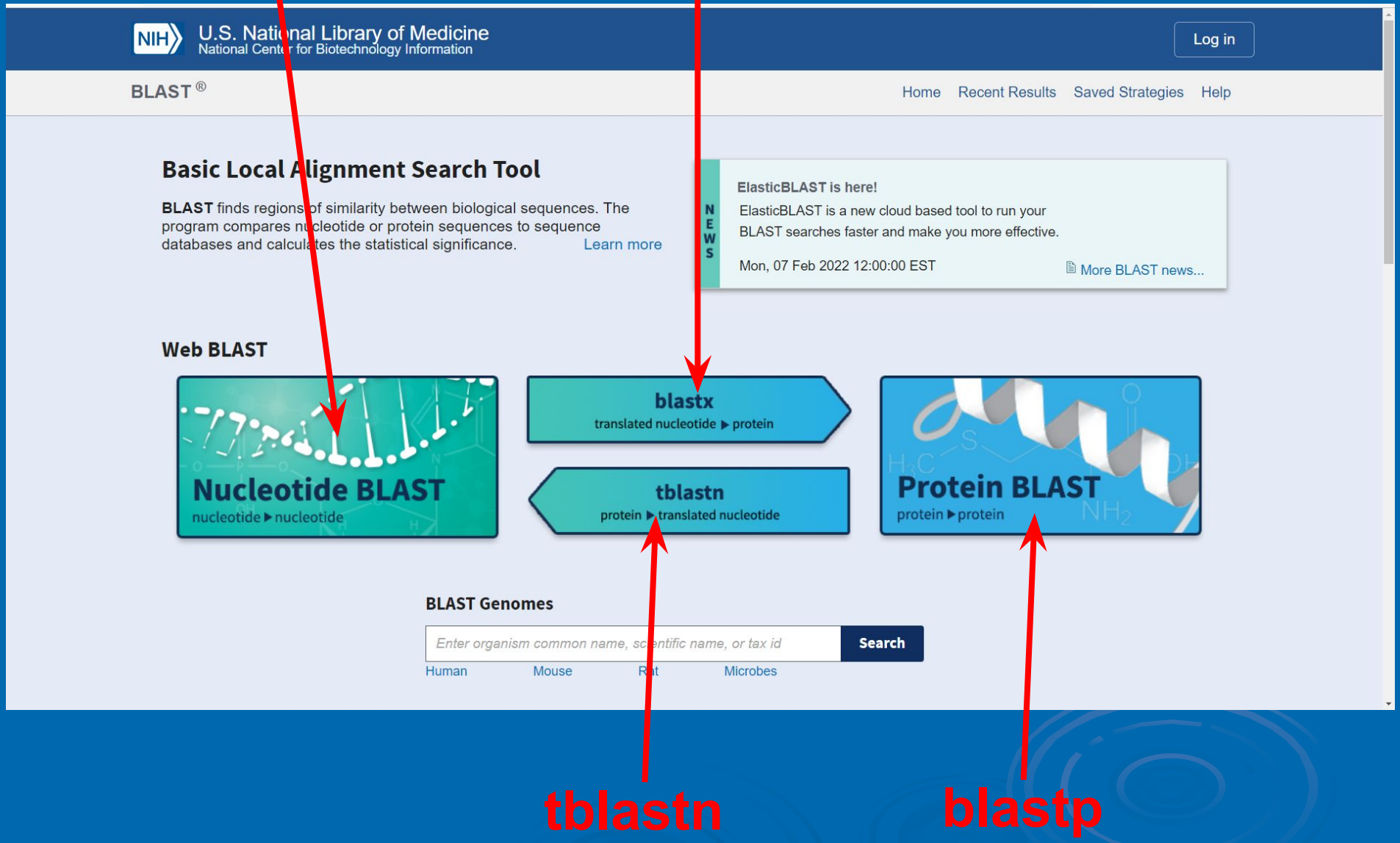
Геномные

- предназначены для сравнения изучаемой нуклеотидной последовательности с базой данных секвенированного генома какого-либо организма (человека, мыши и др.)

- magicblast — картирует прочтения (риды) на полный геном или транскриптом.

- **Переходим в сервис BLAST Национального центра биотехнологической информации США (NCBI) по ссылке:**
- **https://blast.ncbi.nlm.nih.gov/Blast.cgi**



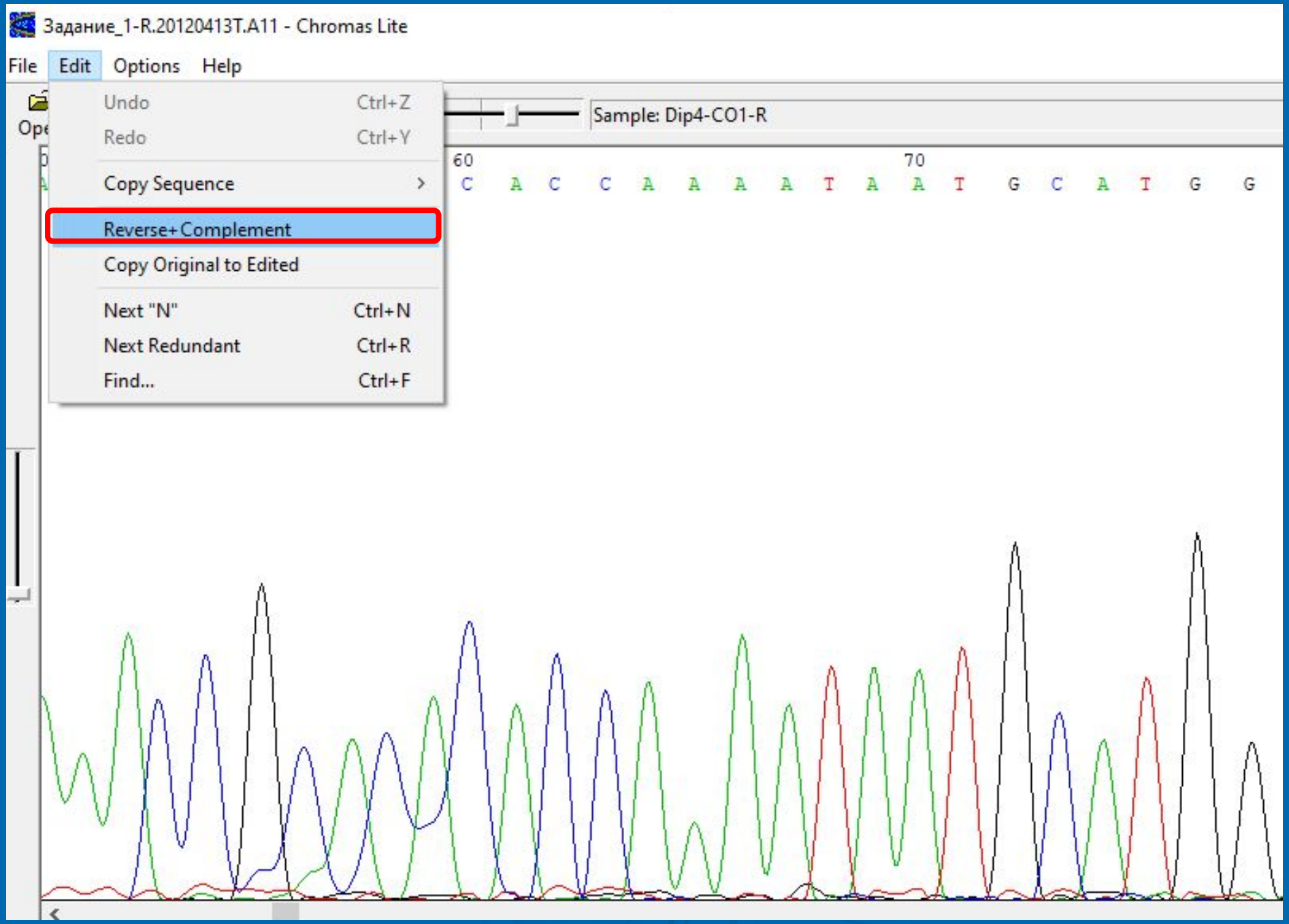

blastn
blastx
U.S. National Library of Medicine
National Center for Biotechnology Information
Log in
BLAST ®
Home
Recent Results
Saved Strategies
Help
Basic Local Alignment Search Tool
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.
Learn more
NEWS
ElasticBLAST is here!
ElasticBLAST is a new cloud based tool to run your BLAST searches faster and make you more effective.
Mon, 07 Feb 2022 12:00:00 EST
More BLAST news...
Web BLAST
Nucleotide BLAST
nucleotide ▶ nucleotide
blastx
translated nucleotide ▶ protein
tblastn
protein ▶ translated nucleotide
Protein BLAST
protein ▶ protein
BLAST Genomes
Enter organism common name, scientific name, or tax id
Search
Human
Mouse
Rat
Microbes
tblastn
blastp

Задача 1. Форма отчета

- Каждый лично на своем компьютере делает скриншот/фото (так, чтобы было видно номер компьютера/монитора, время на мониторе) списка гомологов в сервисе BLAST
- Каждый лично отправляет мне в Вайбере в личку: 1) полученное фото, а также 2) название вида, от которого получена последовательность нуклеотидов и 3) название гена, в котором содержится данная последовательность

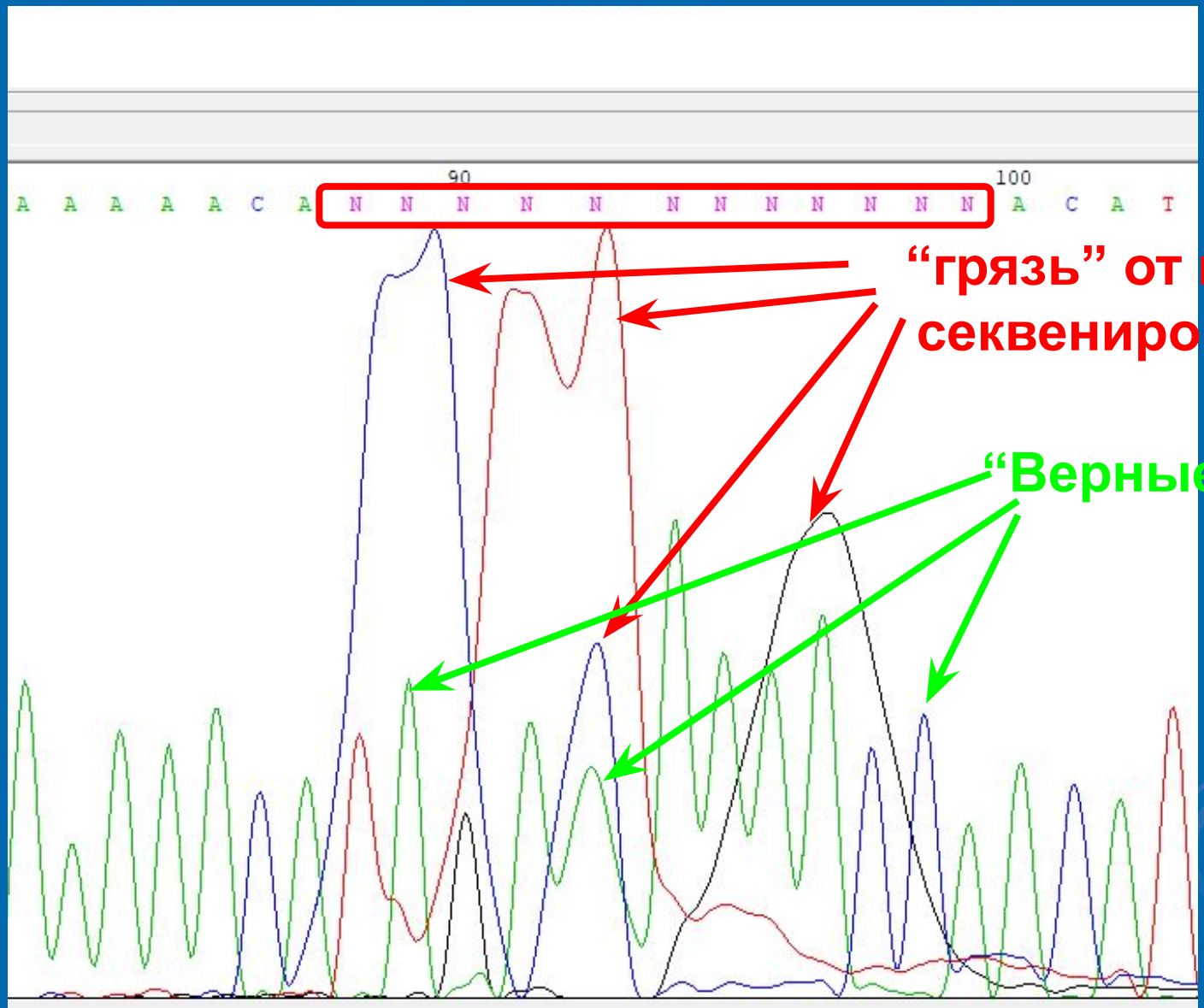
Задача 1

1. Открыть с помощью программы Chromas файл с хроматограммой Задание_1-R.20120413T.A11 из папки Задания
2. Это последовательность нуклеотидов, полученная путем секвенирования ДНК на основании обратного праймера reverse (R)
3. Поэтому для получения прямой последовательности необходимо применить функцию Reverse + Complement (RC) (т.е. перевернуть и получить комплементарную для нее последовательность):


Задание_1-R.20120413T.A11 - Chromas Lite
File
Edit
Options
Help
Undo Ctrl+Z
Redo Ctrl+Y
Copy Sequence
Reverse+Complement
Copy Original to Edited
Next "N" Ctrl+N
Next Redundant Ctrl+R
Find... Ctrl+F
Sample: Dip4-CO1-R
60
70
C A C C A A A A T A A T G C A T G G

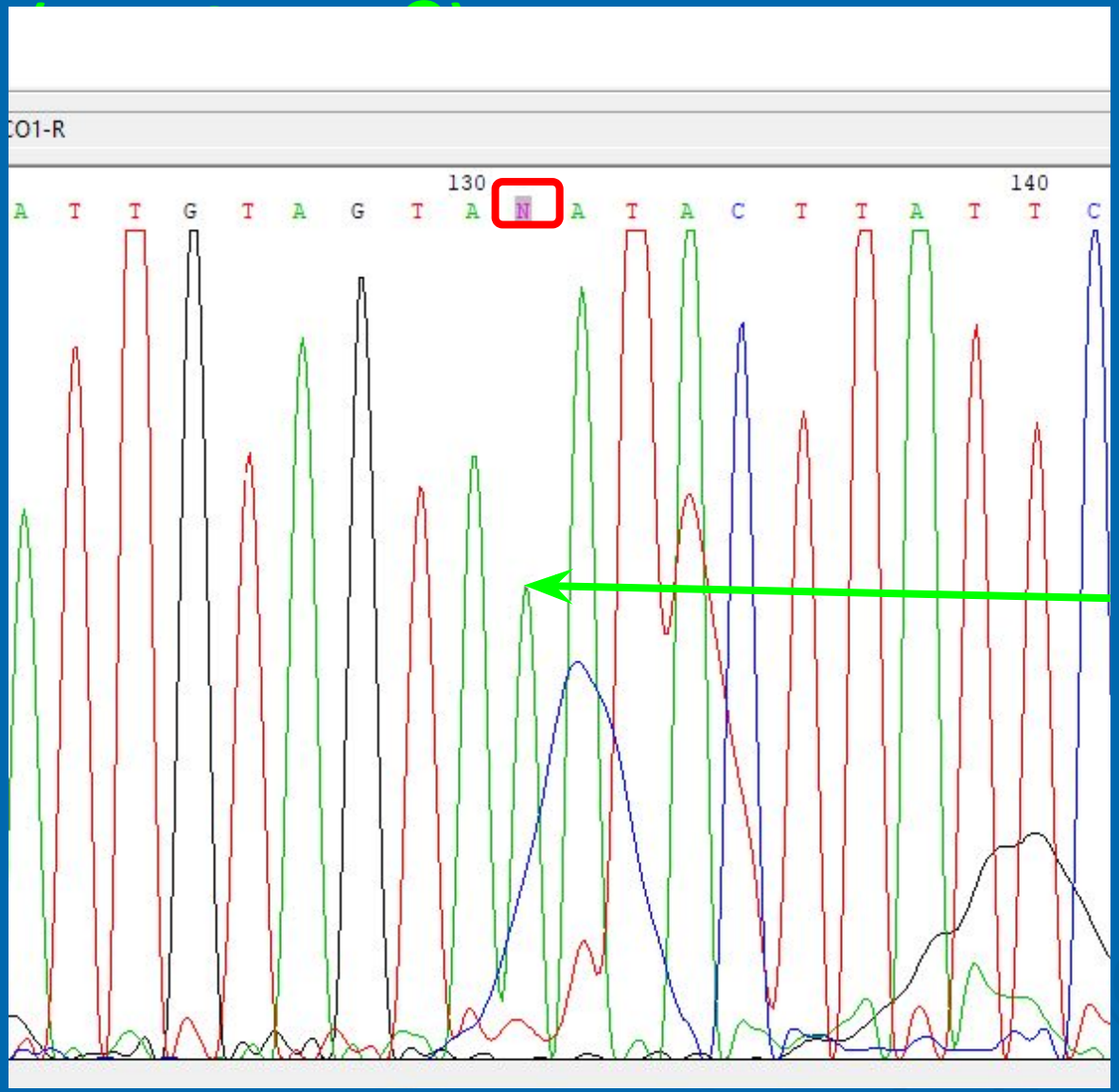
Задача 1

4. Вместо символов N поставить соответствующие хроматограмме нуклеотиды



Задача 1

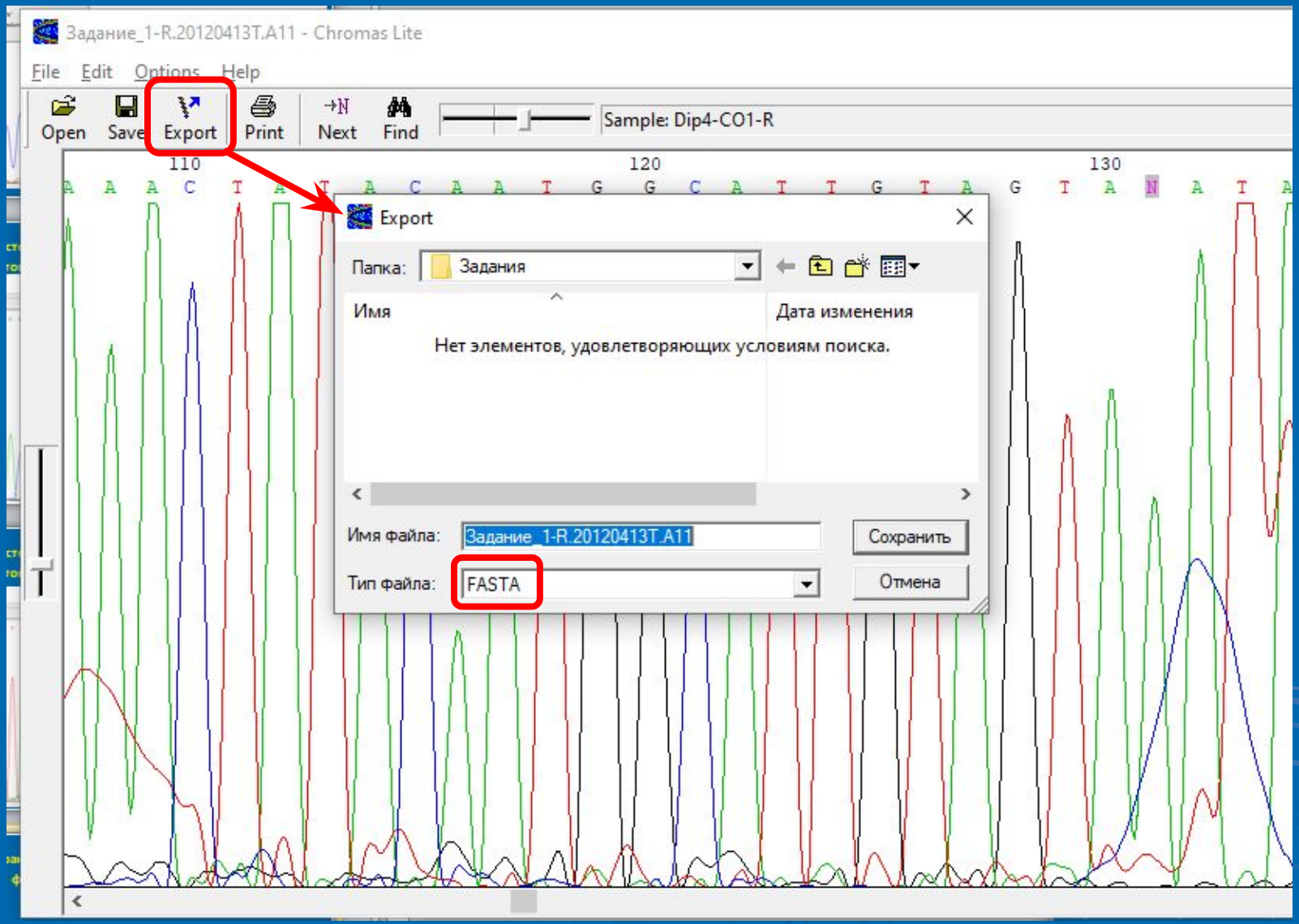
4. Вместо символов N поставить соответствующие хроматограмме нуклеотиды



“Верный нуклеотид”

Задача 1

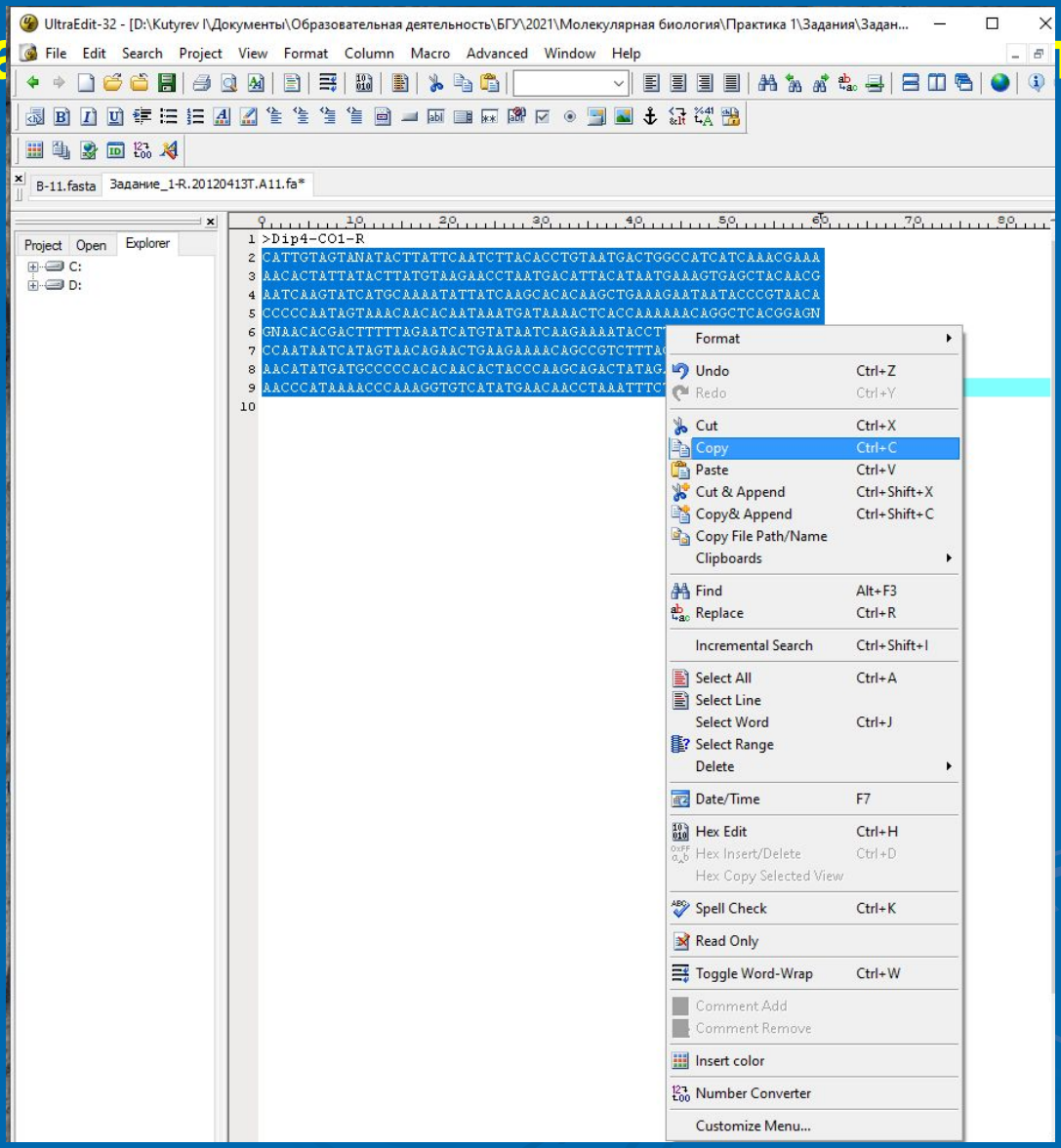
5. Сохранить отредактированную последовательность в новый файл в формате



Задача 1

6. Открыть сохраненный файл в программе UltraEdit-32

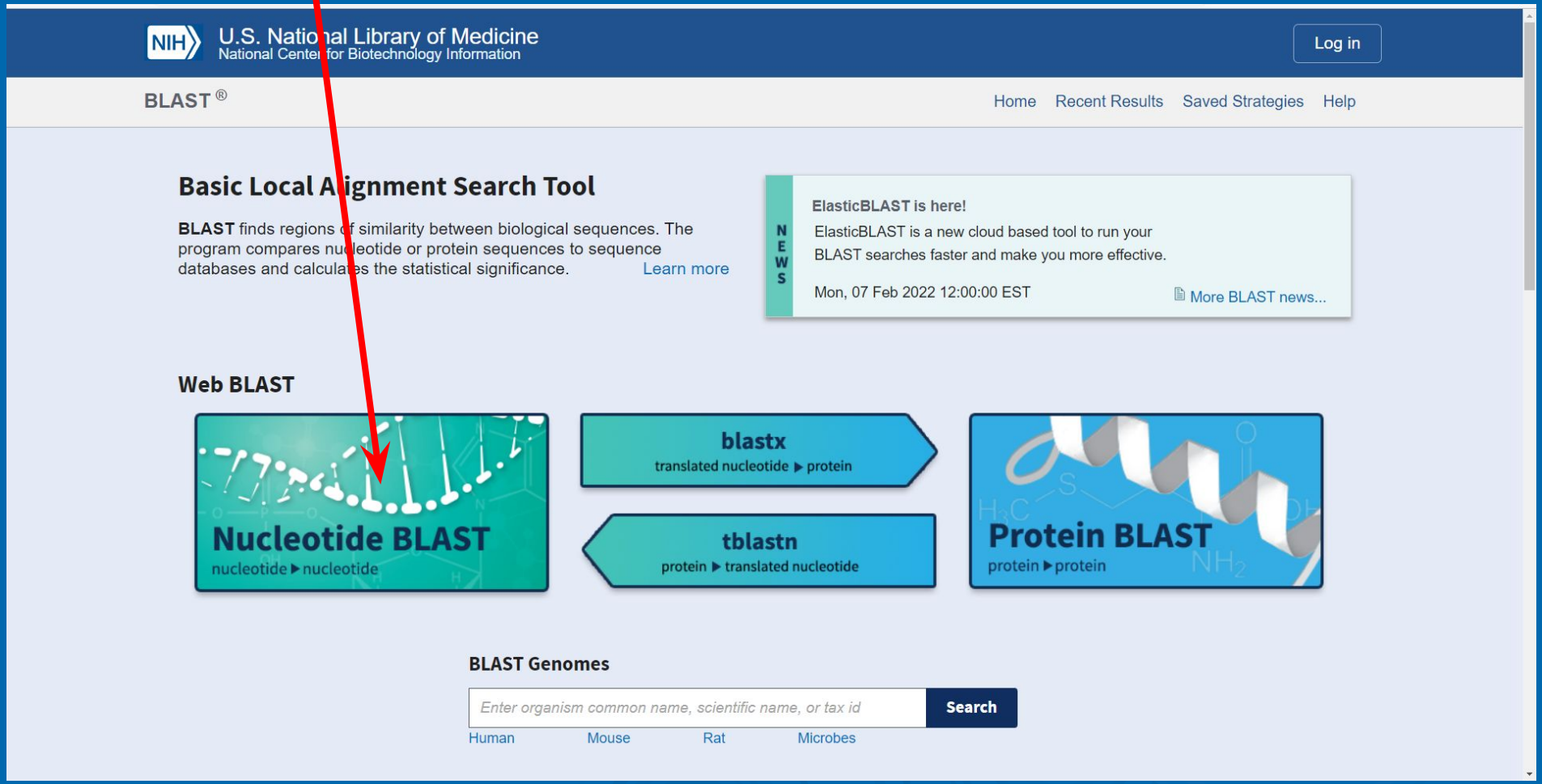
7. Копирова… …леотидов:



Задача 1

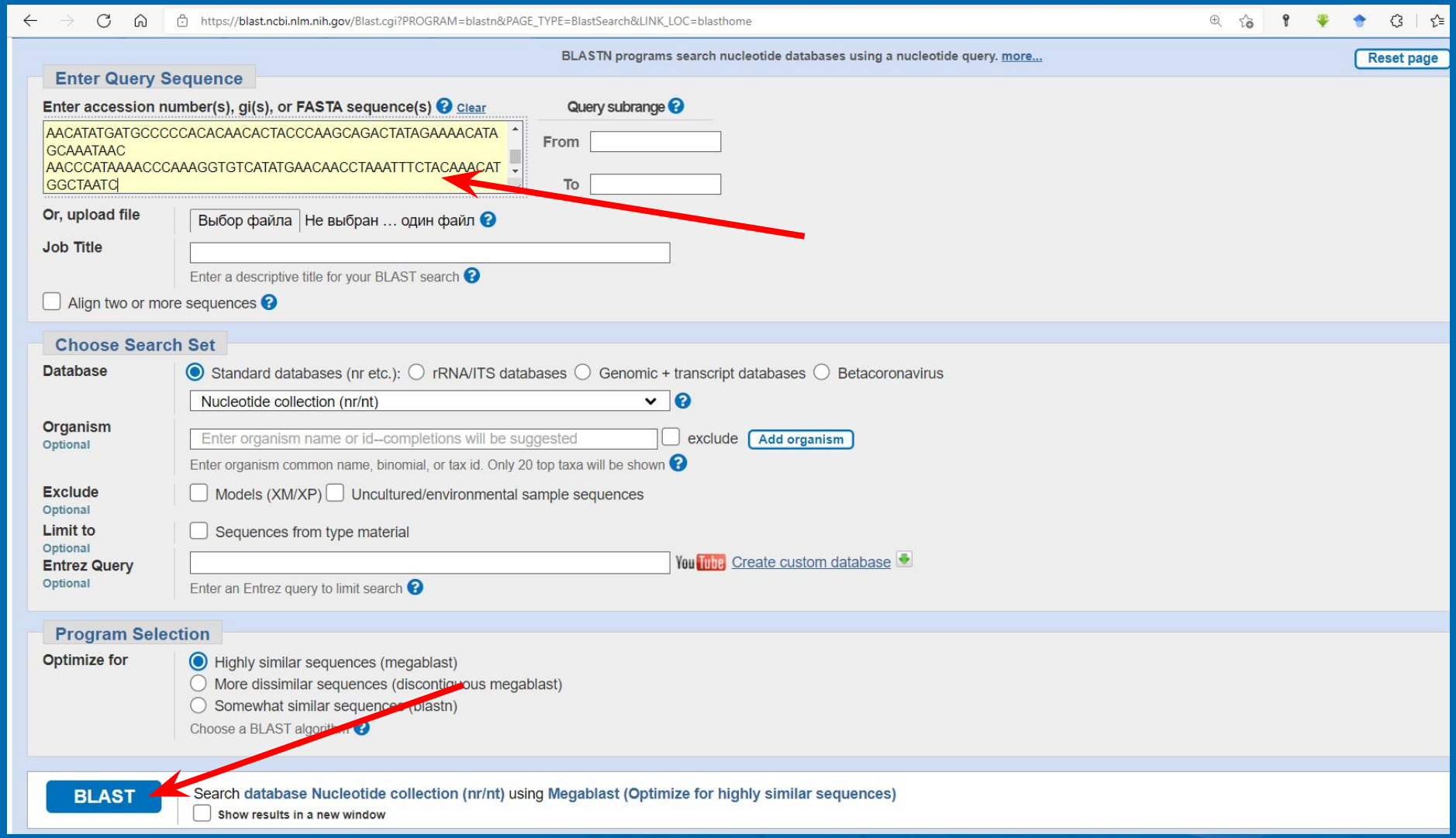
- 8. Переходим в сервис BLAST :
- https://blast.ncbi.nlm.nih.gov/Blast.cgi

blastn



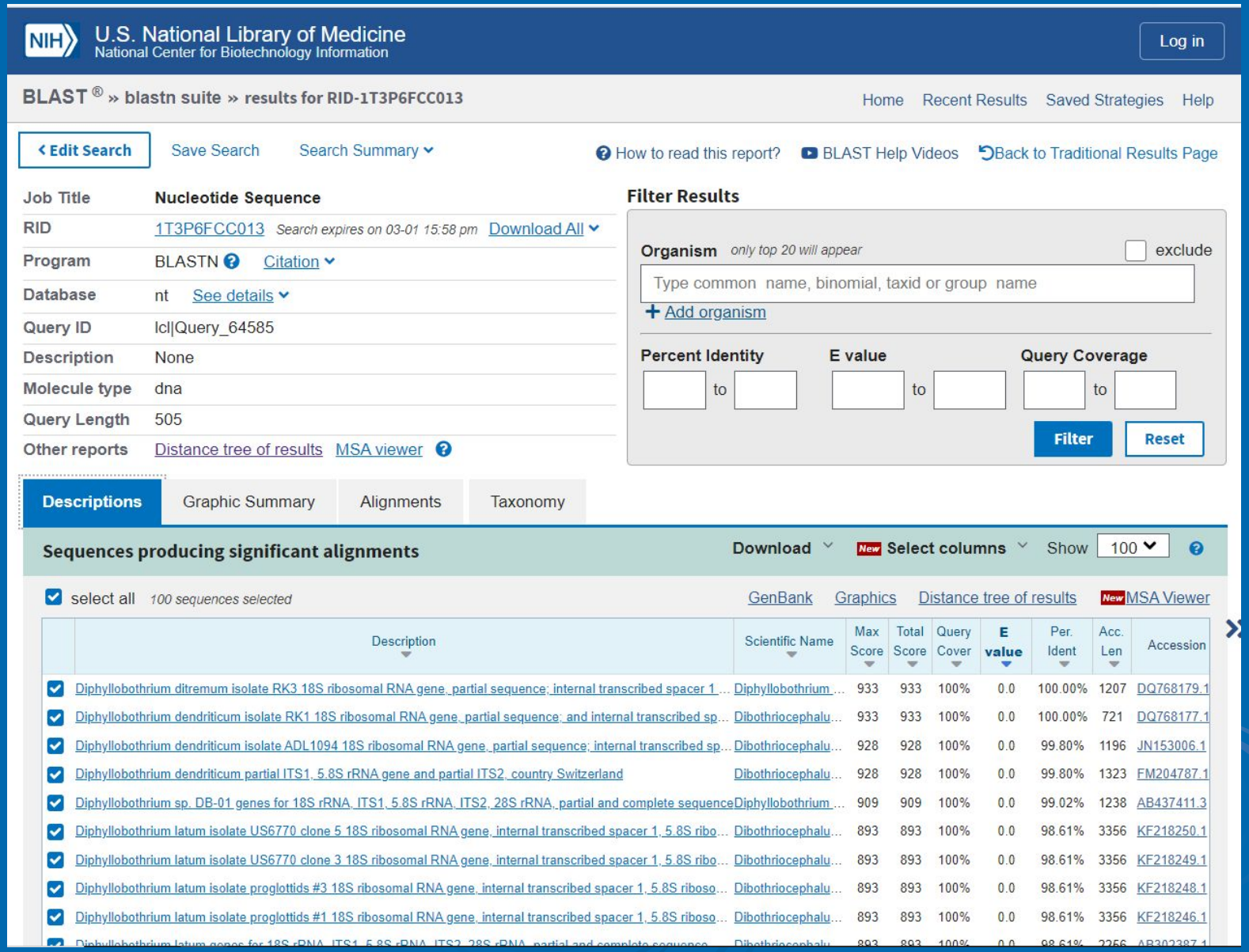
Задача 1

- 9. Вставляем последовательность в окошко
- 10. Нажимаем кнопку BLAST:



Задача 1

- **11. Получаем в итоге список гомологов, близких к нашей последовательности**



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

Log in

BLAST® » blastn suite » results for RID-1T3P6FCC013

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	Nucleotide Sequence
RID	1T3P6FCC013 Search expires on 03-01 15:58 pm Download All
Program	BLASTN Citation
Database	nt See details
Query ID	lcl|Query_64585
Description	None
Molecule type	dna
Query Length	505
Other reports	Distance tree of results MSA viewer

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to

E value to

Query Coverage to

Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments

Download | New Select columns | Show 100

select all 100 sequences selected

GenBank | Graphics | Distance tree of results | New MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Diphyllobothrium ditremum isolate RK3 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1 ...	Diphyllobothrium ...	933	933	100%	0.0	100.00%	1207	DQ768179.1
Diphyllobothrium dendriticum isolate RK1 18S ribosomal RNA gene, partial sequence; and internal transcribed sp...	Dibothriocephalu...	933	933	100%	0.0	100.00%	721	DQ768177.1
Diphyllobothrium dendriticum isolate ADL1094 18S ribosomal RNA gene, partial sequence; internal transcribed sp...	Dibothriocephalu...	928	928	100%	0.0	99.80%	1196	JN153006.1
Diphyllobothrium dendriticum partial ITS1, 5.8S rRNA gene and partial ITS2, country Switzerland	Dibothriocephalu...	928	928	100%	0.0	99.80%	1323	FM204787.1
Diphyllobothrium sp. DB-01 genes for 18S rRNA, ITS1, 5.8S rRNA, ITS2, 28S rRNA, partial and complete sequence	Diphyllobothrium ...	909	909	100%	0.0	99.02%	1238	AB437411.3
Diphyllobothrium latum isolate US6770 clone 5 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribo...	Dibothriocephalu...	893	893	100%	0.0	98.61%	3356	KF218250.1
Diphyllobothrium latum isolate US6770 clone 3 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribo...	Dibothriocephalu...	893	893	100%	0.0	98.61%	3356	KF218249.1
Diphyllobothrium latum isolate proglottids #3 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S riboso...	Dibothriocephalu...	893	893	100%	0.0	98.61%	3356	KF218248.1
Diphyllobothrium latum isolate proglottids #1 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S riboso...	Dibothriocephalu...	893	893	100%	0.0	98.61%	3356	KF218246.1

Задача 1. Форма отчета

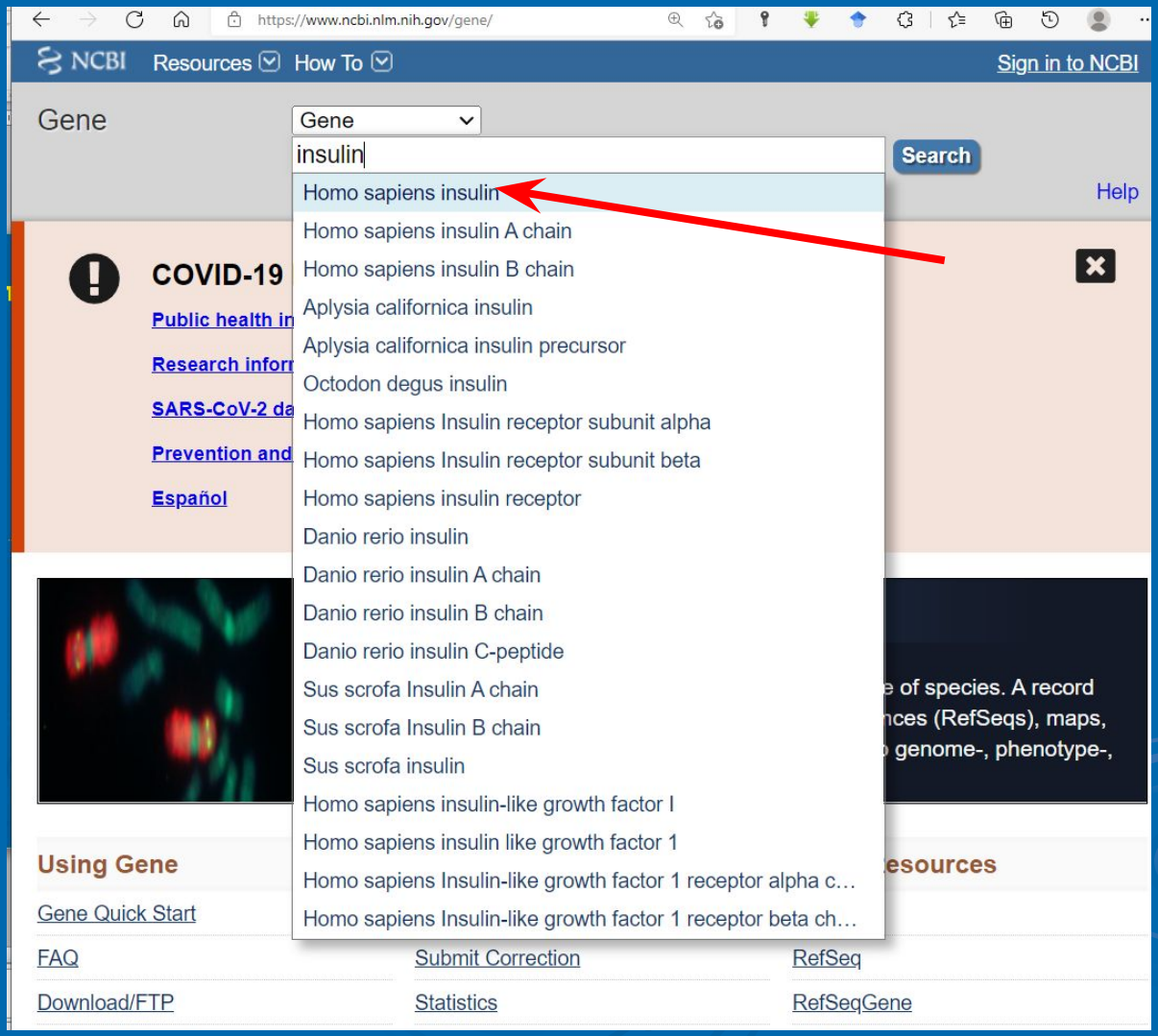
- Каждый лично на своем компьютере делает скриншот/фото (так, чтобы было видно номер компьютера/монитора, время на мониторе) списка гомологов в сервисе BLAST
- Каждый лично отправляет мне в Вайбере в личку: 1) полученное фото, а также 2) название вида, от которого получена последовательность нуклеотидов и 3) название гена, в котором содержится данная последовательность

Задача 2. Форма отчета

- **Каждый лично отправляет мне на почту sankaar@mail.ru :**
- **1) файл insulin.fasta**
- **2) insulin_alignment.fas**
- **3) insulin_best model.xls**
- **4) insulin_tree.mts**

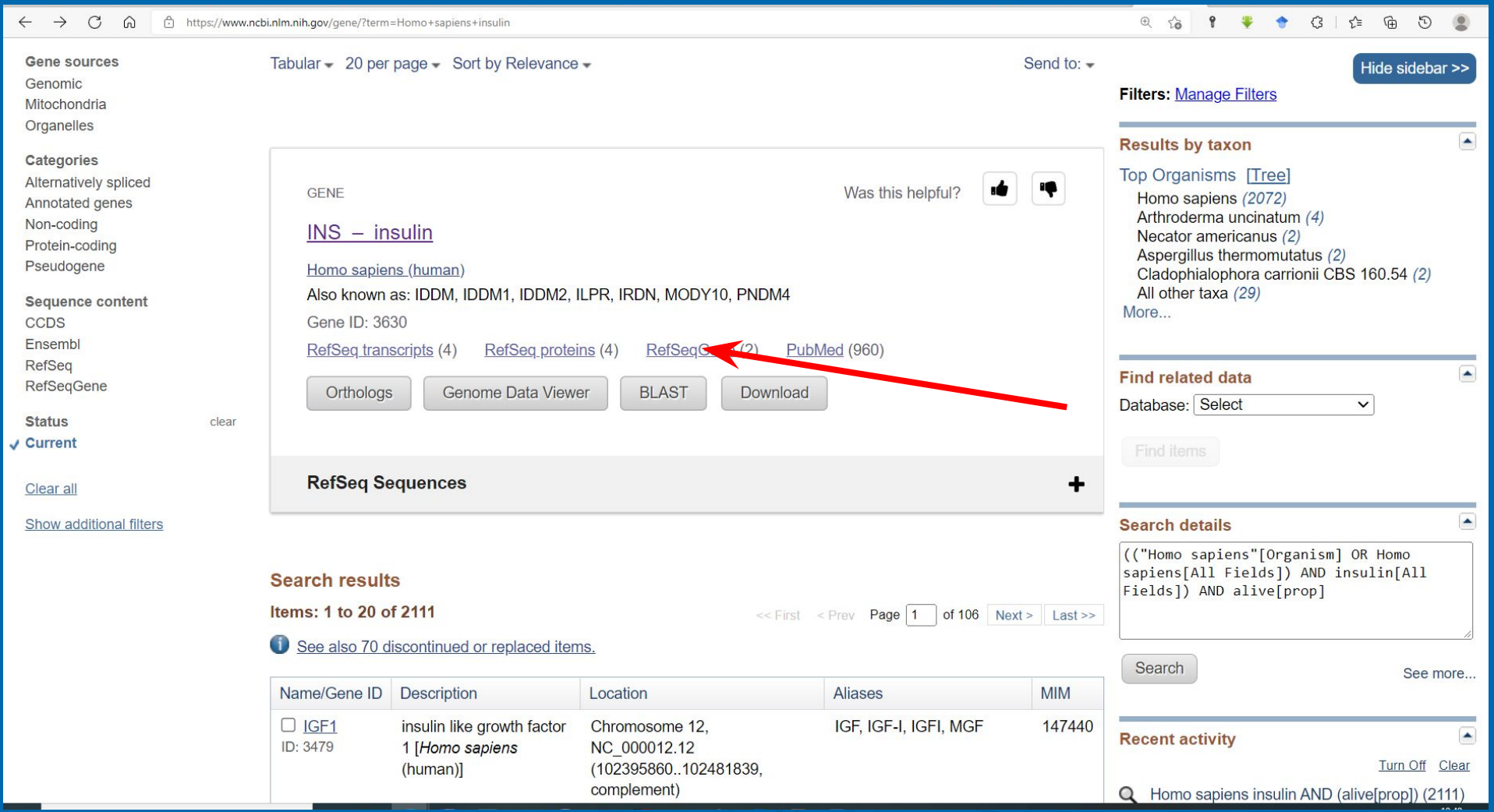
Задача 2

1. Найти в ГенБанке нуклеотидную последовательность гена (части гена) инсулина человека https://www.ncbi.nlm.nih.gov/gene/



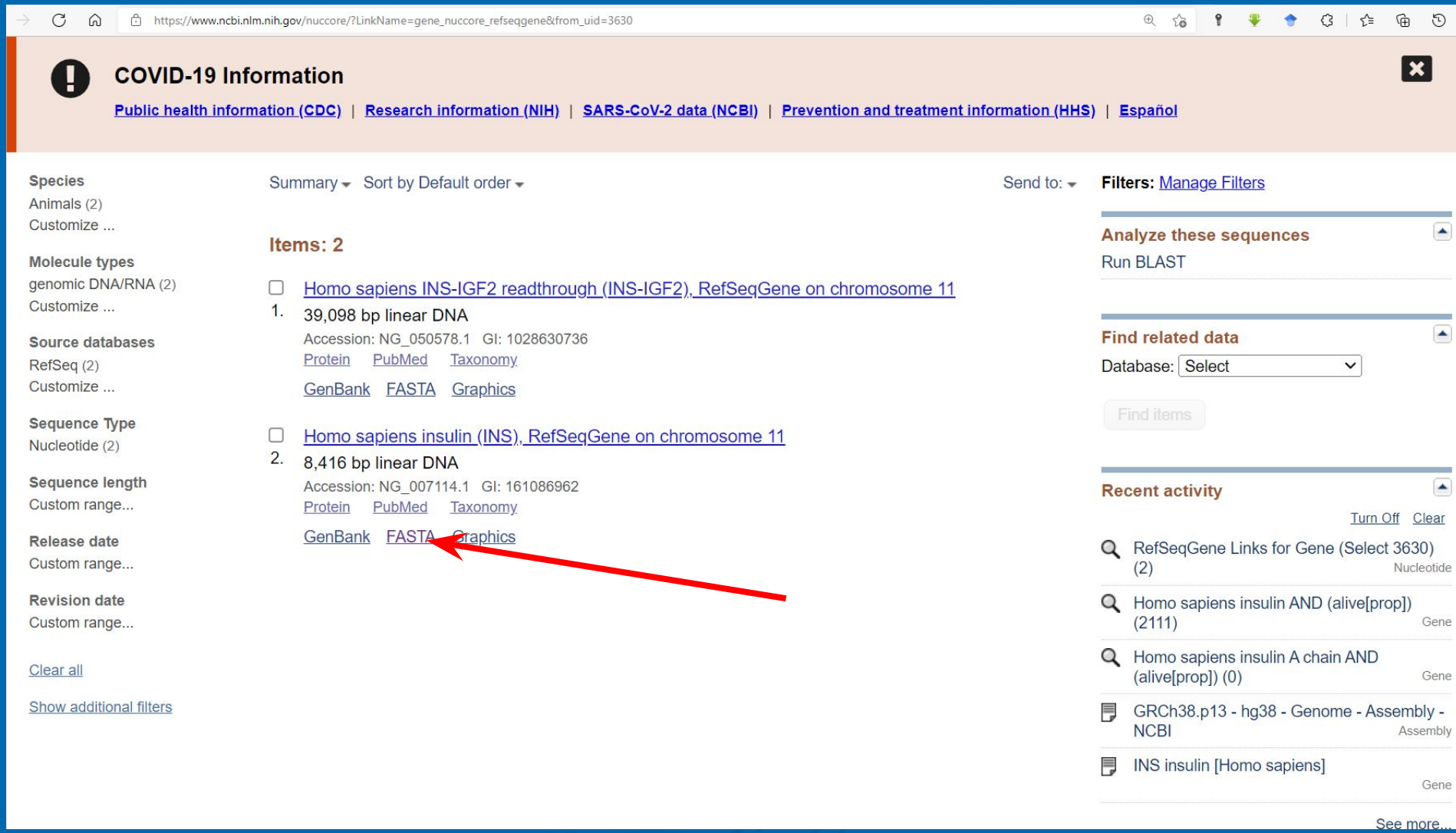
Задача 2

1. Найти в ГенБанке нуклеотидную последовательность гена (части гена) инсулина человека https://www.ncbi.nlm.nih.gov/gene/



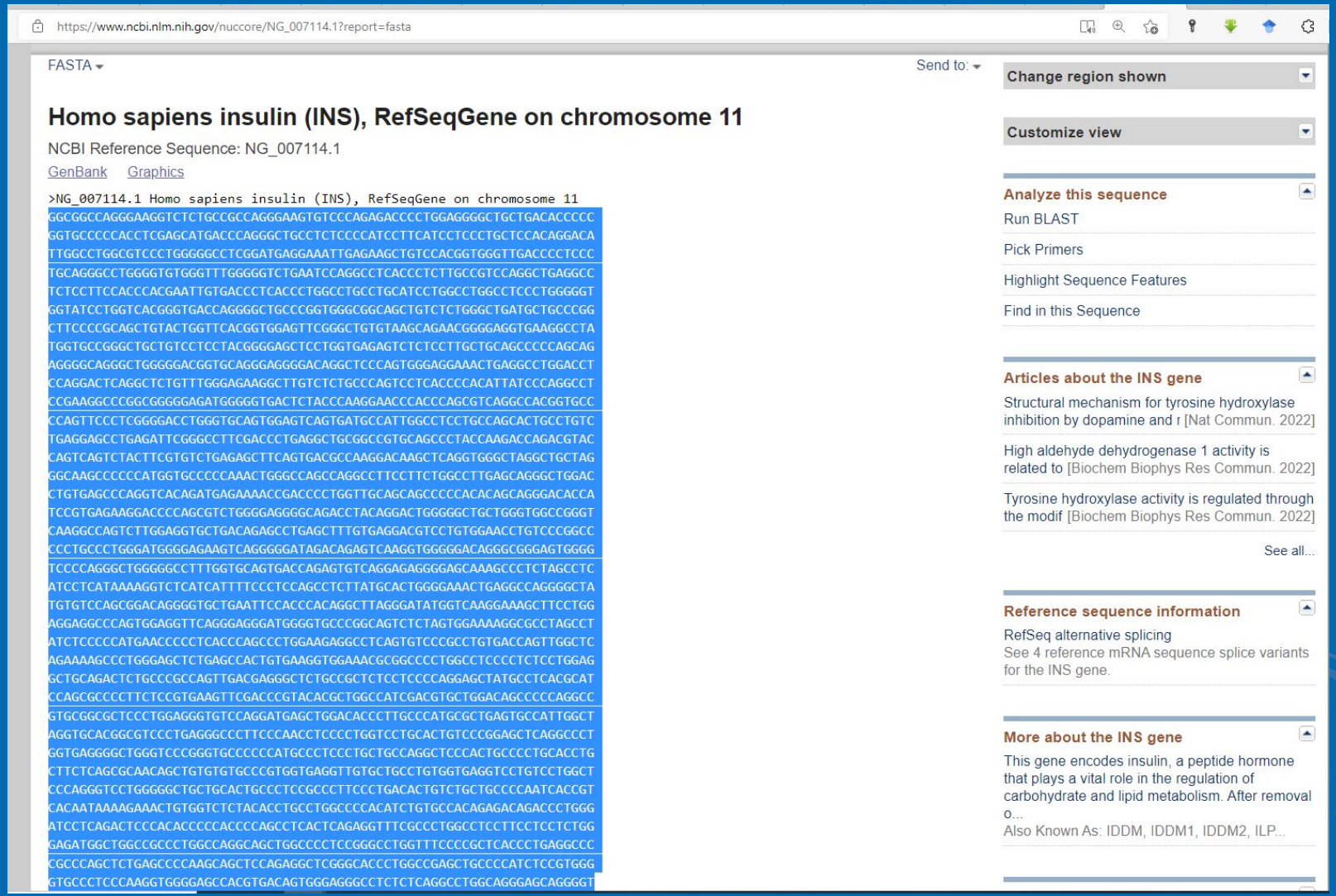
Задача 2

1. Найти в ГенБанке нуклеотидную последовательность гена (части гена) инсулина человека https://www.ncbi.nlm.nih.gov/gene/



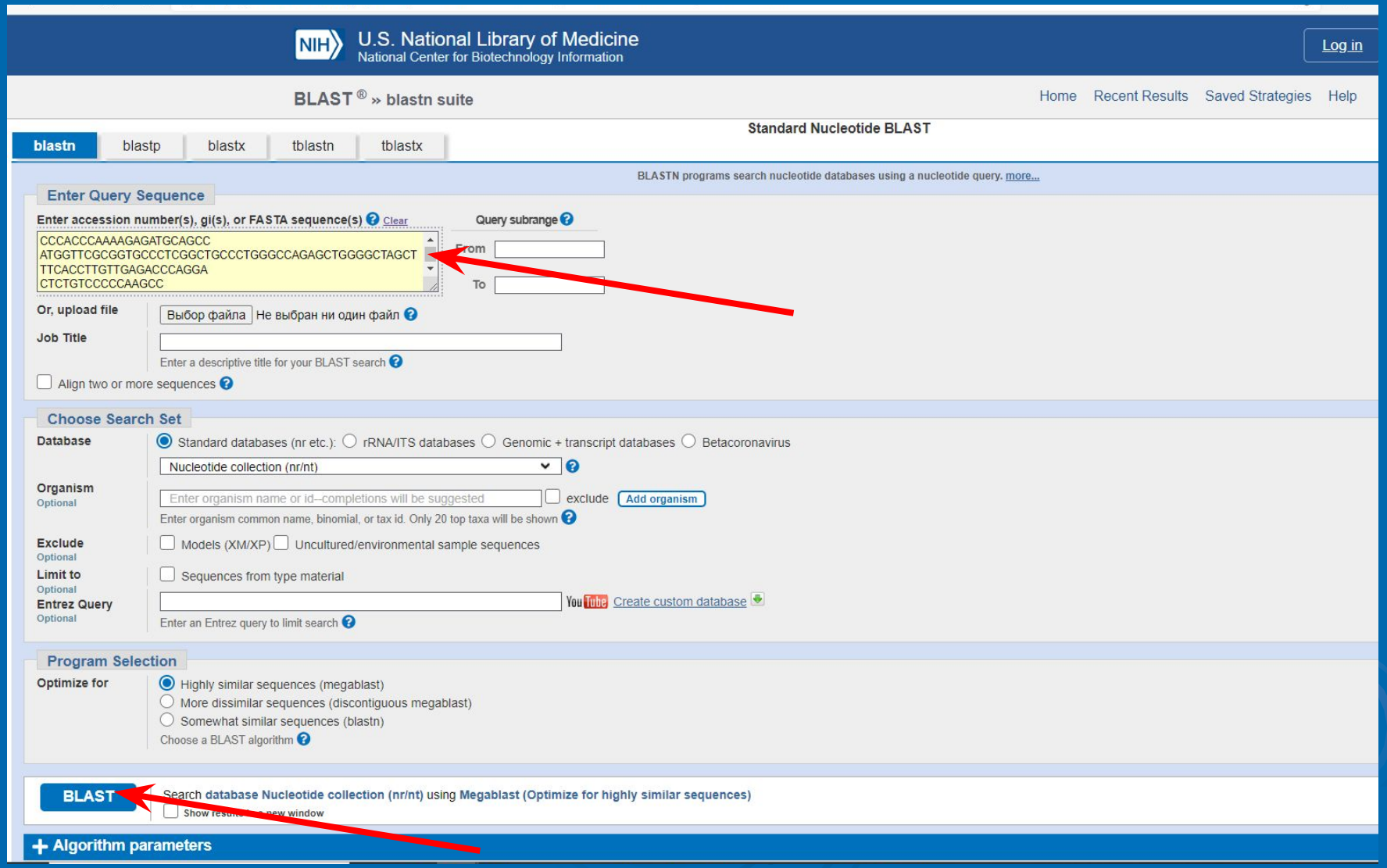
Задача 2

1. Найти в ГенБанке нуклеотидную последовательность гена (части гена) инсулина человека https://www.ncbi.nlm.nih.gov/gene/



Задача 2

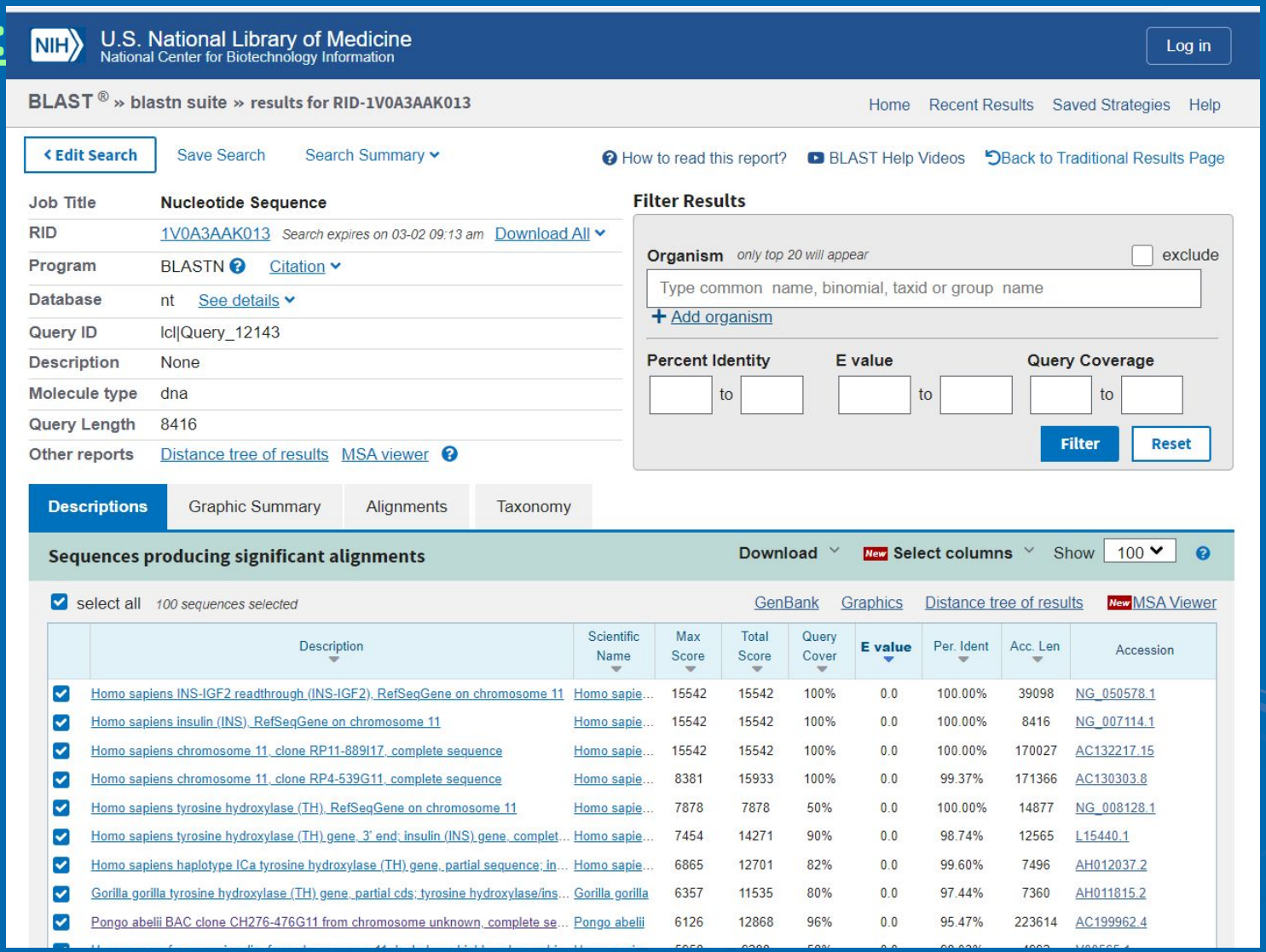
2. С помощью программы blastn найти в нуклеотидном виде гомологи этой последовательности



Задача 2

2. С помощью программы blastn найти в нуклеотидном виде гомологи этой последовательности

https:



NIH U.S. National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastn suite » results for RID-1V0A3AAK013

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	Nucleotide Sequence
RID	1V0A3AAK013 Search expires on 03-02 09:13 am Download All
Program	BLASTN Citation
Database	nt See details
Query ID	lcl|Query_12143
Description	None
Molecule type	dna
Query Length	8416
Other reports	Distance tree of results MSA viewer

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to

E value to

Query Coverage to

Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments

Download | New Select columns | Show 100

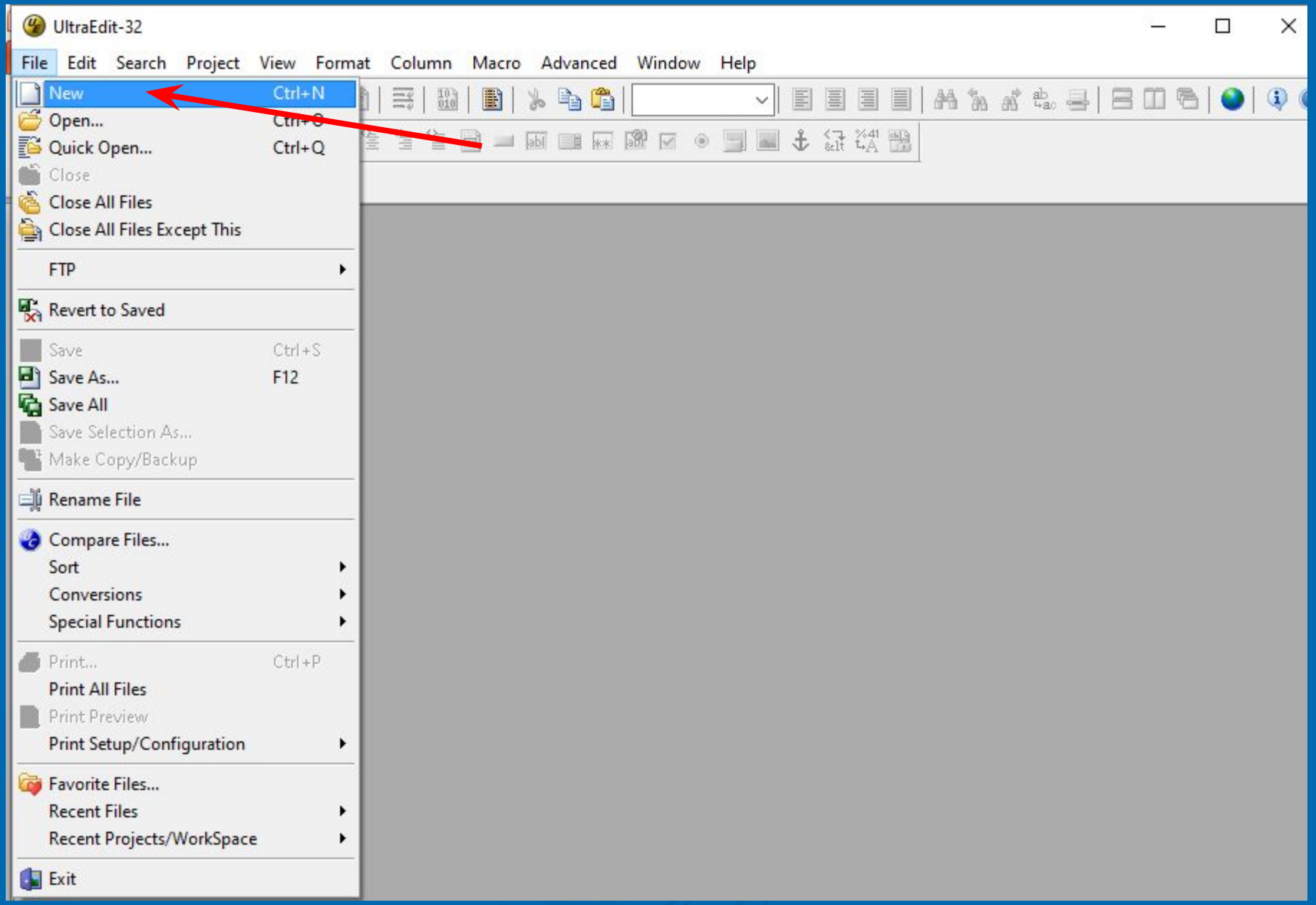
select all 100 sequences selected

GenBank Graphics Distance tree of results New MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11	Homo sapie...	15542	15542	100%	0.0	100.00%	39098	NG_050578.1
Homo sapiens insulin (INS), RefSeqGene on chromosome 11	Homo sapie...	15542	15542	100%	0.0	100.00%	8416	NG_007114.1
Homo sapiens chromosome 11, clone RP11-889I17, complete sequence	Homo sapie...	15542	15542	100%	0.0	100.00%	170027	AC132217.15
Homo sapiens chromosome 11, clone RP4-539G11, complete sequence	Homo sapie...	8381	15933	100%	0.0	99.37%	171366	AC130303.8
Homo sapiens tyrosine hydroxylase (TH), RefSeqGene on chromosome 11	Homo sapie...	7878	7878	50%	0.0	100.00%	14877	NG_008128.1
Homo sapiens tyrosine hydroxylase (TH) gene, 3' end; insulin (INS) gene, complet...	Homo sapie...	7454	14271	90%	0.0	98.74%	12565	L15440.1
Homo sapiens haplotype ICa tyrosine hydroxylase (TH) gene, partial sequence; in...	Homo sapie...	6865	12701	82%	0.0	99.60%	7496	AH012037.2
Gorilla gorilla tyrosine hydroxylase (TH) gene, partial cds; tyrosine hydroxylase/ins...	Gorilla gorilla	6357	11535	80%	0.0	97.44%	7360	AH011815.2
Pongo abelii BAC clone CH276-476G11 from chromosome unknown, complete se...	Pongo abelii	6126	12868	96%	0.0	95.47%	223614	AC199962.4

Задача 2

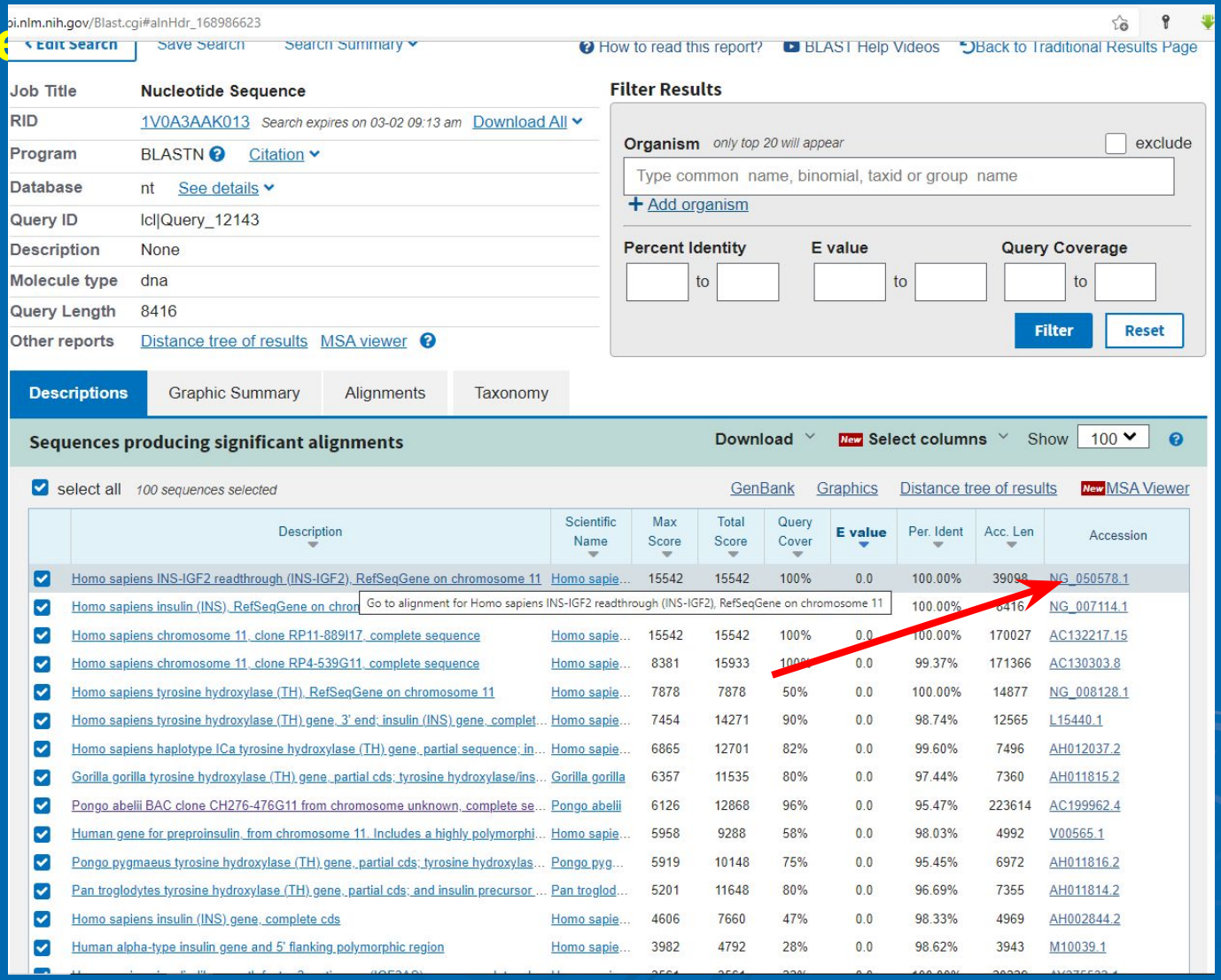
3. Открыть новый файл в программе UltraEdit



Задача 2

4. Выбрать среди найденных гомологов несколько от пяти разных видов организмов (*Homo sapiens, Gorilla gorilla* и т.д.)

5. Открыть их нуклеотидную последовательность в формате



ncbi.nlm.nih.gov/Blast.cgi#alnHdr_168986623

Edit Search | Save Search | Search Summary | How to read this report? | BLAST Help Videos | Back to Traditional Results Page

Job Title	Nucleotide Sequence
RID	1V0A3AAK013 Search expires on 03-02 09:13 am Download All
Program	BLASTN Citation
Database	nt See details
Query ID	lcl|Query_12143
Description	None
Molecule type	dna
Query Length	8416
Other reports	Distance tree of results MSA viewer

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to | E value to | Query Coverage to

Filter Reset

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download | New Select columns | Show 100

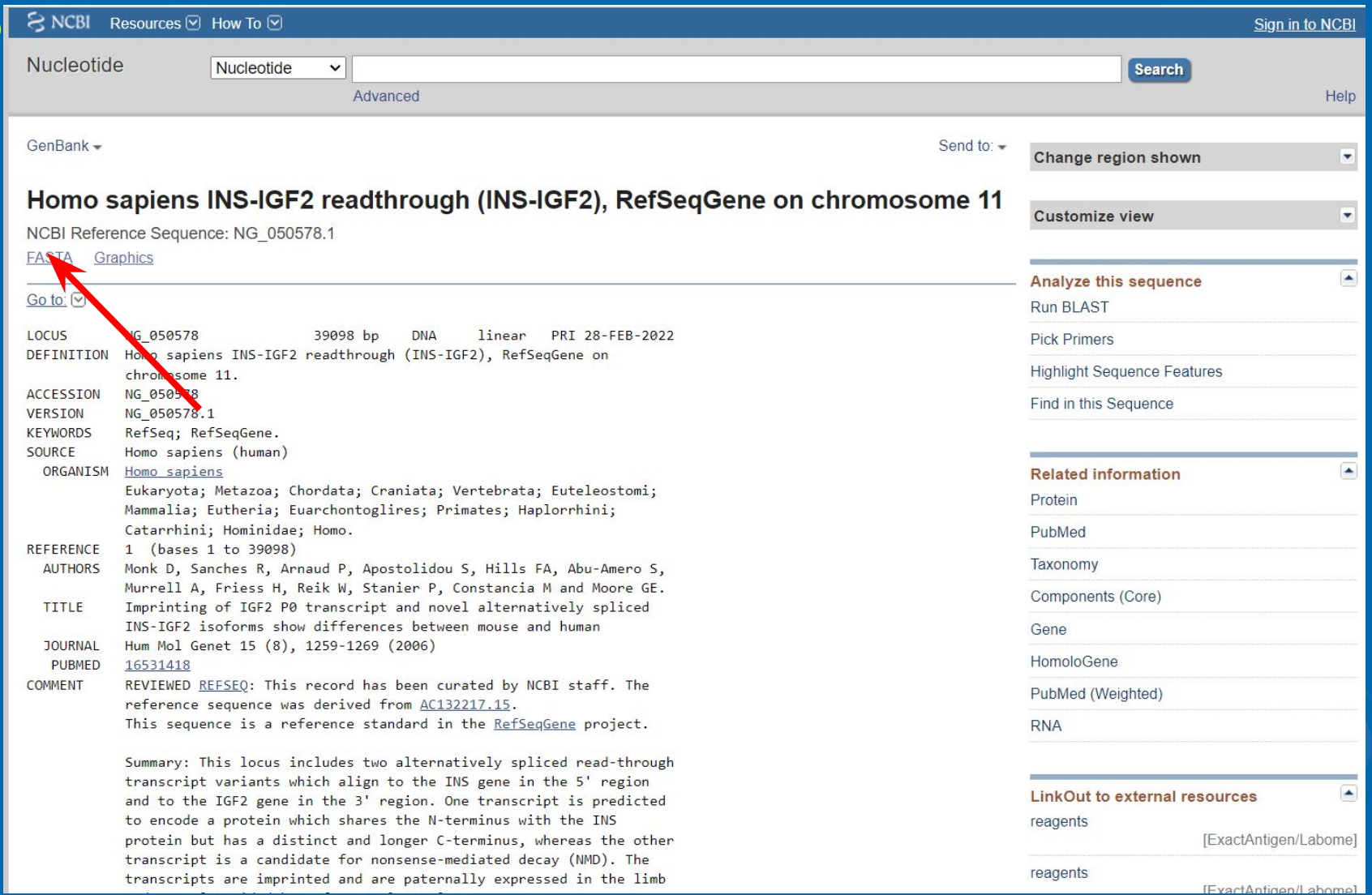
select all 100 sequences selected — GenBank | Graphics | Distance tree of results | New MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11	Homo sapie...	15542	15542	100%	0.0	100.00%	39098	NG_050578.1
Homo sapiens insulin (INS), RefSeqGene on chrom	Go to alignment for Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11					100.00%	8416	NG_007114.1
Homo sapiens chromosome 11, clone RP11-889I17, complete sequence	Homo sapie...	15542	15542	100%	0.0	100.00%	170027	AC132217.15
Homo sapiens chromosome 11, clone RP4-539G11, complete sequence	Homo sapie...	8381	15933	100%	0.0	99.37%	171366	AC130303.8
Homo sapiens tyrosine hydroxylase (TH), RefSeqGene on chromosome 11	Homo sapie...	7878	7878	50%	0.0	100.00%	14877	NG_008128.1
Homo sapiens tyrosine hydroxylase (TH) gene, 3' end; insulin (INS) gene, complet...	Homo sapie...	7454	14271	90%	0.0	98.74%	12565	L15440.1
Homo sapiens haplotype ICa tyrosine hydroxylase (TH) gene, partial sequence; in...	Homo sapie...	6865	12701	82%	0.0	99.60%	7496	AH012037.2
Gorilla gorilla tyrosine hydroxylase (TH) gene, partial cds; tyrosine hydroxylase/ins...	Gorilla gorilla	6357	11535	80%	0.0	97.44%	7360	AH011815.2
Pongo abelii BAC clone CH276-476G11 from chromosome unknown, complete se...	Pongo abelii	6126	12868	96%	0.0	95.47%	223614	AC199962.4
Human gene for preproinsulin, from chromosome 11. Includes a highly polymorphi...	Homo sapie...	5958	9288	58%	0.0	98.03%	4992	V00565.1
Pongo pygmaeus tyrosine hydroxylase (TH) gene, partial cds; tyrosine hydroxylas...	Pongo pyg...	5919	10148	75%	0.0	95.45%	6972	AH011816.2
Pan troglodytes tyrosine hydroxylase (TH) gene, partial cds; and insulin precursor ...	Pan troglod...	5201	11648	80%	0.0	96.69%	7355	AH011814.2
Homo sapiens insulin (INS) gene, complete cds	Homo sapie...	4606	7660	47%	0.0	98.33%	4969	AH002844.2
Human alpha-type insulin gene and 5' flanking polymorphic region	Homo sapie...	3982	4792	28%	0.0	98.62%	3943	M10039.1

Задача 2

4. Выбрать среди найденных гомологов несколько от пяти разных видов организмов (*Homo sapiens, Gorilla gorilla* и т.д.)

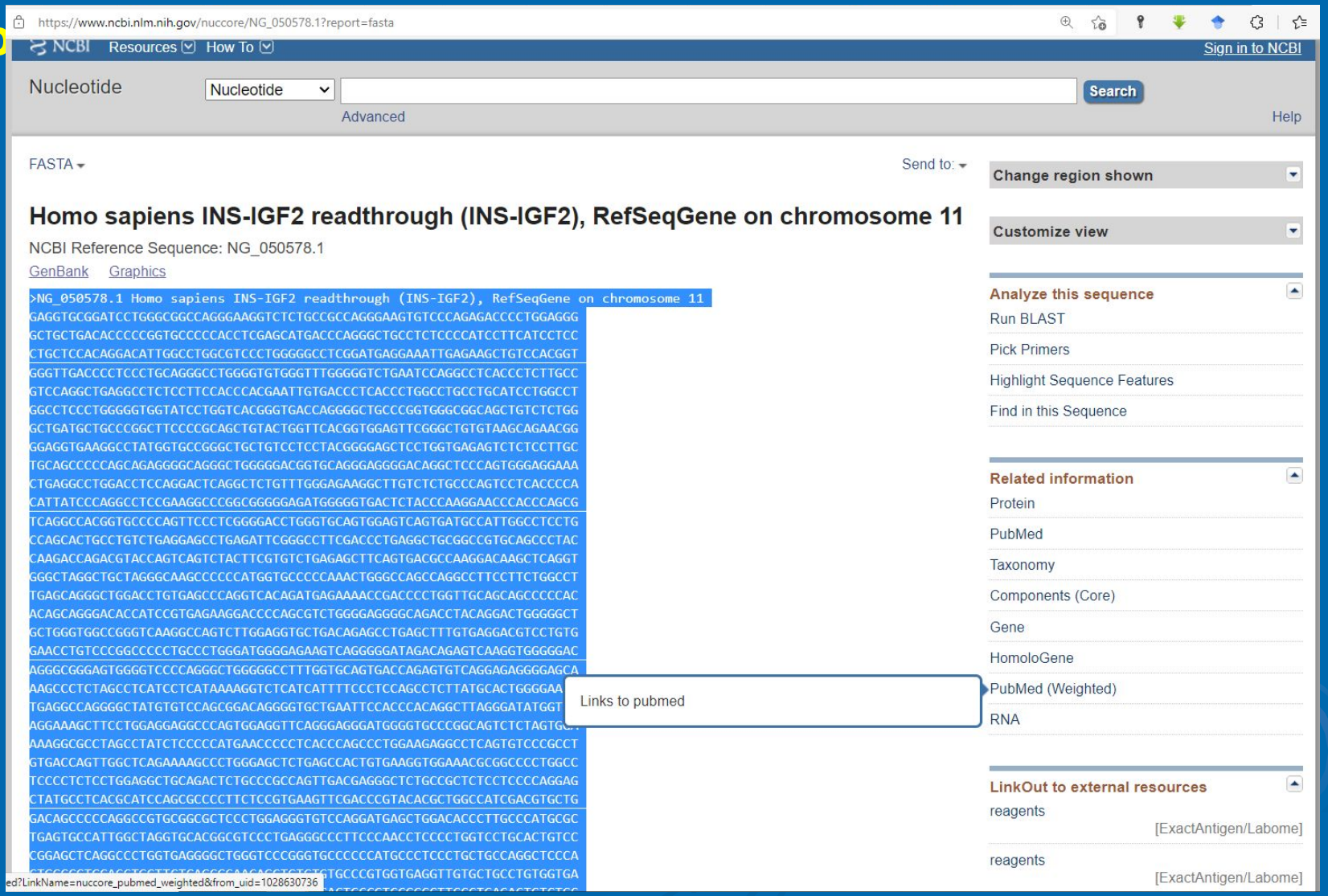
5. Открыть их нуклеотидную последовательность в фо



Задача 2

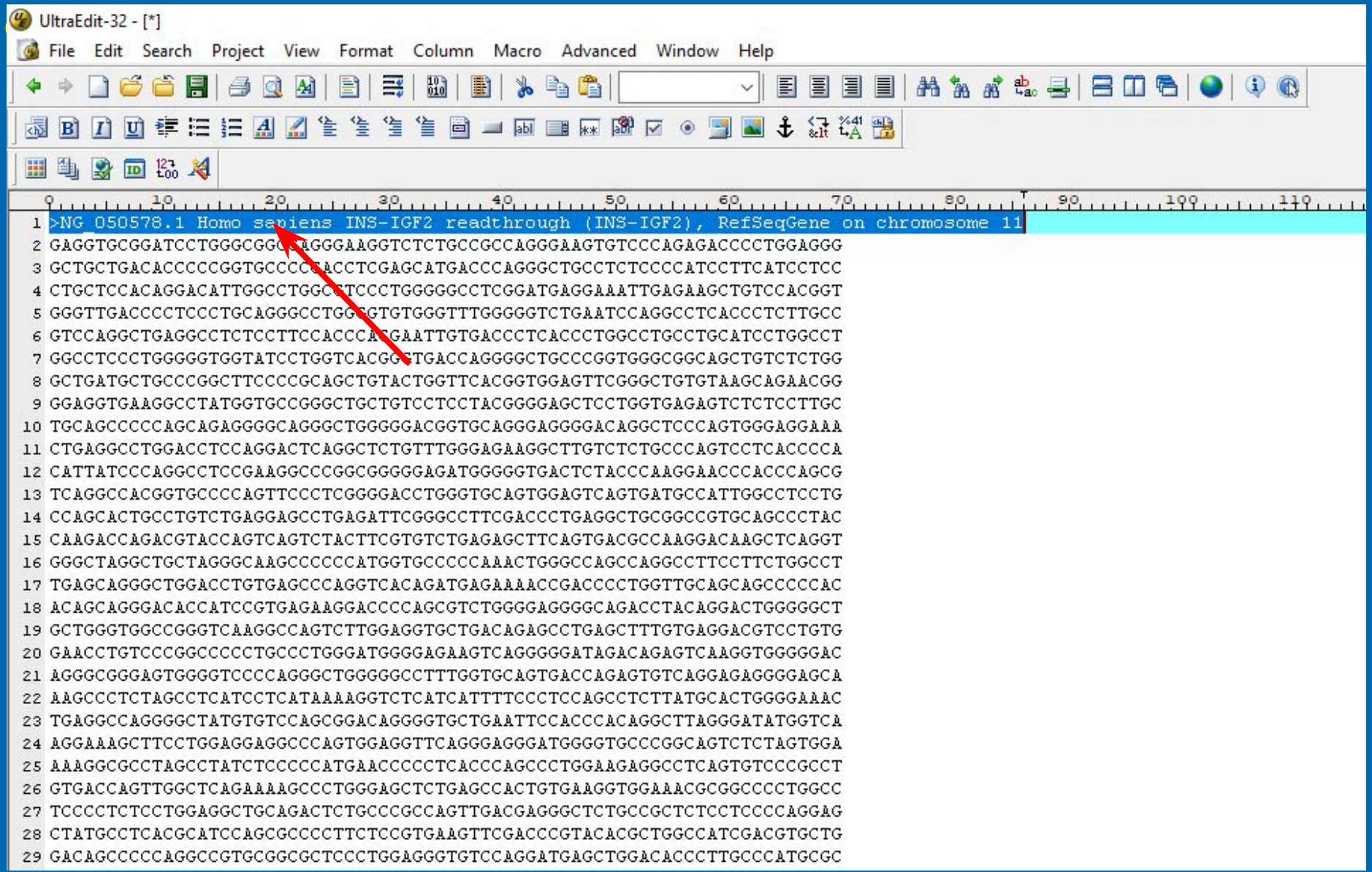
4. Выбрать среди найденных гомологов несколько от пяти разных видов организмов (*Homo sapiens, Gorilla gorilla* и т.д.)

5. Открыть их нуклеотидную последовательность в фор



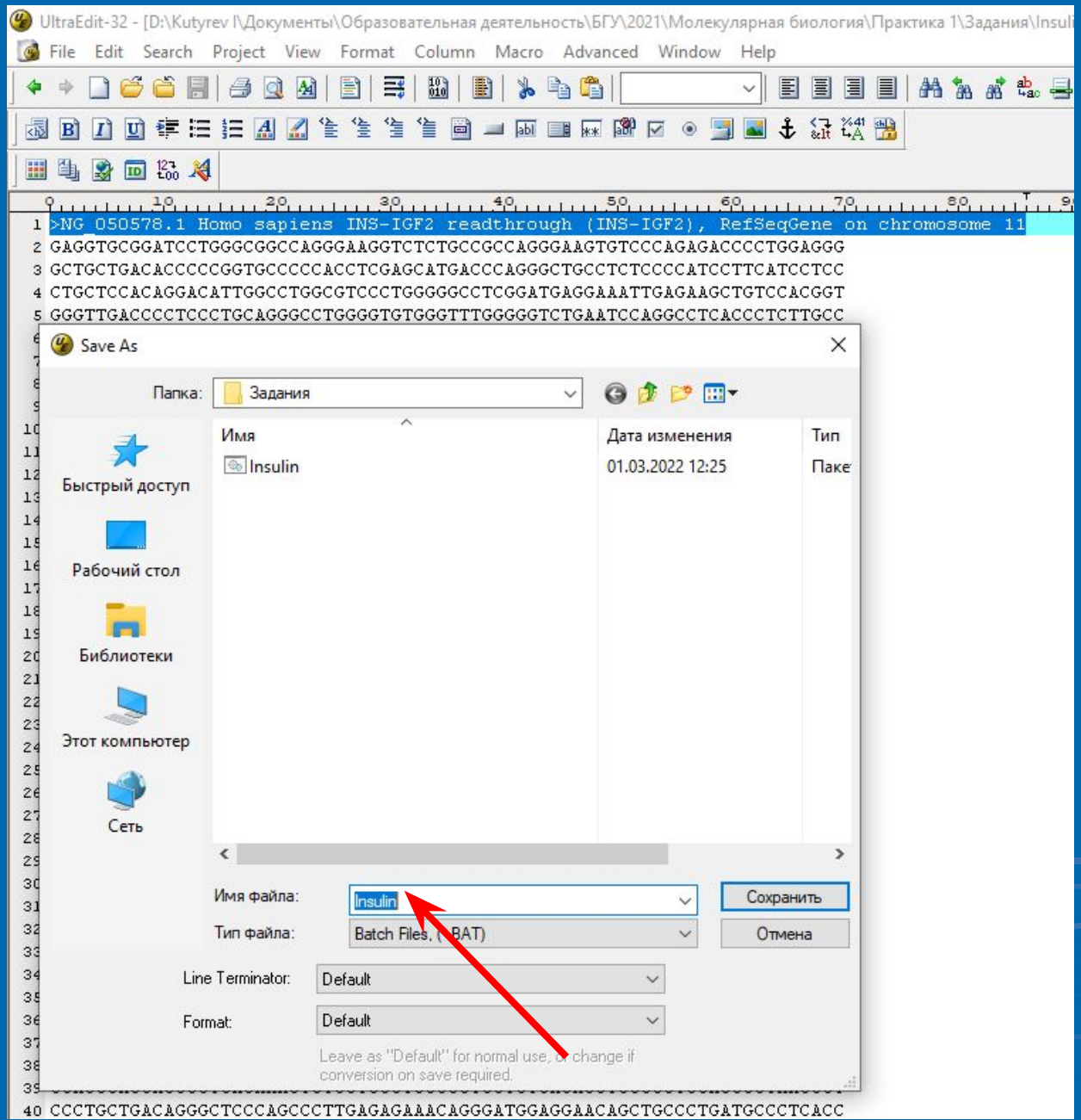
Задача 2

6. Копировать последовательности в созданный файл вместе со спец. символом “>” и названием (требования



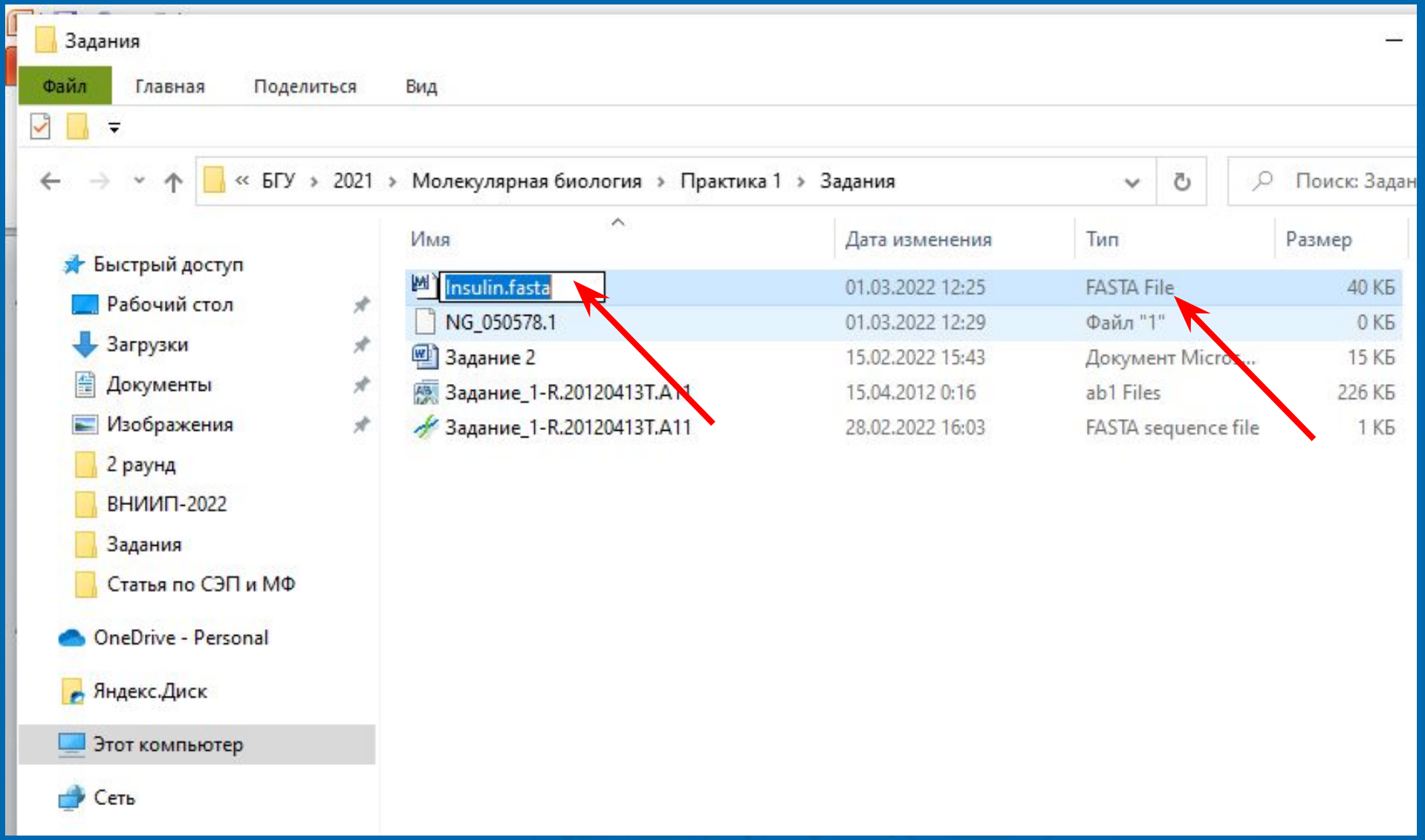
Задача 2

7. Сохранить файл под названием Insulin



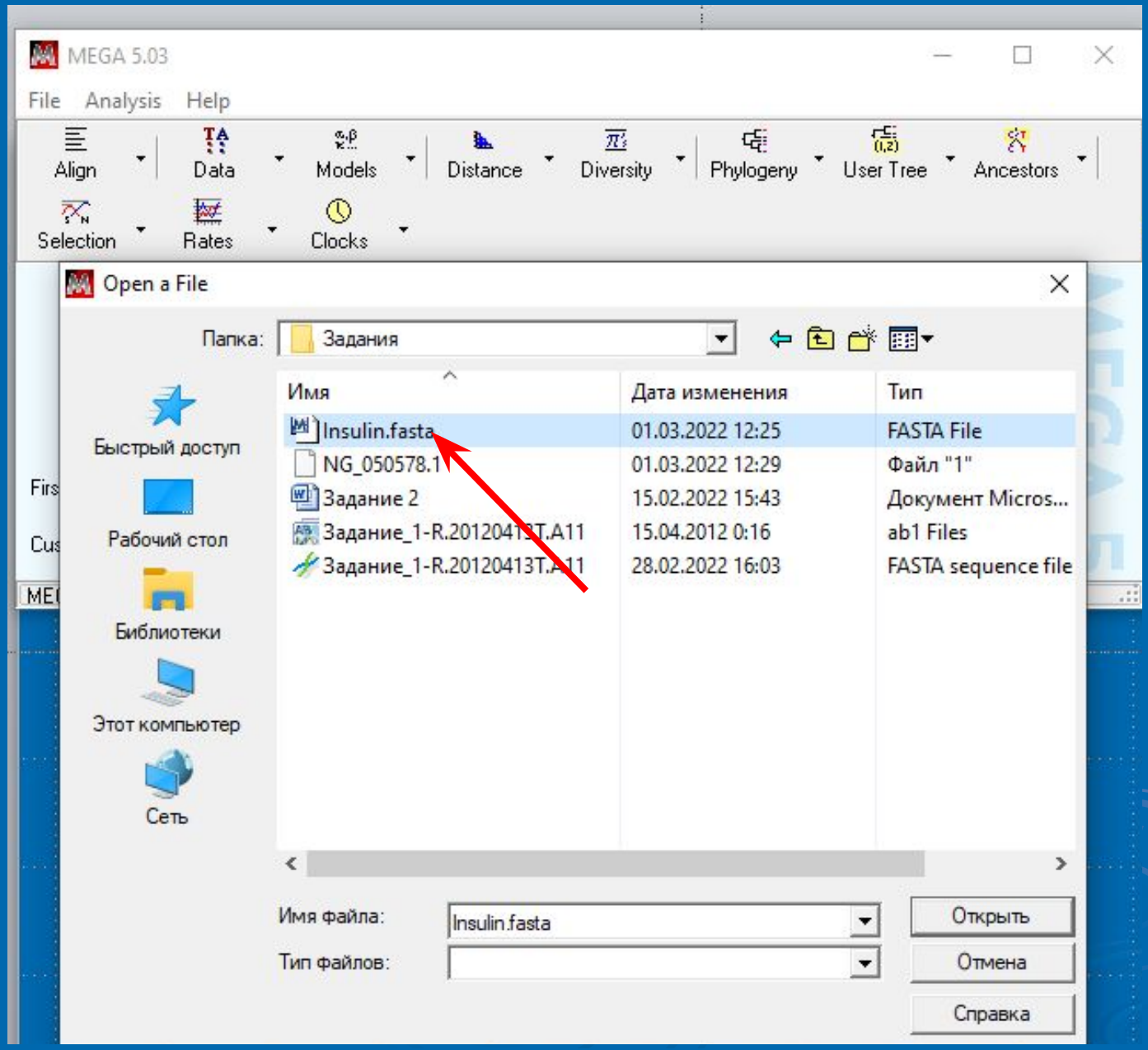
Задача 2

7. Заходим в папку, где сохранили файл, сохраняем его с расширением .fasta, т.е. как insulin.fasta



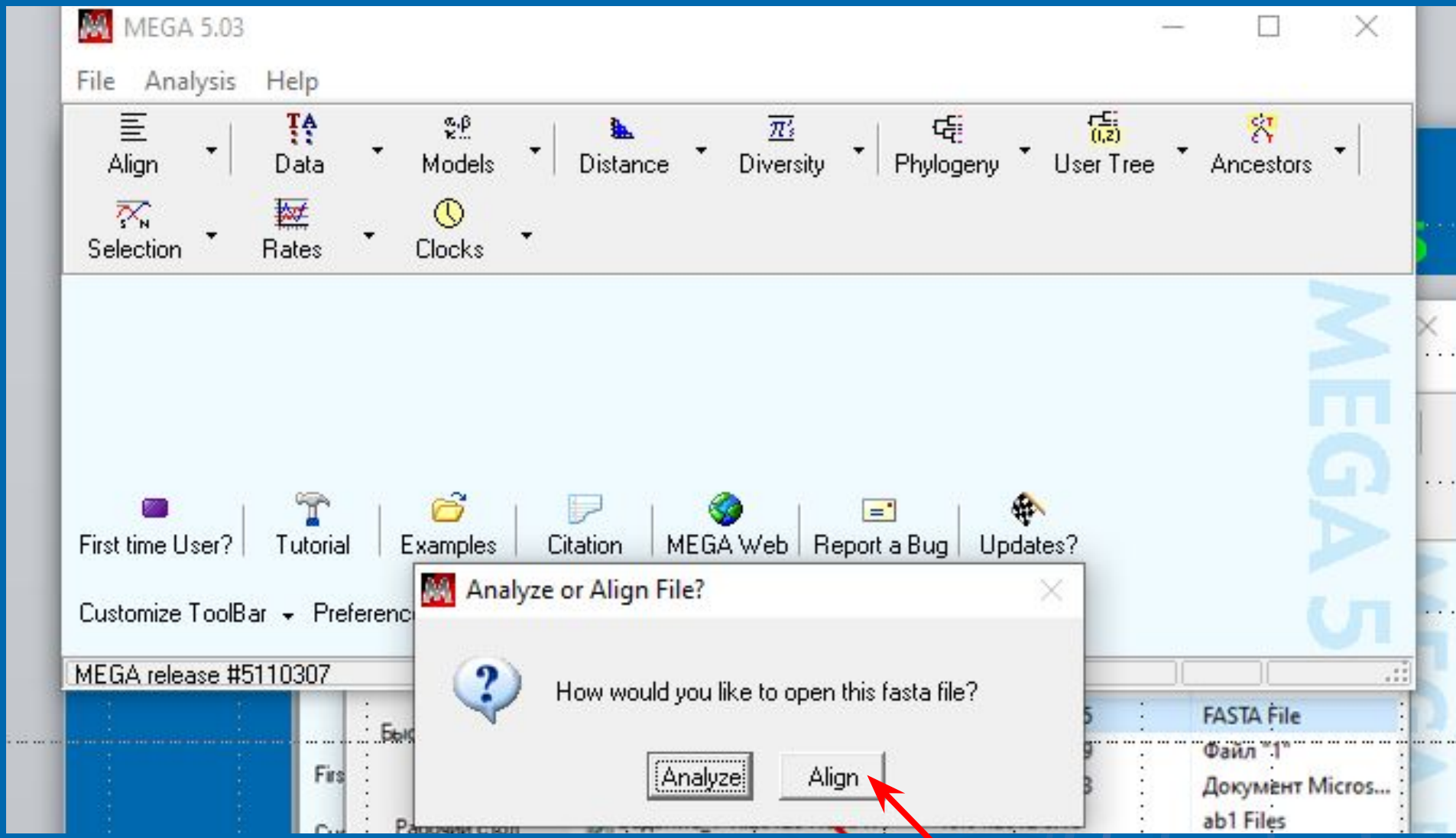
Задача 2

8. Открыть файл с помощью программы MEGA5



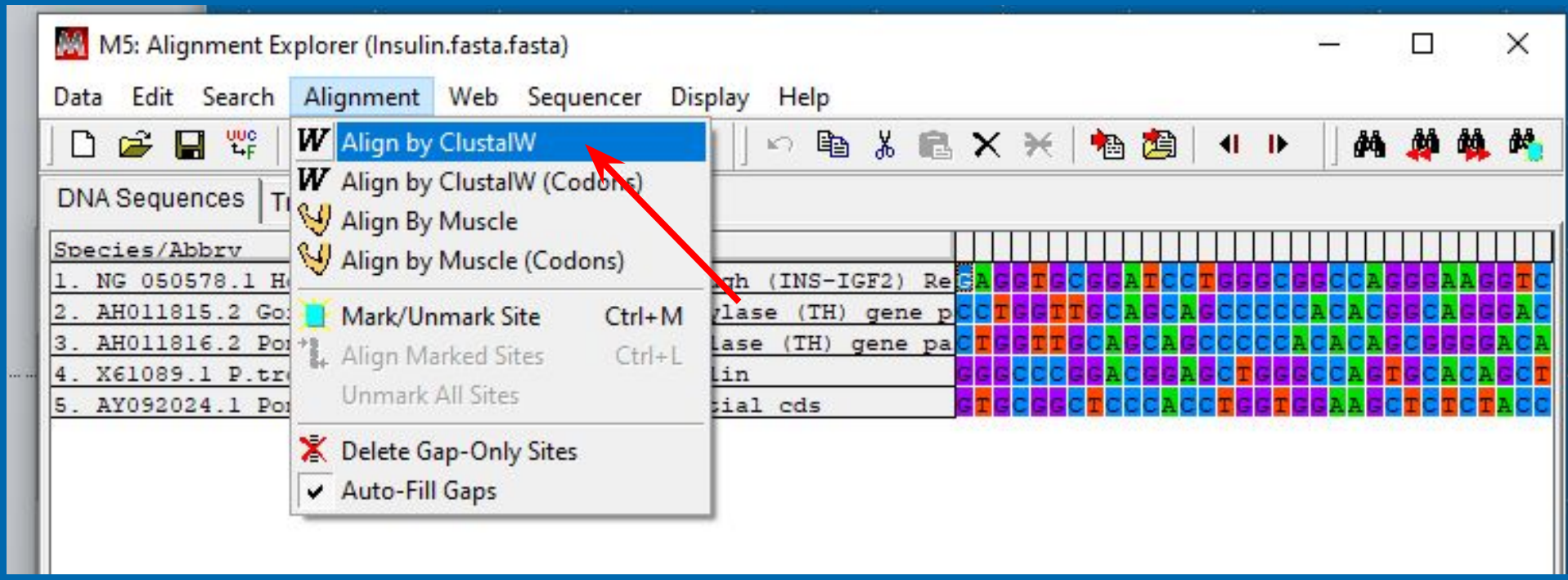
Задача 2

8. Выровнять последовательности в программе



Задача 2

8. Выровнять последовательности в программе MEGA5

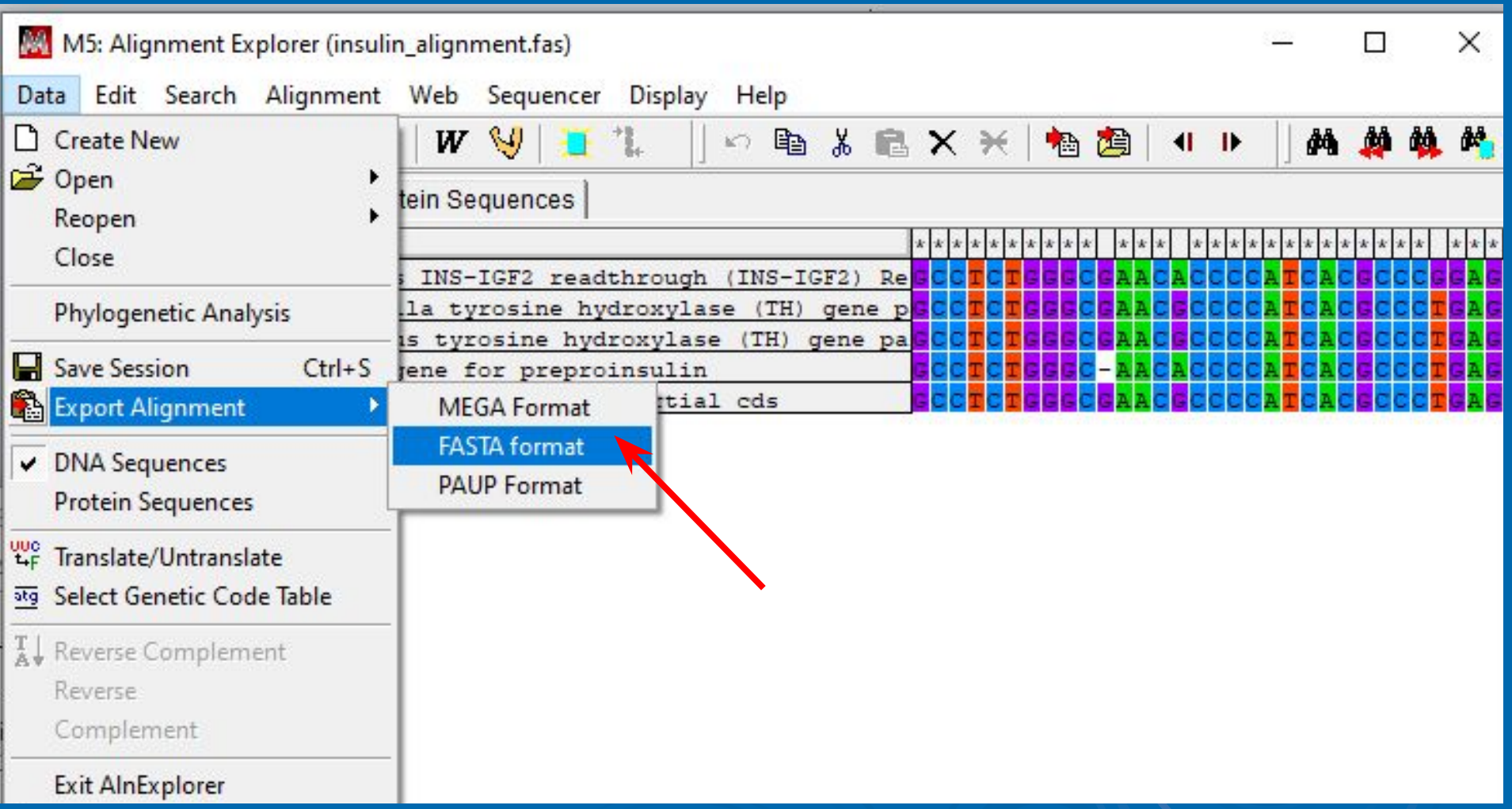


Задача 2

8. Выровнять последовательности в программе MEGA5

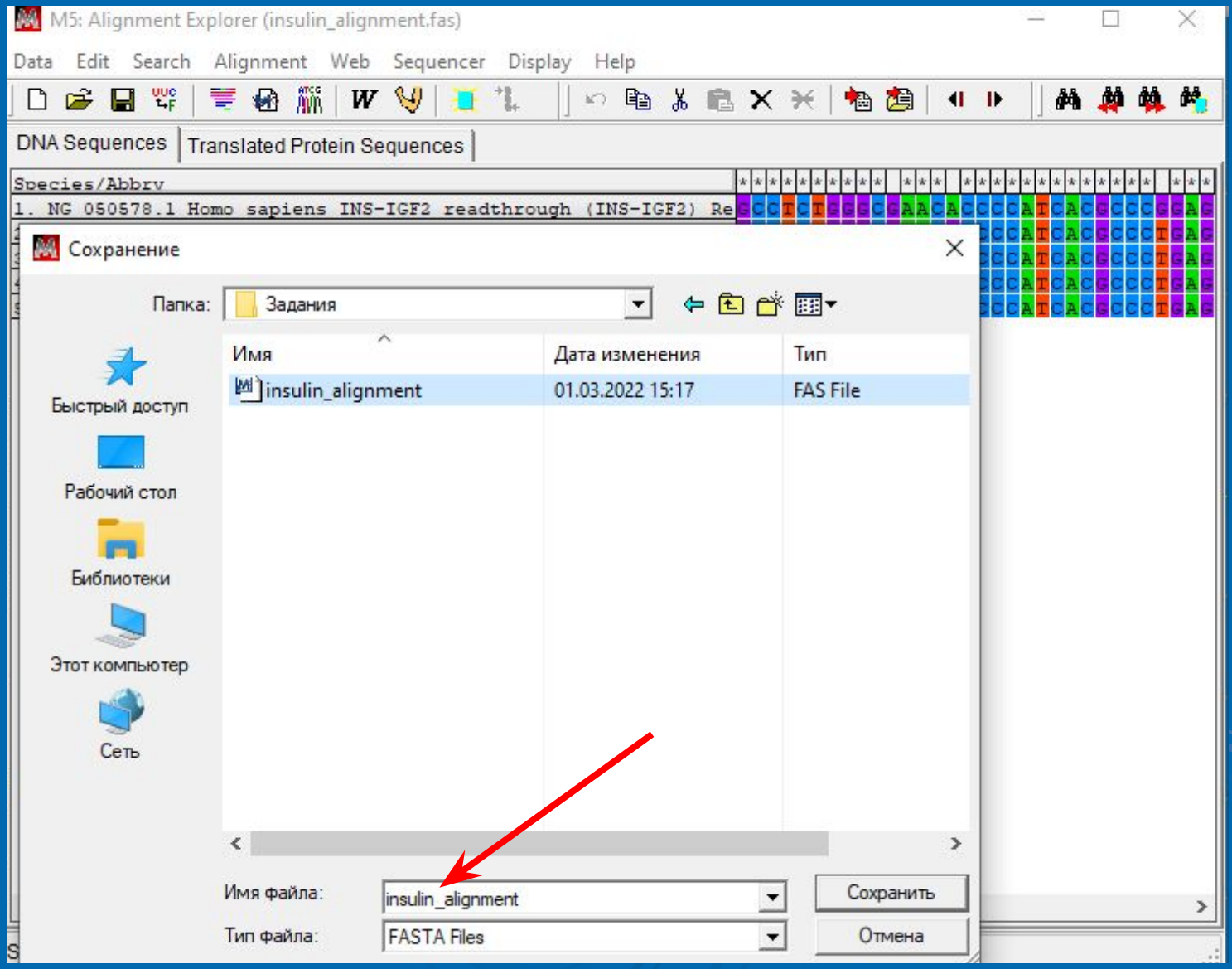
Задача 2

9. Сохранить выравнивание в файл с названием insulin_alignment в формате FASTA



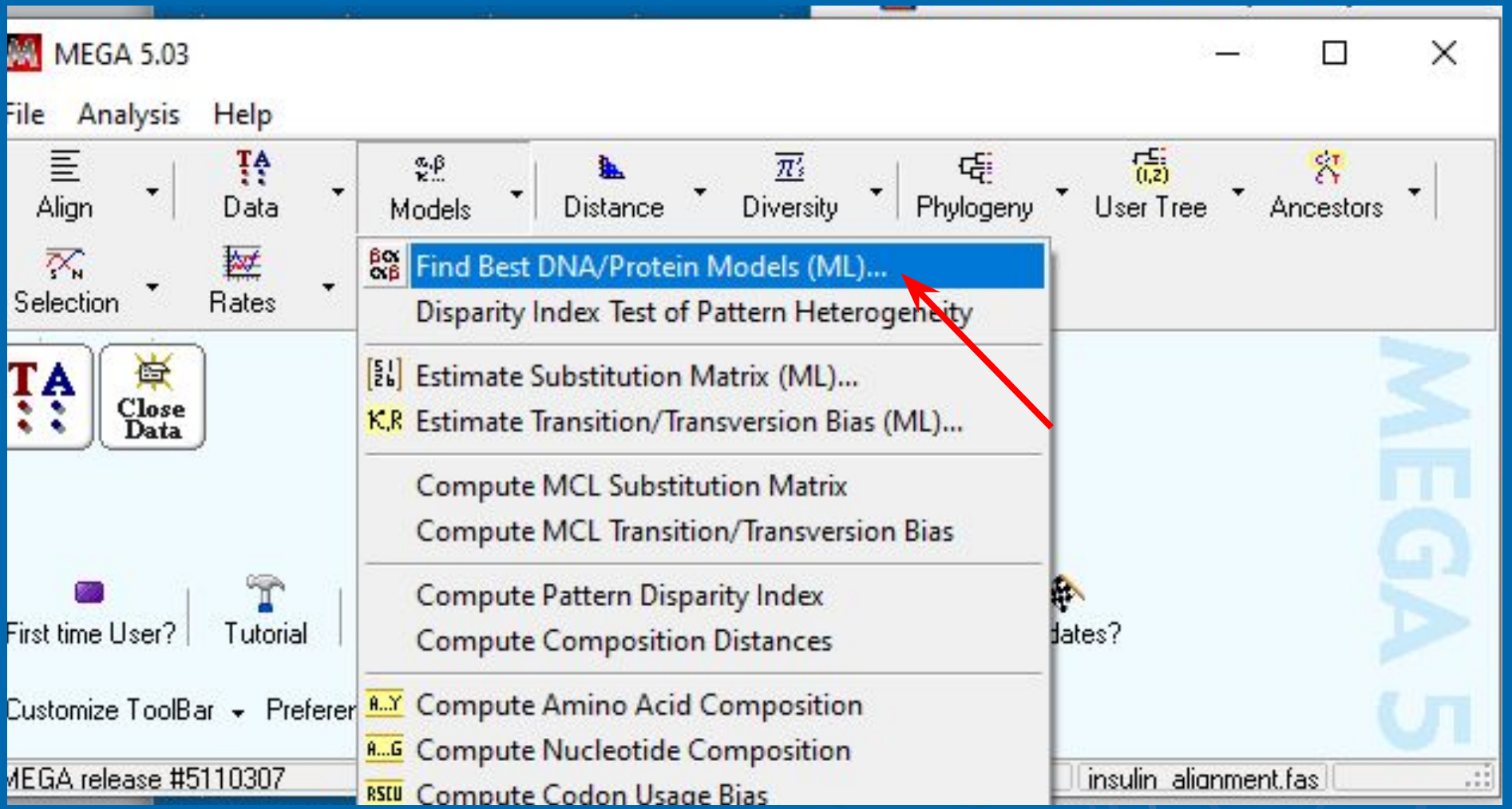
Задача 2

9. Сохранить выравнивание в файл с названием insulin_alignment в формате FASTA



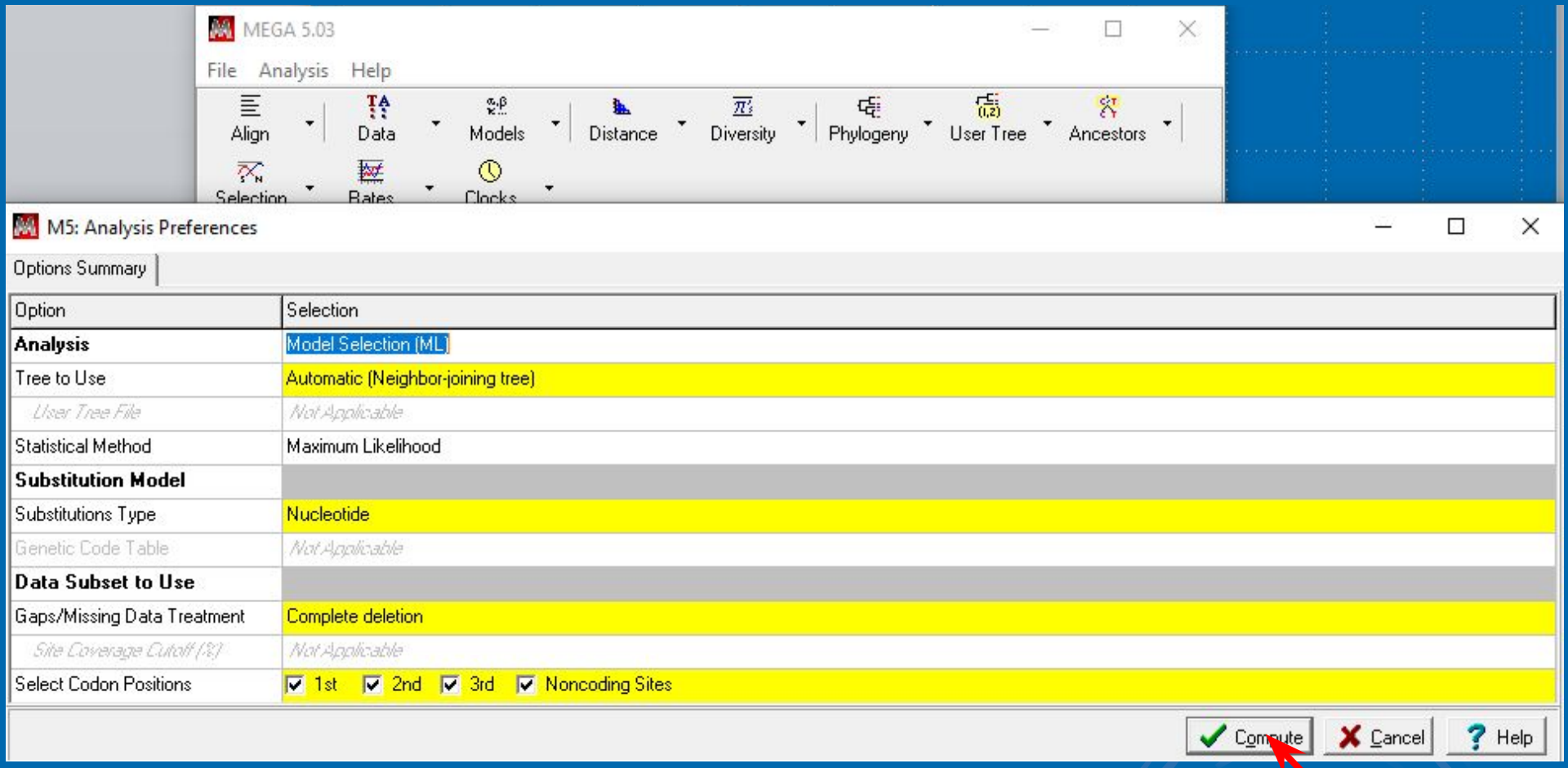
Задача 2

9. Найти наилучшую модель построения филогенетического дерева для полученного



Задача 2

9. Найти наилучшую модель построения филогенетического дерева для полученного



Задача 2

9. Найти наилучшую модель построения филогенетического дерева для полученного выравнив

M5: Find Best-Fit Substitution Model (ML)

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*R*	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG
T92	9	3337.959	3279.832	-1630.897	n/a	n/a	2.44	0.176	0.176	0.324	0.324	0.024	0.044	0.236	0.024	0.236	0.044	0.024	0.128	0.04
T92+G	10	3340.826	3276.245	-1628.099	n/a	0.23	2.60	0.176	0.176	0.324	0.324	0.023	0.042	0.240	0.023	0.240	0.042	0.023	0.130	0.04
T92+I	10	3346.422	3281.841	-1630.897	0.00	n/a	2.44	0.176	0.176	0.324	0.324	0.024	0.044	0.236	0.024	0.236	0.044	0.024	0.128	0.04
T92+G+I	11	3349.091	3278.057	-1628.000	0.00	0.14	2.66	0.176	0.176	0.324	0.324	0.022	0.041	0.242	0.022	0.242	0.041	0.022	0.131	0.04
HKY	11	3354.155	3283.121	-1630.533	n/a	n/a	2.44	0.168	0.184	0.321	0.327	0.025	0.044	0.238	0.023	0.234	0.044	0.023	0.134	0.04
HKY+G	12	3357.024	3279.537	-1627.735	n/a	0.23	2.60	0.168	0.184	0.321	0.327	0.024	0.042	0.242	0.022	0.238	0.042	0.022	0.136	0.04
HKY+I	12	3362.311	3284.825	-1630.379	0.05	n/a	2.45	0.168	0.184	0.321	0.327	0.025	0.044	0.239	0.023	0.234	0.044	0.023	0.134	0.04
TN93	12	3362.456	3284.969	-1630.452	n/a	n/a	2.44	0.168	0.184	0.321	0.327	0.025	0.044	0.253	0.023	0.220	0.044	0.023	0.126	0.04
TN93+G	13	3365.246	3281.307	-1627.615	n/a	0.23	2.61	0.168	0.184	0.321	0.327	0.024	0.042	0.263	0.022	0.219	0.042	0.022	0.125	0.04
HKY+G+I	13	3365.289	3281.351	-1627.637	0.00	0.14	2.66	0.168	0.184	0.321	0.327	0.023	0.041	0.244	0.021	0.239	0.042	0.021	0.137	0.04
TN93+I	13	3370.819	3286.881	-1630.402	0.02	n/a	2.45	0.168	0.184	0.321	0.327	0.025	0.044	0.254	0.023	0.220	0.044	0.023	0.126	0.04
TN93+G+I	14	3373.474	3283.085	-1627.498	0.00	0.13	2.67	0.168	0.184	0.321	0.327	0.023	0.041	0.267	0.021	0.218	0.042	0.021	0.125	0.04
GTR	15	3387.500	3290.661	-1630.280	n/a	n/a	2.44	0.168	0.184	0.321	0.327	0.023	0.034	0.254	0.021	0.221	0.052	0.018	0.126	0.04
GTR+G	16	3390.361	3287.073	-1627.479	n/a	0.23	2.61	0.168	0.184	0.321	0.327	0.023	0.032	0.263	0.021	0.220	0.049	0.017	0.126	0.04
GTR+I	16	3395.652	3292.363	-1630.124	0.05	n/a	2.45	0.168	0.184	0.321	0.327	0.023	0.034	0.254	0.021	0.221	0.052	0.018	0.126	0.04
GTR+G+I	17	3398.609	3288.872	-1627.371	0.00	0.13	2.67	0.168	0.184	0.321	0.327	0.023	0.032	0.266	0.021	0.219	0.048	0.017	0.125	0.04
K2	8	3417.407	3365.735	-1674.852	n/a	n/a	2.44	0.250	0.250	0.250	0.250	0.036	0.036	0.177	0.036	0.177	0.036	0.036	0.177	0.03
K2+G	9	3421.916	3363.790	-1672.876	n/a	0.23	3.53	0.250	0.250	0.250	0.250	0.028	0.028	0.195	0.028	0.195	0.028	0.028	0.195	0.02
K2+I	9	3425.870	3367.743	-1674.852	0.00	n/a	2.44	0.250	0.250	0.250	0.250	0.036	0.036	0.177	0.036	0.177	0.036	0.036	0.177	0.03
K2+G+I	10	3429.087	3364.506	-1672.230	0.00	0.17	2.59	0.250	0.250	0.250	0.250	0.035	0.035	0.180	0.035	0.180	0.035	0.035	0.180	0.03
JC	7	3444.284	3399.069	-1692.522	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.08
JC+G	8	3448.268	3396.596	-1690.283	n/a	0.22	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.08
JC+I	8	3452.638	3400.967	-1692.468	0.02	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.08
JC+G+I	9	3456.730	3398.604	-1690.283	0.00	0.22	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.08

NOTE.-- Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (*lnL*), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+*G*) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (*R*) are shown for each model, as well. They are followed by nucleotide frequencies (*f*) and rates of base substitutions (*r*) for each nucleotide pair. Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. The analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 947 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [2].

Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

Задача 2

10. Смотрим расшифровку наилучшей модели внизу таблицы, необходимую для построения

shown for each model, as well. They are followed by nucleotide frequen
eous *r* should be considered when evaluating them. For simplicity, sum of
matically computed. The analysis involved 5 nucleotide sequences. Codo
ata were eliminated. There were a total of 947 positions in the final datas

o-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-para

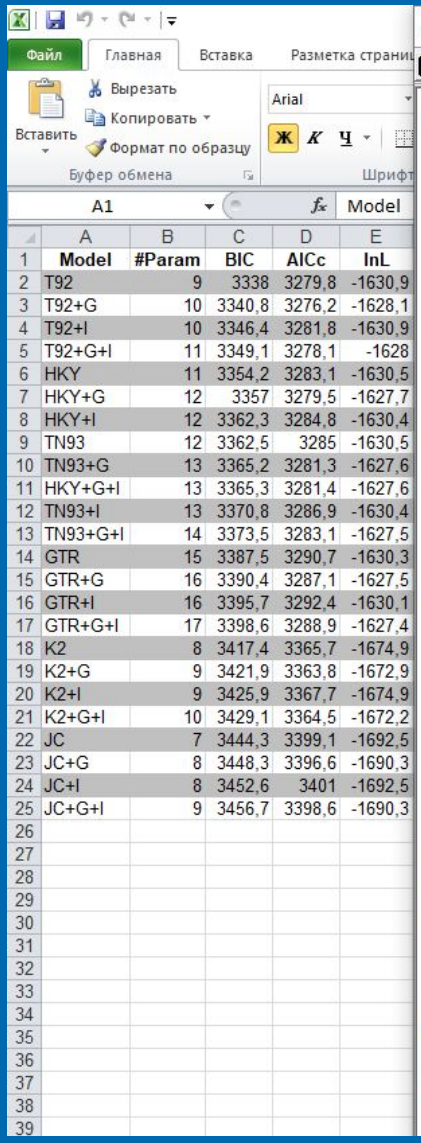
iversity Press, New York.

nary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*

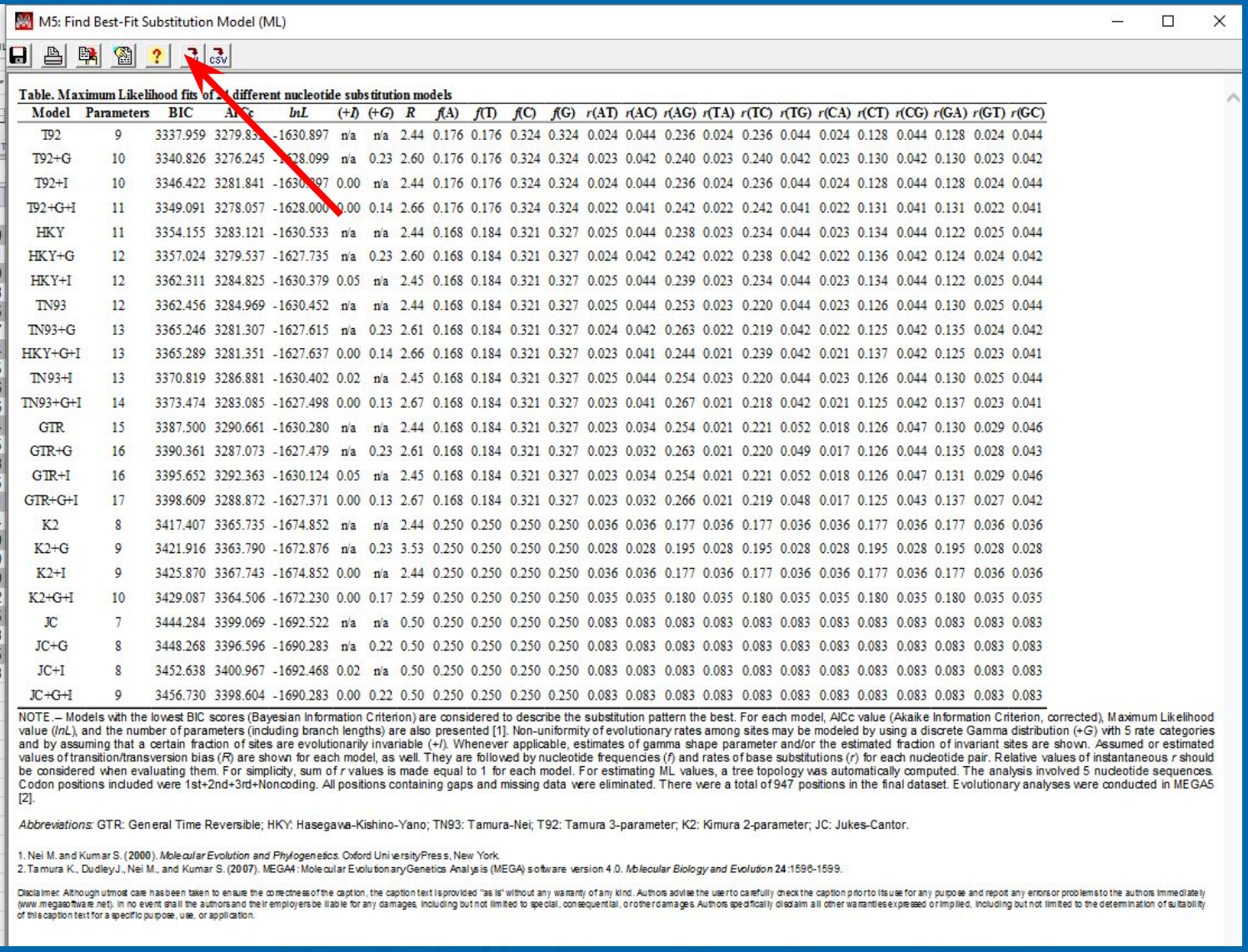
caption text is provided "as is" without any warranty of any kind. Authors advise the user to carefully
are.net). In no event shall the authors and their employers be liable for any damages, including but
r implied, including but not limited to the determination of suitability of this caption text for a spec

Задача 2

11. Сохраняем таблицу в формате Ecxel с именем insulin_best model



	A	B	C	D	E
1	Model	#Param	BIC	AICc	lnL
2	T92	9	3338	3279,8	-1630,9
3	T92+G	10	3340,8	3276,2	-1628,1
4	T92+I	10	3346,4	3281,8	-1630,9
5	T92+G+I	11	3349,1	3278,1	-1628
6	HKY	11	3354,2	3283,1	-1630,5
7	HKY+G	12	3357	3279,5	-1627,7
8	HKY+I	12	3362,3	3284,8	-1630,4
9	TN93	12	3362,5	3285	-1630,5
10	TN93+G	13	3365,2	3281,3	-1627,6
11	HKY+G+I	13	3365,3	3281,4	-1627,6
12	TN93+I	13	3370,8	3286,9	-1630,4
13	TN93+G+I	14	3373,5	3283,1	-1627,5
14	GTR	15	3387,5	3290,7	-1630,3
15	GTR+G	16	3390,4	3287,1	-1627,5
16	GTR+I	16	3395,7	3292,4	-1630,1
17	GTR+G+I	17	3398,6	3288,9	-1627,4
18	K2	8	3417,4	3365,7	-1674,9
19	K2+G	9	3421,9	3363,8	-1672,9
20	K2+I	9	3425,9	3367,7	-1674,9
21	K2+G+I	10	3429,1	3364,5	-1672,2
22	JC	7	3444,3	3399,1	-1692,5
23	JC+G	8	3448,3	3396,6	-1690,3
24	JC+I	8	3452,6	3401	-1692,5
25	JC+G+I	9	3456,7	3398,6	-1690,3



M5: Find Best-Fit Substitution Model (ML)

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
T92	9	3337.959	3279.83	-1630.897	n/a	n/a	2.44	0.176	0.176	0.324	0.324	0.024	0.044	0.236	0.024	0.236	0.044	0.024	0.128	0.044	0.128	0.024	0.044
T92+G	10	3340.826	3276.245	-1628.099	n/a	0.23	2.60	0.176	0.176	0.324	0.324	0.023	0.042	0.240	0.023	0.240	0.042	0.023	0.130	0.042	0.130	0.023	0.042
T92+I	10	3346.422	3281.841	-1630.897	0.00	n/a	2.44	0.176	0.176	0.324	0.324	0.024	0.044	0.236	0.024	0.236	0.044	0.024	0.128	0.044	0.128	0.024	0.044
T92+G+I	11	3349.091	3278.057	-1628.000	0.00	0.14	2.66	0.176	0.176	0.324	0.324	0.022	0.041	0.242	0.022	0.242	0.041	0.022	0.131	0.041	0.131	0.022	0.041
HKY	11	3354.155	3283.121	-1630.533	n/a	n/a	2.44	0.168	0.184	0.321	0.327	0.025	0.044	0.238	0.023	0.234	0.044	0.023	0.134	0.044	0.122	0.025	0.044
HKY+G	12	3357.024	3279.537	-1627.735	n/a	0.23	2.60	0.168	0.184	0.321	0.327	0.024	0.042	0.242	0.022	0.238	0.042	0.022	0.136	0.042	0.124	0.024	0.042
HKY+I	12	3362.311	3284.825	-1630.379	0.05	n/a	2.45	0.168	0.184	0.321	0.327	0.025	0.044	0.239	0.023	0.234	0.044	0.023	0.134	0.044	0.122	0.025	0.044
TN93	12	3362.456	3284.969	-1630.452	n/a	n/a	2.44	0.168	0.184	0.321	0.327	0.025	0.044	0.253	0.023	0.220	0.044	0.023	0.126	0.044	0.130	0.025	0.044
TN93+G	13	3365.246	3281.307	-1627.615	n/a	0.23	2.61	0.168	0.184	0.321	0.327	0.024	0.042	0.263	0.022	0.219	0.042	0.022	0.125	0.042	0.135	0.024	0.042
HKY+G+I	13	3365.289	3281.351	-1627.637	0.00	0.14	2.66	0.168	0.184	0.321	0.327	0.023	0.041	0.244	0.021	0.239	0.042	0.021	0.137	0.042	0.125	0.023	0.041
TN93+I	13	3370.819	3286.881	-1630.402	0.02	n/a	2.45	0.168	0.184	0.321	0.327	0.025	0.044	0.254	0.023	0.220	0.044	0.023	0.126	0.044	0.130	0.025	0.044
TN93+G+I	14	3373.474	3283.085	-1627.498	0.00	0.13	2.67	0.168	0.184	0.321	0.327	0.023	0.041	0.267	0.021	0.218	0.042	0.021	0.125	0.042	0.137	0.023	0.041
GTR	15	3387.500	3290.661	-1630.280	n/a	n/a	2.44	0.168	0.184	0.321	0.327	0.023	0.034	0.254	0.021	0.221	0.052	0.018	0.126	0.047	0.130	0.029	0.046
GTR+G	16	3390.361	3287.073	-1627.479	n/a	0.23	2.61	0.168	0.184	0.321	0.327	0.023	0.032	0.263	0.021	0.220	0.049	0.017	0.126	0.044	0.135	0.028	0.043
GTR+I	16	3395.652	3292.363	-1630.124	0.05	n/a	2.45	0.168	0.184	0.321	0.327	0.023	0.034	0.254	0.021	0.221	0.052	0.018	0.126	0.047	0.131	0.029	0.046
GTR+G+I	17	3398.609	3288.872	-1627.371	0.00	0.13	2.67	0.168	0.184	0.321	0.327	0.023	0.032	0.266	0.021	0.219	0.048	0.017	0.125	0.043	0.137	0.027	0.042
K2	8	3417.407	3365.735	-1674.852	n/a	n/a	2.44	0.250	0.250	0.250	0.250	0.036	0.036	0.177	0.036	0.177	0.036	0.036	0.177	0.036	0.177	0.036	0.036
K2+G	9	3421.916	3363.790	-1672.876	n/a	0.23	3.53	0.250	0.250	0.250	0.250	0.028	0.028	0.195	0.028	0.195	0.028	0.028	0.195	0.028	0.195	0.028	0.028
K2+I	9	3425.870	3367.743	-1674.852	0.00	n/a	2.44	0.250	0.250	0.250	0.250	0.036	0.036	0.177	0.036	0.177	0.036	0.036	0.177	0.036	0.177	0.036	0.036
K2+G+I	10	3429.087	3364.506	-1672.230	0.00	0.17	2.59	0.250	0.250	0.250	0.250	0.035	0.035	0.180	0.035	0.180	0.035	0.035	0.180	0.035	0.180	0.035	0.035
JC	7	3444.284	3399.069	-1692.522	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G	8	3448.268	3396.596	-1690.283	n/a	0.22	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	8	3452.638	3400.967	-1692.468	0.02	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+G+I	9	3456.730	3398.604	-1690.283	0.00	0.22	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083

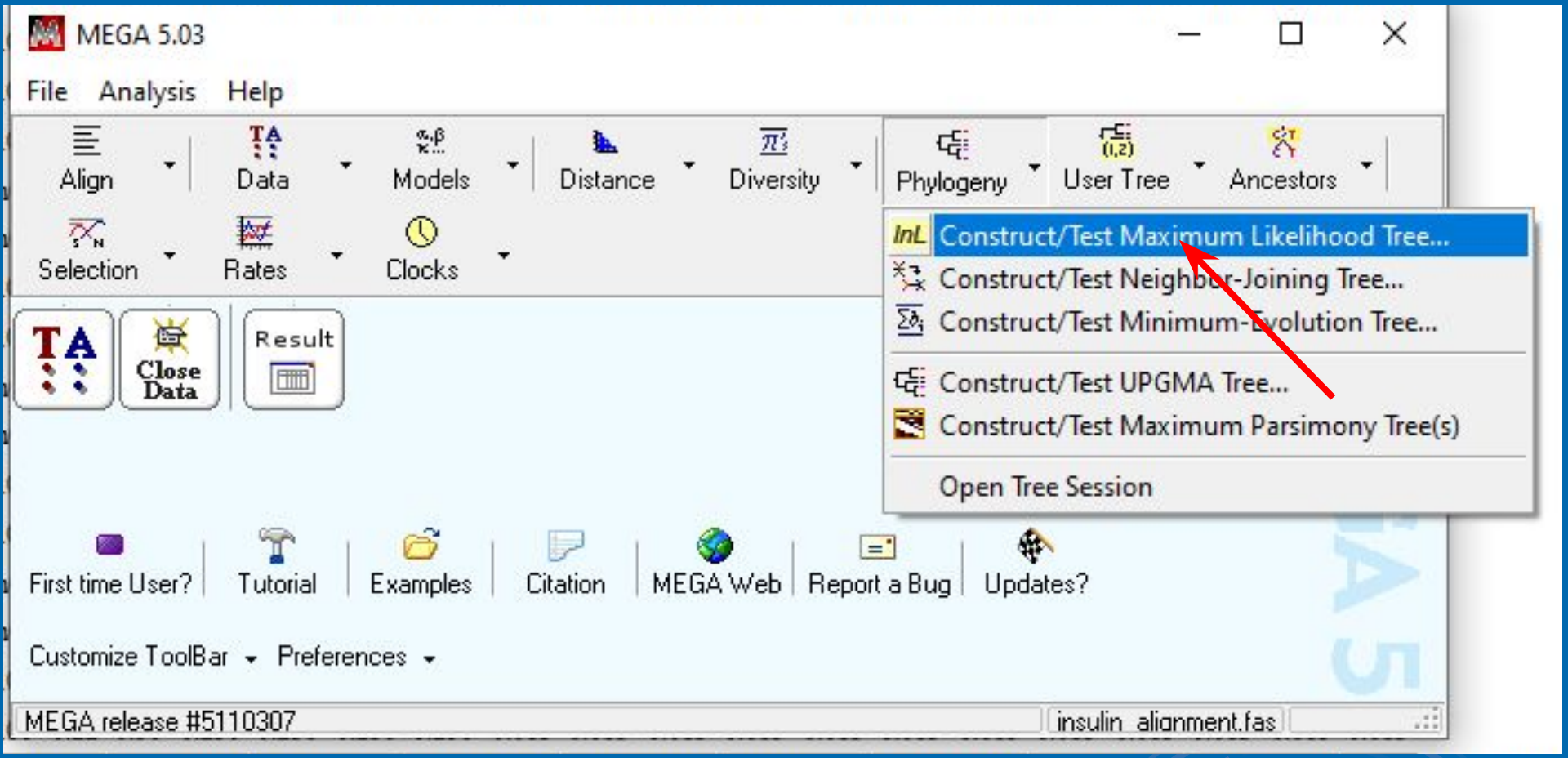
NOTE.– Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (*lnL*), and the number of parameters (including branch lengths) are also presented [1]. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (*+G*) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (*+I*). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (*R*) are shown for each model, as well. They are followed by nucleotide frequencies (*f*) and rates of base substitutions (*r*) for each nucleotide pair. Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. The analysis involved 5 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 947 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [2].

Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

1. Nei M. and Kumar S. (**2000**). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
2. Tamura K., Dudley J., Nei M., and Kumar S. (**2007**). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**:1596-1599.

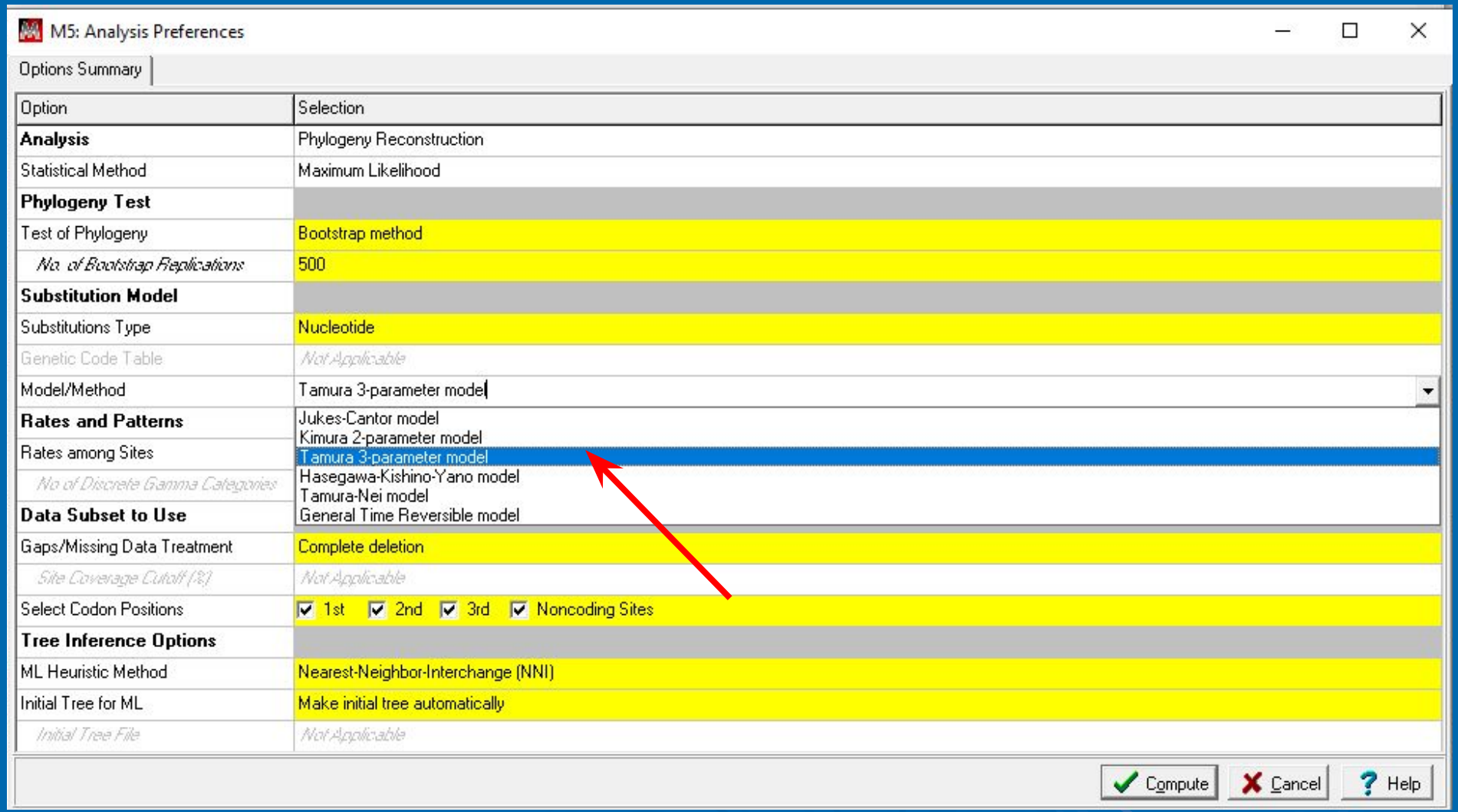
Задача 2

12. Построить филогенетическое дерево для выбранных нуклеотидных последовательностей



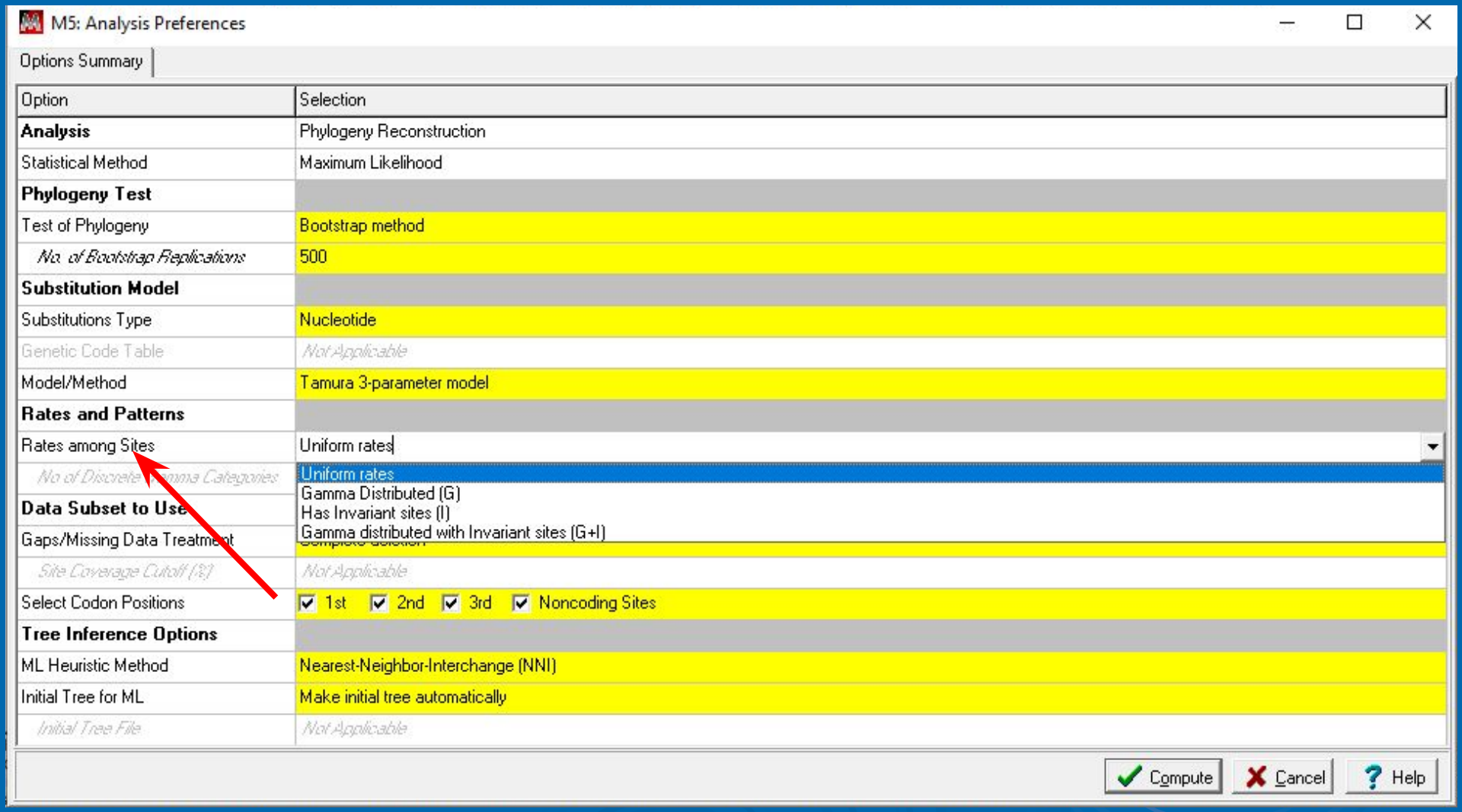
Задача 2

12. Указываем нужную (наилучшую) модель



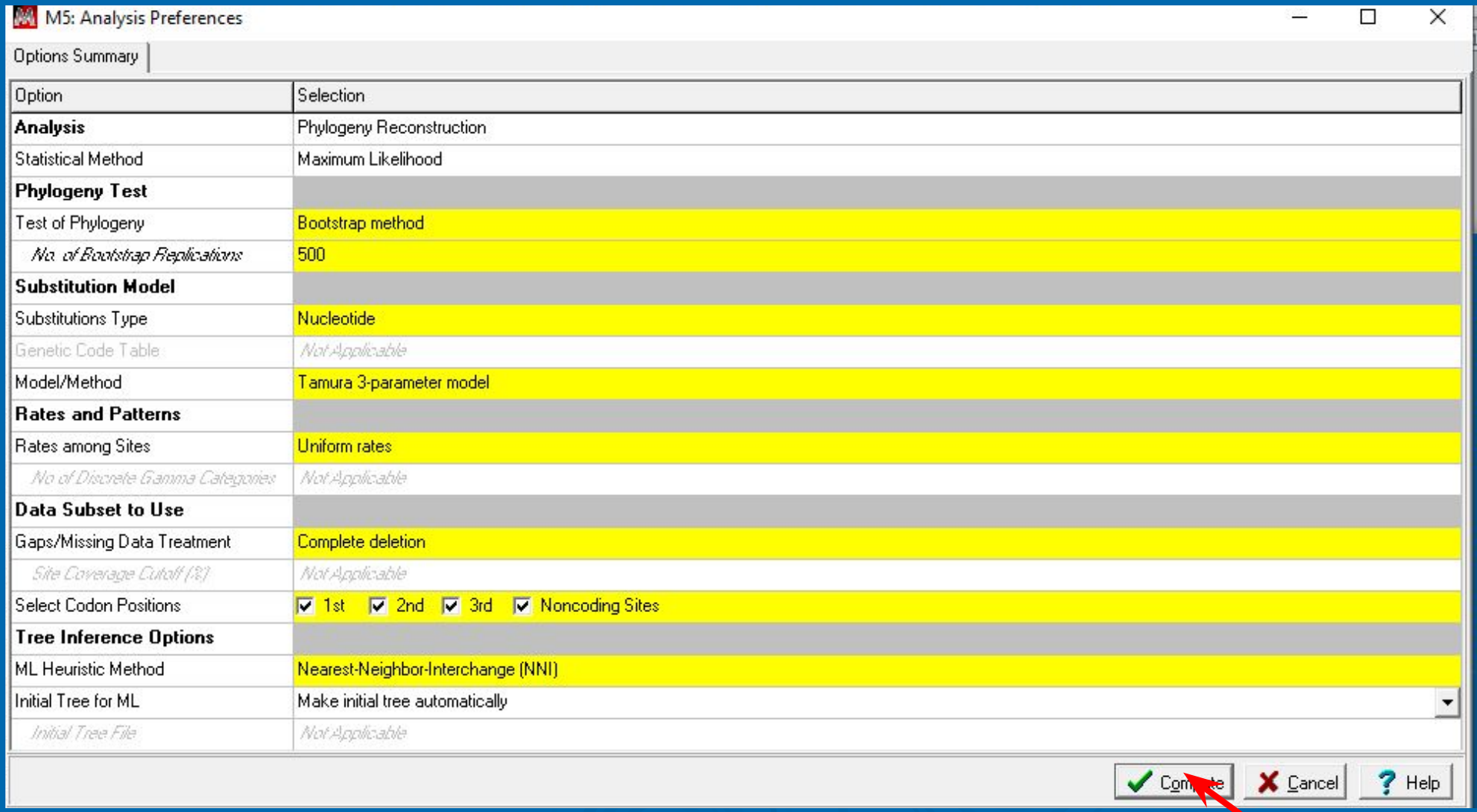
Задача 2

12. При необходимости – другие параметры, указанные в наилучшей модели



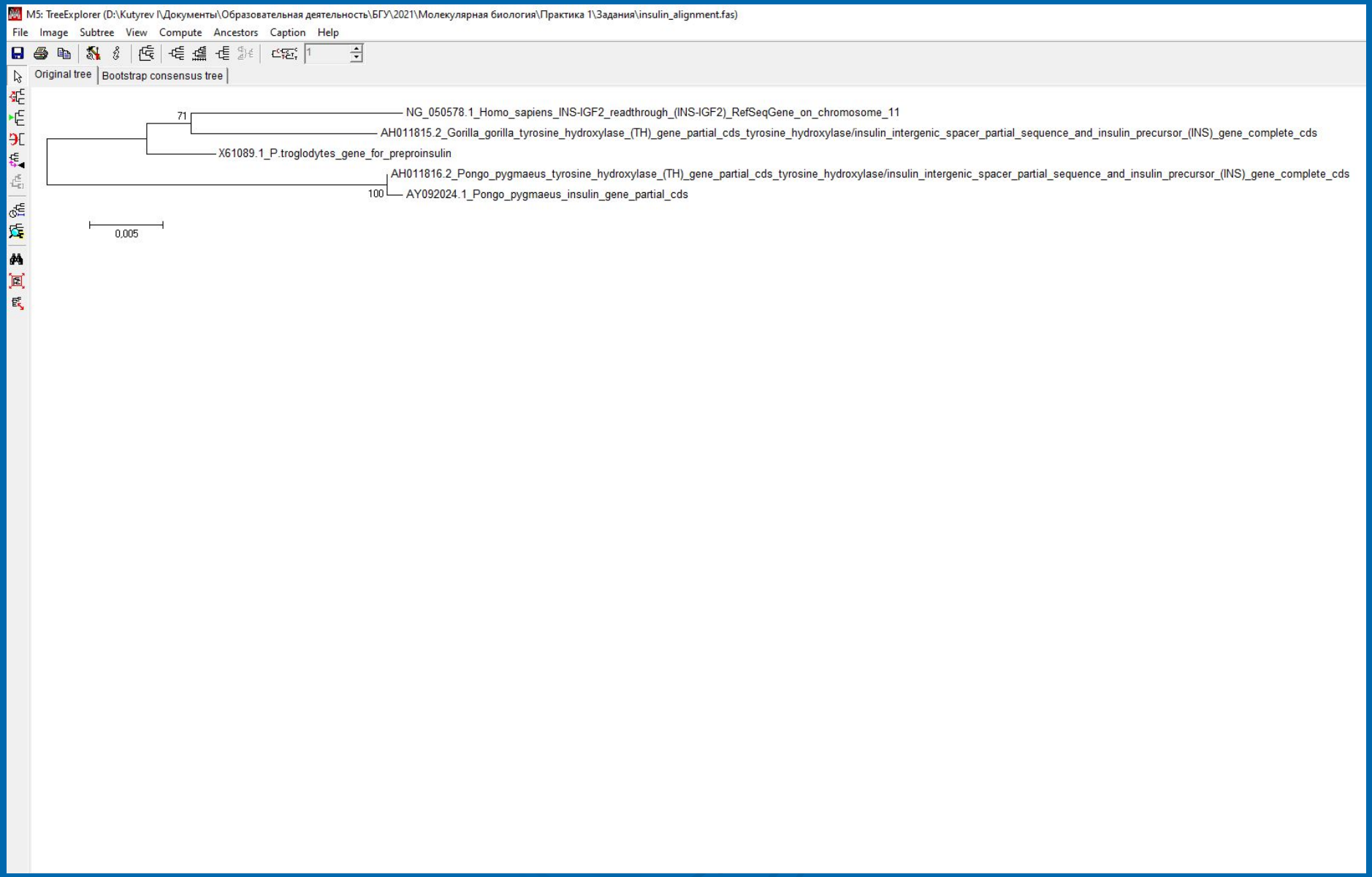
Задача 2

12. Строим филогенетическое дерево для выбранных нуклеотидных последовательностей



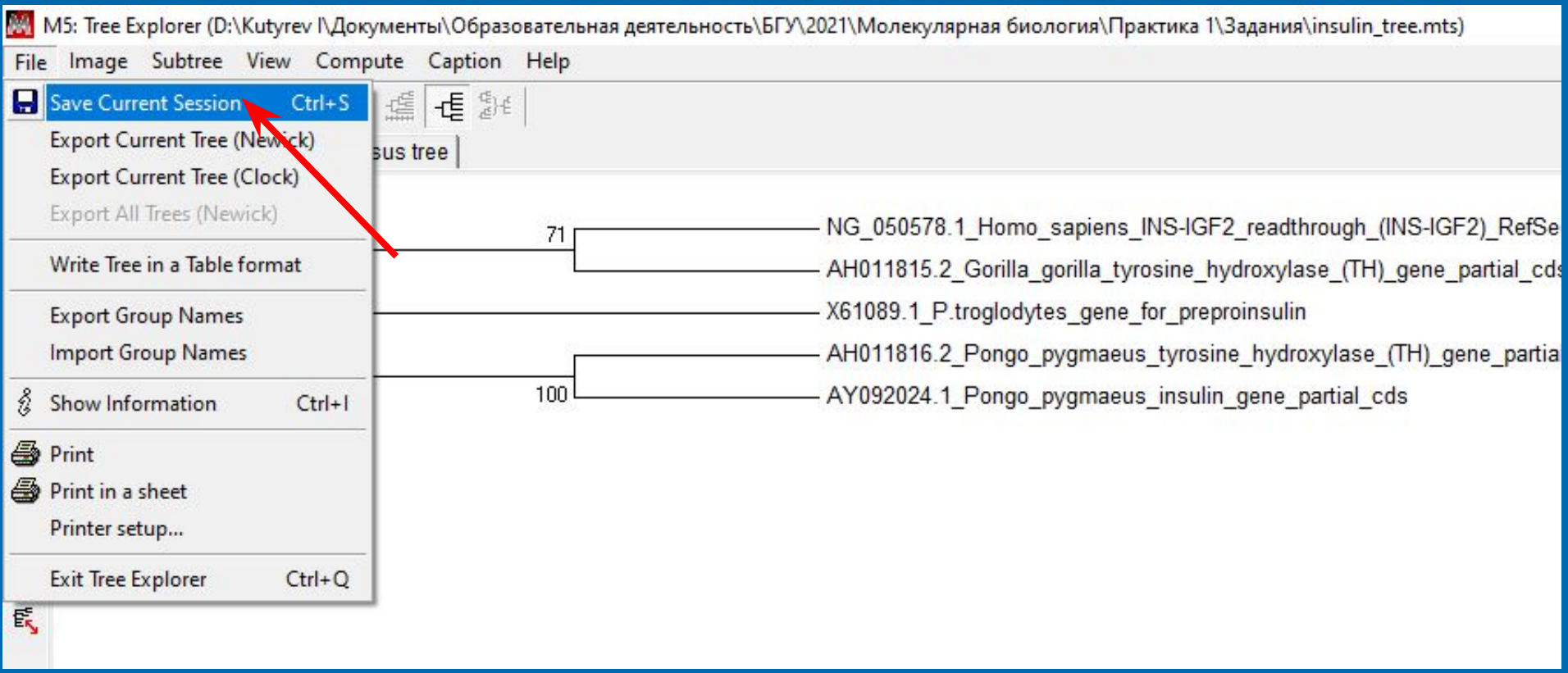
Задача 2

12. Строим филогенетическое дерево для выбранных нуклеотидных последовательностей



Задача 2

13. Сохраняем текущую сессию под названием insulin_tree



Задача 2. Форма отчета

- **Каждый лично отправляет мне на почту sankaar@mail.ru :**
- **1) файл insulin.fasta**
- **2) insulin_alignment.fas**
- **3) insulin_best model.xls**
- **4) insulin_tree.mts**