**Пример отчёта**

**Дата последнего изменения: 14.11.2013**

**Отчёт по первому блоку заданий по практической информатике**

*XXXX YYYYYYY, группа 102*

**1. Описание генома *Allpahuayo virus***

По типу носителя генома вирус *Allpahuayo* (русское название мне неизвестно) классифицируется как ssRNA negative-strand virus, т.е. носителем информации является одноцепочечная РНК негативной полярности. Последнее означает, что транслируется не та РНК, которая входит в состав вируса, а комплементарная ей. Вирус был впервые выделен из древесной крысы, живущей в Бразилии [1].

Геном этого вируса состоит из двух молекул РНК. Первая (сегмент L) содержится в записи NC\_010249.1 банка RefSeq, вторая (сегмент S) - в записи NC\_010253.1. Сегмент S состоит из 3343 нуклеотидов, сегмент L – из 7053 нуклеотидов; всего **10396** нуклеотидов. В геноме закодировано четыре гена, по два в каждом сегменте. Информация о генах приведена в Таблице 1.

*Таблица 1. Гены в геноме вируса Allpahuayo*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№** | **Сегмент** | **Начало** | **Конец** | **Ориентация** | **Длина** | **Продукт** |
| 1 | S | 33 | 1718 | +1 | 1686 | nucleocapsid protein |
| 2 | S | 1789 | 3312 | -1 | 1524 | glycoprotein precursor |
| 3 | L | 31 | 6636 | +1 | 6606 | L protein(полимераза) |
| 4 | L | 6710 | 6997 | -1 | 288 | Z protein |

**Гены не пересекаются. Длина всех генов делится на 3.** Графическое изображение всего L-сегмента вируса ***Allpahuyo* приведено на Рис. 1, а 5’-концевая область гена Z-белка приведена на Рис. 2.**

Хотя в записи указано, что вируса принадлежит к группе вирусов, у которых генетическая информация сохранена в виде одноцепочечной РНК негативной полярности, из таблицы 1 и Рис. 1 следует, что в геноме вируса гены кодируются как на прямой, так и на обратной цепи нуклеиновой кислоты. Разобраться в том, как в таком случае происходит трансляция, мне не удалось.

****

***Рис. 1. Расположение генов в сегменте L вируса Allpahuyo (идентификатор записи RefSeq*** *NC\_010249.1).* ***Рисунок получен с помощью геномного браузера на сайте NCBI (***[*http://www.ncbi.nlm.nih.gov*](http://www.ncbi.nlm.nih.gov/)***)***



***Рис. 2. 5’-концевой фрагмент гена Z-белка из сегмента L вируса Allpahuyo (идентификатор записи RefSeq*** *NC\_010249.1)****. Первый кодон – ATG, так как ген расположен на комплементарной цепи. Он соответствует аминокислотному остатку метионина (M). Рисунок получен с помощью геномного браузера на сайте NCBI (***[*http://www.ncbi.nlm.nih.gov*](http://www.ncbi.nlm.nih.gov/)***)***

**2. Описание генома бактерии *Bacillus infantis* штамм NRRL B-14911**

**Бактерия *Bacillus infantis* была выделена из крови новорожденного, страдающего сепсисом [2]. Авторы статьи не уверены в том, что именно эта бактерия была причиной сепсиса – возможно загрязнение при взятии крови. Установлено, что это аэробная грамм-положительная бактерия. Микрофотографии бактерии или культуры этих бактерий мне найти не удалось.**

**Геном бактерии *B.infantis*, штамм NRRL B-14911, состоит из одной хромосомы, т.е. из одной двухцепочечной молекулы ДНК. Он описан в записи NC\_022524.1 базы данных RefSeq. Длина генома - 4 884 713 пар нуклеотидов. В записи приведена последовательность одной из двух комплементарных цепей ДНК.**

**По данным из записи NC\_022524.1, в геноме закодировано 5 179 генов. Из них 5 066 кодируют белки, остальные 113 кодируют РНК. Указан один псевдоген, т.е. последовательность нуклеотидов, ранее кодировавшая белок, но в процессе эволюции способность бактериальной клетки экспрессировать белки с этой последовательности была утрачена.**

**Описание оперона АТФ-синтазы в геноме *B.anthracis* штамм A0248 (в геноме *B.infantis* АТФ синтаза еще не описана; пришлось заменить на геном другой бациллы. ААл)**

**АТФ-синтаза – белок, состоящий из многих субъединиц, главная функция которого состоит в синтезе АТФ – молекул, аккумулирующих энергию. α-субъединица АТФ-синтазы закодирована на комплементарной цепи, см. рис.3, на котором ее ген обозначен atpA.**

****

***Рис. 3. Окрестность гена atpA в геноме B.anthracis штамм A0248. Приведен участок хромосомы с координатами 5032704..5050876. Рисунок получен с помощью геномного браузера на сайте NCBI (***[*http://www.ncbi.nlm.nih.gov*](http://www.ncbi.nlm.nih.gov/)***)***

**На рис. 3 можно увидеть все со-направленные с atpA гены. Все со-направленные гены описаны также в таблице 2. Из таблицы следует, что возможный оперон включает 19 белков, из них 9 описаны как компоненты АТФ синтазы.**

***Таблица 2. Гены, предположительно, входящие в один оперон с геном atpA. Предыдущий и следующий гены расположены на прямой цепи (т.е. в ориентации +1), в таблицу они не включены. Мое недоумение вызвало расхождение координат одного и того же гена в геномном браузере (см. рис. 3) и в записи GenBank. Однако у заметил, что запись, которую я открыл, содержит не весь геном, а только приведенные гены и еще несколько. Наверное, полученные мной координаты не в полном геноме, а в этой записи. Для удовлетворения собственного интереса я рассчитал также длины межгенных промежутков. Координаты генов я разнес по столбикам, а не так, как указавалось в заданиях практикума 4.***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **№** | **Начало** | **Конец** | **Ориен-тация** | **Меж-генное****Рас-стояние** | **Поле /product** | **Перевод** |
| 1 | 19542 | 20039 | -1 |  | putative PTS system, glucose-specific IIA |  |
| 2 | 18854 | 19294 | -1 | 247 | low molecular weight phosphotyrosine protein |  |
| 3 | 18314 | 18757 | -1 | 96 | sugar-phosphate isomerase, RpiB/LacA/LacB family |  |
| 4 | 17675 | 18265 | -1 | 48 | putative lipoprotein |  |
| 5 | 16217 | 17458 | -1 | 216 | glycine hydroxymethyltransferase |  |
| 6 | 15316 | 15945 | -1 | 271 | uracil phosphoribosyltransferase |  |
| 7 | 14831 | 14944 | -1 | 371 | hypothetical protein | Гипотетический белок |
| 8 | 14580 | 14804 | -1 | 26 | conserved hypothetical protein | Консервативный гипотетический белок |
| 9 | 14188 | 14580 | -1 | -1 | ATP synthase protein I | АТФ синтаза, белок I |
| 10 | 13461 | 14180 | -1 | 7 | ATP synthase F0, A subunit | АТФ синтаза FО, субъединица A |
| 11 | 13186 | 13404 | -1 | 56 | ATP synthase F0, C subunit | АТФ синтаза FО, субъединица C |
| 12 | 12552 | 13058 | -1 | 127 | ATP synthase F0, B subunit | АТФ синтаза FО, субъединица B |
| 13 | 12013 | 12555 | -1 | -4 | ATP synthase F1, delta subunit | АТФ синтаза F1, субъединица δ |
| 14 | 10493 | 12001 | -1 | 11 | ATP synthase F1, alpha subunit | АТФ синтаза F1, субъединица α |
| 15 | 9297 | 10157 | -1 | 335 | ATP synthase F1, gamma subunit | АТФ синтаза F1, субъединица γ |
| 16 | 7768 | 9177 | -1 | 119 | ATP synthase F1, beta subunit | АТФ синтаза F1, субъединица β |
| 17 | 7346 | 7747 | -1 | 20 | ATP synthase F1, epsilon subunit | АТФ синтаза F1, субъединица ε |
| 18 | 6433 | 7224 | -1 | 121 | conserved hypothetical protein | Консервативный гипотетический белок |
| 19 | 5732 | 6385 | -1 | 47 | conserved hypothetical protein | Консервативный гипотетический белок |

**3. Описание белка глютамин амидотрансфераза (glutamine amidotransferase) из генома бактерии *Bacillus infantis*, штамм NRRL B-14911**

**Информация о белке и его гене, полученная из записи NC\_022524.1 базы данных RefSeq с полным геномом бактерии, приведена в таблице 3. На рис. 4 изображена окрестность в геноме гена описываемого белка.**

***Таблица 3. Информация о глютамин амидотрансферазе из генома Bacillus infantis, штамм NRRL B-14911 (идентификатор YP\_008606782.1 в базе RefSeq)***

|  |  |  |
| --- | --- | --- |
| **Вид информации** | **Квалификатор в записи генома** | **Значение** |
| Локус гена в геноме | /locus\_tag | N288\_00095 |
| Имя гена | /gene | отсутствует |
| дентификатор гена в базе данных NCBI Gene | /db\_xref=”GeneID:…” | отсутствует |
| Начало гена в геноме | — | 22479 |
| Конец гена в геноме | — | 23066 |
| На какой цепи кодируется ген (прямая или обратная) | — | Обратная |
| Длина гена (в парах нуклеотидов) | — | 588 |
| Идентификатор белка в базе данных NCBI Protein | /protein\_id | YP\_008606782.1 |
| Длина белка (в аминокислотных остатках) | — | 195 |



***Рис. 4. Геномная окрестность гена белка глютамин амидотрансфераза из генома бактерии Bacillus infantis, штамм NRRL B-14911. Ген отмечен черной рамочкой. Рисунок получен с помощью геномного браузера на сайте NCBI (***[*http://www.ncbi.nlm.nih.gov*](http://www.ncbi.nlm.nih.gov/)***)***

**Из рисунка следует, что соседние гены со-направлены с данным. Расстояния между геном и соседом с 5’-конца - 13 пар нуклеотидов, с 3’-конца – 127 пар нуклеотидов. Я думаю, что эти три гена могут быть в одном опероне поскольку они со-направлены. Не исключено, что в оперон входят еще два гена с 5’-конца, см. рис. 4.**

**Источники:**

**[1] Moncayo et al., Allpahuayo virus: a newly recognized arenavirus (arenaviridae) from arboreal rice rats (oecomys bicolor and oecomys paricola) in northeastern peru. Virology, 2001, 284(2):277-86.**

**[2] Kwan Soo Ko et al., Bacillus infantis sp. nov. and Bacillus idriensis sp. nov., isolated from a patient with neonatal sepsis, International Journal of Systematic and Evolutionary Microbiology, 2006, vol.56, no.11, pgs 2541-2544 (**<http://ijs.sgmjournals.org/content/56/11/2541.full>**)**