

Эпидемия вируса Эболы в Западной Африке является крупнейшей из зарегистрированных. Она унесла жизни 11 299 человек из более чем 28 599 заболевших. Секвенирование генома во время вспышек вирусных заболеваний очень важно для характеристики инфекционного агента и наблюдения за его эволюцией. Секвенирование генома также позволяет идентифицировать механизмы адаптации организма к вторжению вируса, провести мониторинг диагностических мероприятий, характеризовать влияние вакцины и лечения. Скорость эволюции вируса Эболы (EBOV) штамма Макона была оценена в пределах  $0.87 \times 10^{-3}$  и  $1.42 \times 10^{-3}$  мутаций на сайт в год, что эквивалентно 16-27 мутациям в каждом геноме. Это означает, что последовательности меняются достаточно быстро, чтобы отчетливо выделить субпопуляции вируса во время эпидемии. Секвенирование генома позволяет наблюдать в высоком разрешении эволюцию патогена. Данные изучения последовательности могут быть использованы для контроля за принимаемыми мерами, но только в том случае, если результаты секвенирования поступают достаточно быстро, чтобы успевать проводить вмешательство. Ранее наблюдение за геномом во время эпидемии было не достаточно эффективным из-за отсутствия возможности провести анализ на месте или из-за сложностей с транспортировкой образцов к стационарным секвенаторам. Для решения этой проблемы мы разработали систему секвенирования ДНК. В апреле 2015 года эта система была доставлена в ручной клади в Гвинею, где была использована для исследования генома в режиме реального времени во время эпидемии. Мы представляем данные секвенирования и анализ 142 образцов EBOV, собранных в период с марта по октябрь 2015. Мы получали результат в виде последовательности менее чем через 24 часа после получения образцов, положительных на наличие вируса Эболы, а процесс секвенирования занимал 15-60 минут. Мы продемонстрировали, что исследование генома в реальном времени возможно в условиях с ограниченными возможностями для установки лаборатории и может быть использован для мониторинга вспышек заболеваемости.

Классические технологии исследования генома достаточно сложно развернуть в развивающихся странах, где этому мешает отсутствие соответствующего оборудования, современных лабораторий и обученного персонала. К тому же некоторые инструменты для сиквенса требуют использования высокоточной оптики наличия сложного микроскопического оборудования, что требует калибровки высококвалифицированными инженерами. В настоящее время появился портативный секвенатор MinION (Oxford Nanopore Technologies, Oxford, UK), который весит меньше 100 г. Данные считываются с MinION на ноутбук с помощью USB порта, через который секвенатор также и подзаряжается. MinION регистрирует изменения электрического тока в процессе прохождению одноцепочечной ДНК через белковые нанопоры со скоростью 30 оснований в секунду. Поскольку MinION идентифицирует отдельные молекулы, у него существенно более высокий коэффициент ошибок (10-20%) в сравнении с устройствами, детектирующими клоны-копии молекул ДНК. Преимущество одномолекулярного сиквенса заключается в возможности прочтения очень длинных фрагментов ДНК (50 kb или более). Чтобы создать точные последовательности, геномные области должны быть прочитаны много раз, что приводит к устранению ошибок через консенсусное усреднение. Такая система ранее использовалась для исследования бактериальных вспышек, но не для верификации вирусного генома.

Мы разработали протокол для проведения секвенирования EBOV на MinION, который включает в себя обратную транскрипцию с помощью соответствующего фермента (RT-PCR), чтобы выделить достаточное количество ДНК для исследования. Мы рассмотрели возможность и отвергли идею сиквенса тотальной РНК, так как в этом случае

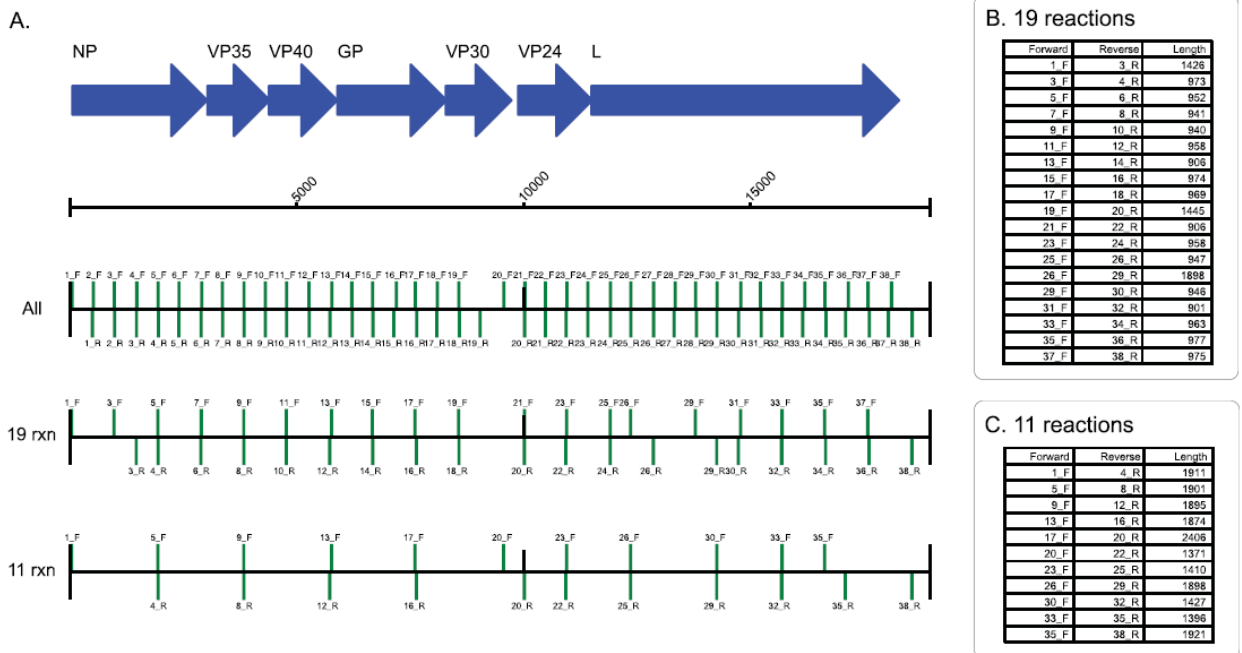
будут проанализирована еще и человеческие транскрипты, которые «заглушат» вирусный сигнал. Мы разработали панель из 38 пар праймеров, что было достаточно для исследования всего генома EBOV. В пилотных исследованиях в Defence Science and Technology Laboratory (Dstl) Porton Down, UK мы секвенировали заирский EBOV используя MinION и Illumina MiSeq. Из-за сложностей получения равного баланса каждой из 38 пар ампликонов, только 65,7% генома EBOV было прочитано 25 раз, в то же время у Illumina этот показатель был 87,4%. Однако нуклеотидные последовательности, полученные на обоих секвенаторах полностью согласуются, за исключением одного варианта в гомополимерной области. На данный момент секвенирование на MinION не позволяет определять гомополимеры длиной 5 оснований или более.

Далее мы разработали систему анализа генома, которую можно было бы отправить в Западную Африку. В общей сложности система состояла из трех MinION, четырех ноутбуков, термоцилера, с термоблоком, пипеток и необходимых реагентов для проведения. Нам удалось упаковать это в менее чем пятидесятикилограммовый стандартный багаж. (рис.1а). Сначала мы установили системы в European Mobile Laboratory в Donka Hospital в Конакри, Гвинея. (рис 1б). Затем, оборудование было перемещено в специализированную лабораторию, располагающуюся в Coyah Ebola Treatment Unit (рис 1в, г).



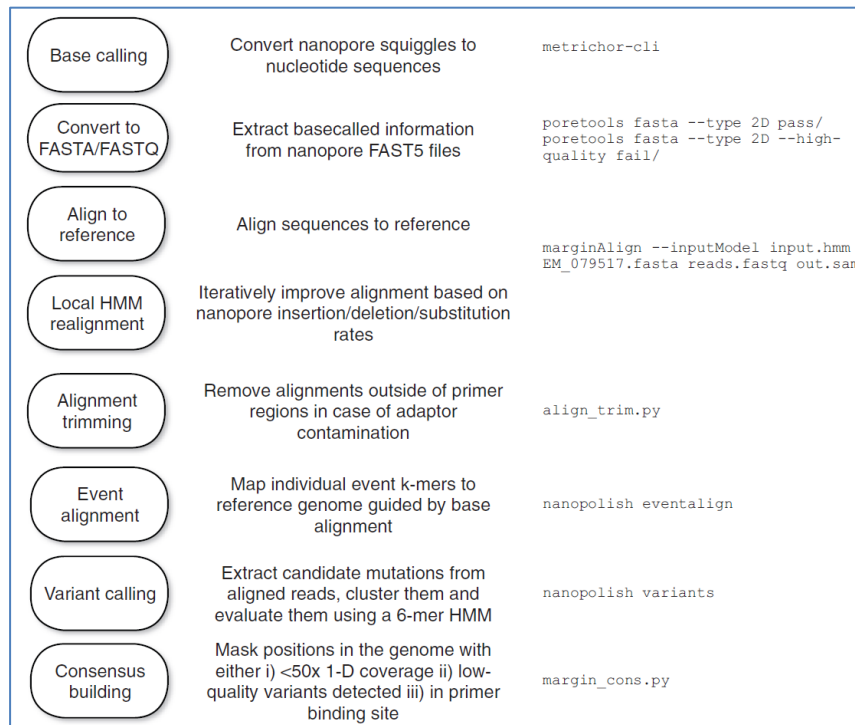
**Рисунок 1. Размещение портативной системы анализа генома вируса Эбола.**

Мы приступили к секвенированию менее чем через 2 дня после прибытия в Гвинею. На ранних этапах мы обнаружили, что нам удалось достоверно генерировать длинные ампликоны (около 2 kb в длину) используя пары праймеров в разных комбинациях (рис 2). Использование как можно большего количества ампликонов значительно облегчает усилия при подготовке образцов. Мы обнаружили комбинацию из 11 ампликонов, которые позволяли достоверно амплифицировать более 97% генома EBOV.



**Рисунок 2. Схема праймеров, использованных в исследовании.**

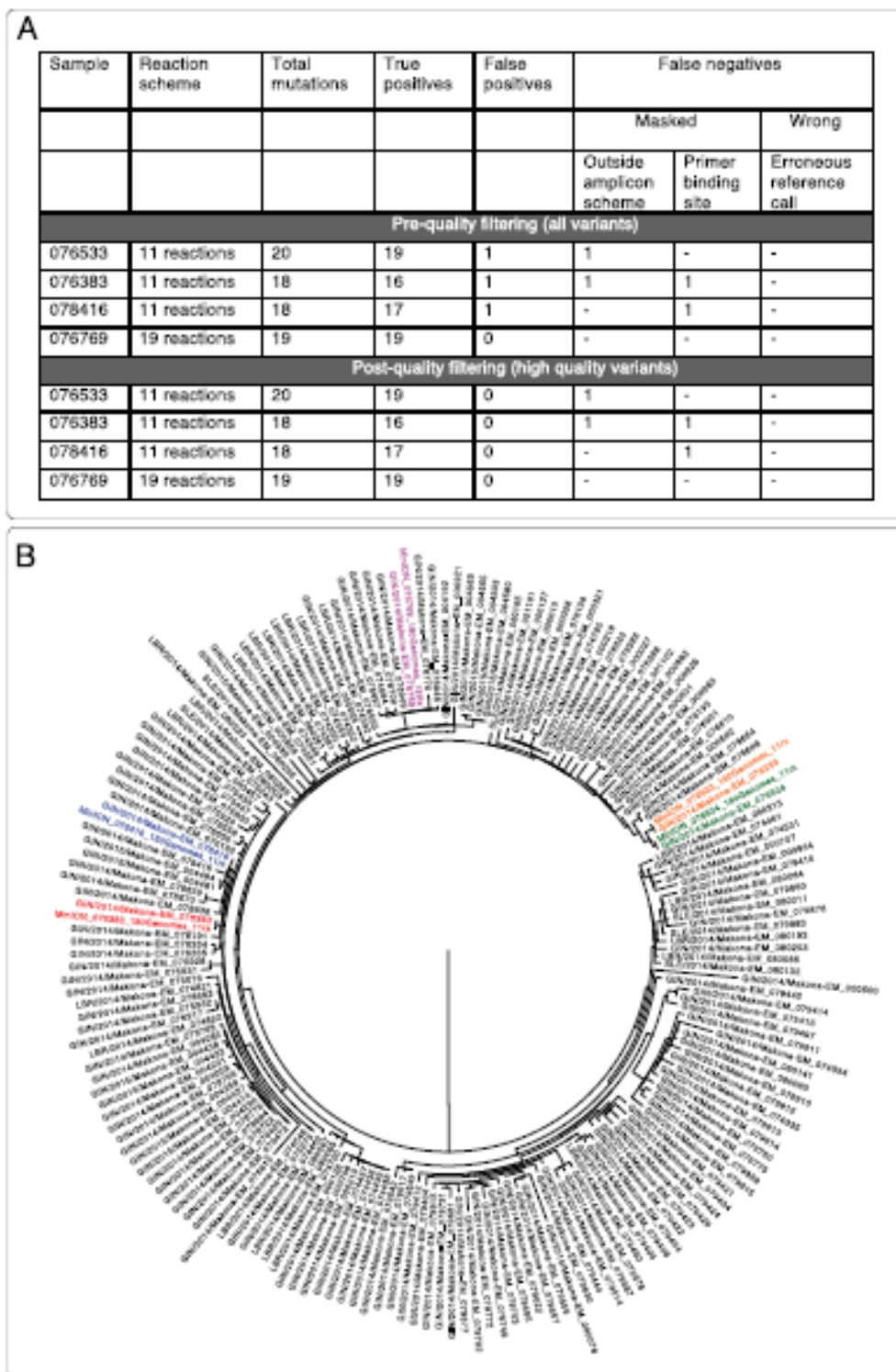
Мы разработали биоинформатический механизм, который позволял получить хороший результат генотипирования и подтвердили это, используя образцы вируса Макона из предыдущих исследований. Этот механизм подробно описан в разделе «методы» и обобщен в «расширенных данных» (рис 3). Эта проверка продемонстрировала, что данный подход к обработке данных и их анализу является надежным.



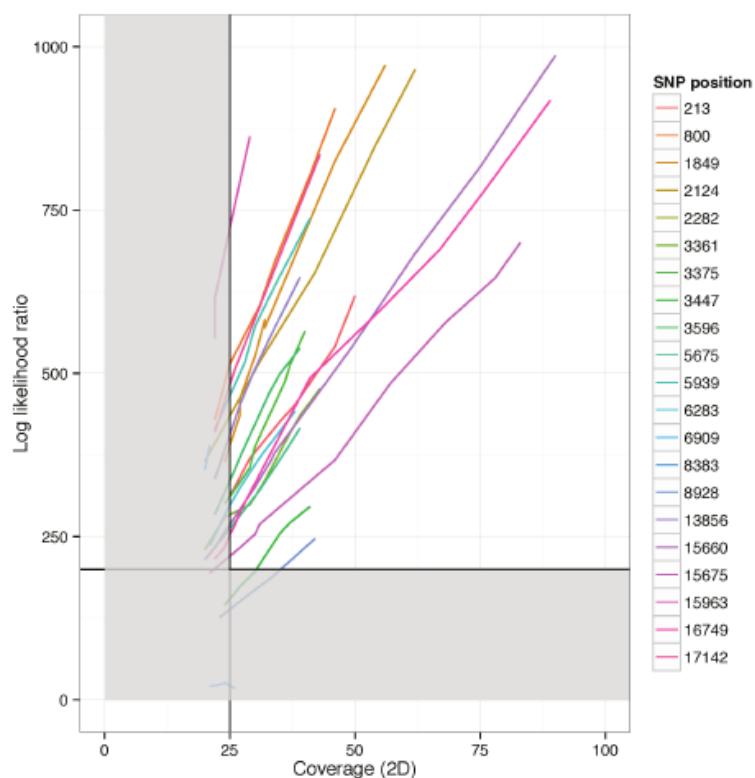
**Рисунок 3. Схема биоинформатического анализа**

Мы сравнили наши консенсусные последовательности с теми, которые были получены с использованием Pilonia и получили результаты, подтверждающие правильность нашего алгоритма. В некоторых случаях нам не удалось определить варианты, так как некоторые

из них находились в области связывания праймера или были за пределами области генома EBOV, покрываемых нашим набором ампликонов (рис 4a). Эти позиции представлены неопределенными нуклеотидами в конечной последовательности. Несмотря на эти неопределенные последовательности, филогенетический анализ показал, что образцы сгруппированы одинаково (рис 4b). Мы определили, что несмотря на высокую частоту инструментальных ошибок, использование информации от 25 прочтений последовательности позволяет точно определить генотип (рис.5).

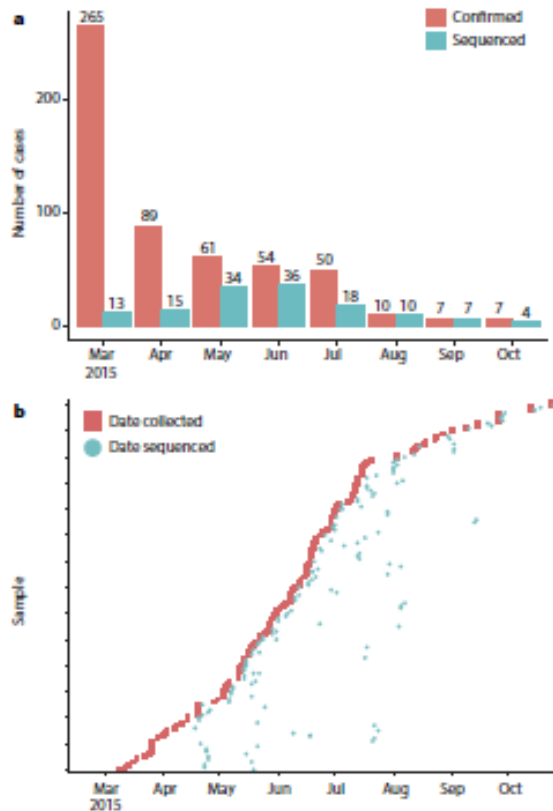


**Рисунок 4. Результаты валидации результатов, полученных на MinION**



**Рисунок 5. Зависимость достоверности результатов от количества покрытий**

После размещения системы исследования генома мы работали совместно с диагностической лабораторией в Гвинее для обеспечения отправки данных в режиме реального времени центру координации в Гвинее в ВОЗ. Сотрудничающие лаборатории предоставляли выделенную РНК для секвенирования. Работа по секвенированию генома, включающая в себя амплификацию, подготовку библиотек последовательностей и сам процесс секвенирования может быть выполнена в течении одного рабочего дня. В одном из случаев, работа, включая биоинформатический анализ, была выполнена менее, чем за 24 часа, хотя стандартное время на выполнение данного протокола – 2 рабочих дня. Мы обнаружили, что в половине случаев мы смогли сгенерировать достаточное количество прочтений на MinION (от 5000 до 10000) менее чем за час. Всего 142 образца было исследовано на MinION в течение 6-ти месяцев, что обеспечило достаточное представление о случаях вспышек (рис б). Полные детали об образцах представлены в дополнении. Нам не удалось идентифицировать ампликоны из некоторых образцов, вследствие утраты некоторых областей генома. Мы предположили, что связанные с этим трудности, вызваны низким содержанием длинных молекул в исходном образце. Мы исключили 17 образцов, в которых было больше всего погрешностей и обнаружили, что производительность системы была одинаковой во всей серии экспериментов.

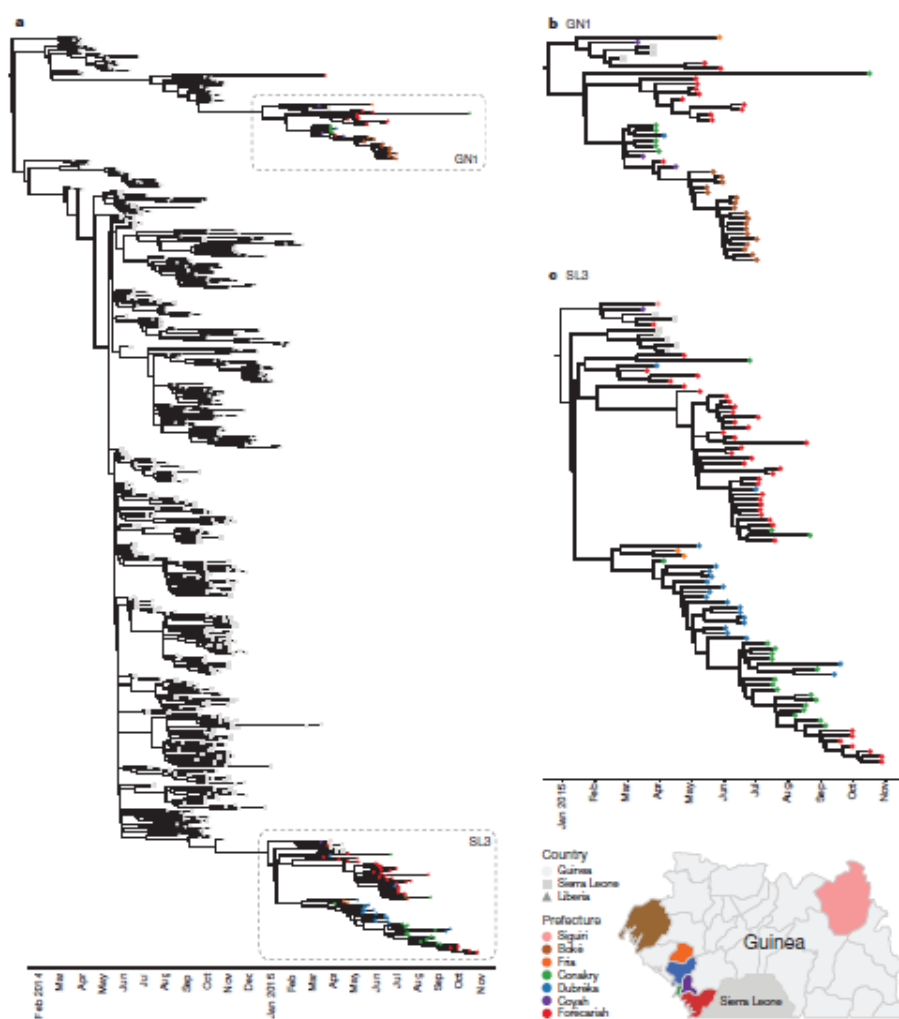


**Рисунок 6. Геномный мониторинг вспышки лихорадки Эбола в Гвинее в реальном времени.**

Мы объединили данные секвенирования, полученные в нашем исследовании еще с 603 образцами из других лабораторий и, используя программное обеспечение, вывели филогенетическое дерево (рис 7). Анализ показал хорошее согласование между данными выборки и отклонением от корня к концу (рис 8). Мы оценили скорость замены как  $1,9 \times 10^{-3}$  (95% интервал,  $1,09 \times 10^{-3}$ ,  $1,29 \times 10^{-3}$ ) в объединенном наборе (рис 8). Эти результаты сочетаются с результатами предыдущих исследований. Результаты, полученные в первые 10 дней от начала секвенирования, указывают на то, что во время вспышки заболевания в Гвинее преобладали две основные линии, названные GN1 и SL3, которые и были идентифицированы в начале эпидемии. Родословная GN1 изначально идет из Гвинеи и не встречается в Сьерра Леоне, что позволяет предположить, что эта линия в основном ограничивается присутствием в Гвинее. Вторая родословная SL3 была детектирована в Сьерра Леоне Гайром и др. (Geri et al) и позже была замечена в Корнаки в конце 2014 года. Благодаря объединению данных, полученных нами и группой, работавшей в Сьерра Леоне мы обнаружили, что и GN1 и SL3 в 2015 были замечены в Сьерра Леоне, что наводит на мысль о передаче модификаций вируса между странами.

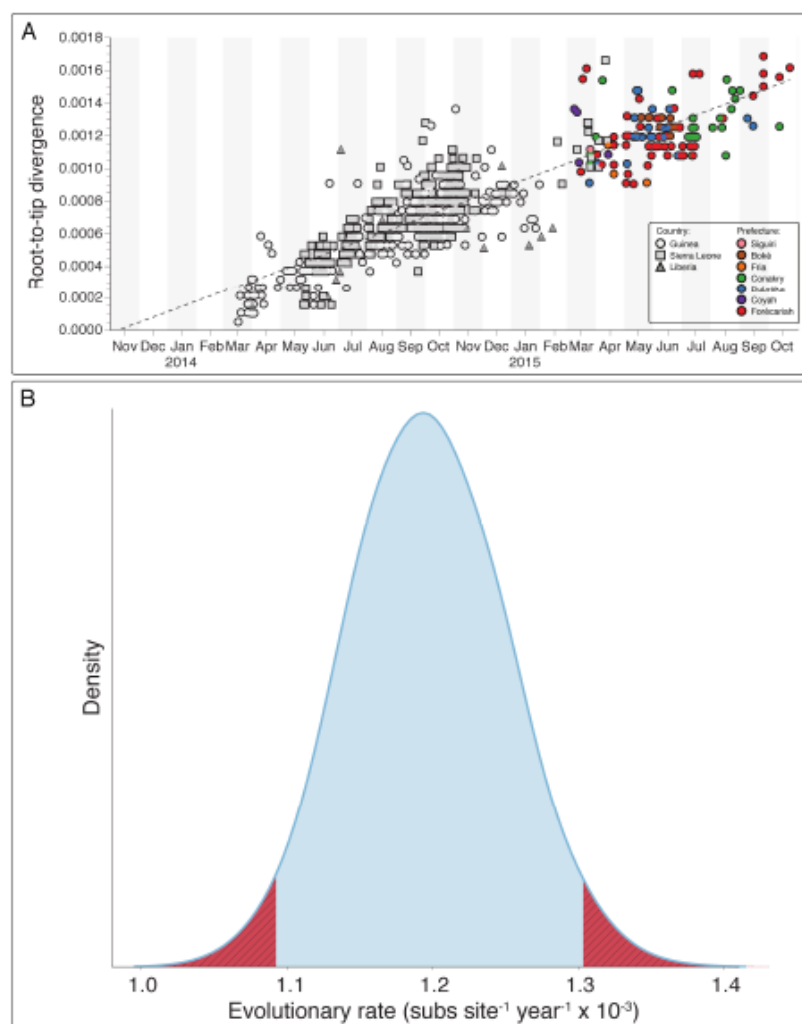
Данная работа демонстрирует постепенное изменение наших возможностей в исследовании генома во время вспышек в условиях ограниченных ресурсов. Однако на данный момент существует все еще достаточно много препятствий для рутинного исследования. На практике мы столкнулись со значительным количеством материально-технических проблем при выполнении этой работы, в частности, с отсутствием надежных систем энергообеспечения, что приводило к необходимости пользования блоками бесперебойного питания, что неудобно, особенно в связке с громоздким оборудованием для проведения ПЦР. Однако переносные термоциклеры с батарейным питанием уже в разработке. Секвенатор MinION не был подвергнут перебоям с питанием и скачкам напряжения. В данном случае мы столкнулись с перебоями с выходом в интернет, что было необходимым условием для проведения этой работы. Таким образом, существует потребность в создании полностью автономной системы, что существенно снизит

зависимость от высокоскоростных соединений. Однако вполне вероятно, что филогенетический анализ будет также проводиться удаленно. В данном случае мы сконцентрировались на разных вариантах подхода. Предпочтительным был бы анализ *de novo*, но в настоящее время это приводит к ошибкам во время прохождения через поры гомополимеров. Наш подход опирается на амплификацию генетического материала перед секвенированием. Это может быть недостатком в том случае, когда необходимо секвенировать геном во время эпидемии. Тогда можно использовать непосредственно клинический материал, однако, в этом случае будут присутствовать проблемы с чувствительностью.



**Рисунок 7. Эволюция вируса Эбола в течение вспышки заболевания.**

Секвенирование в режиме реального времени является новым инструментом в нашем арсенале методов исследования эпидемиологической обстановки, а также прогнозирования новых вспышек. Полученные результаты могут повысить эффективность распределения ресурсов. Такое исследование может пролить свет на еще не известные составляющие патогенеза заболевания. Объединяя наши исследования с коллегами из Сьерра-Леоне мы заметили свидетельства частых передач заболевания из Гвинеи. Так же мы регулярно публиковали данные наших исследований на GitHub, где обменивались с коллегами мнениями по результатам. Мы использовали web-форум для общения по поводу наиболее сложных случаев. Такая система поможет держать под контролем надзор за вспышками эпидемий в Западной Африке и может стать шаблоном для использования подобной схемы в других регионах.



**Рисунок 8. График дивергенции и оценка среднего значения скорости эволюции.**

Эпидемия Эболы была официально объявлена 14 января 2016 года, спустя некоторое время был подтвержден новый случай вспышки заболевания в Сьерра-Леоне, а также предупреждения о возможном появлении вспышек в других областях. В таком случае актуально определить источник инфицирования, в чем может помочь исследование генома. Получив эти данные легко ответить на все поставленные вопросы.

## МЕТОДЫ

**Этические аспекты:** Национальный комитет по этике медицинских исследований в Гвинее (разрешение №11 CNERS/14) одобрил исследования диагностических образцов и обработку данных пациентов. Поскольку образцы были собраны в рамках мер в области общественного здравоохранения и борьбы с эпидемией в Западной Африке, согласие на обработку данных пациентов не требовалось.

**Транспортировка:** Все оборудование было погружено в корпус Pelican 1610 (Pelican, Торранс, США), холодные реагенты были упакованы в две коробки из полистирола со льдом или охлаждающими пакетами. Они были запечатаны и помещены в контейнер с расходными материалами из пластика. Обе единицы багажа были отправлены воздушным транспортом и зарегистрированы как обычный багаж.

**РНК-экстракция:** РНК экстрагировали из 50 мкл цельной крови, 140 мкл сыворотки, 140 мкл ресуспендированного ватного тампона или 140 мкл мочи с использованием набора QIAamp Viral RNA Mini Kit (Qiagen, Manchester, UK), в соответствии с инструкциями



производителя. Образцы инактивировали добавлением 560 мкл буфера AVL (Qiagen) и 560 мкл 100% этанола, показано, что этот метод полностью инактивирует EBOV в образцах крови. После инактивации образцы обрабатывались с использованием стандартных лабораторных мер предосторожности.

**RT-PCR:** Индивидуальные реакции проводили в объеме 25 мкл с использованием SuperScript III One-Step RT-PCR с Platinum Taq ДНК-полимеразой (Life Technologies Ltd, Пейсли, Великобритания). Каждую реакционную смесь готовили добавлением 12,5 мкл 2-кратной реакционной смеси, 1 мкл ферментной смеси, 1 мкл праймеров (10 мкМ), 0,5 мкл экстрагированной РНК и воду, не содержащую нуклеаз. RT-PCR проводили на Eppendorf Master Cycler со следующей программой: 60° С в течение 30 минут, 94° С в течение 2 минут, затем 45 циклов при 94° С в течение 15 с, 55° С в течение 30 с, 68° С в течение 2 мин и конечной 68° С в течение 5 мин.

**Приготовление библиотек для MinION:** Каждая реакция была количественно определена на флуориметре Qubit 3.0 с использованием dsDNA HS (Life Technologies). Эквивалентные количества каждого ампликонного продукта до полной массы ДНК 1 мкг объединяли в одну трубку и очищали с помощью MAGBIO HighPrep PCR (AutoQ Biosciences, Reading, UK). Объединенные ампликоны разбавляли до 85 мкл, и дополняли до общего объема в 100 мкл с использованием модуля восстановления NEBNext End End (New England Biolabs, Hitchin, UK) перед очисткой с использованием равного объема HighPrep PCR и элюирования в 25 мкл воды без нуклеазы. 3'dA-хвост был присоединен с использованием модуля NEBNext dA-Tailing (New England Biolabs) в объеме 30 мкл, перед очисткой с использованием равного объема HighPrep ПЦР-капсул и элюированием в 30 мкл воды, свободной от нуклеазы. 10 мкл «переходной смеси» и 10 μl 'HP adaptor', входящий в комплект для подготовки библиотеки SQK-MAP005 (Оксфорд Nanopore Technologies, Oxford, UK) были добавлены к dA-хвостовым ампликонам вместе с 50 мкл Blunt / TA Ligase Master Mix (New England Biolabs) в белке LoBind (Eppendorf UK) и инкубированы в течение 10 мин. Получившаяся последовательность была очищена с использованием Dynabeads His-Tag Isolation и Pulldown beads (Life Technologies, Stevenage, UK) в соответствии с протоколом SQK-MAP005, предоставленным Oxford Nanopore Technologies в рамках программы доступа MinION. Итоговая библиотека была количественно определена с помощью Qubit, чтобы подтвердить, что процесс прошел успешно. 6 мкл полученного раствора разводили с использованием 75 мкл «2x Running Buffer», 66 мкл воды, очищенной от нуклеазы (Promega UK, Chilworth, UK) и 3 мкл, «Fuel Mix».

**Секвенирование на MinION:** Новая проточная ячейка была распакована и установлена на MinION. Проточная ячейка была заполнена пустым образцом, созданным, как описано выше и оставлена для инкубации в течение 10 мин. Процесс заполнения повторяли второй раз перед загрузкой образца. Запуск версии MinKNOW 0.49.2.9 и запуск протокола «MAP\_48Hr\_Sequencing\_Run.py» инициировал запуск программы. Автономная версия MinKNOW, без подключения к сети интернет и отключенным интернет-обновлением была предоставлена нам Oxford Nanopore Technologies специально для проекта (предоставляется по запросу от Oxford Nanopore Technologies).

**Передача данных:** В условиях отсутствия автономного анализа, доступного во время вспышки заболевания, существует зависимость от локального интернет-соединения. Разнообразные методы использовались в зависимости от местоположения и обстоятельств. Подавляющее большинство данных, загружалось из места работы на точки мобильных операторов в Европе в Койе, Гвинея, через точку доступа мобильного интернета 3G (TP-LINK M5350 3G hotspot в мобильной сети MTN). Время от времени из-за неизвестных факторов скорость загрузки была ограничена до 2G и загрузка занимала значительно больше времени. Использование Cygwin версии 2.0.0 и Linux tar со сжатым архивом, содержащий первые 5 000 - 10 000. Fast5 считывает файлы, сгенерированные каждым прогоном. Все эти данные были загружены на Google Диск в общий каталог. В

конечном счете в Койе нам был предоставлен доступ к широкополосной сети соединения (сеть MTN, установленная Мировой продовольственной программой), которая оказалась более надежной, чем мобильный интернет.

**Обработка данных:** Данные были загружены на Linux-сервер в облаке MRC. Инфраструктура для микробиологической биоинформатики, расположена в Бирмингеме, Великобритания. Файлы были распакованы с использованием интерфейса командной строки Metrichor и 2D Basecalling для MAP-005 (версии 1.14, 1.24 и 1.34). Такое программное обеспечение было предоставлено Oxford Nanopore Technologies (доступно по запросу) для проекта, чтобы позволить базовое выполнение входа через Linux. MinION позволяет осуществить чтение в одном (1D) и двух направлениях (2D). 2D-чтение является более качественными и именно этот формат использовался для анализа.

**Биоинформатический анализ:** Мы используем метод ссылочного сопоставления (reference mapping approach) для обнаружения однонуклеотидных вариантов путем сопоставления с эталонным штаммом, идентифицированным ранее (номер для поиска GenBank EM\_079517). Варианты были определены с использованием модуля вариантов программного пакета. Начальное выравнивание нуклеотидного потока было осуществлено с помощью ПО MarginAlign. Затем Nanopolish использует данные уровня событий (squiggle), созданные MinION для оценки возможных вариантов, найденных в выровненных прочтениях, как было описано в предыдущем разделе. Варианты с логарифмическим коэффициентом вероятности  $> 200$  и глубиной покрытия  $> 50 \times$  (охват  $25 \times 2D$ ), принимались как достоверные для каждого была создана консенсусная последовательность. Регионы неопределенности (например, в труднодоступных гомополимерных областях или сайтах связывания праймеров) или с низким охватом ( $< 50 \times$  или  $25 \times 2D$  охват) маскировали символом N.

**SNP сигнал:** Однонуклеотидные мутации выявляли с помощью специализированного пакета «variants» (рукопись в подготовке: <https://github.com/jts/nanopolish>, branch snp\_calling\_alternative\_models, commit ID 25ea7bac3ab9e1d266079ac105ab2005cfa39a14).

Программное обеспечение сначала находит кандидатные позиции на мутации SNP, обнаруживая несоответствие между результатом прохождения через нанопоры и контрольным образцом. Мутации комбинируются в группы и из них выделяется исчерпывающий набор кандидатов, выведенный из возможных комбинаций SNP и гаплотипа, который максимизирует вероятность уровня событий, характерных для данного региона последовательности. Ниже мы подробно описываем каждый шаг.

**Выбор SNP-кандидата:** Мы обрабатываем весь референсный геном и наблюдаем позиции с не менее чем 20 прочтениями через нанопоры. На этих, хорошо проверенных позициях, мы рассматривали любые основания, не представленные в референсном геноме, присутствующие не менее чем в 20% прочтений, и рассматривали их как кандидаты в SNP. Эти кандидаты далее переходят на следующий этап.

**Отбор гаплотипов:** Поскольку секвенатор MinION не идентифицирует одиночные основания, а детектирует сигналы, зависящие от короткой последовательности нуклеотидов, находящихся в поре, мы не могли оценить каждую SNP индивидуально. Чтобы решить этот вопрос мы разделили набор потенциальных SNP на группы, сигналы которых могут взаимодействовать и перекрываться. Мы определили, что SNP, разделенные по меньшей мере на 10 пар оснований, могут обрабатываться независимо; поэтому мы разделили набор SNP кандидатов на подмножества SNP, которые составляют 10 п.о. Для каждого подмножества кандидатов SNP мы генерировали все возможные последовательности гаплотипов, включая и исключая отдельные SNP в подмножестве. Поскольку число возможных комбинаций  $n$  SNPs равно  $2^n$ , нам пришлось отказаться от подмножеств, содержащих более 10 кандидатов SNP, или охватить референтную область больше 100 пар оснований. Для каждой производной последовательности гаплотипов S

мы вычисляем вероятность  $S$ , используя модифицированную версию марковской модели (НММ).

**Вероятность гаплотипа:** Программное обеспечение Марковской модели рассчитывает вероятность наблюдения последовательности событий, детектируемых при прохождении через нанопоры, которую мы обозначаем как  $D$ , учитывая произвольную последовательность  $S$ . Структура НММ такая же, как описано ранее, но теперь позволяет событиям быть «soft-clipped», чтобы лучше справляться с неопределенностью относительно того, где начинается и заканчивается выравнивание событий в последовательности. Кроме того, мы включили новую модель из оксфордского программного обеспечения, которая моделирует сигналы событий, которые будут зависеть от подпоследовательностей из шести пар оснований, а не пяти. Чтобы использовать эту модель для данных SQK-MAP-005, мы вычислили глобальный параметр сдвига (`shift_offset`), который повторяет данные SQK-MAP-005 в функции излучения при шести парах оснований.

**Варианты сигналов:** SNP, содержащиеся в рассматриваемом гаплотипе (если они есть), выводятся в формате VCF. Отношение правдоподобия логарифма между полученным гаплотипом и эталонным гаплотипом (без SNP) выводилось как оценка для каждого варианта для облегчения фильтрации по нисходящему потоку. Также выводятся промежуточные данные, такие как общая глубина области и количество считываний, которые поддерживают названный гаплотип по эталонной последовательности.

**Эксперименты по валидации:** Dstl amplicons. Архивный вирус Эбола Заир был амплифицирован с использованием 38 пар праймеров и давал, согласно протоколу исследования ампликоны размером приблизительно 500 пар оснований. Поскольку эта работа выполнялась до секвенирования в поле, были использованы разные версии программного обеспечения MinKNOW и базового инструмента Metrichor. Ампликоны были секвенированы как на MinION, так и на Illumina. Была разработана библиотека Illumina из того же пула ампликонов с использованием набора для подготовки библиотеки Nextera XT. Библиотека была секвенирована на Illumina MiSeq. Из-за огромных объемов охвата, этот набор данных был сокращен до 400 000 парных прочтений перед выравниванием с эталонной последовательностью EM\_079517 с использованием BWA-MEM. После сортировки и преобразования полученного выравнивания в BAM с использованием samtools, варианты определялись с помощью FreeBayes. Консенсусная последовательность была сгенерирована с использованием компонента vcf2fasta vcflib (<https://github.com/ekg/vcflib>). Данные с MinION анализировались в соответствии с методами исследования, за исключением модификации программного обеспечения, чтобы позволить ей рассматривать до 15 вариантов на сегмент. Результаты с MinION

и последовательные конспекты Illumina были выровнены с использованием нуклонного компонента MUMMER и вариантов, определенных с помощью модуля show-snps24. Скрипты и документация для этого анализа представлены в `Gitub-laptop Dstl validation.ipynb`. Исследованные образцы остаточной РНК из ранее выполненного исследования были обработаны в Public Health England Porton Down, по методам, описанным в рукописи. Один образец не дал результатов, так что пять геномов (EM\_076534, EM\_076533, EM\_076383, EM\_078416, EM\_076769) были упорядочены на MinION в PHE Porton Down. Использовали 11-реакционную схему, за исключением образца EM\_076769, когда использовалась 19-реакционная схема. Эти последовательности сравнивали с результатом Illumina из ранее опубликованных данных, установленных в Carroll et al. Варианты были сопоставлены с эталонным геномом (EM\_079517) и каждый из успешно выбранных добавлен show-snps из MUMMER. Определенные варианты по нашей программе были сопоставлены с ожидаемыми вариантами до и после фильтрации с использованием пользовательских программ Python, заложенных в репозитории Github и задокументированные в IPython Notebook. Филогенез был выведен с использованием RaXML включая консенсусные последовательности из

набора валидаций, а также всех консенсусных последовательностей от Carroll et al. Показатели точности последовательности с MinION для двунаправленного (2D) чтения были определены с использованием счетчика A. Quinlan count-errors.py скрипт (<http://github.com/arq5x/nanopore-scripts>), описан в Quick et al. Сценарии и документация для этого анализа приведены Github.

**Анализ чувствительности SNP-сигнала:** прочтения были поделены на субобразцы по интервалам времени отбора с помощью poretools times command, с симуляцией порядка прочтений, полученного с помощью секвенирования в реальном времени на нанопоровой единице, чтобы продемонстрировать эффект покрытия SNP-сигнала и логарифмического коэффициента вероятности.

**Анализ образцов одного и того же пациента.** Образцы были проанализированы как часть наблюдения в режиме реального времени. Были созданы консенсусные последовательности из четырех пар образцов. Каждая пара сравнивалась индивидуально с использованием модуля show-snps MUMmer для исследования различий.

**Обнаружение предполагаемых путей передачи из Сьерра-Леоне:** Мы скачали 74 геномных последовательности, доступных на <http://virological.org> (<http://virological.org/t/direct-deep-sequencing-in-sierra-leone-yields-73-new-ebov-genomes-fromfebruary-may-2015/134>) и выровняли их по последовательностям из нашего анализа с использованием MUSCLE. Затем мы создали филогенетическое дерево, используя FastTree 2 и модели GTR. Любые последовательности, попавшие в линии GN1 или SL3, были включены в анализ.

**Филогенетические выводы:** Консенсусные последовательности из секвенирования в режиме реального времени были согласованы с ранее опубликованными последовательностями генома из Гвинеи, Сьерра-Леоне и Либерии. Чтобы устранить чрезмерное количество образцов из Сьерра-Леоне в этом наборе, мы случайным образом сбрасывали выборку доступных образцов, в результате чего в общей сложности было проанализировано 313 последовательностей из Сьерра-Леоне. Филогенетические деревья выполнялись с использованием RAxML 8.2.3 с использованием модели GTRGAMMA и 200 копий бутстрапа. Деревья с временными шкалами были получены с помощью BEAST v1.8.2 с использованием модели замещения гамма НКУ +, разделенной позициями первого, второго и третьего кодонов и межгенных областей, дерева Skygrid и некоррелированными логнормальными часами, а также неинформативным ранее по среднему значению молекулярной тактовой частоты (XML-файл в Github). Филогенетические деревья были созданы с использованием пакета ete3 Python.

**Доступность кода:** Воспроизводимые рабочие процессы для представленного здесь анализа и консенсусные последовательности можно найти по ссылке <http://github.com/nickloman/ebovand> доступной по лицензии MIT. Полный набор биоинформатических скриптов доступен в Github с соответствующими программами IPython для работы с цифрами и таблицами. Их можно найти по ссылке <http://github.com/nickloman/ebov>.