Предмет: Прикладная биоинформатика

Исполнительница: Смолкина Ю.А

Группа: Адбм

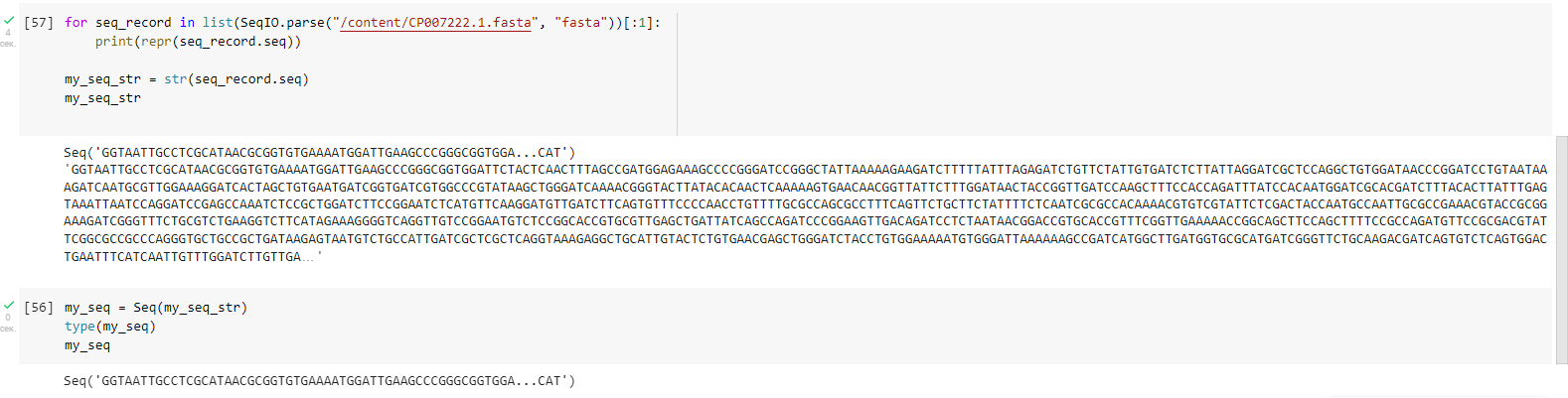
Домашнее задание №1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

| Смолкина Ю. А. | CP007222.1 | ATP synthase subunit alpha |
| --- | --- | --- |

**ЧАСТЬ 1**

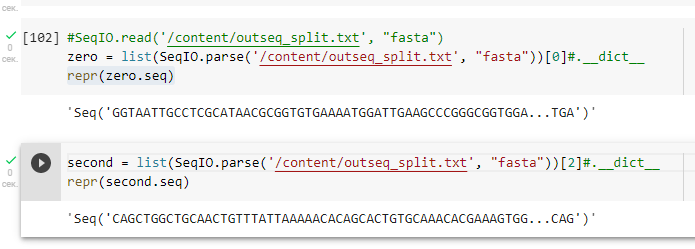
После чтения файла :



* 1. Разрезать геном на 10 частей. (команда **splitter - seqretsplit**)
  2. Посчитать число слов AAAA, ATAT, ATTA, AATT (команда **compseq**)
  3. Перемешать каждую из 10 частей (команда **shuffleseq)**
  4. Посчитать число слов AAAA, ATAT, ATTA, AATT в перемешанных последовательностях

Команды выполнялись с помощью : <https://www.bioinformatics.nl/cgi-bin/emboss>

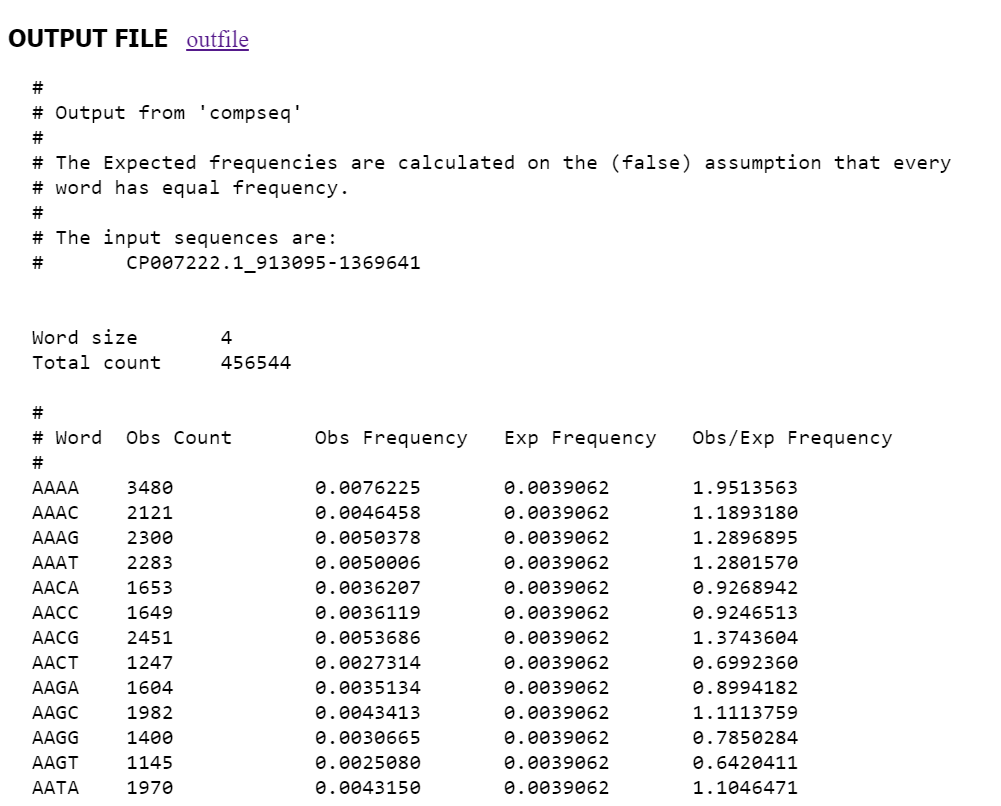
Проверим, действительно исходный файл распарсился, используя **splitter**



Разбивая на 10 файлов:

Получаем файлы 1-10.fasta

Посчитаем в новых файлах число слов AAAA, ATAT, ATTA, AATT с помощью **compseq**

****

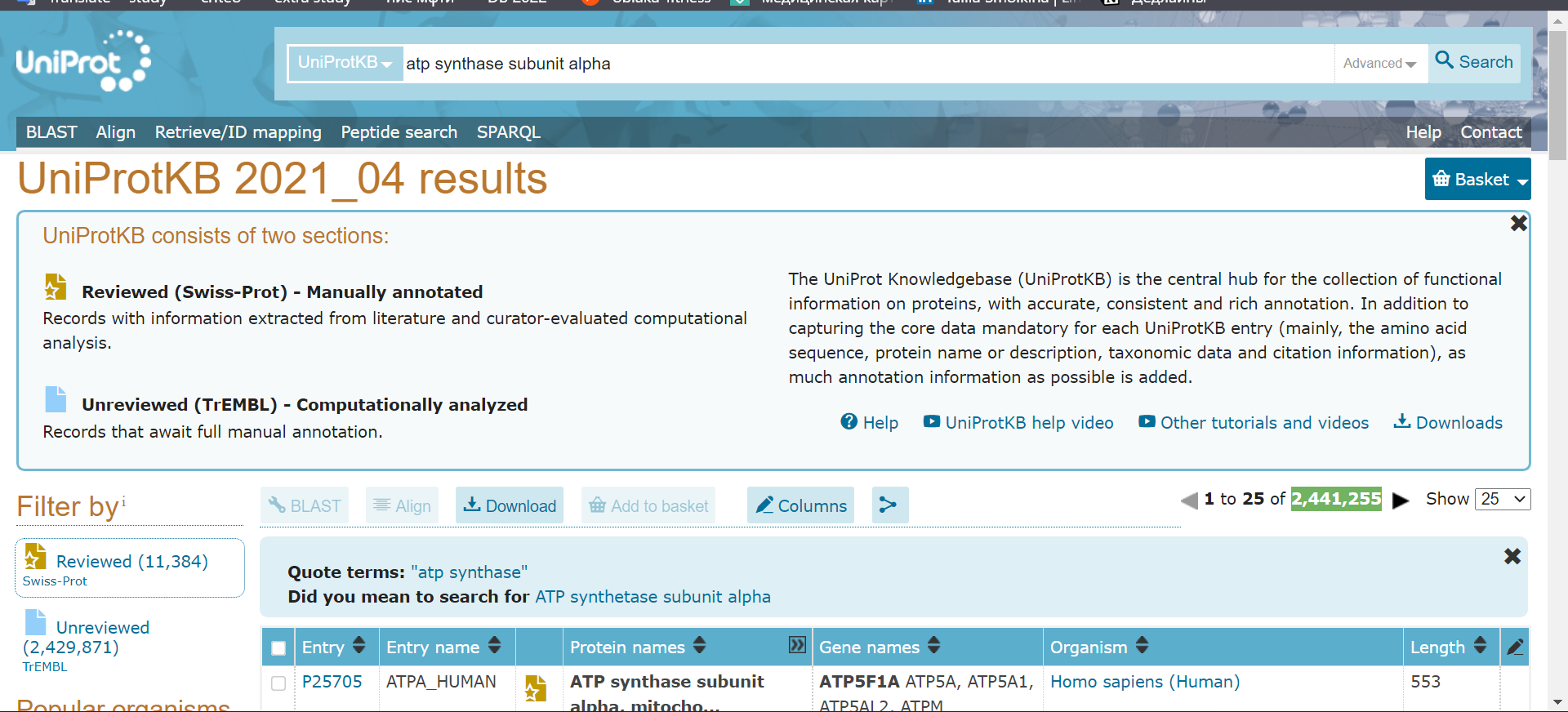
|  | AAAA | ATAT | ATTA | AATT |
| --- | --- | --- | --- | --- |
| **1** | **3681** | **1870** | **1774** | **1723** |
| **2** | **3814** | **2092** | **2009** | **1882** |
| **3** | **3306** | **1831** | **1779** | **1620** |
| **4** | **3455** | **1877** | **1786** | **1676** |
| **-** | AAAA | ATAT | ATTA | AATT |
| **5** | **3431** | **1786** | **1765** | **1673** |
| **6** | **3492** | **1780** | **1733** | **1523** |
| **7** | **3480** | **1941** | **1806** | **1636** |
| **8** | **3816** | **1768** | **1824** | **1656** |
| **9** | **3490** | **1876** | **1891** | **1722** |
| **10** | **4036** | **2189** | **2244** | **1899** |

Затем применяем **shuffleseq** каждому файлу и пересчитываем последовательности:

|  | AAAA | ATAT | ATTA | AATT |
| --- | --- | --- | --- | --- |
| **1** | **1387** | **1410** | **1431** | **1330** |
| **2** | **1573** | **1580** | **1674** | **1601** |
| **3** | **1477** | **1456** | **1488** | **1488** |
| **4** | **1429** | **1490** | **1470** | **1402** |
| **-** | AAAA | ATAT | ATTA | AATT |
| **5** | **1467** | **1396** | **1447** | **1445** |
| **6** | **1478** | **1344** | **1395** | **1349** |
| **7** | **1281** | **1384** | **1386** | **1438** |
| **8** | **1388** | **1492** | **1465** | **1411** |
| **9** | **1594** | **1490** | **1486** | **1436** |
| **10** | **1636** | **1707** | **1722** | **1709** |

**ЧАСТЬ 2**

Сколько результатов найдено в UniprotKB по запросу в виде названия белка **ATP synthase subunit alpha**?В Swiss-Prot? В TrEMBL?



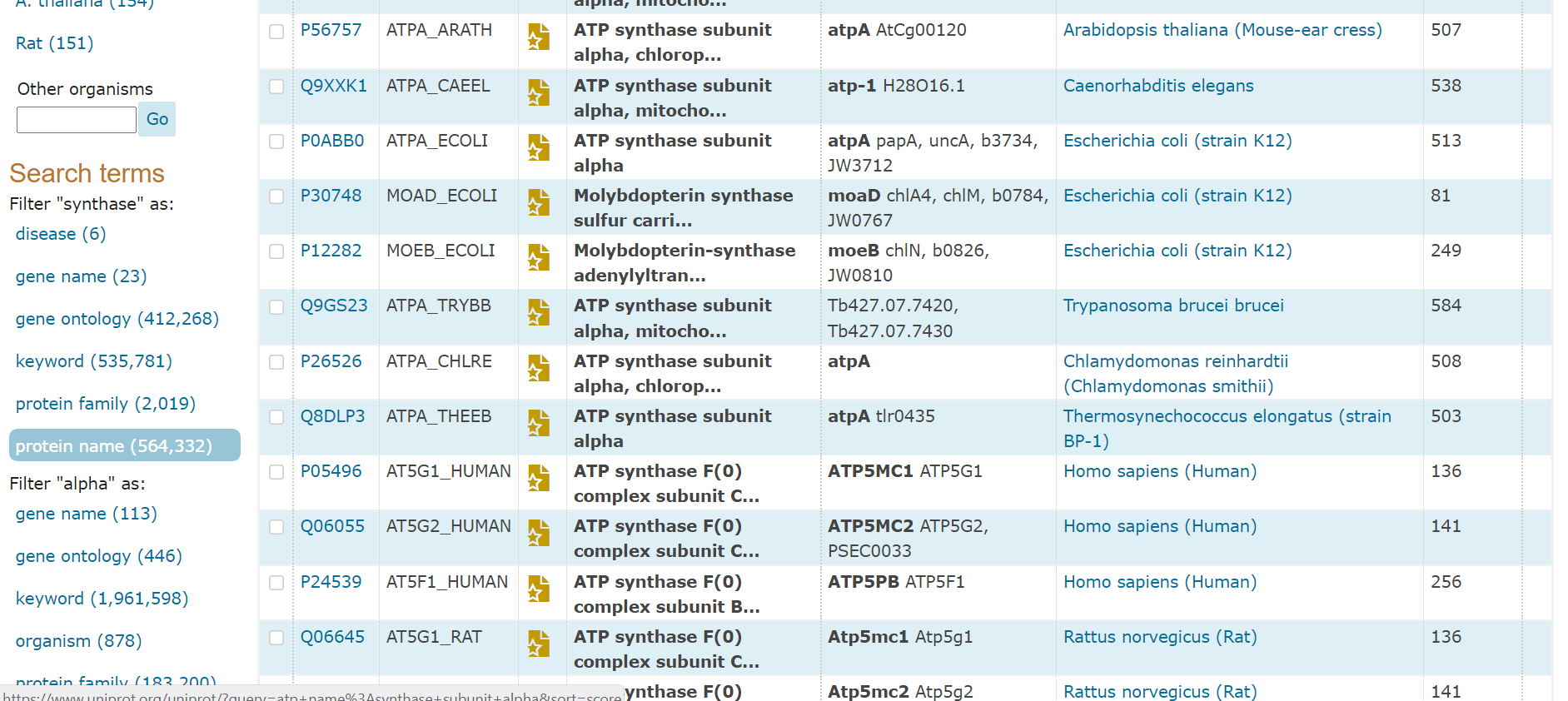
общее кол-во для

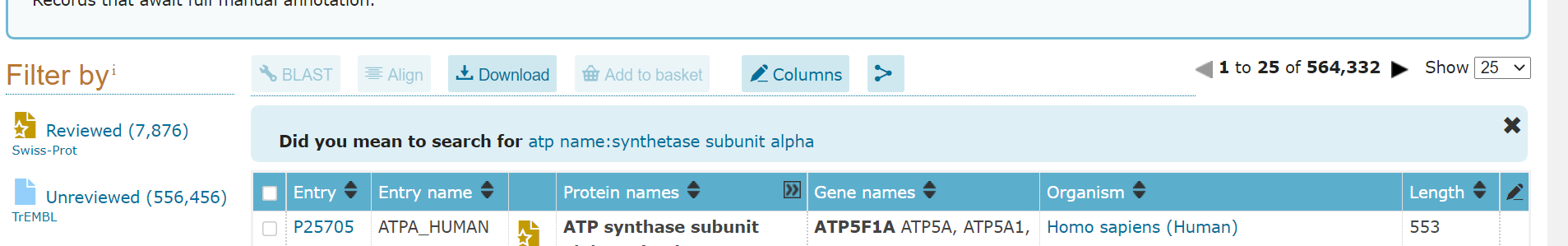
UniprotKB - 2441255

Swiss-Prot - 11384

TrEMBL (unreviewed) - 2429871

Сколько результатов найдено при использовании расширенного поиска в поле protein name? В Swiss-Prot? В TrEMBL?





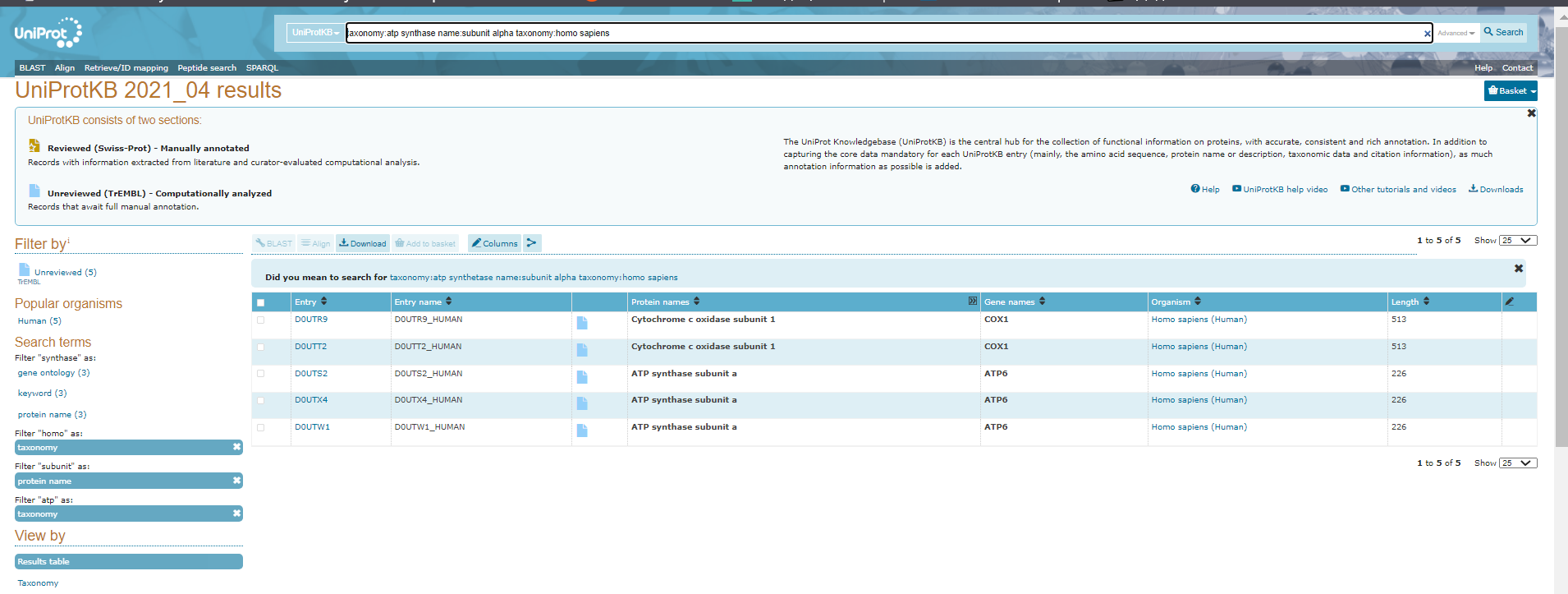
то есть уже

UniprotKB - 564322

Swiss-Prot - 7876

TrEMBL (unreviewed) - 556456

Сколько результатов остается при добавлении фильтра Homo sapiens в поле Taxonomy? В Swiss-Prot? В TrEMBL?



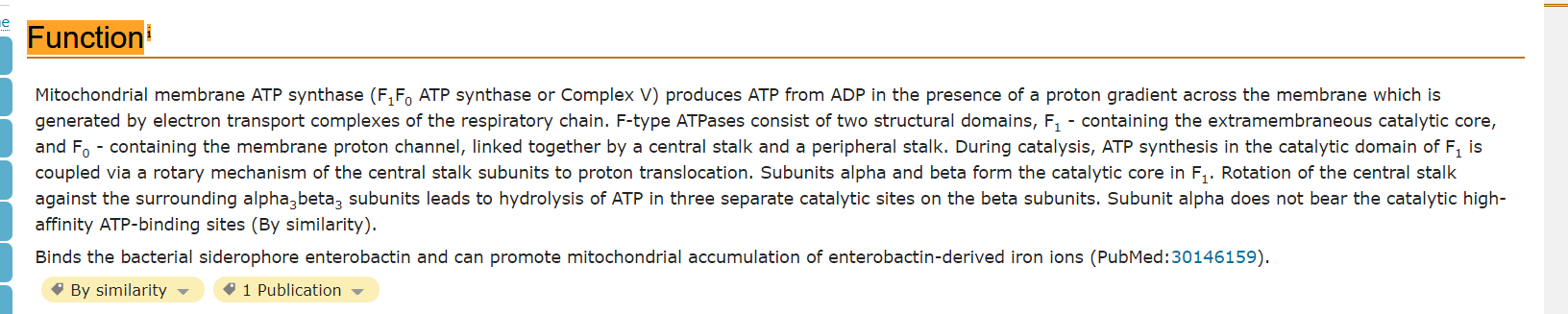
UniprotKB - 5

Swiss-Prot - 1

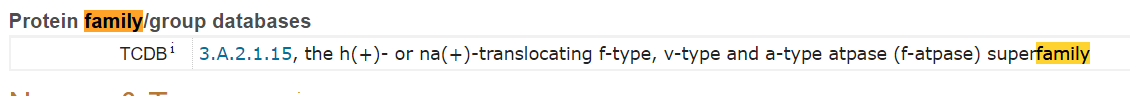
TrEMBL (unreviewed) - 3

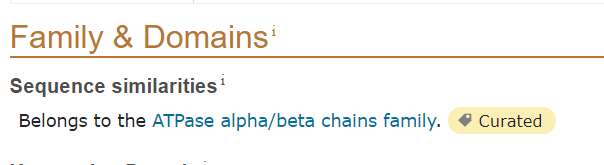
Откройте запись о вашем белке и ответьте на следующие вопросы:

Какова функция белка?



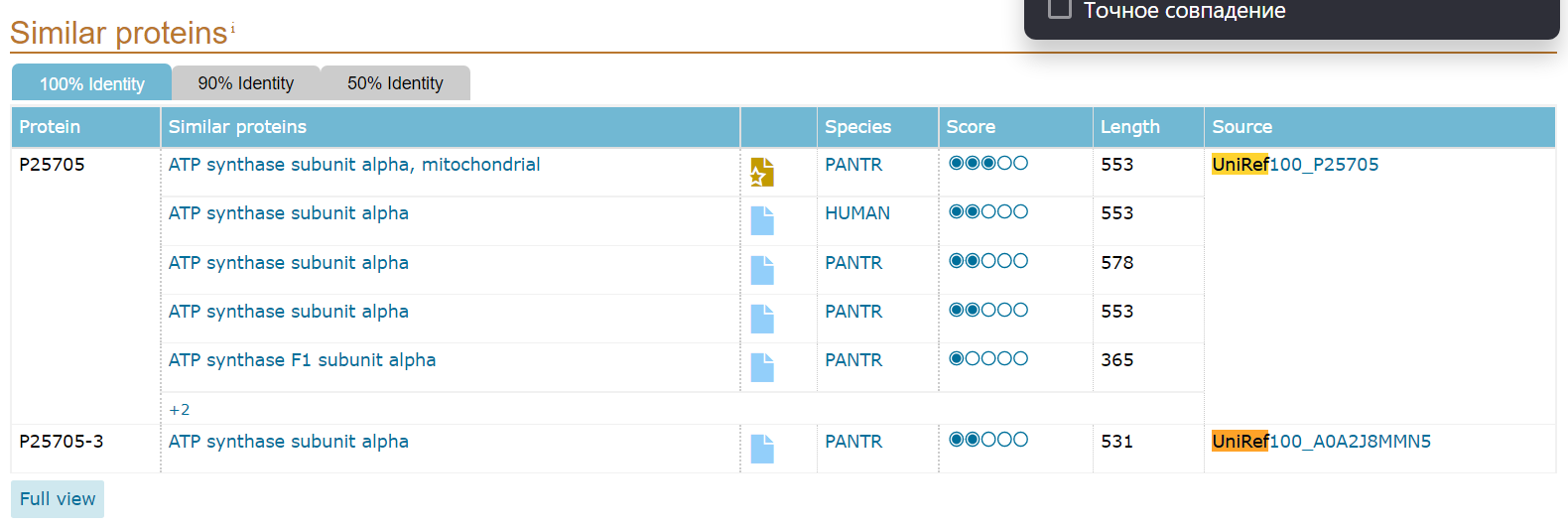
К какому семейству он принадлежит?



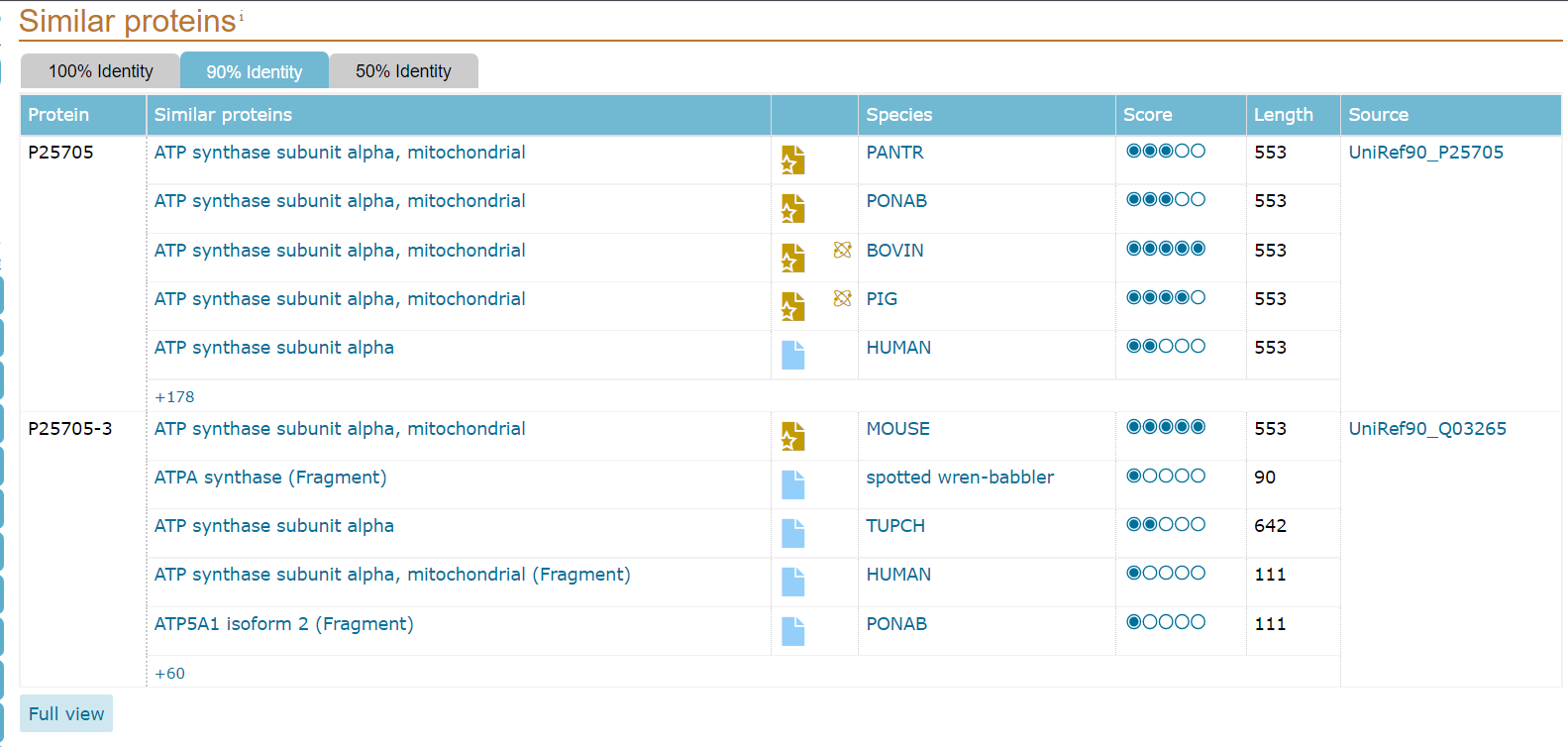


К скольки кластерам UniRef с идентичностью 1.0, 0.9, 0.5 принадлежит белок (включая изоформы)?

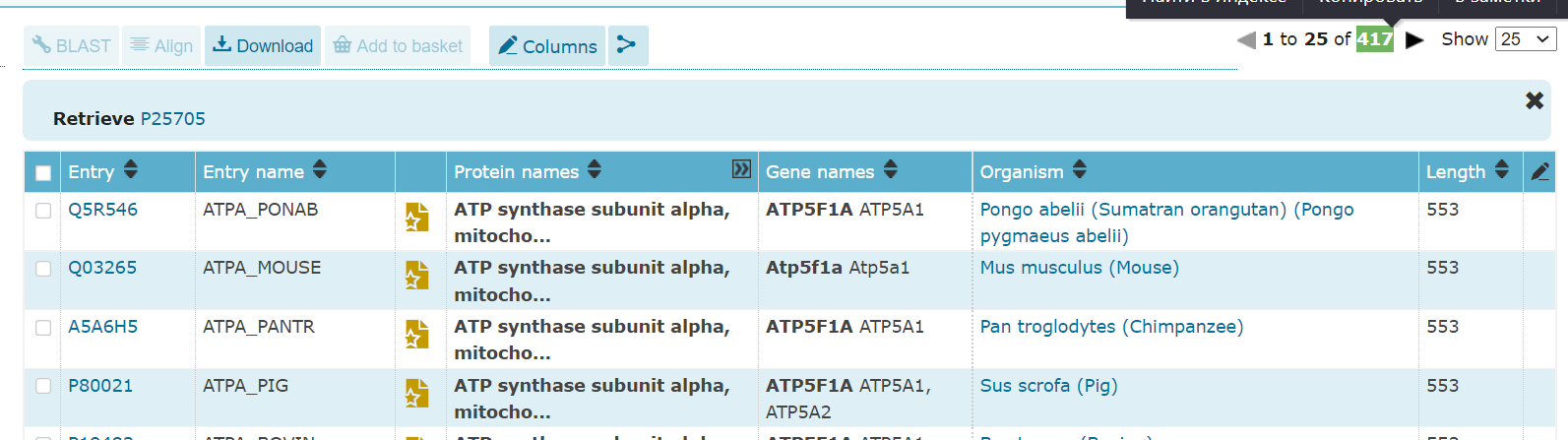
Для 1.0 **7**



Для 0.9 **188**



Для 0.5 **417**

****