

# Молекулярная биология



# Некоторые параметры молекул ДНК и белка:

- Один шаг это полный виток спирали ДНК-поворот на  $360^\circ$
- 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:  **$A+G=T+C$**  (Правило Чаргаффа)
- Комплементарность нуклеотидов:  **$A=T$ ;  $G=C$**
- Цепи ДНК удерживаются водородными связями, которые образуются между комплементарными азотистыми основаниями: аденин с тиминном 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

```

                                     111111111122
123456789012345678901
1  ATACCTGCGATAGCTTCTGAT
   ||||| ||| |*****
2  ATACCTGCGAAGCTTCTGAT.
```

# Выравнивание генетических последовательностей

```
                111111111122
123456789012345678901
1  ATACCTGCGATAGCTTCTGAT
   ||| ||| ||| ||| ||| ||| ||| |||
2  ATACCTGCGA . AGCTTCTGAT
```



# 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/>

## 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

PROTEIN

PROTEIN

DNA

RNA

Or, [upload a file](#):  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.

*(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)*

1. Выбираем  
нужный тип  
данных (белок,  
ДНК или РНК)  
В нашем случае -  
DNA

# 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:

```
>UcCmIs_JQ245480
ATTCCAGCATATGTTTGGATTTTTGGTACCCTGAGGTATATGTTTAAATTTACCTGGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCAT
ATACACACCTTTGGGTTTTATGGGTTGTTATTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTGTGTGGGGGCATCATATGTTTACAGTAGGTTTAGATGTA
AAGACGGCTGTTTTCTTCAGTTCGTACTATGATTATTGGTGTTCCTACCGGTATAAAGGTATTTCTTGATTATACATGATTCATAAAAGTCGGTTTTCACTC
CGTGAGCCTGTTTTGGTGAGTTTTATCATTTATTGTGTTGTTACTATTGGGGGTGTACGGGTATTATCTTTTCAGCTTGTGTGCTTGATAATATTTGTCAT
GATACTTGATTGTTGTTAGCTCACTTTCATTATGTAATGTCATTAGGTTCTTACATAAGTATAATAGTGTTTTTCGTTTGATGATGGCCAGTCATTACAGGTGTA
GATTGAATAAGTATTTACTA
>SpCmKo_JQ245482
```

Or, upload a file:  Не выбран ни один файл

[Use an 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.

(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)

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

## 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

```
TTATTCAGCATATGTTTGGATTTTTGGTCACCCTGAGGTATATGTTTAAATTTACCTGGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTCAT
ATGACACCTTTGGGTTTTATGGGTTGTTATTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTGTGTGGGGGCATCATATGTTTACAGTAGGTTTAGATGTA
AAGACGGCTGTTTTCTTCAGTTCGTACTATGATTATTGGTGTTCCTACCGGTATAAAGGTATTTCTTGATTATACATGATTCTAAAAAGTCGTGTTTCACTC
CGTGAGCCTGTTTTTGGTGAGTTTTATCATTTATTGTGTTTACTATTGGGGGTGTACGGGTATTATCTTTTCAGCTTGTGTGCTTGATAATATTTGTCAT
GATACTTGATTCGTTGTAGCTCACTTTCATTATGTAATGTCATTAGGTTCTTACATAAAGTATAATAGTGTTTTTCGTTTGATGATGGCCAGTCATTACAGGTGTA
GATTGAATAAGTATTTACTA
```

>SpCmKo\_JQ245482

Or, upload a file:  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.

(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)

3. Нажимаем Submit



# 4. Получаем результат

Input form	Web services	Help & Documentation	Bioinformatics Tools FAQ
DeCmCh1.2-Hap1		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
SpTaKo_JQ245484		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeClUK_KC812045		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
SpTaKo_JQ245483		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo11-Hap18		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeOkCL_AB623150		GGTTTTGGAAATGATTAGTCATGTTTGTAGAAACTTAGGTTGTTTCATATGATACCTTTGGG	120
SpCmKo_JQ245482		GGTTTTGGAAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeOkCL_AB623149		GGTTTTGGAAATGATTAGTCATGTTTGTAGAAACTTAGGTTGTTTCATATGATACCTTTGGG	120
KlUaKa_AB375661		GGTTTTGGAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
NiUaKm_JQ245471		GGTTTTGGAAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DiClbKa1-HapB		GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACCTTTGGG	120
DiClbKa11-HapH		GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACCTTTGGG	120
DiClbKa8.9-HapF		GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACCTTTGGG	120
DiClbKa16-HapL		GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACCTTTGGG	120
DiClbKa10-HapG		GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACCTTTGGG	120
SpCmCh12-HapA		GGGTTTGGTATGATTAGTCATGTTTGTAGAAATTTGGGTTGCTCATATGACACCTTTGGG	120
Dipl.grandis_AB425840		GGGTTTGGAAATGATTAGACATGTTTGTAGTAACTTAGGTTGTTTCATATGATACCTTTGGA	120
Dipl.balaenopterae_AB822370		GGGTTTGGAAATGATTAGACATGTTTGTAGTAACTTAGGTTGTTTCATATGATACCTTTGGA	120
DeSlAz_AB374223		GGTTTTGGAAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo13-Hap20		GGTTTTGGAAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo7.8-Hap15		GGTTTTGGAAATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo10-Hap17		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
SpCmCh9-Hap12		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeCmSe10.11-Hap7		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeCmCh5-Hap4		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeHsNE_KC812048		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo15-Hap22		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeBlHo14-Hap21		GGTTTTGGGATGATTAGCCATGCTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo18-Hap25		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
DeTnHo16-Hap23		GGTTTTGGGATGATTAGCCATGTTTGTAGAAATTTAGGTTGTTTCATATGACACCTTTGGG	120
		** ***** **	
DiOtUS_KY552872		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGGGGG	180
DiClbKa5.6-HapD		TTTTATGGGTGTTGTTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGGGGG	180
DiCsTy_JQ245472		TTTTATGGGTGTTGTTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGGGGG	180
SpOmKi_JQ245477		TTTTATGGGTGTTGTTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGGGGG	180
DiHpJP_AB979518		TTTTATGGGTGTTGTTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGGGGG	180
DiClbKa12-HapI		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGTGTGATGAGGG	180
DiClbKa14.15-HapK		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGAGGG	180
DiClbKa7-HapE		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGAGGA	180
DiClbKa13-HapJ		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGAGGA	180
DiClbKa2.3.4-HapC		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGAGGA	180
DiCsTy_JQ245474		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGAGGA	180
DiSaUK_FM209182		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTTTGTAGGTAGCGTTGATGAGGA	180
SpHsJP_AB488497		TTTTACGGGTGTTATTTGCTATGTTTTCTATAGTTTGTAGGTAGTGTGTTGTTGGGGG	180
UrUaUS_AB605762		TTTTATGGGTATTTATTTGCTATGTTTTCTATAGTCTGTTTGGGTAGTGTGTTGTTGGGGG	180
UrUaUS_AB605763		TTTTATGGGTATTTATTTGCTATGTTTTCTATAGTCTGTTTGGGTAGTGTGTTGTTGGGGG	180
NiOkJP_AB521677		TTTTACGGATGTTATTTGCTATGTTTTCTATAGTTTGTAGGTAGTGTGTTGTTGGGGG	180
DeTnHo4.5.6-Hap14		TTTTATGGGTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTGTTGTTGGGGG	180
DeTnHo2.3-Hap13		TTTTATGGGTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTGTTGTTGGGGG	180
DeHsCH_AM412738		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTCTGCTTGGGTAGTGTGTTGTTGGGGG	180
DeHsCZ_KC812047		TTTTATGGGTGTTATTTGCCATGTTTTCTATAGTCTGCTTGGGTAGTGTGTTGTTGGGGG	180
DeCmSe14-Hap10		TTTTATGGGTGTTATTTGCTATGTTTTCTATAGTTTGTAGGTAGTGTGTTGTTGGGGG	180
DeCmIs_JQ245480		TTTTATGGGTGTTATTTGCTATGTTTTCTATAGTCTGCTTGGGTAGTGTGTTGTTGGGGG	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-l20220228-023416-0897-41339244-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      TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiClbKa5.6-HapD     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAAAGTGTATGTCTTAATTTTACCG    60
DiCsTy_JQ245472     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAAAGTGTATGTCTTAATTTTACCG    60
SpOmKi_JQ245477     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAAAGTGTATGTCTTAATTTTACCG    60
DiHpJP_AB979518     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAAAGTGTATGTCTTAATTTTACCG    60
DiClbKa12-HapI      TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiClbKa14.15-HapK   TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiClbKa7-HapE       TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiClbKa13-HapJ      TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiClbKa2.3.4-HapC   TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiCsTy_JQ245474     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
DiSaUK_FM209182     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCG    60
SpHsJP_AB488497     TTATTTCCAGCATATGTTTGGATTTTTGGCCATCCCGAAGTATATGTTTTAATTTTACCT    60
UrUaUS_AB605762     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCC    60
UrUaUS_AB605763     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTGATGTCTTAATTTTACCC    60
NiOkJP_AB521677     TTATTTCCAGCATATGTTTGGATTTTTGGACACCCCGAAGTATATGTTTTAATTTTACCT    60
DeTnHo4.5.6-Hap14   TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeTnHo2.3-Hap13    TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeHsCH_AM412738     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATCTTACCT    60
DeHsCZ_KC812047     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATCTTACCT    60
DeCmSe14-Hap10     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeCmIs_JQ245480     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeCmBa_KC812046     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeCmSe13-Hap9      TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeTnHo12-Hap19     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeCmSe12-Hap8      TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeCmCh3-Hap2       TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
DeTnHo17-Hap24     TTATTTCCAGCATATGTTTGGATTTTTGGTCACCCGAGGTATATGTTTTAATTTTACCT    60
```

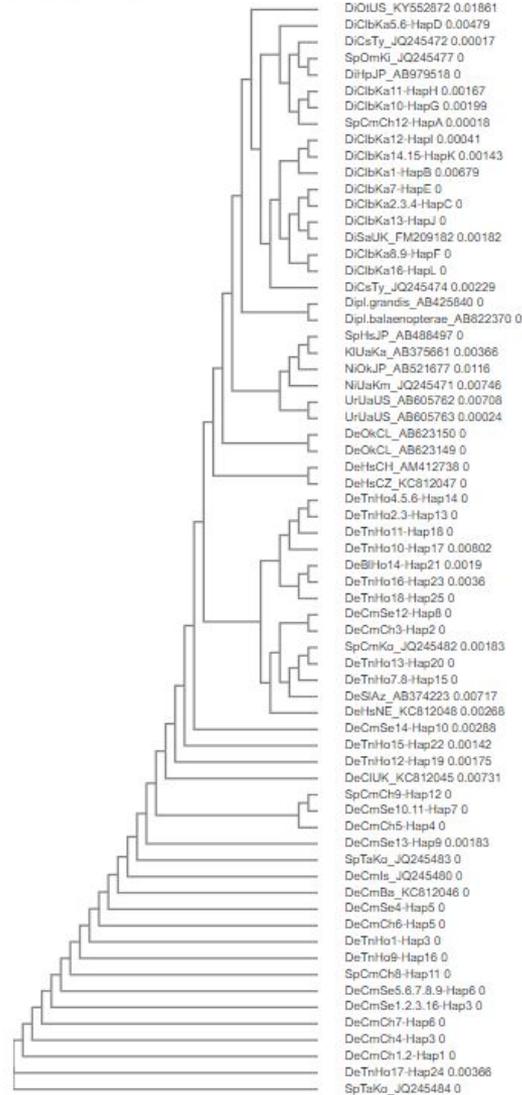


Download Phylogenetic Tree Data

## Phylogenetic Tree

Distance corrections

Branch length:  Cladogram  Real



# BLAST

- **BLAST** (англ. **B**asic **L**ocal **A**lignment **S**earch **T**ool — 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>



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

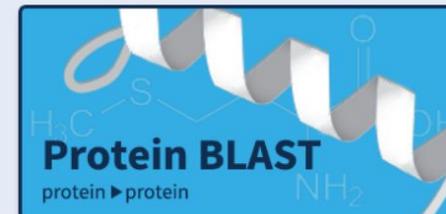
ElasticBLAST is here!

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



## BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Human

Mouse

Rat

Microbes

**blastn**

**blastx**



### 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

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**

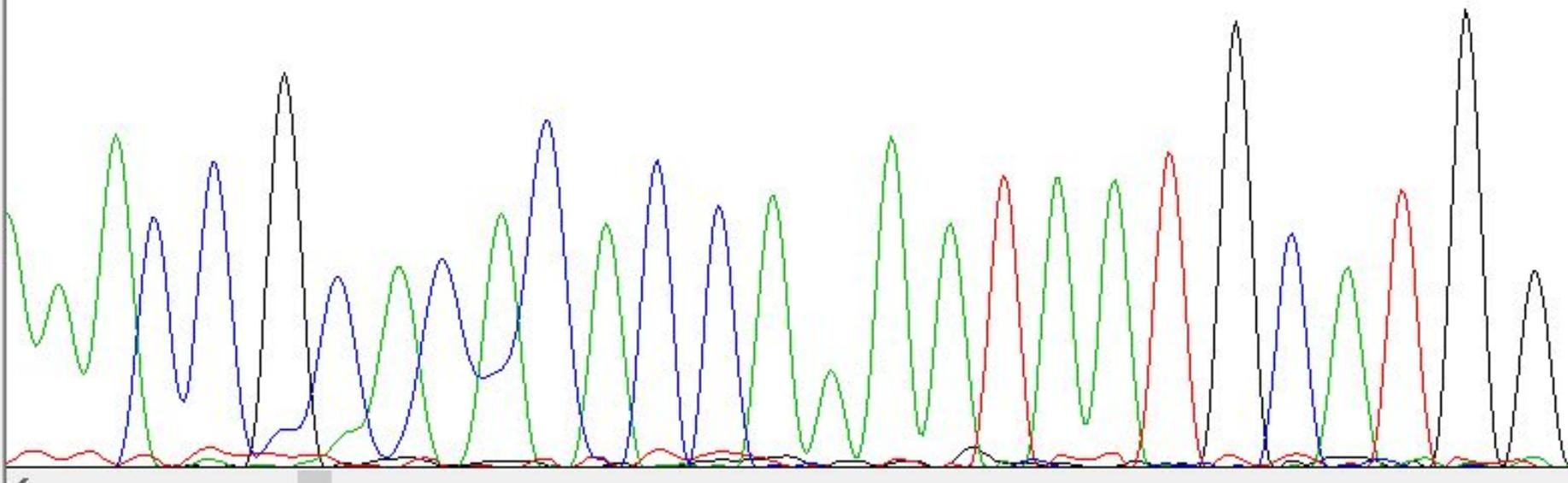
## Задача 1

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):

- 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

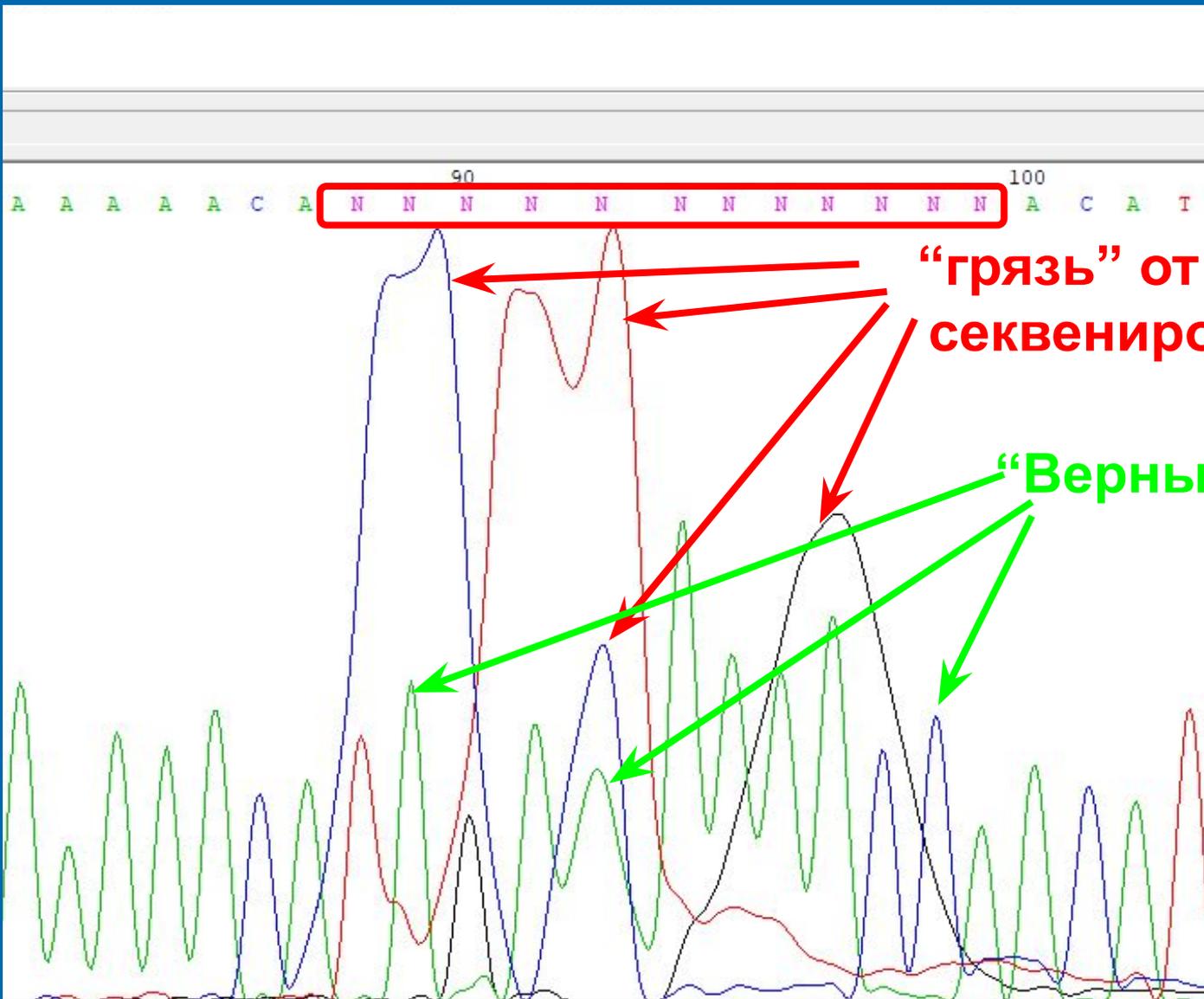
Sample: Dip4-CO1-R

60 70  
C A C C A A A A T A A T G C A T G G



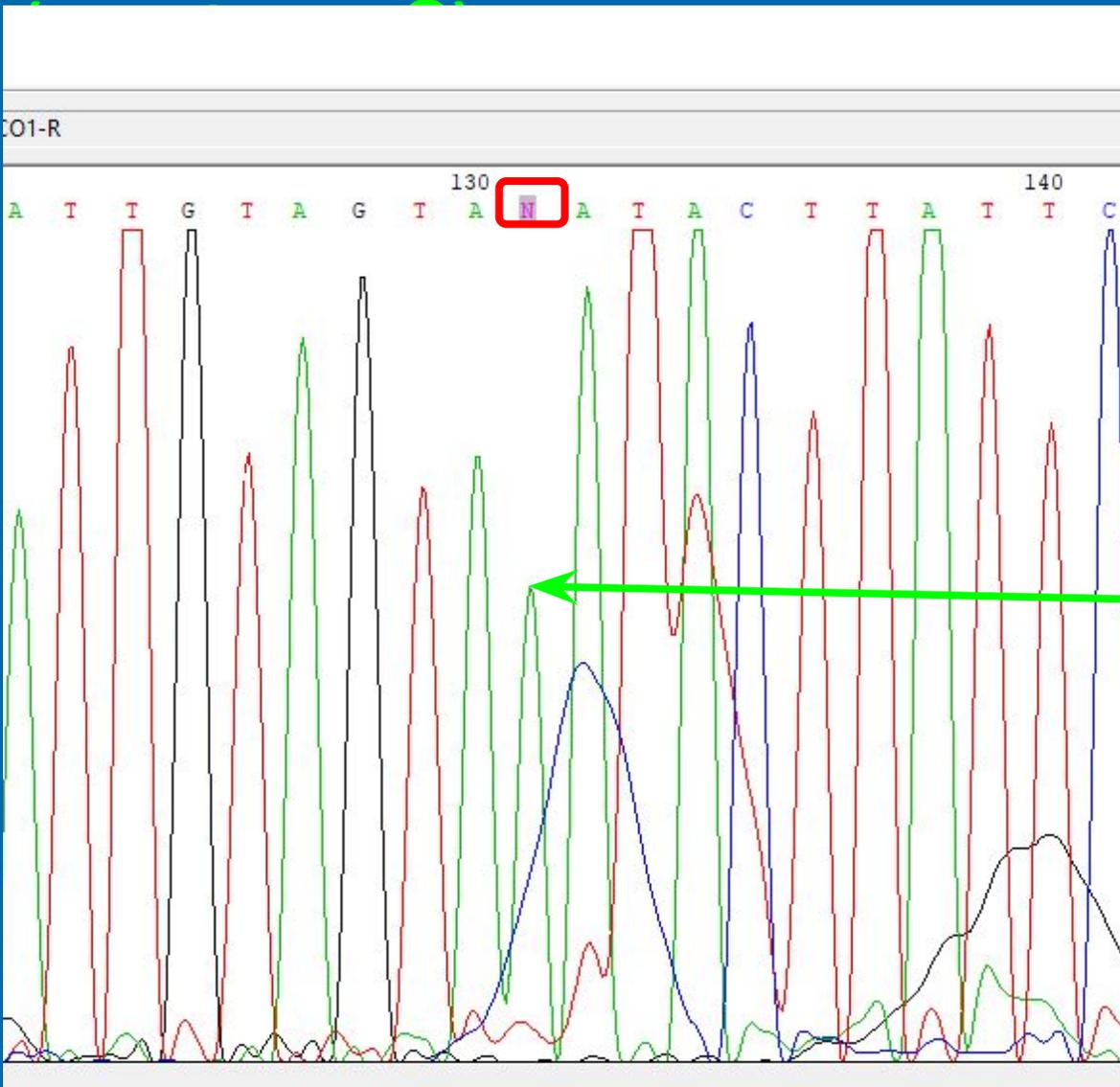
# Задача 1

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

The screenshot displays the Chromas Lite software interface. The main window shows a DNA sequence chromatogram with peaks and a sequence view above it. The sequence view shows the following nucleotides: A A A C T A C A A T G G C A T T G T A G T A N A T. The 'Export' button in the top toolbar is highlighted with a red box, and a red arrow points to the 'Export' dialog box. The dialog box is titled 'Export' and shows the following fields and options:

- Папка: Задания
- Имя: (empty)
- Дата изменения: (empty)
- Имя файла: Задание\_1-R.20120413T.A11
- Тип файла: FASTA
- Buttons: Сохранить, Отмена

The 'FASTA' option in the file type dropdown is also highlighted with a red box.

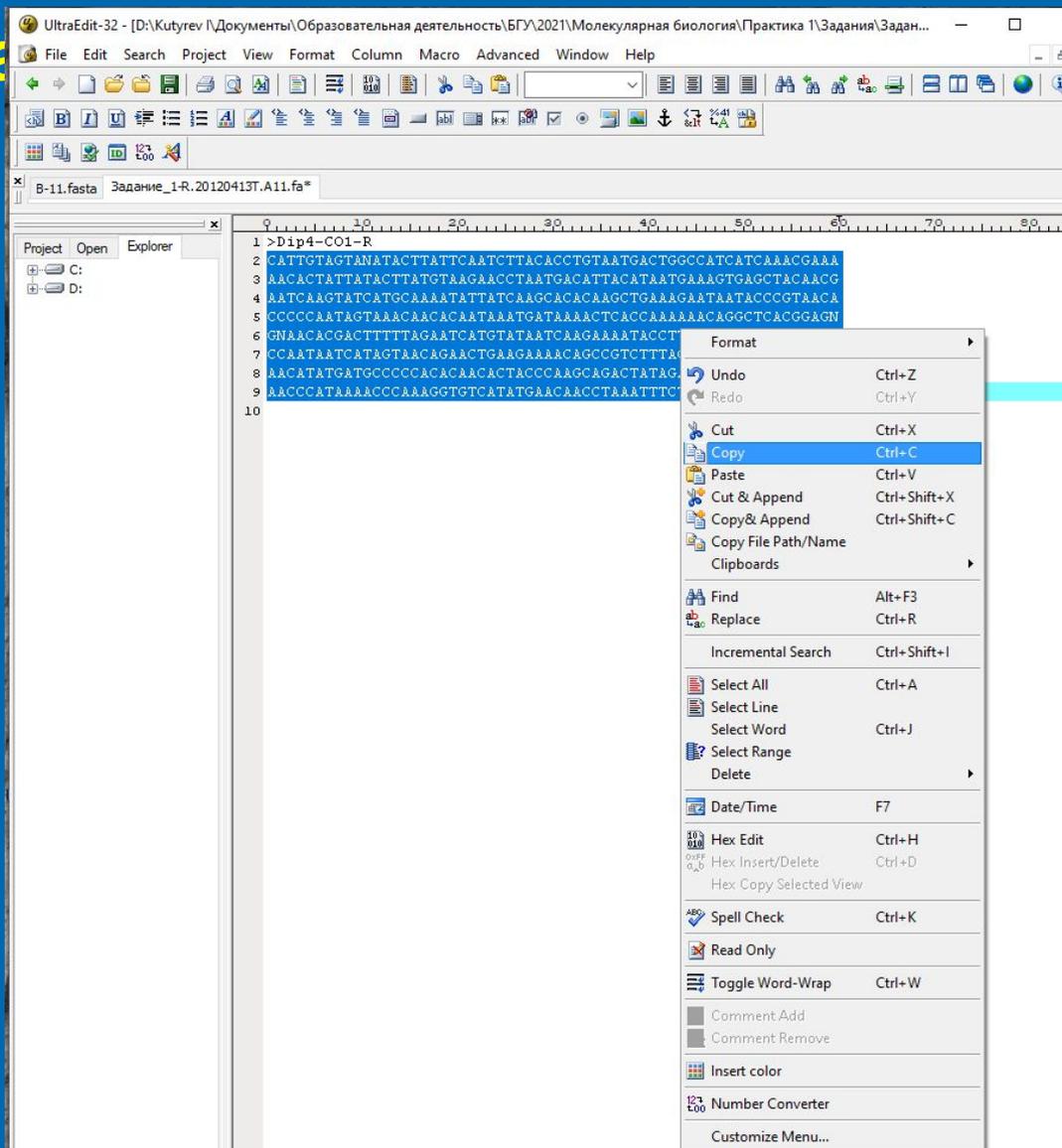
# Задача 1

## 6. Открыть сохраненный файл в программе

## UltraEdit-32

## 7. Копировать

## ПЕПТИДОВ:



# Задача 1

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

**blastn**



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

# Задача 1

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

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE\_TYPE=BlastSearch&LINK\_LOC=blasthome

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) Reset page

### Enter Query Sequence

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

```
AACATATGATGCCCCACACAACTACCCAAGCAGACTATAGAAAACATA
GCAAATAAC
AACCCATAAAACCCAAAGGTGTCATATGAACAACCTAAATTTCTACAACAT
GGCTAATC
```

Query subrange [?](#)

From

To

Or, upload file  Не выбран ... один файл [?](#)

Job Title

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

Align two or more sequences [?](#)

### Choose Search Set

Database  Standard databases (nr etc.):  rRNA/ITS databases  Genomic + transcript databases  Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism Optional   exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Limit to Optional  Sequences from type material

Entrez Query Optional  [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

### Program Selection

Optimize for  Highly similar sequences (megablast)

More dissimilar sequences (discontinuous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

**BLAST** Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in a new window

# Задача 1

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

NIH 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	<b>Nucleotide Sequence</b>
RID	<a href="#">1T3P6FCC013</a> <small>Search expires on 03-01 15:58 pm</small> <a href="#">Download All</a> <span>▼</span>
Program	BLASTN <a href="#">?</a> <a href="#">Citation</a> <span>▼</span>
Database	nt <a href="#">See details</a> <span>▼</span>
Query ID	lcl Query_64585
Description	None
Molecule type	dna
Query Length	505
Other reports	<a href="#">Distance tree of results</a> <a href="#">MSA viewer</a> <a href="#">?</a>

**Filter Results**

Organism only top 20 will appear  exclude

[+ 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  [?](#)

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
<input checked="" type="checkbox"/>	Diphylobothrium ditremum isolate RK3 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1 ...	Diphylobothrium ...	933	933	100%	0.0	100.00%	1207	DQ768179.1
<input checked="" type="checkbox"/>	Diphylobothrium dendriticum isolate RK1 18S ribosomal RNA gene, partial sequence; and internal transcribed sp ...	Dibothriocephalu...	933	933	100%	0.0	100.00%	721	DQ768177.1
<input checked="" type="checkbox"/>	Diphylobothrium dendriticum isolate ADL1094 18S ribosomal RNA gene, partial sequence; internal transcribed sp ...	Dibothriocephalu...	928	928	100%	0.0	99.80%	1196	JN153006.1
<input checked="" type="checkbox"/>	Diphylobothrium dendriticum partial ITS1, 5.8S rRNA gene and partial ITS2, country, Switzerland	Dibothriocephalu...	928	928	100%	0.0	99.80%	1323	FM204787.1
<input checked="" type="checkbox"/>	Diphylobothrium sp. DB-01 genes for 18S rRNA, ITS1, 5.8S rRNA, ITS2, 28S rRNA, partial and complete sequence	Diphylobothrium ...	909	909	100%	0.0	99.02%	1238	AB437411.3
<input checked="" type="checkbox"/>	Diphylobothrium 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
<input checked="" type="checkbox"/>	Diphylobothrium 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
<input checked="" type="checkbox"/>	Diphylobothrium latum isolate proglottids #3 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribo ...	Dibothriocephalu...	893	893	100%	0.0	98.61%	3356	KF218248.1
<input checked="" type="checkbox"/>	Diphylobothrium latum isolate proglottids #1 18S ribosomal RNA gene, internal transcribed spacer 1, 5.8S ribo ...	Dibothriocephalu...	893	893	100%	0.0	98.61%	3356	KF218246.1
<input checked="" type="checkbox"/>	Diphylobothrium latum genes for 18S rRNA, ITS1, 5.8S rRNA, ITS2, 28S rRNA, partial and complete sequence	Dibothriocephalu...	893	893	100%	0.0	98.61%	2256	AB302387.1

## 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](mailto: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/>

The screenshot shows the NCBI Gene search page. The search bar contains the text 'insulin'. A dropdown menu is open, displaying a list of search results. A red arrow points to the first result, 'Homo sapiens insulin'. The page also features a 'COVID-19' sidebar with links to 'Public health in', 'Research infor', 'SARS-CoV-2 da', 'Prevention and', and 'Español'. At the bottom, there are links for 'Using Gene', 'Gene Quick Start', 'FAQ', 'Download/FTP', 'Submit Correction', 'Statistics', 'RefSeq', and 'RefSeqGene'.

Gene

insulin

Search

Help

COVID-19

Public health in

Research infor

SARS-CoV-2 da

Prevention and

Español

Homo sapiens insulin

Homo sapiens insulin A chain

Homo sapiens insulin B chain

Aplysia californica insulin

Aplysia californica insulin precursor

Octodon degus insulin

Homo sapiens Insulin receptor subunit alpha

Homo sapiens Insulin receptor subunit beta

Homo sapiens insulin receptor

Danio rerio insulin

Danio rerio insulin A chain

Danio rerio insulin B chain

Danio rerio insulin C-peptide

Sus scrofa Insulin A chain

Sus scrofa Insulin B chain

Sus scrofa insulin

Homo sapiens insulin-like growth factor I

Homo sapiens insulin like growth factor 1

Homo sapiens Insulin-like growth factor 1 receptor alpha c...

Homo sapiens Insulin-like growth factor 1 receptor beta ch...

Using Gene

Gene Quick Start

FAQ

Download/FTP

Submit Correction

Statistics

RefSeq

RefSeqGene

# Задача 2

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

The screenshot shows the NCBI Gene database page for the **INS** gene in *Homo sapiens* (human). The page includes navigation options like 'Tabular', '20 per page', and 'Sort by Relevance'. A sidebar on the left lists 'Gene sources', 'Categories', 'Sequence content', and 'Status'. The main content area displays the gene name, its human origin, and various identifiers. A red arrow points to the 'RefSeq' link in the 'Also known as' section. Below this, there are buttons for 'Orthologs', 'Genome Data Viewer', 'BLAST', and 'Download'. The 'RefSeq Sequences' section is currently collapsed. On the right, there are sections for 'Filters', 'Results by taxon', 'Find related data', and 'Search details'. The 'Search results' section shows 1 to 20 of 2111 items, with a table listing the first result: IGF1 (insulin like growth factor 1).

Gene sources: Tabular ▾ 20 per page ▾ Sort by Relevance ▾ Send to: ▾ Hide sidebar >>

Gene sources: Genomic, Mitochondria, Organelles

Categories: Alternatively spliced, Annotated genes, Non-coding, Protein-coding, Pseudogene

Sequence content: CCDS, Ensembl, RefSeq, RefSeqGene

Status:  Current, Clear all, Show additional filters

GENE Was this helpful?

**INS – insulin**

[Homo sapiens \(human\)](#)

Also known as: IDDM, IDDM1, IDDM2, ILPR, IRDN, MODY10, PNDM4

Gene ID: 3630

[RefSeq transcripts \(4\)](#) [RefSeq proteins \(4\)](#) [RefSeq](#) (2) [PubMed \(960\)](#)

**RefSeq Sequences** +

Search results

Items: 1 to 20 of 2111

See also 70 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> <a href="#">IGF1</a> ID: 3479	insulin like growth factor 1 [ <i>Homo sapiens</i> (human)]	Chromosome 12, NC_000012.12 (102395860..102481839, complement)	IGF, IGF-I, IGFI, MGF	147440

Results by taxon

Top Organisms [\[Tree\]](#)

- Homo sapiens (2072)
- Arthroderma uncinatum (4)
- Necator americanus (2)
- Aspergillus thermomutatus (2)
- Cladophialophora carrionii CBS 160.54 (2)
- All other taxa (29)

More...

Find related data

Database:

Search details

```
((("Homo sapiens"[Organism] OR Homo sapiens[All Fields]) AND insulin[All Fields]) AND alive[prop])
```

See more...

Recent activity

Homo sapiens insulin AND (alive[prop]) (2111)

## Задача 2

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

COVID-19 Information  
[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Species: Animals (2), Customize ...  
Molecule types: genomic DNA/RNA (2), Customize ...  
Source databases: RefSeq (2), Customize ...  
Sequence Type: Nucleotide (2)  
Sequence length: Custom range...  
Release date: Custom range...  
Revision date: Custom range...  
[Clear all](#)  
[Show additional filters](#)

Summary ▾ Sort by Default order ▾  
Send to: ▾ Filters: [Manage Filters](#)

**Items: 2**

[Homo sapiens INS-IGF2 readthrough \(INS-IGF2\), RefSeqGene on chromosome 11](#)  
1. 39,098 bp linear DNA  
Accession: NG\_050578.1 GI: 1028630736  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

[Homo sapiens insulin \(INS\), RefSeqGene on chromosome 11](#)  
2. 8,416 bp linear DNA  
Accession: NG\_007114.1 GI: 161086962  
[Protein](#) [PubMed](#) [Taxonomy](#)  
[GenBank](#) [FASTA](#) [Graphics](#)

Analyze these sequences  
Run BLAST

Find related data  
Database:   
[Find items](#)

Recent activity  
[Turn Off](#) [Clear](#)

- RefSeqGene Links for Gene (Select 3630) (2) Nucleotide
- Homo sapiens insulin AND (alive[prop]) (2111) Gene
- Homo sapiens insulin A chain AND (alive[prop]) (0) Gene
- GRCh38.p13 - hg38 - Genome - Assembly - NCBI Assembly
- INS insulin [Homo sapiens] Gene

See more...



# Задача 2

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

FASTA

Send to:

Change region shown

Customize view

## Homo sapiens insulin (INS), RefSeqGene on chromosome 11

NCBI Reference Sequence: NG\_007114.1

[GenBank](#) [Graphics](#)

>NG\_007114.1 Homo sapiens insulin (INS), RefSeqGene on chromosome 11  
GGCGGCCAGGGAAGGCTCTGCGCGCAGGGAAGTGTCCAGAGACCCCTGGAGGGGCTGCTGACACCCCGGTCGCCCCACCTCGAGCATGACCCAGGGCTGCCCTCCCCATCCTCATCCTCCTGCTCCACAGACA  
TTGGCCTGGCGTCCCTGGGGGCTCGGATGAGGAAATGAGAAGTGTCCACGGTGGGTGACCCCTCCCT  
TGACAGGGCTGGGGTGTGGGTTGGGGTCTGAATCCAGGCCTCACCTCTTGCCGTCCAGGCTGAGGCC  
TCTCCTCCACCACGAATGTGACCCCTCACCTGGCCTGCATCCTGGCCTGGCCTCCTGGGGGT  
GGTATCCTGGTACAGGGTGACAGGGGCTGCCGGTGGCGGCAGCTGTCTGGGCTGATGCTGCCCGG  
CTCCCGCAGCTGTACTGGTTCACGGTGGAGTTCGGGCTGTGTAAGCAGAACGGGAGGTGAAGCCATA  
TGGTGCCGGGCTGCTCCTCCTACGGGAGCTCCTGGTGAAGTCTCCTTGTGTCAGCCCCAGCAG  
AGGGCAGGGCTGGGGACGGTGCAGGAGGGGACAGGCTCCAGTGGGAGGAAACTGAGGCCTGACCT  
CCAGGACTCAGGCTGTGTTGGGAGAAGGCTGTCTCTGCCAGTCTCACCCACATATCCAGGCCT  
CCGAAGGCCCGCGGGGAGATGGGGTGACTTACCAAGGAACCCACCCAGCTCAGGCCACGGTGC  
CCAGTCCCTCGGGACCTGGGTGCAGTGGAGTCAAGTATGTCATTGGCCTCTGCCAGCACTGCCGTCT  
TGAGGAGCTGAGATTCGGGCTTCCACCTCAGGCTGGGGCTGCAAGCCCTACCAAGACAGAGTAC  
CAGTCACTACTTCGTGCTGAGAGCTTCAAGTCAAGGACAGCTCAGTGGGCTAGGCTGCTAG  
GGCAAGCCCCATGGTCCGCCAAATGGGGCAGCCAGGCCTCTTCTGGCCTTGAGCAGGGCTGGAC  
CTGTGAGCCAGGTCACAGATGAGAAAACCGACCCCTGGTTGACAGAGCCCAACACAGCAGGGACACCA  
TCCGTGAGAAGGACCCAGCTCTGGGAGGGGACAGCTACAGGACTGGGGCTGCTGGTGGCCGGT  
CAAGGCCAGTCTGGAGGCTGACAGAGCTGAGCTTGTGAGGAGCTCCTGTGGAACCTGCTCCGGCC  
CCCTGCCCTGGGATGGGGAAGTCAAGGAGGGGATAGACAGAGTCAAGGTGGGGACAGGGCGGGAGTGGG  
TCCCCAGGGCTGGGGCTTGGTGCAGTGACAGAGTGTGAGGAGGGGAGCAAGGCCCTTAGCCTC  
ATCCTCATAAAGGCTCATCATTTTCCCTCCAGCTCTTATGCACTGGGGAACCTGAGGCCAGGGGCTA  
TGTGTCCAGCGGACAGGGTGTGAATCCACCCACAGGCTTAGGATATGGTCAAGGAAGCTTCTGTGG  
AGGAGGCCCAGTGGAGGTTCAAGGAGGGATGGGGTCCCGGCAGTCTCTAGTGGAAAAGGCGCCTAGCCT  
ATCTCCCCATGAACCCCTCACCCAGCCCTGGAAGAGGCTCAGTGTCCCGCTGTGACCACTGGCTC  
AGAAAAGCCCTGGGAGCTCTGAGCCACTGTGAAGTGGAAACCGGGCCCTGGCCTCCCTCTCTGGAG  
GCTGCAGACTCTGCCCGCAGTTGACAGGGCTTGCCTCCTCCTCCAGGAGCTATGCTCAGCAT  
CCAGCGCCCTTCTCCGTGAAGTTCGACCCGTACAGCTGGCCATCGACGTGCTGGACAGCCCCAGGCC  
GTGCGGGCTCCTGGAGGGTGTCCAGGATGAGCTGGACACCTTGGCCATGGCTGAGTGCATTTGGCT  
AGGTGCAGGGCTCCTGAGGGCCCTTCCCAACCTCCCTGGTCTGCACTGTCCGGAGCTCAGGCCCT  
GGTGAAGGGCTGGTCCCGGTCGCCCCATGCCCTCCTGCTGCCAGGCTCCCAGTCCCTGACCTG  
CTTCTCAGCGCAACAGCTGTGTGTGCGCTGGTGAAGTTGTGCTGCCGTGGTGAAGTCTGTCTGCTG  
CCCAGGTCCTGGGGCTGCTGCACTGCCCTCCGCCCTTCCCTGACACTGTCTGTGCCCAATCACCT  
CACAAATAAAGAAACTGTGGTCTTACACCTGCTGGCCCACTCTGTGCCACAGAGACAGACCTGGG  
ATCCTCAGACTCCCAACCCACCCAGCTCAGTCAAGGTTTGGCCCTGGCTCTTCTCCTCTGG  
GAGATGGCTGGCCGCTTGGCCAGGAGCTGGCCCTCCGGGCTGGTTCGCCGCTCAGCTGAGGCC  
CGCCAGCTCTGAGCCCAAGCAGCTCAGAGGCTCGGGCAGCTGGCCGAGCTGCCCATCTCCGTTGG  
GTGCCCTCCCAAGTGGGAGCCAGCTGACAGTGGGAGGCCCTCTCAGGCTGGCAGGGAGCAGGGT

Analyze this sequence

- Run BLAST
- Pick Primers
- Highlight Sequence Features
- Find in this Sequence

Articles about the INS gene

- Structural mechanism for tyrosine hydroxylase inhibition by dopamine and r [Nat Commun. 2022]
- High aldehyde dehydrogenase 1 activity is related to [Biochem Biophys Res Commun. 2022]
- Tyrosine hydroxylase activity is regulated through the modif [Biochem Biophys Res Commun. 2022]

See all...

Reference sequence information

- RefSeq alternative splicing
- See 4 reference mRNA sequence splice variants for the INS gene.

More about the INS gene

This gene encodes insulin, a peptide hormone that plays a vital role in the regulation of carbohydrate and lipid metabolism. After removal o...

Also Known As: IDDM, IDDM1, IDDM2, ILP...

# Задача 2

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

blastn blastp blastx tblastn tblastx

BLASTn programs search nucleotide databases using a nucleotide query. [more...](#)

### Enter Query Sequence

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

Query subrange [?](#)

```
CCCACCCAAAAGAGATGCAGCC
ATGGTTCGCGGTGCCCTCGGCTGCCCTGGGCCAGAGCTGGGGCTAGCT
TTCACCTTGTGAGACCCAGGA
CTCTGTCCCCAAGCC
```

From

To

Or, upload file

Не выбран ни один файл [?](#)

Job Title

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

Align two or more sequences [?](#)

### Choose Search Set

Database  Standard databases (nr etc.);  rRNA/ITS databases  Genomic + transcript databases  Betacoronavirus

Nucleotide collection (nr/nt) [?](#)

Organism

Optional

Enter organism name or id—completions will be suggested  exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude

Optional

Models (XM/XP)  Uncultured/environmental sample sequences

Limit to

Optional

Sequences from type material

Entrez Query

Optional

[?](#)

Enter an Entrez query to limit search [?](#)

### Program Selection

Optimize for

- Highly similar sequences (megablast)  
 More dissimilar sequences (discontiguous megablast)  
 Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

**BLAST**

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

Show results in new window

+ Algorithm parameters

# 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](#)

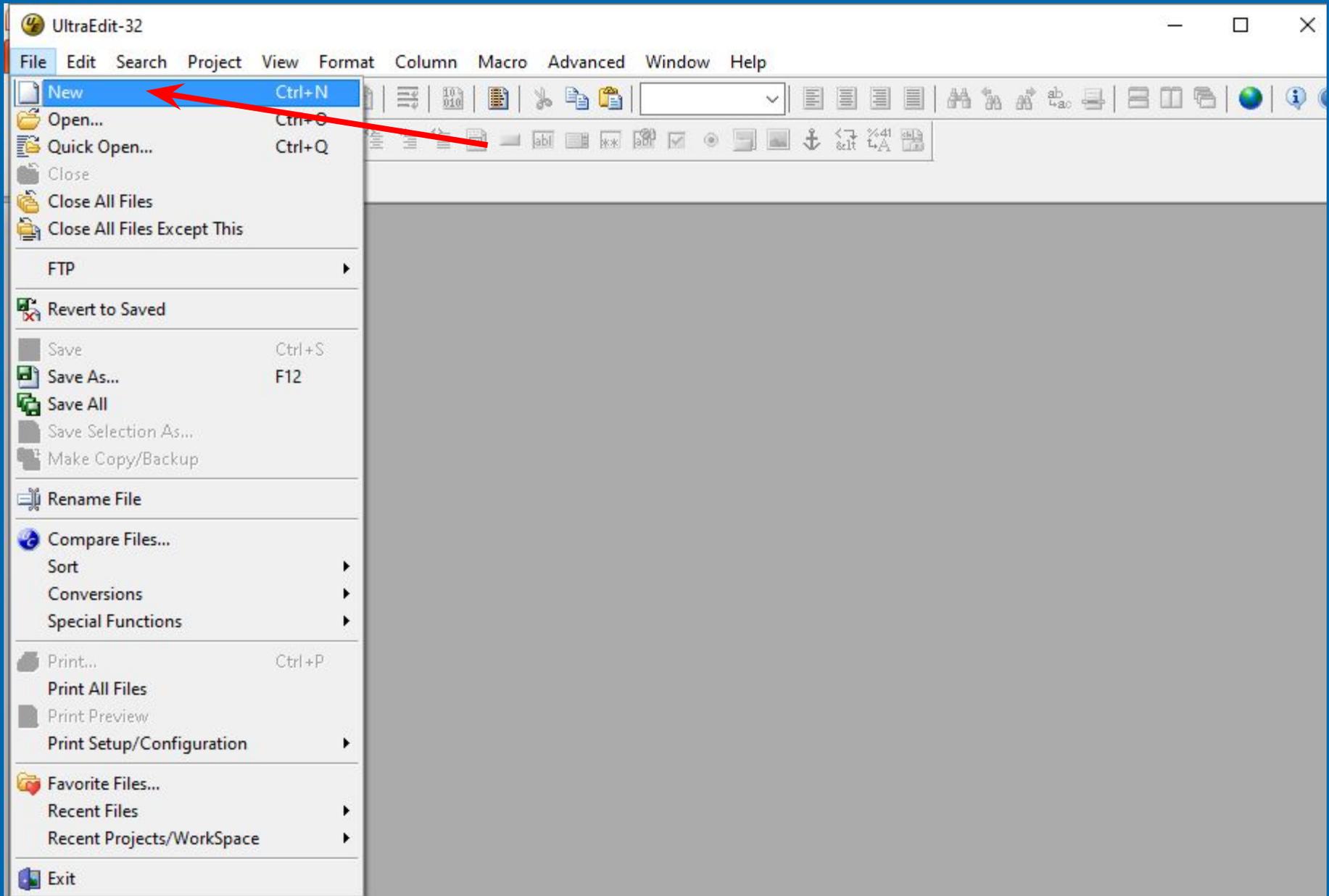
**Sequences producing significant alignments** [Download](#) [New Select columns](#) [Show](#) 100 [?](#)

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
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11</a>	Homo sapie...	15542	15542	100%	0.0	100.00%	39098	NG_050578.1
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens insulin (INS), RefSeqGene on chromosome 11</a>	Homo sapie...	15542	15542	100%	0.0	100.00%	8416	NG_007114.1
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens chromosome 11, clone RP11-889I17, complete sequence</a>	Homo sapie...	15542	15542	100%	0.0	100.00%	170027	AC132217.15
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens chromosome 11, clone RP4-539G11, complete sequence</a>	Homo sapie...	8381	15933	100%	0.0	99.37%	171366	AC130303.8
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tyrosine hydroxylase (TH), RefSeqGene on chromosome 11</a>	Homo sapie...	7878	7878	50%	0.0	100.00%	14877	NG_008128.1
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens tyrosine hydroxylase (TH), gene, 3' end, insulin (INS), gene, complet...</a>	Homo sapie...	7454	14271	90%	0.0	98.74%	12565	L15440.1
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens haplotype ICA tyrosine hydroxylase (TH), gene, partial sequence; in...</a>	Homo sapie...	6865	12701	82%	0.0	99.60%	7496	AH012037.2
<input checked="" type="checkbox"/>	<a href="#">Gorilla gorilla tyrosine hydroxylase (TH), gene, partial cds; tyrosine hydroxylase/ins...</a>	Gorilla, gorilla	6357	11535	80%	0.0	97.44%	7360	AH011815.2
<input checked="" type="checkbox"/>	<a href="#">Pongo abelii BAC clone CH276-476G11 from chromosome unknown, complete se...</a>	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

The screenshot shows a BLAST search results page. The top section includes a 'Filter Results' panel with options for 'Organism', 'Percent Identity', 'E value', and 'Query Coverage'. Below this is a table of sequences producing significant alignments. A red arrow points to the first row of the table, which is highlighted in grey.

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

# 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

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Search Advanced Help

GenBank Send to: Change region shown Customize view Analyze this sequence Run BLAST Pick Primers Highlight Sequence Features Find in this Sequence Related information Protein PubMed Taxonomy Components (Core) Gene HomoloGene PubMed (Weighted) RNA LinkOut to external resources reagents [ExactAntigen/Labome] reagents [ExactAntigen/Labome]

## Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11

NCBI Reference Sequence: NG\_050578.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS NG\_050578 39098 bp DNA linear PRI 28-FEB-2022

DEFINITION Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11.

ACCESSION NG\_050578

VERSION NG\_050578.1

KEYWORDS RefSeq; RefSeqGene.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 39098)

AUTHORS Monk D, Sanches R, Arnaud P, Apostolidou S, Hills FA, Abu-Amero S, Murrell A, Friess H, Reik W, Stanier P, Constanica M and Moore GE.

TITLE Imprinting of IGF2 P0 transcript and novel alternatively spliced INS-IGF2 isoforms show differences between mouse and human

JOURNAL Hum Mol Genet 15 (8), 1259-1269 (2006)

PUBMED [16531418](#)

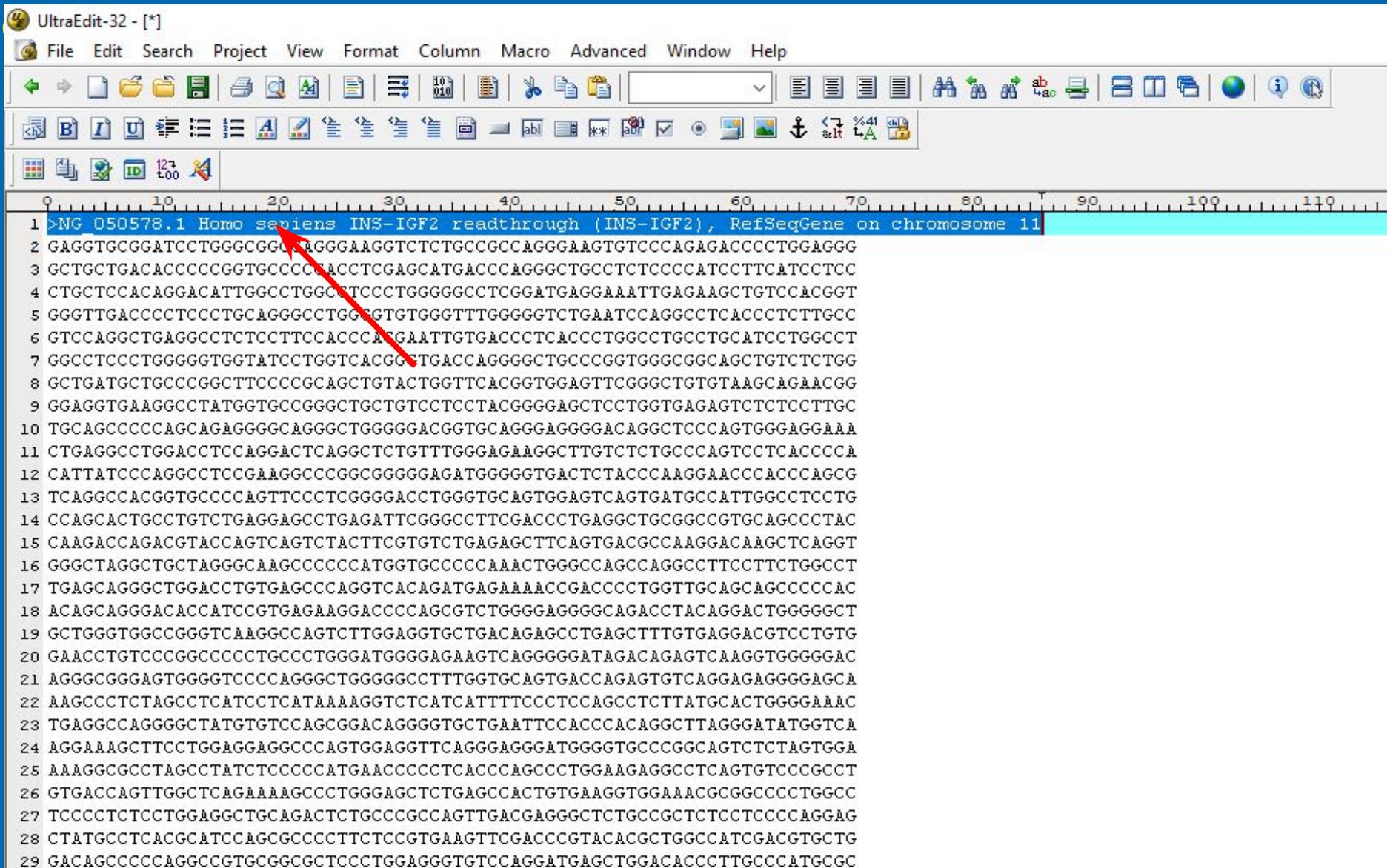
COMMENT REVIEWED [REFSEQ](#): This record has been curated by NCBI staff. The reference sequence was derived from [AC132217.15](#). This sequence is a reference standard in the [RefSeqGene](#) project.

Summary: This locus includes two alternatively spliced read-through transcript variants which align to the INS gene in the 5' region and to the IGF2 gene in the 3' region. One transcript is predicted to encode a protein which shares the N-terminus with the INS protein but has a distinct and longer C-terminus, whereas the other transcript is a candidate for nonsense-mediated decay (NMD). The transcripts are imprinted and are paternally expressed in the limb



# Задача 2

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



UltraEdit-32 - [\*]

File Edit Search Project View Format Column Macro Advanced Window Help

1 >NG\_050578.1 Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11  
2 GAGGTGCGGATCCTGGGCGGCAAGGGAAGGTCTCTGCCGCCAGGGAAGTGTCCAGAGACCCCTGGAGGG  
3 GCTGCTGACACCCCGGTGCCCCACCTCGAGCATGACCCAGGGCTGCCCTCCTCCCATCCTTCATCCTCC  
4 CTGCTCCACAGGACATTGGCCTGGCTCCCTGGGGCCCTCGGATGAGGAAAATTGAGAAAGCTGTCCACGGT  
5 GGGTTGACCCCTCCCTGCAGGGCCTGGAGTGTGGGTTTGGGGTCTGAAATCCAGGCCTCACCCCTTTGCC  
6 GTCCAGGCTGAGGCCTCTCCTTCCACCCATGAATTGTGACCCCTACCCTGGCCTGCCTGCATCCTGGCCT  
7 GGCCTCCCTGGGGTGGTATCCTGGTCACGGTGTGACCAGGGGCTGCCCGGTGGGCGGCAGCTGTCTCTGG  
8 GCTGATGCTGCCCGCTTCCCCGACGTGTACTGGTTACGGTGGAGTTCGGGCTGTGTAAGCAGAACGG  
9 GGAGGTGAAGGCCTATGGTGCCGGGCTGCTGTCTCTACGGGGAGCTCCTGGTGAGAGTCTCTCCTTGC  
10 TGCAGCCCCCAGCAGAGGGGCAGGGCTGGGGACGGTGCAGGGAGGGACAGGCTCCCAGTGGGAGGAAA  
11 CTGAGGCCTGGACCTCCAGGACTCAGGCTCTGTGTTGGGAGAAGGCTTGTCTCTGCCAGTCTCACCCCA  
12 CATTATCCCAGGCCTCCGAAGCCCGCGGGGAGATGGGGGTGACTCTACCCAAGGAACCCACCCAGCG  
13 TCAGGCCACGGTGCCTTCCCTCCGCGGACCTGGGTGCAGTGGAGTCACTGATGCCATTGGCCTCCTG  
14 CCAGCACTGCCTGTCTGAGGAGCCTGAGATTGGGCCTTCGACCCTGAGGCTGCGGCCGTGCAGCCCTAC  
15 CAAGACCAGACGTACCAGTCACTTCTGCTGTCTGAGAGCTTCAGTGACGCCAAGGACAAGCTCAGGT  
16 GGGCTAGGCTGCTAGGGCAAGCCCCCATGGTGGCCCCAAAAGTGGGCCAGCCAGGCCTTCTTCTGGCCT  
17 TGAGCAGGGCTGGACCTGTGAGCCCAGGTCACAGATGAGAAAAACCGACCCCTGGTTGCAGCAGCCCCAC  
18 ACAGCAGGGACACCATCCGTGAGAAGGACCCACGCTCTGGGGAGGGGCAGACCTACAGGACTGGGGCT  
19 GCTGGGTGGCCGGGTCAGGCCAGTCTTGGAGGTGCTGACAGAGCCTGAGCTTTGTGAGGACGTCCTGTG  
20 GAACTGTCCCGGCCCTGCCCTGGGATGGGGAGAAGTCAGGGGGATAGACAGAGTCAAGGTGGGGGAC  
21 AGGGCGGGAGTGGGGTCCCCAGGGCTGGGGCCTTTGGTGCAGTGACCAGAGTGTCAAGGAGGGGAGCA  
22 AAGCCCTCTAGCCTCATCCTCATAAAAAGGTCTCATCATTTTTCCCTCCAGCCTCTTATGCACTGGGGAAC  
23 TGAGGCCAGGGCTATGTGTCCAGCGGACAGGGGTGCTGAATTCCACCCACAGGCTTAGGGATATGGTCA  
24 AGGAAAAGCTTCTGGAGGAGGCCAGTGGAGGTTGAGGGAGGGATGGGGTGGCCGCACTCTAGTGGAA  
25 AAAGGCGCCTAGCCTATCTCCCCATGAACCCCTCACCCAGCCCTGGAAGAGGCCTCAGTGTCCCGCCT  
26 GTGACCAGTTGGCTCAGAAAAGCCCTGGGAGCTCTGAGCCACTGTGAAAGGTGGAACCGGGCCCTGGCC  
27 TCCCTCTCCTGGAGGCTGCAGACTCTGCCCGCCAGTTGACGAGGGCTCTGCCGCTCTCCTCCCCAGGAG  
28 CTATGCCTCACGCATCCAGCGCCCTTCTCCGTGAAGTTCGACCCGTACACGCTGGCCATCGACGTGCTG  
29 GACAGCCCCCAGGCCGTGCGGCCTCCCTGGAGGGTGTCCAGGATGAGCTGGACACCCTTGCCCATGCGC

# Задача 2

## 7. Сохранить файл под названием Insulin

The screenshot shows the UltraEdit-32 interface with a DNA sequence loaded in the main window. The sequence is as follows:

```
1 >NG_050578.1 Homo sapiens INS-IGF2 readthrough (INS-IGF2), RefSeqGene on chromosome 11
2 GAGGTGCGGATCCTGGGCGGCCAGGGAAGGTCTCTGCCGCCAGGGAAGTGTCCAGAGACCCCTGGAGGG
3 GCTGTGACACCCCCGGTGCCCCACCTCGAGCATGACCAGGGCTGCCTCTCCCCATCCTTCATCCTCC
4 CTGCTCCACAGGACATTGGCCTGGCGTCCCTGGGGCCCTCGGATGAGGAAATTGAGAAGCTGTCCACGGT
5 GGGTTGACCCCTCCCTGCAGGGCCTGGGGTGTGGGTTTGGGGGTCTGAATCCAGGCCTCACCCCTCTTGCC
```

A 'Save As' dialog box is open, showing the current directory as 'Задания'. The file name is 'Insulin' and the file type is 'Batch Files (\*.BAT)'. A red arrow points to the 'Insulin' text in the file name field.

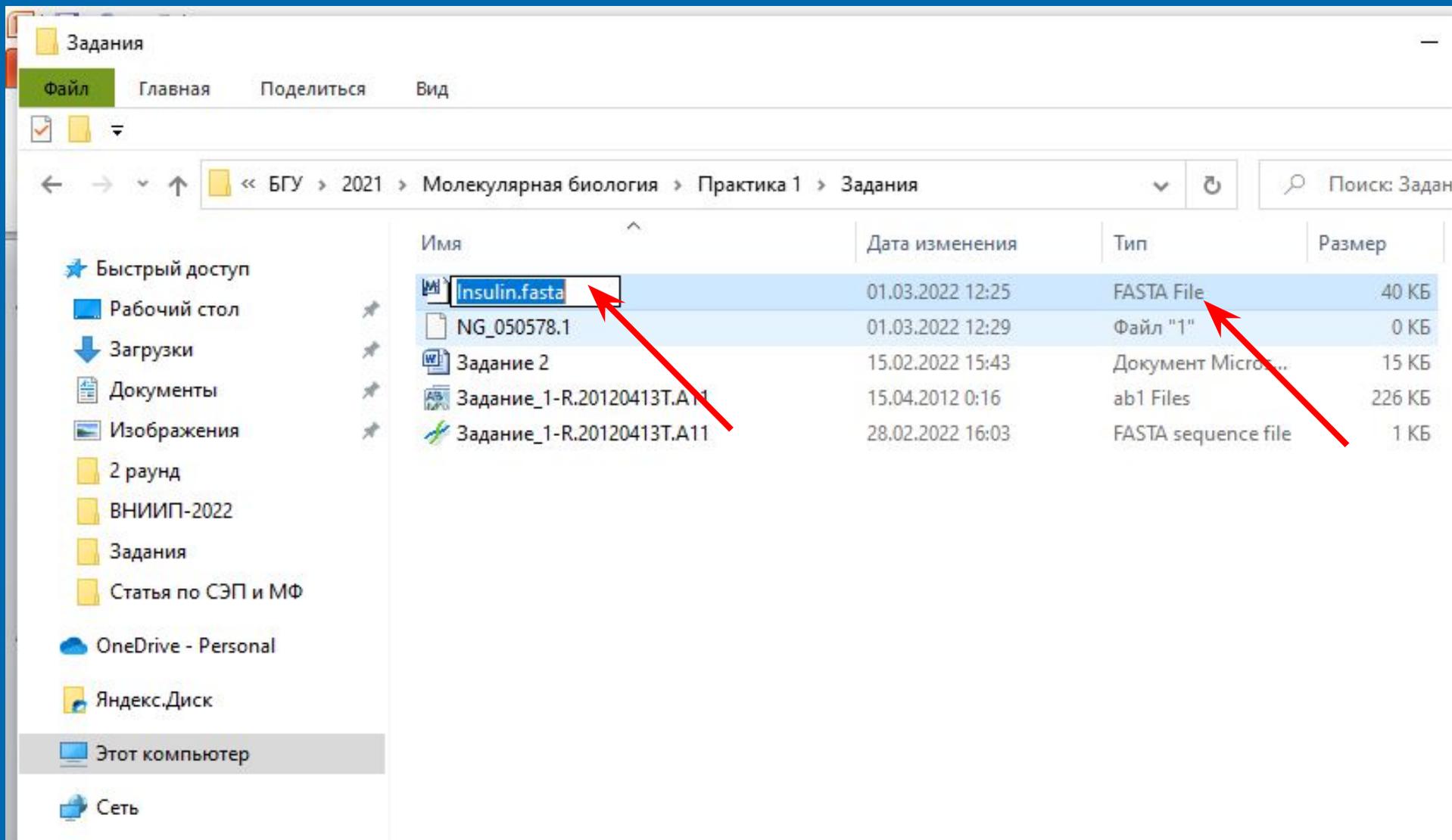
Имя	Дата изменения	Тип
Insulin	01.03.2022 12:25	Паке

Имя файла: Insulin  
Тип файла: Batch Files (\*.BAT)  
Line Terminator: Default  
Format: Default

Leave as "Default" for normal use, or change if conversion on save required.

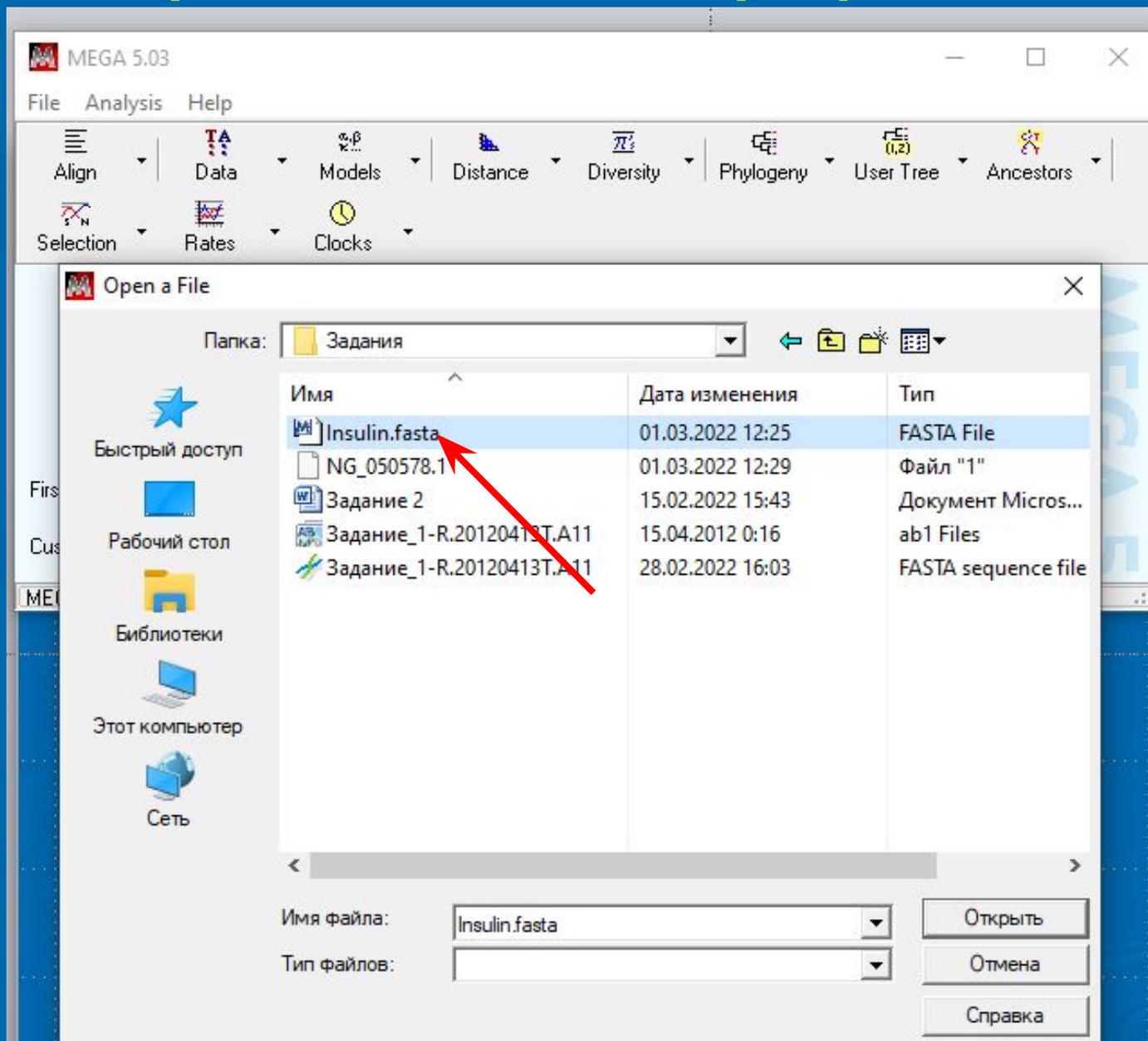
## Задача 2

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



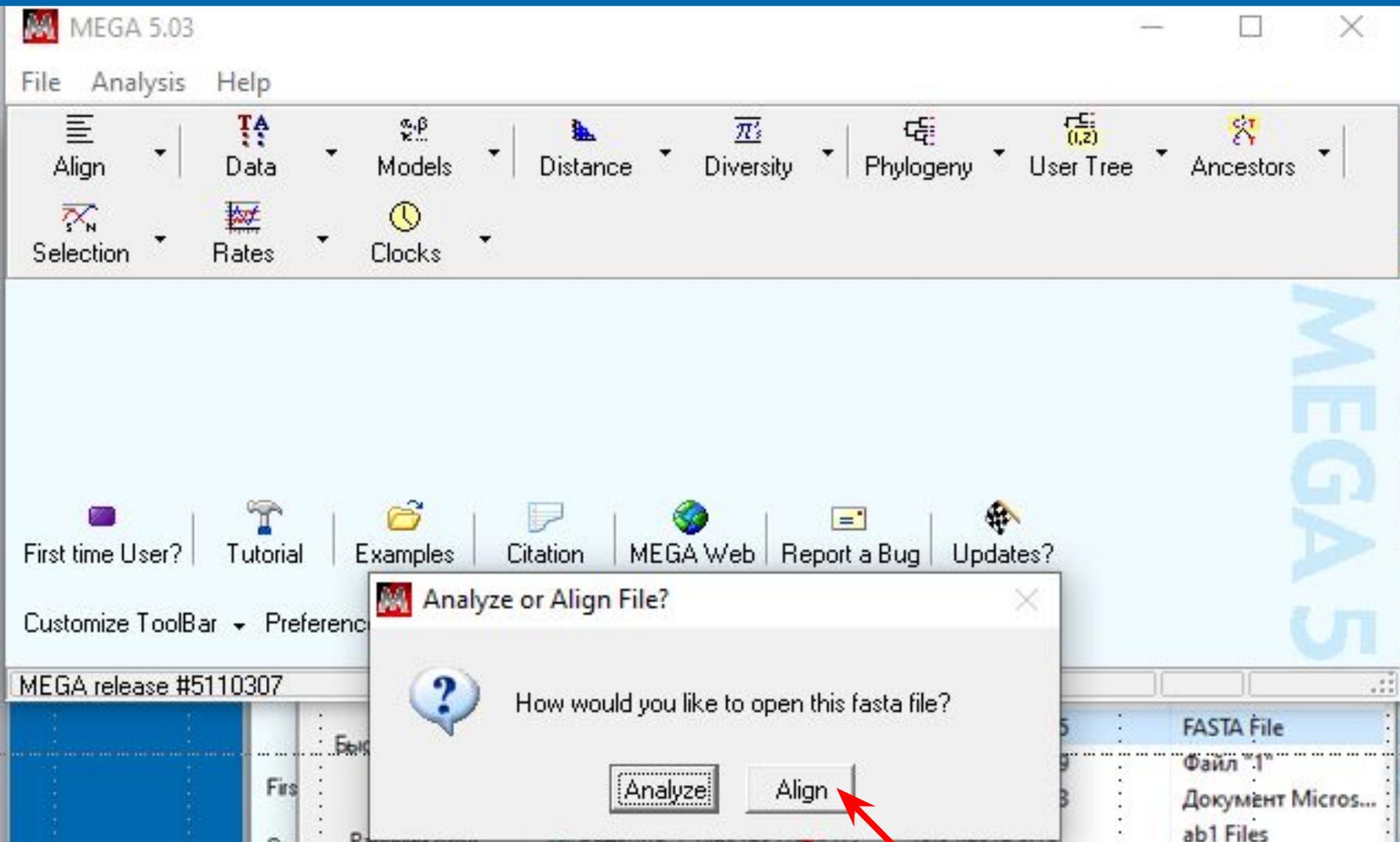
# Задача 2

## 8. Открыть файл с помощью программы MEGA5



## Задача 2

# 8. Выровнять последовательности в программе



## Задача 2

# 8. Выровнять последовательности в программе MEGA5

The screenshot shows the MEGA5 Alignment Explorer window for a file named 'Insulin.fasta.fasta'. The 'Alignment' menu is open, and the 'Align by ClustalW' option is highlighted with a red arrow. The interface includes a menu bar (Data, Edit, Search, Alignment, Web, Sequencer, Display, Help), a toolbar with various icons, and a main display area showing a sequence alignment grid. The grid contains several rows of DNA sequences, with the first row being 'h (INS-IGF2) Re' and the second row being 'lase (TH) gene p'. The sequences are aligned column by column, with gaps represented by dashes.

Menu items in the 'Alignment' menu:

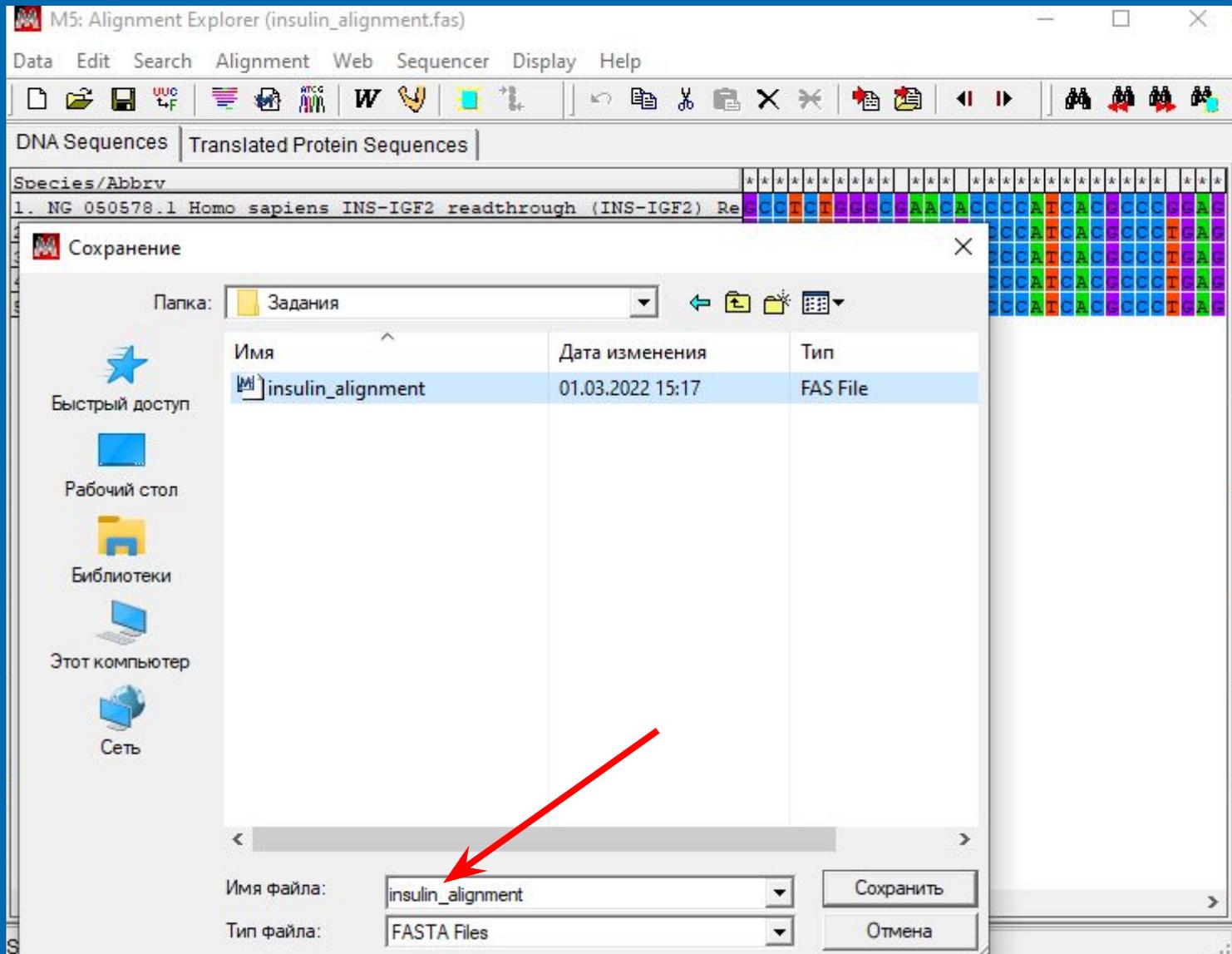
- Align by ClustalW
- Align by ClustalW (Codons)
- Align By Muscle
- Align by Muscle (Codons)
- Mark/Unmark Site (Ctrl+M)
- Align Marked Sites (Ctrl+L)
- Unmark All Sites
- Delete Gap-Only Sites
- Auto-Fill Gaps (checked)





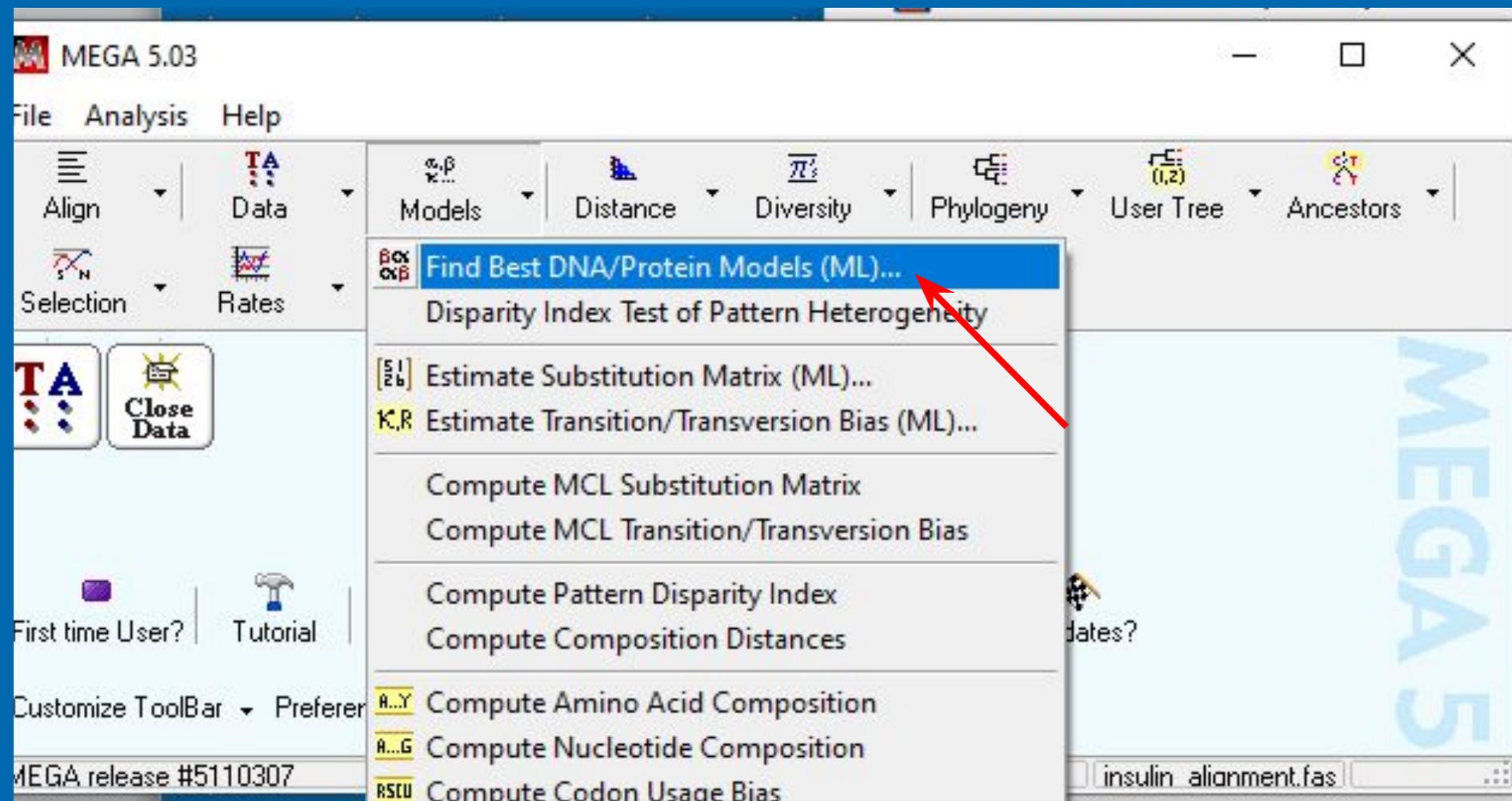
## Задача 2

# 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

The screenshot shows the MEGA 5.03 software interface. The main window displays the 'M5: Analysis Preferences' dialog box. The 'Options Summary' tab is active, showing a table of analysis options. The 'Analysis' section is selected, and the 'Model Selection [ML]' option is highlighted. The 'Compute' button is indicated by a red arrow.

Option	Selection
<b>Analysis</b>	Model Selection [ML]
Tree to Use	Automatic (Neighbor-joining tree)
<i>User Tree File</i>	<i>Not Applicable</i>
Statistical Method	Maximum Likelihood
<b>Substitution Model</b>	
Substitutions Type	Nucleotide
Genetic Code Table	<i>Not Applicable</i>
<b>Data Subset to Use</b>	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	<i>Not Applicable</i>
Select Codon Positions	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites

Buttons:  Compute  Cancel  Help

# Задача 2

## 9. Найти наилучшую модель построения филогенетического дерева для полученного

выравнившего

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(CC)
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 frequencies.  $r$  should be considered when evaluating them. For simplicity, sum of  $r$  is automatically computed. The analysis involved 5 nucleotide sequences. Codons were eliminated. There were a total of 947 positions in the final dataset.

no-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter

University Press, New York.

Genetic 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 read the terms of use (www.mega-bioinformatics.org). In no event shall the authors and their employers be liable for any damages, including but not limited to, or implied, including but not limited to the determination of suitability of this caption text for a specific purpose.

# Задача 2

## 11. Сохраняем таблицу в формате Excel с именем insulin\_best model

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].

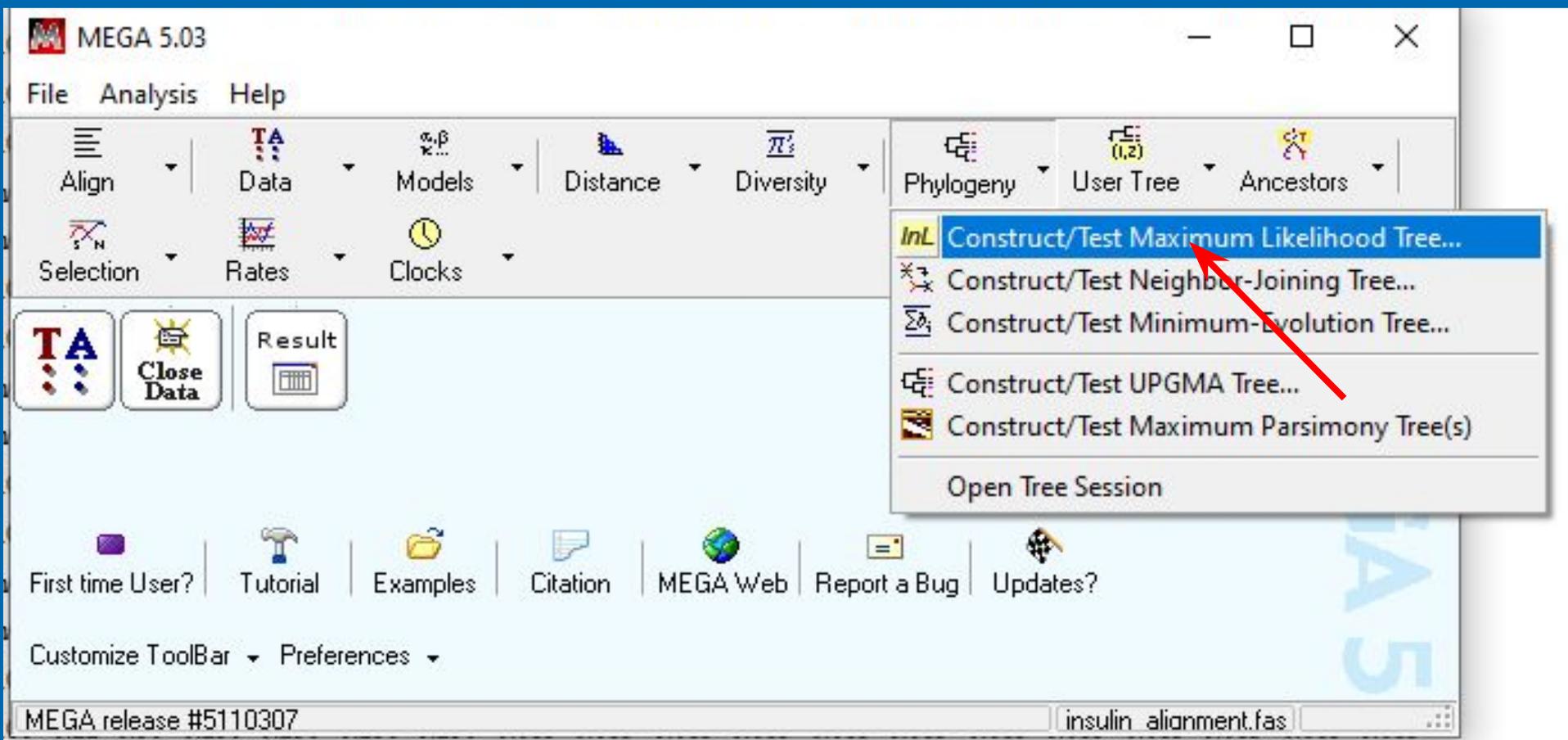
Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

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.

Disclaimer: Although utmost care has been taken to ensure the correctness of the caption, the caption text is provided "as is" without any warranty of any kind. Authors advise the user to carefully check the caption prior to its use for any purpose and report any errors or problems to the authors immediately ([www.megasoftware.net](http://www.megasoftware.net)). In no event shall the authors and their employers be liable for any damages, including but not limited to special, consequential, or other damages. Authors specifically disclaim all other warranties expressed or implied, including but not limited to the determination of suitability of this caption text for a specific purpose, use, or application.

## Задача 2

# 12. Построить филогенетическое дерево для выбранных нуклеотидных последовательностей



# Задача 2

## 12. Указываем нужную (наилучшую) модель

M5: Analysis Preferences

Options Summary

Option	Selection
<b>Analysis</b>	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
<b>Phylogeny Test</b>	
Test of Phylogeny	Bootstrap method
<i>No. of Bootstrap Replications</i>	500
<b>Substitution Model</b>	
Substitutions Type	Nucleotide
Genetic Code Table	<i>Not Applicable</i>
Model/Method	Tamura 3-parameter model
<b>Rates and Patterns</b>	
Rates among Sites	Jukes-Cantor model Kimura 2-parameter model Tamura 3-parameter model
<i>No. of Discrete Gamma Categories</i>	Hasegawa-Kishino-Yano model Tamura-Nei model General Time Reversible model
<b>Data Subset to Use</b>	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	<i>Not Applicable</i>
Select Codon Positions	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
<b>Tree Inference Options</b>	
ML Heuristic Method	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML	Make initial tree automatically
<i>Initial Tree File</i>	<i>Not Applicable</i>

Compute Cancel Help



# Задача 2

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

M5: Analysis Preferences

Options Summary

Option	Selection
<b>Analysis</b>	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
<b>Phylogeny Test</b>	
Test of Phylogeny	Bootstrap method
<i>No. of Bootstrap Replications</i>	500
<b>Substitution Model</b>	
Substitutions Type	Nucleotide
Genetic Code Table	Not Applicable
Model/Method	Tamura 3-parameter model
<b>Rates and Patterns</b>	
Rates among Sites	Uniform rates
<i>No. of Discrete Gamma Categories</i>	Uniform rates Gamma Distributed (G) Has Invariant sites (I) Gamma distributed with Invariant sites (G+I)
<b>Data Subset to Use</b>	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	Not Applicable
Select Codon Positions	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
<b>Tree Inference Options</b>	
ML Heuristic Method	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML	Make initial tree automatically
<i>Initial Tree File</i>	Not Applicable

Compute Cancel Help



# Задача 2

## 12. Строим филогенетическое дерево для выбранных нуклеотидных последовательностей

M5: Analysis Preferences

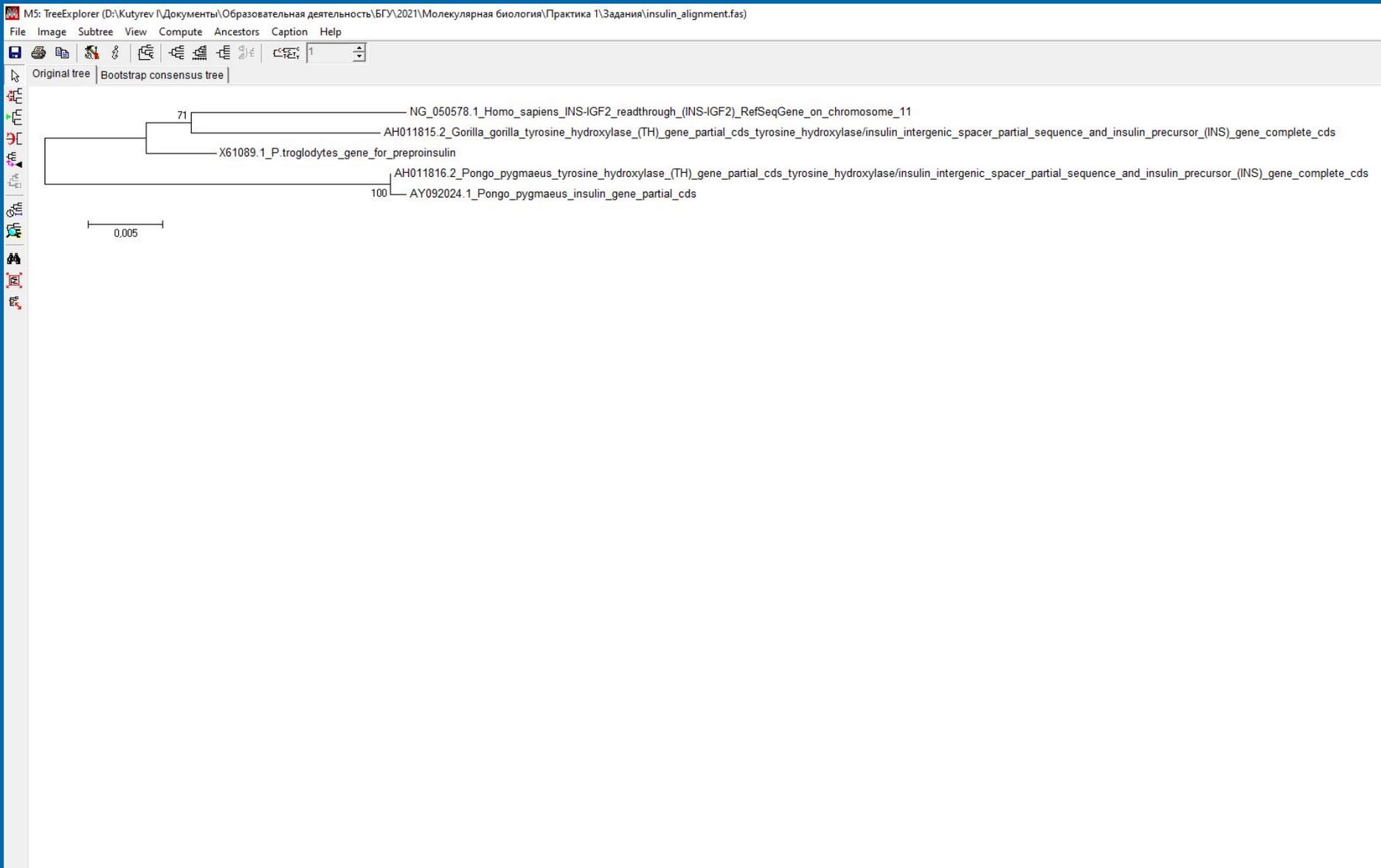
Options Summary

Option	Selection
<b>Analysis</b>	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
<b>Phylogeny Test</b>	
Test of Phylogeny	Bootstrap method
<i>No. of Bootstrap Replications</i>	500
<b>Substitution Model</b>	
Substitutions Type	Nucleotide
Genetic Code Table	<i>Not Applicable</i>
Model/Method	Tamura 3-parameter model
<b>Rates and Patterns</b>	
Rates among Sites	Uniform rates
<i>No. of Discrete Gamma Categories</i>	<i>Not Applicable</i>
<b>Data Subset to Use</b>	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	<i>Not Applicable</i>
Select Codon Positions	<input checked="" type="checkbox"/> 1st <input checked="" type="checkbox"/> 2nd <input checked="" type="checkbox"/> 3rd <input checked="" type="checkbox"/> Noncoding Sites
<b>Tree Inference Options</b>	
ML Heuristic Method	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML	Make initial tree automatically
<i>Initial Tree File</i>	<i>Not Applicable</i>

Complete  Cancel  Help

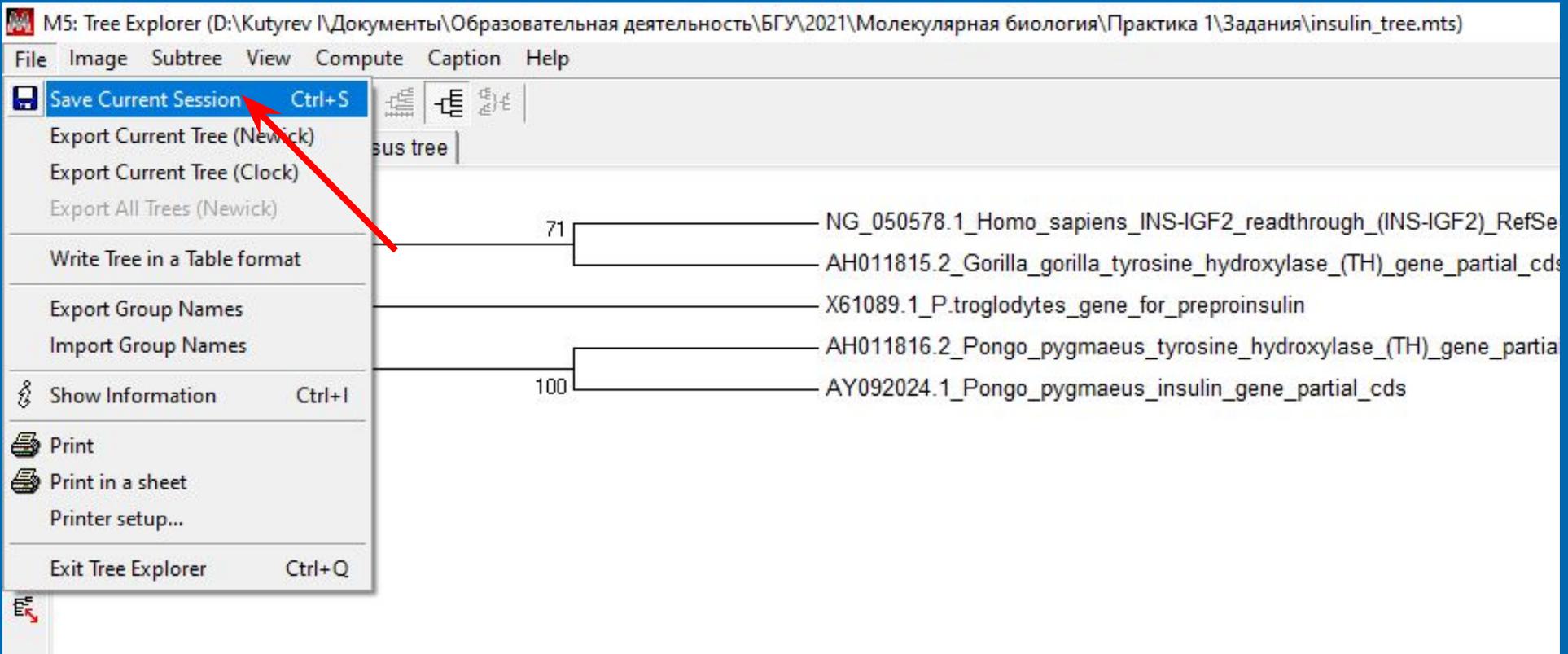
# Задача 2

## 12. Строим филогенетическое дерево для выбранных нуклеотидных последовательностей



## Задача 2

# 13. Сохраняем текущую сессию под названием insulin\_tree



The screenshot shows the Tree Explorer application window. The title bar reads "M5: Tree Explorer (D:\Kutyrev I\Документы\Образовательная деятельность\БГУ\2021\Молекулярная биология\Практика 1\Задания\insulin\_tree.mts)". The menu bar includes "File", "Image", "Subtree", "View", "Compute", "Caption", and "Help". The "File" menu is open, and the "Save Current Session" option is highlighted in blue, with a red arrow pointing to it. The menu items are: "Save Current Session Ctrl+S", "Export Current Tree (Newick)", "Export Current Tree (Clock)", "Export All Trees (Newick)", "Write Tree in a Table format", "Export Group Names", "Import Group Names", "Show Information Ctrl+I", "Print", "Print in a sheet", "Printer setup...", and "Exit Tree Explorer Ctrl+Q".

The background shows a phylogenetic tree with several branches. The tree is rooted at a node labeled "71". The branches lead to the following sequences: "NG\_050578.1\_Homo\_sapiens\_INS-IGF2\_readthrough\_(INS-IGF2)\_RefSe", "AH011815.2\_Gorilla\_gorilla\_tyrosine\_hydroxylase\_(TH)\_gene\_partial cds", "X61089.1\_P.troglodytes\_gene\_for\_preproinsulin", "AH011816.2\_Pongo\_pygmaeus\_tyrosine\_hydroxylase\_(TH)\_gene\_partia", and "AY092024.1\_Pongo\_pygmaeus\_insulin\_gene\_partial cds". A node labeled "100" is also visible on the tree.

## Задача 2. Форма отчета

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