

Сведения о выполненных работах в 2021 году
по проекту «Механизмы хромосомной пластичности малярийных комаров»,
поддержанному Российским научным фондом

Соглашение № 21-14-00182

Руководитель: Шарахов Игорь Валентинович, д-р биол. наук
Коханенко Алина Андреевна, канд. биол. наук

Целью нашего проекта является изучение механизмов возникновения и функционального значения структурной изменчивости генома, которая может модифицировать способность комаров быстро адаптироваться к природным условиям обитания. Для этого необходимо получить высококачественные референсные геномные сборки для *An. messeae*, *An. beklemishevi* и *An. quadrimaculatus*, *An. freeborni* с помощью секвенирования Oxford Nanopore, Illumina и создания сверхдлинных скэффолдов, соответствующих по длине хромосомам, используя Hi-C.

1. Секвенирование и сборка генома малярийных комаров.

На первом этапе работы была отработана методика выделения высокомолекулярной геномной ДНК в большой концентрации фенол-хлороформной экстракцией. В качестве материала были использованы определенные по видовому составу и полу замороженные на -70 С куколки *An. beklemishevi* и *An. messeae*. Сбор имаго малярийных комаров проводили в населенных пунктах Колпашевского района Томской области в местах дневок самок, стайках для крупного рогатого скота и других домашних животных. Всего было получено 4 образца геномной ДНК по два для каждого вида *An. beklemishevi* и *An. messeae* (самцы и самки). Полученные образцы соответствуют всем требованиям (количество, чистота), предъявляемым к ДНК-библиотекам для секвенирования. Было проведено секвенирование Oxford Nanopore комаров *An. freeborni* и *An. quadrimaculatus* и подготовлена Hi-C библиотеку для *An. quadrimaculatus*. *An. quadrimaculatus* - наиболее близко-родственный вид к *An. freeborni* и изучаемых видов. Комары *An. quadrimaculatus* были взяты из культуры, геномная ДНК была выделена и отсеквенирована Oxford Nanopore. Параллельно были получены прочтения Hi-C из имаго *An. quadrimaculatus*. Сборка генома *An. quadrimaculatus* осуществлялась посредством выравнивая прочтений Hi-C на геномные последовательности *An. quadrimaculatus* определённые методом Oxford Nanopore. Далее, с помощью свободно доступного программного обеспечения JuicerTools проводилось построение карт Hi-C. Оценка библиотек Hi-C согласно стандартам ENCODE позволяет говорить о хорошем качестве проведения эксперимента Hi-C и возможности использования его результатов для сборки генома. Полученные выравнивания были поданы для обработки свободно доступному программному обеспечению 3D-DNA. Для хромосомы 3L был обнаружен особый паттерн контактов, сходный с паттерном локальных полиморфизмов на других хромосомах, однако он обладал рядом исключительных особенностей. Во-первых, паттерн локальных полиморфизмов характерен для регионов с низким покрытием прочтениями Hi-C, в то время как паттерн хромосомы 3L характерен для регионов с

высоким покрытием. Во-вторых, локальные полиморфизмом покрывают не более 20 % генома, в то время как паттерн хромосомы 3L покрывает 90-95 % прочтений, входящих в хромосому 3L. Из литературных данных известно, что хромосома 3L у *An. quadrimaculatus* представлена в двух альтернативных вариантах, отличающихся массой перестроек. Мы предположили, что данный паттерн контактов (далее – паттерн глобальной глобального полиморфизма) наблюдается для прочтений, относящихся к разным вариантам хромосомы 3L (именуемых далее 3L1 и 3L2). Стоит отметить, что из-за особенностей организации хромосом 3R и 3L у *An. quadrimaculatus* оказалось невозможно собрать эти хромосомы в полном объёме. Хромосома 3R представлена 5 фрагментами. Фрагмент 3Ri – инвариантный длиной ~54 млн. п.о. захватывает большую часть хромосомы от теломеры к центромере. Фрагмент 3Ra представлен двумя альтернативными вариантами: 3Ra1 (~1.9 млн. п.о.) и 3Ra2 (~1.65 млн. п.о.). Данные фрагменты показывают сходство около 99 % и главным образом различаются делецией в районе 1.2-1.5млн. п.о, в координатах 3Ra1. Фрагмент 3Rb непосредственно примыкает к центромере и представлен также двумя альтернативными вариантами: 3Rb1 (1.1 млн. п.о) и 3Rb2 (1.08 млн. п.о.), показывающие сходство в 95%. Установить взаимосвязь между разными вариантами 3Ra и 3Rb не удалось из-за крайне незначительных отличий в числе контактов. 3L хромосома была разделена на два альтернативных варианта: 3L1 (~40.8 млн. п.о.) и 3L2 (38.1 млн. п.о.). Эти варианты хромосомы 3L демонстрируют 96% сходства при взаимном выравнивании. На текущий момент получены сборки для хромосом X, 2R, 2L и инвариантного участка хромосомы 3R. Для хромосомы 3L и переменная части хромосомы 3R дальнейшее улучшение сборки с использованием прочтений Hi-C не представляется возможным. Тем не менее, результаты построения карт Hi-C демонстрируют, что, не смотря на все трудности, проведённым нами манипуляции не привели к существенному искажению геномных расстояний и позволили сохранить большую часть генома.

2. Разработка цитогенетической карты политенных хромосом клеток слюнных желез *An. messeae*.

В качестве материала при разработке хромосомных карт высокого разрешения использовали слюнные железы личинок 4-го возраста и трофоциты яичников имаго *An. messeae* s. l., отловленных в природных популяциях Томской области. Видовую принадлежность личинок и имаго определяли с помощью анализа полиморфизма длин рестрикционных фрагментов региона ITS2. Сразу после приготовления препараты микрофотографировали и получали микрофотографии политенных хромосом. Препараты хромосом микрофотографировали с помощью AxioImager A1 (Carl Zeiss) в фазовом контрасте с объективом 100x. Изображения хромосом высокого разрешения были получены с помощью программного обеспечения AxioVision 4.9.1 (Carl Zeiss), а затем выпрямлены и обрезаны по методике, разработанной ранее для создания цитогенетических карт малярийных комаров. Для обработки изображений хромосом и конструирования карт использовали программы Helicon focus lite и Adobe Photoshop CC 2019. В ходе работы приготовлено более 30 препаратов неокрашенных политенных хромосом слюнных желез и более 20 препаратов политенных хромосом трофоцитов яичников. Получено 130 изображений хромосом слюнных желез и

35 изображений политенных хромосом трофоцитов яичников высокого качества, которые далее использованы для создания карт. Была создана карта хромосом слюнных желез *An. messeae*, которая включающая наиболее распространенные в инверсионные генотипы: XL11, 2R11, 2L00, 3R00, 3L00. Для трофоцитов яичников *An. messeae*, разработаны карты хромосомных плеч 2R11, 2L00, а также фрагмента 3L00.

3. Физическое картирование точек разрывов фиксированных перестроек и полиморфной инверсий в X хромосоме *Anopheles messeae*.

В данной работе проводилось геномное картирование и с описанием окрестностей точек разрывов фиксированных перестроек и полиморфной инверсий X2 в X хромосоме *An. messeae* с использованием в качестве референсного геном *An. atroparvus* (AatrE3). С помощью физического картирования ортологов генов *An. atroparvus* на цитогенетической карте *An. messeae* и проводилось уточнение координат точек разрывов фиксированных перестроек и полиморфной инверсии *An. messeae* на геномной карте *An. atroparvus*.

1) Фиксированные перестройки в X хромосоме *An. messeae* – 2 вложенные одна в другую инверсии.

X хромосомы трех близкородственных видов *An. atroparvus*, *An. labranchiae* и *An. maculipennis* не несут крупных (протяжённостью более 1 млн. п.н.) фиксированных хромосомных перестроек. В ходе настоящей работы было картировано 39 генов для определения точек разрывов фиксированных хромосомных перестроек у *An. messeae* в сравнении с порядком генов у *An. atroparvus*. Были картированы гены, фланкирующие четыре точки разрыва фиксированных хромосомных перестроек в X хромосоме *An. messeae* на геномной карте *An. atroparvus* с разрешением от 7 до 15 тыс. п.н. Анализ перестроек синтенных блоков ограниченных этими точками разрыва с учётом их положения и ориентации у *An. atroparvus* и *An. messeae* в программе GRIMM показало, что межвидовые различия в порядке генов в X хромосомах у этих видов могли быть вызваны двумя вложенными одна в другую хромосомными инверсиями. Эта хромосомная перестройка приводит к изменению локализации и ориентации двух районов хромосом протяжённостью 500 тыс. п. н. и около 2 млн. п.н.

2) В окрестностях точек разрывов фиксированных хромосомных перестроек X хромосомы находятся мобильные генетические элементы.

С помощью базы данных Gene Ontology была проанализирована онтология генов находящихся в окрестностях четырёх точек разрывов инверсий в окне около 100-200 тыс. п.н. Проанализированные данные демонстрируют, что точки разрывов находятся в окружении генов «домашнего хозяйства». Вблизи точек разрывов удалось обнаружить мобильные генетические элементы принадлежащих классам Gypsy, R1, Tc-Mariner, Helitron и некоторым другим. Подтверждением уникальности события, в результате которого возникла одна из двух фиксированных инверсий, может быть принадлежность МГЭ в окрестностях обеих точек разрывов одному классу - Gypsy.

3) Окрестности точек разрывов фиксированных инверсий в X хромосоме *An. messeae* находятся в одной группе сцепления у других видов малярийных комаров.

С помощью инструментов базы VectorBase было проанализировано: возникают ли разрывы в точках, фланкированных ортологами (обнаруженных нами) генов или их окрестностей, у других видов малярийных комаров. При анализе групп сцепления генов вблизи точек разрывов *An. messeae* и *An. atroparvus* у *An. albimanus*, *An. arabiensis*, *An. coluzzii*, *An. funestus*, *An. gambiae*, *An. merus*, *An. sinensis*, *An. stephensi* не было выявлено случаев, когда те же гены находились бы в других группах сцепления (синтенных блоках).

4) Одна из точек разрыва полиморфной инверсии X2 *An. messeae* расположена в окрестностях точки разрыва одной из фиксированных хромосомных перестроек.

Было физически картировано 11 генов с целью определить гены, фланкирующие точки разрыва полиморфной инверсии X2 у *An. messeae*. Картирование проведено с разрешением около 15 тыс. п.н. для первой второй точек. Обнаружено, что вторая точка разрыва полиморфной инверсии фланкируется одним из генов, что и одна из фиксированных инверсий в X хромосоме. Таким образом, обе точки разрыва расположены предположительно в диапазоне около 15 тыс. п.н. Первая же точка разрыва полиморфной инверсии локализуется в пределах одного из двух участков, положение которого меняется в ходе фиксированной перестройки X хромосомы *An. messeae*. Таким образом, положение точек разрыва трех хромосомных инверсий, которые сопровождали эволюцию X хромосомы *An. messeae* укладывается в диапазон не более 2.5 млн. п.н., что составляет немногим более 10 % длины.