Отчет по выполнению Задания №1 - Выравнивание генов

# I. Описание

В качестве изучаемого фенотипа был взят фенотип:

[Blood group, P1PK system, P(k) phenotype]

Ссылка: <https://omim.org/entry/111400>

Фенотип P(k) – один из двух фенотипов (P1 и P(k)), отвечающих за определение 5 различных фенотипов системы Р-групп крови по системе P1PK (P(1), P(2), P(1)(k), P(2)(k), и p).

P(k) является результатом добавления альфа-Gal альфа-1,4-галактозилтрансферазой (A4GALT). P(k) является субстратом для бета-3-галактозилтрансферазы-3 (B3GALT3), которая добавляет N-ацетилглюкозамин для образования P-антигена.

С помощью иммунохимических исследований Найки и Маркус (1975) идентифицировали антигены Р(к) и Р-группы крови как тригексозид керамида (CTH) и глобозид, соответственно, и предложили структуру антигена Р1. Хотя детерминанты P1 и P(k) имели идентичные концевые дисахариды, ХТГ не ингибировал анти-P1, а антигены P1 не реагировали с анти-P(k) сывороткой. Дальнейшие исследования не выявили перекрестных реакций между антигенами P, P1 или P(k).

Мутация в фенотипе P(k) приводит к развитию анемии.

# II - III. Выбранные гены

Таким образом, выбранными генами являются:

* A4GALT
  + human (Homo sapiens)
  + MODEL - dog (Canis lupus familiaris)
* B3GALT3
  + human (Homo sapiens)
  + MODEL - sheep (Ovis aries)

# IV. Выравнивание

Было решено использовать методы выравнивания Needle и Water.

Результаты выравнивания:

|  |  |  |
| --- | --- | --- |
| - | NEEDLE | WATER |
| A4GALT | Length: 1326  # Identity: 623/1326 (47.0%)  # Similarity: 623/1326 (47.0%)  # Gaps: 552/1326 (41.6%)  # Score: 1235.0 | Length: 1394  # Identity: 685/1394 (49.1%)  # Similarity: 685/1394 (49.1%)  # Gaps: 537/1394 (38.5%)  # Score: 1376.0 |
| B3GALT3 | Length: 1298  # Identity: 744/1298 (57.3%)  # Similarity: 744/1298 (57.3%)  # Gaps: 356/1298 (27.4%)  # Score: 2183.0 | Length: 1241  # Identity: 742/1241 (59.8%)  # Similarity: 742/1241 (59.8%)  # Gaps: 305/1241 (24.6%)  # Score: 2183.0 |

**Вывод**: разница между количеством символов последовательностей во всех измерениях не настолько велика, чтобы значительно повлиять на результат. Значит алгоритм **WATER в среднем лучше** на (2,1 + 1,5) / 2 = 1,8% (среднее от разницы в пункте Similarity).

# V. Приложение

Файлы:

* папка full\_genes содержит полные гены A4GALT и B3GALT3
* папка gene\_chunks содержит первые 17 строк генов A4GALT и B3GALT3, по которым делались расчеты. Их пришлось использовать из-за того, что алгоритм не принимал полные последовательности.
* ncbi\_NEEDLE\_A4GALT\_humanANDdog.txt, ncbi\_NEEDLE\_B3GALNT1\_humanANDsheep.txt, ncbi\_WATER\_A4GALT\_humanANDdog.txt и ncbi\_WATER\_B3GALNT1\_humanANDsheep.txt

содержат результаты алгоритмов NEEDLE и WATER.