

Биоинформатика: Домашнее задание 2

Выполнил: Козолий Михаил

Группа: 22214

Взять человеческий ген (Gene symbol)

Выбранный фенотип

Gene Symbol: FOXP2

Полное название: forkhead box P2

Ссылка: [OMIM #605317](#) 

Описание:

Роль: транскрипционный фактор, критически важен для нейроразвития и речевой функции.

Описание с OMIM:

Ген **FOXP2** кодирует **forkhead box P2** — предполагаемый транскрипционный фактор, содержащий **полиглутаминовый тракт** и **домен связывания с ДНК типа forkhead** (Lai и др., 2001).

Получение последовательности FOXP2 в формате FASTA

Была получена нуклеотидная последовательность человеческого гена **FOXP2** (транскрипт **NM_014491.4**) из базы данных **NCBI Nucleotide**.

Последовательность сохранена в формате **FASTA** в виде текстового файла с именем:

FOXP2.fasta

С помощью NCBI BLAST найти не менее 10 (десяти) гомологичных генов в других видах.

Использование протеиновой версии

Для поиска гомологичных генов был использован протеиновый BLAST (protein BLAST). В качестве исходного белковой последовательности взята белковая версия транскрипта **NM_014491.4**, а именно — **NP_055306.1**.

Параметры:

Источник и файл для поиска:

NCBI: <https://www.ncbi.nlm.nih.gov/protein/17017963> 

Приложение: FOXP2_protein_isoform.fasta

Query subrange

From:

To:

Database: Non-redundant protein sequences (nr)

Organism: __

Exclude: __

Entrez Query: __

Algorithm: Quick BlastP

Проведение поиска protein BLAST

Поиск выполнен с использованием базы данных "**non-redundant protein sequences (nr)**" при сохранении остальных параметров по умолчанию.

В результате фильтруем значение "Homo sapiens", получаем результат:

Sequences producing significant alignments

Download Select columns Show 100 ?

☒ select all 89 sequences selected

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

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X4 [Aotus nancymaae]	Aotus na...	1435	1435	100%	0.0	99.72%	715	XP_064230320.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X3 [Piliocolobus tephrosc...	Piliocolo...	1432	1432	100%	0.0	99.58%	715	XP_023084003.2
<input checked="" type="checkbox"/>	forkhead box protein P2 [Pan troglodytes]	Pan trogl...	1431	1431	100%	0.0	99.58%	716	NP_001009020.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X5 [Aotus nancymaae]	Aotus na...	1428	1428	100%	0.0	99.58%	714	XP_064230332.1
<input checked="" type="checkbox"/>	forkhead box protein P2 [Macaca mulatta]	Macaca...	1428	1428	100%	0.0	99.58%	714	NP_001028193.1
<input checked="" type="checkbox"/>	forkhead box protein P2 [Gorilla gorilla gorilla]	Gorilla g...	1427	1427	100%	0.0	99.44%	713	NP_001266466.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X4 [Piliocolobus tephrosc...	Piliocolo...	1426	1426	100%	0.0	99.44%	714	XP_023084004.2
<input checked="" type="checkbox"/>	FOXP2 protein [Pongo pygmaeus]	Pongo p...	1426	1426	100%	0.0	99.44%	714	AAN60059.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform 1 [Mus musculus]	Mus mus...	1426	1426	100%	0.0	99.44%	714	NP_444472.2
<input checked="" type="checkbox"/>	FOXP2 isoform 12 [Pan troglodytes]	Pan trogl...	1426	1426	100%	0.0	99.44%	717	PNJ01009.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X10 [Callithrix jacchus]	Callithrix...	1425	1425	100%	0.0	99.30%	713	XP_035110073.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X3 [Aotus nancymaae]	Aotus na...	1424	1424	100%	0.0	97.40%	732	XP_064230316.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X7 [Pan paniscus]	Pan pani...	1424	1424	100%	0.0	99.44%	715	XP_034819287.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X9 [Cavia porcellus]	Cavia po...	1423	1423	100%	0.0	98.61%	721	XP_063098043.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X2 [Sigmodon hispidus]	Sigmodo...	1423	1423	100%	0.0	98.89%	718	KAL1787227.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X2 [Piliocolobus tephrosc...	Piliocolo...	1422	1422	100%	0.0	97.27%	732	XP_023084002.2
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X3 [Eulemur rufifrons]	Eulemur...	1422	1422	100%	0.0	99.16%	714	XP_069318623.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X2 [Nomascus leucogenys]	Nomasc...	1422	1422	100%	0.0	99.44%	713	XP_003261263.1
<input checked="" type="checkbox"/>	forkhead-related transcription factor 2 [Mus musculus]	Mus mus...	1422	1422	100%	0.0	99.16%	714	AAK69651.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X2 [Arvicanthis niloticus]	Arvicanth...	1422	1422	100%	0.0	99.02%	711	XP_034374981.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X1 [Psammomys obesus]	Psammo...	1422	1422	100%	0.0	99.16%	711	XP_055468656.1
<input checked="" type="checkbox"/>	unnamed protein product [Mus musculus]	Mus mus...	1421	1421	100%	0.0	99.16%	714	BAC38477.1
<input checked="" type="checkbox"/>	forkhead box protein P2 isoform X11 [Pongo pygmaeus]	Pongo p...	1420	1420	100%	0.0	99.30%	713	XP_054351476.1

Выбор организмов для анализа

В итоговую выборку включил три вида птиц, три наиболее схожих варианта из верхних строк, а также четыре случайных варианта из различных классов.

Номер Accession	Вид / Класс	Краткое описание
XP_066175109.1	Sylvia atricapilla (Птица)	Певчий дрозд (класс Aves)
XP_059672351.1	Gavia stellata (Птица)	Морская гагара (класс Aves)

Номер Accession	Вид / Класс	Краткое описание
XP_061873476.1	Colius striatus (Птица)	Вид птицы (класс Aves)
XP_006882717.1	Elephantulus edwardii (Насекомоядное)	Африканский слонокрот
XP_064230320.1	Aotus nancymaae	Ночной обезьяна Нэнси (род Aotus)
XP_023084003.2	Ptilocolobus tephrosceles	Красный колобус (вид мартышковых обезьян)
NP_001009020.1	Pan troglodytes	Шимпанзе обыкновенный (Common chimpanzee)
KAL2777572.1	Daubentonia madagascariensis (Лемур)	Лемур (отряд Lemuriformes)
XP_012576466.1	Condylura cristata (Насекомоядное)	Крот (отряд Soricomorpha)
AFN11569.1	Eospalax fontanierii baileyi (Грызун)	Подземный хомяк (отряд Rodentia)

Построить единое множественное выравнивание полученных последовательностей

Объединяем все последовательности в одну и воспользовавшись несколькими программами получаем выравнивания для всех 10 видов:

Clustal Omega (EMBL-EBI):

Приложение: CLUSTAL_RES.aln-clustal_num

MAFFT online

Приложение: MAFFT_RES.aln-clustalw

MUSCLE online

Проанализировать полученное выравнивание с точки зрения консервативных участков

Анализ выравнивания, полученного с помощью Muscle

1. Высококонсервативный N-конец:

- ◆ Первые ~60 аминокислот (начиная с **MMQESATETISNSSMN ...**) полностью идентичны у всех последовательностей.
- ◆ В строке консенсуса:
*******:*******
****** — почти все символы «*», что означает полное совпадение.
- ◆ Вероятно, этот регион важен для запуска трансляции или начального сворачивания белка.

2. Сильная консервативность центральной части:

- ◆ Участки с повторами **QQQ ...** сохраняются почти полностью. Повторяющиеся глутамины (Q) часто встречаются в транскрипционных факторах и могут быть важны для связывания ДНК или белок-белковых взаимодействий.
- ◆ Несмотря на присутствие отдельных вариаций (например, замен $S \leftrightarrow K$ или $R \leftrightarrow E$), общая структура сохраняется.

3. Гипервариабельный участок (~позиции 120–140):

- ◆ В некоторых последовательностях есть делеции или вставки, например, последовательность **XP_012576466.1** содержит дополнительный фрагмент **QQVMTFWDSGLENFRAALEK**.
- ◆ Это может свидетельствовать о изоформе с дополнительным доменом или вставкой, влияющей на функцию.

4. Консервативность С-конца:

- ◆ Последние 100+ аминокислот (начиная с **MEDNGIKHGGLDLTTN ...**) практически идентичны.
- ◆ Здесь видно множество полных совпадений (строка с ******* ...**), что говорит о важной функциональной доменной области.

Анализ выравнивания, полученного с помощью MAFFT

1. Начальные 60 остатков

```
MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELLHLQQQQALQ
A
*****:*****
*
```

Почти полная идентичность: вариации только у 3-х из 10 белков и только в одной позиции. Это свидетельствует о высоко консервативном сигнале-пептиде или домене, важном для начальной сборки белка.



2. Участок с мотивами Q и L

```
ARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQQLQ
A
***** .. :*****
*
```

Большая часть остатков — глутамин (Q) и лейцин (L) — характерны для регуляторных белков, участвующих во взаимодействиях с другими белками. Небольшие вариации в позициях 17–19 (KG/KS) не затрагивают общую консервативность.



3. Центральный участок с Q-мотивами

```
QQQQQQQQQQQQQ—————
QQQQQQQQQQQQQQQQQHPGKQAKEQQQQQQQQQQQQQLAA *****
*****. ** *****
```

Здесь прослеживается длинный глутамин-обогащённый участок (polyQ-мотив), характерный для белков, вовлечённых в транскрипционную регуляцию. Это типичный **функционально значимый регион**, чувствительный к длине и составу повтора.



4. Участок с PGL и HL-мотивами

```
QQLVFQQQLLQMQLQQQHLHLSLQRQGLISIPPGQAALPVQSLPQAGLSPAETQQQLWK
E *****: **
*****:*****
```

Устойчивость глутамина (Q), пролина (P), глицина (G) и лейцина (L) указывает на важность этого участка. Пролин-глициновые мотивы часто участвуют в формировании гибких петель.



5. С-концевой участок

```
VTGVHSMEDNGIKHGGLDLTTNNSSTTSSTTSKASPPITHHSIVNGQSSVLNARRDSS
S
*****
*
```

Полностью консервативен — что крайне редко для С-концевых областей, что может указывать на участие в важной клеточной функции, например, в ядерной локализации или взаимодействиях.

С помощью баз данных биологической систематики определить ближайший таксон, объединяющий виды, полученные в результате анализа.

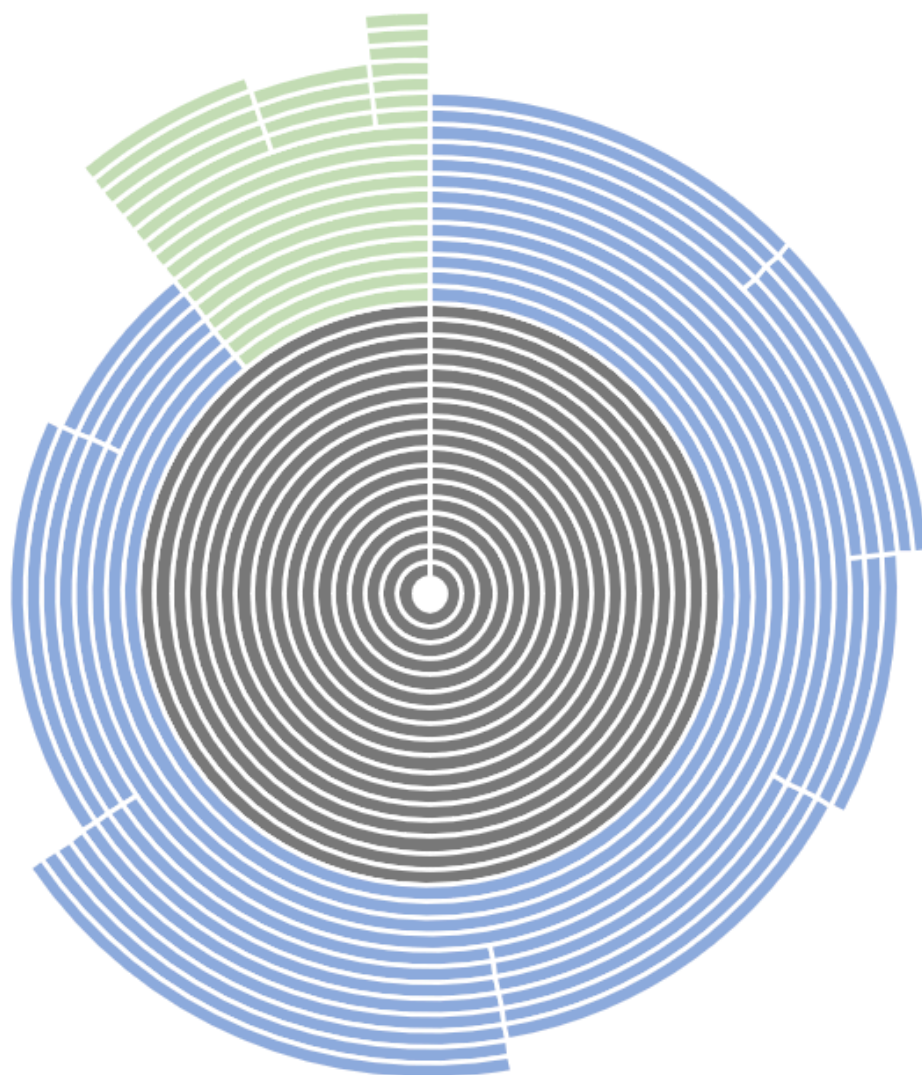
Вот таксономические линии (lineage) для каждого вида, оформленные в структурированном, читаемом виде для удобного сравнения.

В PDF документ таблица войдет не полностью

1. Sylvia atricapilla (Птица)	2. Gavia stellata (Птица)	3. Colius striatus (Птица)	4. ... (Н...
Cellular organisms	Cellular organisms	Cellular organisms	C
Eukaryota	Eukaryota	Eukaryota	E
Opisthokonta	Opisthokonta	Opisthokonta	O
Metazoa	Metazoa	Metazoa	M
Eumetazoa	Eumetazoa	Eumetazoa	E
Bilateria	Bilateria	Bilateria	B
Deuterostomia	Deuterostomia	Deuterostomia	D
Chordata	Chordata	Chordata	C
Craniata	Craniata	Craniata	C
Vertebrata	Vertebrata	Vertebrata	V
Gnathostomata	Gnathostomata	Gnathostomata	G
Teleostomi	Teleostomi	Teleostomi	Te
Euteleostomi	Euteleostomi	Euteleostomi	E
Sarcopterygii	Sarcopterygii	Sarcopterygii	S
Dipnotetrapodomorpha	Dipnotetrapodomorpha	Dipnotetrapodomorpha	D
Tetrapoda	Tetrapoda	Tetrapoda	Te
Amniota	Amniota	Amniota	A
Sauropsida	Sauropsida	Sauropsida	M

1. <i>Sylvia atricapilla</i> (Птица)	2. <i>Gavia stellata</i> (Птица)	3. <i>Colius striatus</i> (Птица)	4. е (Н
Sauria	Sauria	Sauria	TI
Archelosauria	Archelosauria	Archelosauria	E
Archosauria	Archosauria	Archosauria	A
Dinosauria	Dinosauria	Dinosauria	M
Saurischia	Saurischia	Saurischia	M
Theropoda	Theropoda	Theropoda	E
Coelurosauria	Coelurosauria	Coelurosauria	
Aves	Aves	Aves	
Neognathae	Neognathae	Neognathae	
Neoaves	Neoaves	Neoaves	
Telluraves	Aequornithes	Telluraves	
Australaves	Gaviiformes	Coraciimorphae	
Passeriformes	Gaviidae	Coliiformes	
Sylvioidea	Gavia	Coliidae	
Sylviidae		Colius	
Sylviinae			
Sylvia			

График показывающий где расходятся виды.



Ответ: ближайший общий таксон — ***Amniota* Амниоты**

Amniota — это таксон позвоночных животных, объединяющий всех животных, которые откладывают яйца с амнионом — специализированной оболочкой, которая защищает эмбрион и позволяет развиваться вне водной среды.

Проще говоря, амниотами называют группу животных, включающую:

- ◆ Рептилий (ящерицы, змеи, черепахи, крокодилы, динозавры и птицы)
- ◆ Млекопитающих

Результаты

Название выбранного гена (Gene Symbol)

Gene Symbol: FOXP2

Полное название: forkhead box P2

Ссылка: [OMIM #605317](#) 

Параметры BLAST, использованные для поиска гомологичных генов

Источник и файл для поиска:

NCBI: <https://www.ncbi.nlm.nih.gov/protein/17017963> 

Приложение: FOXP2_protein_isoform.fasta

Query subrange

From:

To:

Database: Non-redundant protein sequences (nr)

Organism: __

Exclude: __

Entrez Query: __

Algorithm: Quick BlastP

Таблицу с названиями полученных гомологичных генов (Gene Symbol), систематических названий видов и русских названий видов

Номер Accession	Вид / Класс	Краткое описание
XP_066175109.1	<i>Sylvia atricapilla</i> (Птица)	Певчий дрозд (класс Aves)
XP_059672351.1	<i>Gavia stellata</i> (Птица)	Морская гагара (класс Aves)
XP_061873476.1	<i>Colius striatus</i> (Птица)	Вид птицы (класс Aves)
XP_006882717.1	<i>Elephantulus edwardii</i> (Насекомоядное)	Африканский слонокрот

Номер Accession	Вид / Класс	Краткое описание
XP_064230320.1	Aotus nancymaae	Ночной обезьяна Нэнси (род Aotus)
XP_023084003.2	Ptilocolobus tephrosceles	Красный колобус (вид мартышковых обезьян)
NP_001009020.1	Pan troglodytes	Шимпанзе обыкновенный (Common chimpanzee)
KAL2777572.1	Daubentonia madagascariensis (Лемур)	Лемур (отряд Lemuriformes)
XP_012576466.1	Condylura cristata (Насекомоядное)	Крот (отряд Soricomorpha)
AFN11569.1	Eospalax fontanierii baileyi (Грызун)	Подземный хомяк (отряд Rodentia)

Краткий анализ консервативности полученного выравнивания в произвольной форм

Анализ выравнивания, полученного с помощью MAFFT

1. Начальные 60 остатков

```
MMQESATETISNSSMNQNGMSTLSSQLDAGSRDGRSSGDTSSSEVSTVELLHLQQQALQ
A
*****:*****
*
```

Почти полная идентичность: вариации только у 3-х из 10 белков и только в одной позиции. Это свидетельствует о высоко консервативном сигнале-пептиде или домене, важном для начальной сборки белка.



2. Участок с мотивами Q и L

```
ARQLLLQQQTSGLKSPKSSDKQRPLQVPVSVAMMTPQVITPQQMQQILQQQVLSPQQQLQ
A
***** .. :*****
*
```

Большая часть остатков — глутамин (Q) и лейцин (L) — характерны для регуляторных белков, участвующих во взаимодействиях с другими белками. Небольшие вариации в позициях 17–19 (KG/KS) не затрагивают общую консервативность.



3. Центральный участок с Q-мотивами

```
QQQQQQQQQQQQQ—————
QQQQQQQQQQQQQQQHNPGKQAKEQQQQQQQQQQQQQLAA *****
*****. ** *****
```

Здесь прослеживается длинный глутамин-обогащённый участок (polyQ-мотив), характерный для белков, вовлечённых в транскрипционную регуляцию. Это типичный **функционально значимый регион**, чувствительный к длине и составу повтора.



4. Участок с PGL и HL-мотивами

```
QQLVFQQQLLQMQLQQQHLLSLQRQGLISIPPGQAALPVQSLPQAGLSPAIEIQQQLWK
E *****: **
*****:*****
```

Устойчивость глутамина (Q), пролина (P), глицина (G) и лейцина (L) указывает на важность этого участка. Пролин-глициновые мотивы часто

участвуют в формировании гибких петель.



5. С-концевой участок

```
VTGVHSMEDNGIKHGGDLTTNNSSTTSSTTSKASPPITHHSIVNGQSSVLNARRDSS
S
*****
*
```

Полностью консервативен — что крайне редко для С-концевых областей, что может указывать на участие в важной клеточной функции, например, в ядерной локализации или взаимодействиях.

Полученное название объединяющего таксона на латыни и на русском и краткое описание таксона в произвольной форме

Ответ: ближайший общий таксон — ***Amniota*** | **Амниоты**

Amniota — это таксон позвоночных животных, объединяющий всех животных, которые откладывают яйца с амнионом — специализированной оболочкой, которая защищает эмбрион и позволяет развиваться вне водной среды.

Проще говоря, амниотами называют группу животных, включающую:

- ◆ Рептилий (ящерицы, змеи, черепахи, крокодилы, динозавры и птицы)
- ◆ Млекопитающих