

# Кольцевые последовательности ДНК в геноме *Dicyema* sp.

Исаев С. В.<sup>1,2</sup>, Зверков О. А.<sup>2</sup>, Михайлов К. В.<sup>2,3</sup>, Алёшин В. В.<sup>2,3</sup> и Панчин Ю. В.<sup>2,3</sup>

<sup>1</sup> Московский государственный университет им. М. В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, Россия

<sup>2</sup> Институт проблем передачи информации им. А. А. Харкевича РАН, Москва, Россия

<sup>3</sup> Московский государственный университет им. М. В. Ломоносова, НИИ физико-химической биологии им. А. Н. Белозерского, Москва, Россия  
ypanchin@yahoo.com

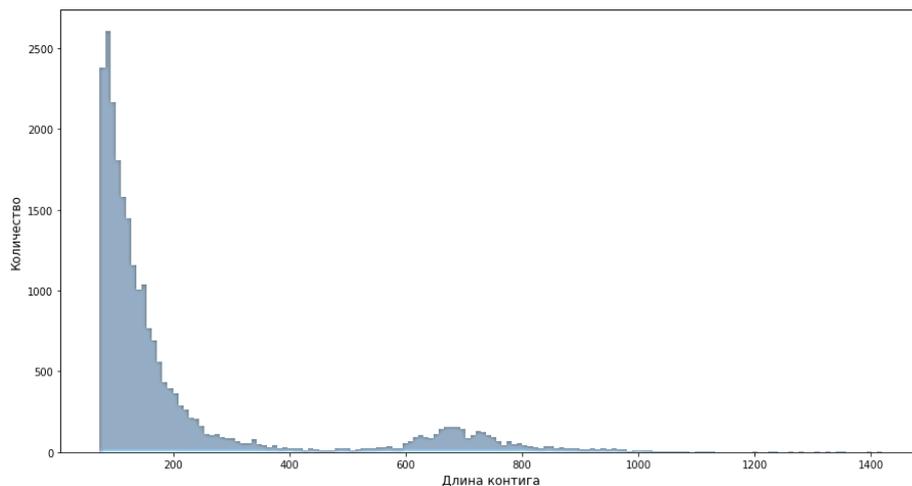
**Аннотация.** Дициемиды — клада беспозвоночных животных, являющихся паразитами головоногих моллюсков. Ранее эта группа животных рассматривалась как базальная для дерева Metazoa, однако позже было показано, что дициемиды — это вторично упрощённые представители Lophotrochozoa [1], которые в связи с паразитизмом утратили большинство типичных для билатерий систем: кровеносную, нервную, мышечную [2]. Ещё одна интересная особенность дициемид заключается в том, что в них были обнаружены экстрахромосомные кольцевые ДНК ядерного и митохондриального происхождения [3, 4]. В ходе работы впервые был произведён анализ генома *Dicyema* sp., который показал наличие кольцевых ДНК двух типов («длинных» и «коротких»). Характеристики обнаруженных кольцевых ДНК согласуются с приведёнными ранее данными электронной микроскопии и ПЦР. Показано, что митохондриальные гены, кодирующие белки и рРНК, содержатся в кольцевых ДНК по правилу «одна молекула — один ген». Нами не было обнаружено свидетельств того, что последовательности из кольцевых митохондриальных ДНК содержатся в больших линейных участках, что ставит под сомнение образование колец из каких-либо некольцевых предшественников. Таким образом, в данной работе было показано, что для дициемид характерно нетипичные для билатерий особенности устройства митохондриального генома.

**Ключевые слова:** Mesozoa, Dicyemida, диминуция хроматина, митохондриальная ДНК.

## 1 Кольцевые контиги

В качестве материала для анализа были использованы секвенированные и собранные геном и транскриптом *Dicyema* sp. Идентификация кольцевых ДНК основывается на поиске таких геномных контигов, у которых первые и последние 77 нуклеотидов совпадают — такой критерий является следствием параметров сборки генома (77 — длина использованного  $k$ -мера) [5]. Стоит отметить, что

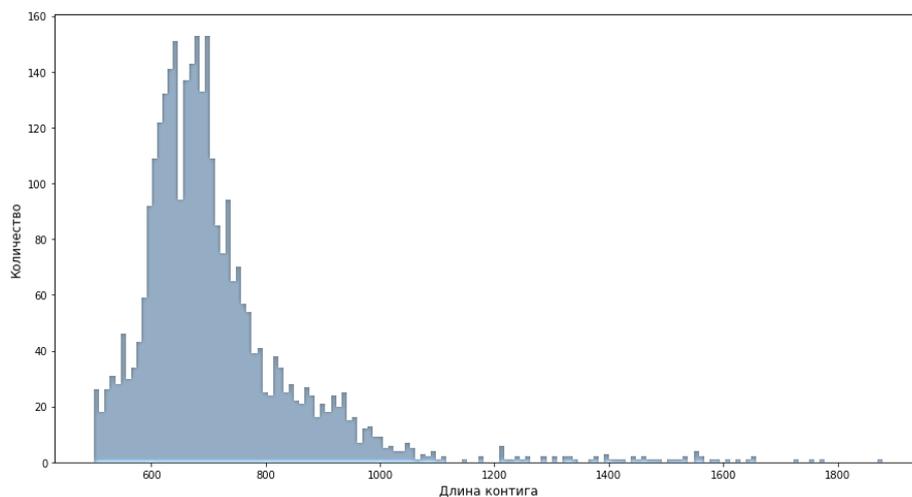
при помощи методов анализа последовательностей сложно отличить кольцевые молекулы ДНК от прямых повторов. Однако существует ряд исследований, подтверждающих наличие колец ДНК у дициемид [3, 4]. В результате поиска было обнаружено, что в сборке присутствует 24 065 контигов с 77-нуклеотидными повторами по концам. Распределение их длин представлено на рис. 1.



**Рис. 1.** Гистограмма распределения длин контигов с 77-нуклеотидными повторами по концам. Взятые последовательности длиной более порога 77+77 нуклеотидов. На гистограмме представлены длины контигов за вычетом концевых повторов. Видны два максимума — в начале и в районе длины в 700 нуклеотидов.

На гистограмме с рис. 1 видно, что распределение получается бимодальным — чётко видны два кластера, различающиеся по длине. Экспериментальные данные, в которых при помощи электронной микроскопии и методов базовой молекулярной биологии было показано наличие фракции кольцевых ДНК длиной около 700 нуклеотидов (с учётом поправки на то, что молекулы кольцевые) [4], верифицируют полученные результаты. С учётом вышесказанного в ходе работы было решено сосредоточить внимание именно на фракции ранее описанных «тяжёлых» колец (длиной более 500 нуклеотидов) — их оказалось 3220. Распределение длин контигов «тяжёлых» колец представлено на рис. 2.

Для подтверждения того, что это именно кольцевые последовательности, а не прямые повторы, был произведён анализ парных чтений секвенирования, которые картируются на рассматриваемые кольца. Нами не было обнаружено таких случаев, когда только одно чтение из пары ложится на кольцо, а другое — на какой-то фланкирующий фрагмент ДНК.



**Рис. 2.** Гистограмма распределения длин кольцевых контигов «тяжёлой» фракции.

Сравнением частот вхождений коротких слов в кольцевые и некольцевые контиги сборки генома обнаружено, что последовательность 5'-AGGGAGCACCCCT-3' (и комплементарная ей 5'-AGGGTGCTCCCT-3') встречается в «длинных» «кольцевых» контигах значительно чаще, чем в других участках генома. Так, в длинных контигах мы видим одно вхождение на 26905 нуклеотидов, по всему геному — одно вхождение на 8162571 нуклеотид. По критерию хи-квадрат уровень значимости отличия частоты встречаемости  $p < 0.001$ . Стоит отметить, что практически во всех кольцах сигнал встречается исключительно один раз (за исключением одного случая). Несмотря на обнаруженную статистически значимую закономерность, сигнал в кольцах встречается редко: на 3220 колец приходится 104 кольца с сигналом.

## 2 Митохондриальная ДНК

Ранее было показано, что у дициемид гены *cox1*, *cox2* и *cox3* находятся в индивидуальных кольцевых молекулах ДНК [4]. Авторами упоминаемой работы была предложена гипотеза «один ген — один репликон», согласно которой все гены митохондриального генома закодированы в одиночных молекулах ДНК.

С целью проверки этой гипотезы последовательности, соответствующие генам митохондриальной ДНК, были идентифицированы на кольцах «тяжёлой» фракции с использованием веб-сервиса MITOS [6], а также дополнительно верифицированы при помощи программ blastn и blastx [7].

В ходе анализа на кольцах были обнаружены последовательности белок-кодирующих генов *cox1*, 2, 3, *cob*, *nad1*, 2, 3, 4, 5, а также *atp6*. Все они находятся на контигах с концевыми поавторами. На таких же контигах найдены гены рибосомных РНК (*rnl* и *rns*), однако достоверность находок рРНК с помощью

других подходов подтвердить сложно: выравнивание нуклеотидных последовательностей не подтверждает гомологии, в то время как в основе алгоритмов поиска MITOS лежит структурный подход. Стоит упомянуть, что распознать ген *rrnL* по последовательности трудно, например, в мтДНК *Intoshia linei*, имеющих мтДНК типичного для животных строения. Мы предполагаем, что трудности распознавания связаны с тем, что *Dicyema* sp., как и *I. linei*, является организмом с «высокой скоростью эволюции», что видно на филогенетическом дереве как длинные ветви [9]. В линейных контигах копии исследуемых генов обнаружены не были.

Все найденные митохондриальные гены присутствуют в транскриптом. Интересными представляются случаи, когда длина обнаруженного транскрипта больше, чем длина кольца, с которого данная последовательность транскрибируется. Такое наблюдается, например, в случае с парой кольцо-транскрипт для гена *cox2*. Точечная матрица локальных выравниваний показала, что транскрипт является прямым повтором и копирует кольцо. Это же справедливо и для случаев с генами *nad1* и *rrnS*. Такие ситуации можно объяснить ошибками сборки; другое же объяснение заключается в том, что транскрипция может проходить «по второму кругу» кольца. Есть ли в этом какая-то биологическая значимость, неизвестно.

Таким образом, геномный анализ подтверждает ранее высказанную гипотезу о нетипичном устройстве митохондриального генома у дициемид.

## Список литературы

1. Lu, T.-M., Kanda, M., Satoh, N. & Furuya, H. The phylogenetic position of dicyemid mesozoans offers insights into spiralian evolution. *Zoological letters* 3, 6 (2017).
2. Furuya, H. & Tsuneki, K. Biology of dicyemid mesozoans. *Zool. Sci.* 20, 519–532 (2003).
3. Noto, T., Yazaki, K. & Endoh, H. Developmentally regulated extrachromosomal circular DNA formation in the mesozoan *Dicyema japonicum*. *Chromosoma* 111, 359–368 (2003).
4. Watanabe, K. I., Bessho, Y., Kawasaki, M. & Hori, H. Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. *Journal of molecular biology* 286, 645–650 (1999).
5. Compeau, P. E., Pevzner, P. A. & Tesler, G. How to apply de Bruijn graphs to genome assembly. *Nature biotechnology* 29, 987 (2011).
6. Bernt, M. et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* 69(2), 313–9 (2017).
7. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25, 3389–3402 (1997).
8. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25, 3389–3402 (1997).
9. Robertson, H. et al. The mitochondrial genomes of the mesozoans *Intoshia linei*, *Dicyema* sp., and *Dicyema japonicum*. *bioRxiv* (2018).
10. Schiffer, P. H. et al. Orthonectids Are Highly Degenerate Annelid Worms. *Current Biology* 28, 1970–1974.e3 (2018).