

ГЕНОМ ЯЩЕРИЦЫ *DAREVSKIA VALENTINI*

Очкалова С.Д. (Университет ИТМО), Дикая В.А. (Университет ИТМО),

Ряховский С.С. (Университет ИТМО)

Научный руководитель – в.н.с, к.б.н Комиссаров А.С. (Университет ИТМО)

С наступлением эпохи NGS появилась возможность сравнивать последовательности геномов самых разнообразных видов, изменившая наше понимание эволюции. Для изучения видообразования и молекулярной эволюции перспективными животными-моделями являются рептилии, однако в настоящее время по сравнению с млекопитающими и другими амниотами этому классу посвящено пренебрежительно мало геномных проектов. В данной работе мы собрали и проаннотировали первый полный геном ящерицы *Darevskia valentini*, по которому планируется произвести популяционный и филогенетический анализ этого вида, а также изучить состав и сравнительную геномику повторов между ним и ближайшими родственными видами рептилий.

Введение. В XXI веке произошла смена парадигмы эволюционной биологии, ставшая возможной благодаря основательному изучению сотен геномов различных организмов. Закрепившийся за повторами ДНК, занимающими почти половину всей длины генома высших позвоночных, термин «мусорная ДНК» на многие годы отложил важные открытия о их биохимической активности. Так, существует гипотеза, по которой повторяющиеся последовательности ДНК катализируют хромосомные перестановки, обеспечивая вариабельность кариотипа и подталкивая эволюцию организмов. Одним из лучших объектов исследования для поиска ответов на вопросы о влиянии повторов на биологическое разнообразие являются геномы рептилий. Среди амниот, которые включают в себя большинство высокоорганизованных позвоночных, рептилии обладают исключительным геномным разнообразием и видовым богатством. Однако даже теперь, в эпоху NGS, число исследований геномов млекопитающих значительно превышает таковое для рептилий, хотя у последних число видов почти в четыре раза больше. С момента завершения первых геномных проектов млекопитающих (мышь и человека) до публикации первого генома рептилии, зеленого анолиса (*Anolis carolinensis*), прошло десятилетие. На текущий момент в базе геномов National Center of Biotechnology Information (NCBI) опубликовано только 64 геномных сборки рептилий (*Reptilia*) и еще меньше для семейства ящериц (*Lacertidae*) – 5 геномов. Целью данной работы являлись сборка и анализ секвенированного генома кавказской ящерицы *Darevskia valentini*.

Основная часть. Секвенирование ДНК осуществлялось по технологии 10x GENOMICS, сборка коротких прочтений – с помощью программы Supernova2. 20727 кодирующих участка ДНК были определены *de novo* с помощью скрытых Марковских моделей (СММ) программы AUGUSTUS по базе данных человека. Для их функциональной аннотации была использована база генов EggNOG и программа EggNOG-mapper, таким образом, были найдены 16629 генов с известной функцией.

Для того чтобы оценить внутривидовое разнообразие *Darevskia valentini*, было сделано генотипирование 3-х образцов ДНК от разных индивидов. После выравнивания коротких прочтений на собранный геном полиморфизмы были идентифицированы с помощью DeepVariant.

Для сравнительного геномного анализа были выбраны 5 видов ящериц-лацертид (*Podarcis muralis*, *Lacerta viridis*, *Lacerta bilineata*, *Zootoca vivipara*, *Lacerta agilis*), а также 3 вида рептилий из сестринской клады *Toxicofera* (*Python bivittatus*, *Thamnophis elegans*, *Anolis carolinensis*), для которых опубликованы качественные репрезентативные геномные сборки. Филогенетический анализ между этими видами был выполнен методом максимального правдоподобия по последовательностям генов-ортологов из базы данных sauropsida_odb10 и

пайплайна BUSCO Phylogenomics. На полученном дереве ветви имеют апостериорную вероятность не ниже 0.99, полученная кладограмма полностью согласуется с ранее опубликованными филогениями *Lacertidae* по митохондриальным геномам. *Darevskia valentini* формирует самостоятельную кладу, ближайшей к которой является род зеленых ящериц *Lacerta*.

Для поиска и описания геномных повторов в выбранном наборе видов и созданной нами новой сборке *Darevskia valentini* были использованы RepeatModeler и RepeatMasker. В результате у *Darevskia valentini* было замаскировано 582503742 п.н. или 41.53 % от всей длины генома. Половина из этого числа была классифицирована как ретротранспозоны. В совокупном результате для других видов рептилий было замаскировано 39.48% и 19% удалось классифицировать. Большая часть из них идентифицирована как длинные диспергированные повторы, типичные для *Squamate*, такие как L2, CR1 и Rex семейства (5.92%), а также семейства RTE (1.83%) и L1 (1.54%). Также были обнаружены немногочисленные (менее >1%) элементы семейства Penelope, ретротранспозоны с длинными концевыми повторами (LTR) и простые повторы.

Валидация целостности сборки была произведена с помощью Benchmarking Universal Single-Copy Orthologs (BUSCO) v2.0 по базам данных eukaryota_odb10 и sauropsida_odb10, содержащей 255 и 76 ультраконсервативных белковых семейств для эукариот и рептилий соответственно. Анализ показал, что геномная сборка *Darevskia valentini* содержит 97.3% белков из базы данных эукариот и 87.8% – рептилий. Метрики качества сборки были рассчитаны с помощью QUAST: длина генома – 1 457 658 600 п.н., число скаффолдов – 32348, N50 – 3861830, доля пробелов (%N) – 3.79%. Эти параметры показывают, что полученную сборку можно считать репрезентативной для генома *Darevskia valentini* и использовать для дальнейших исследований этого вида.

Выводы. Таким образом, был собран и проаннотирован первый полный геном ящерицы *Darevskia valentini* длиной 1 457 658 600 п.н, качество и аккуратность сборки которого были подтверждены. Также в результате работы было произведено большое количество данных, вдумчивый анализ которых еще предстоит произвести для изучения внутривидового разнообразия, геномики и истории этого вида в контексте клады *Squamata*.