

DOI: 10.21055/0370-1069-2024-1-176-181

УДК 616.98:579.842.23

А.В. Федоров, Я.М. Краснов, Е.А. Нарышкина, Е.А. Соседова, А.Д. Катышев, Г.А. Ерошенко,
Л.М. Куклева, Н.А. Осина, В.В. Кутырев

Гибридная сборка полных геномов штаммов *Yersinia pestis*

ФКУН «Российский научно-исследовательский противочумный институт «Микроб», Саратов, Российская Федерация

Цель исследования – сборка полноразмерных нуклеотидных последовательностей хромосомы и плазмид для 13 штаммов *Yersinia pestis* из 11 природных очагов чумы, находящихся на территории Российской Федерации, используя данные двух технологий секвенирования. **Материалы и методы.** Штаммы *Y. pestis* выращивали на агаре Хоттингера (pH 7,2) при 37 °C. Выделение ДНК проводили методом фенол-хлороформной экстракции. Для генетического анализатора MinIon (Oxford Nanopore) подготовку ДНК-фрагментов проводили методом лигирования по модифицированному протоколу. Для генетического анализатора Ion S5 (IonTorrent) подготовку образцов проводили по стандартному протоколу получения библиотеки с размером фрагментов ДНК 400 пар нуклеотидов (п.н.). Полученные единичные прочтения отфильтровывались по среднему качеству Q30 для IonTorrent и Q7 для Oxford Nanopore. **Результаты и обсуждение.** Проведена подготовка фрагментов ДНК, содержащих 50000 и более пар нуклеотидов, для последующего секвенирования с использованием технологии секвенирования через нанопоры (Oxford Nanopore). Использован алгоритм TruSeq для гибридной сборки генома штаммов *Y. pestis* и коррекции возникающих при этом процессе ошибок, позволяющий собрать полноразмерные нуклеотидные последовательности хромосомы и плазмид для каждого генома штамма. В международную генетическую базу данных NCBI GenBank депонированы нуклеотидные последовательности хромосом геномов 13 штаммов *Y. pestis* из 11 природных очагов чумы, находящихся на территории Российской Федерации. Установлено, что для сборки полноразмерных геномов штаммов *Y. pestis* необходимо значительное количество прочтений размером 50000 п.н. и более, а использование алгоритма TruSeq позволяет получить более точную сборку полных геномов бактерий.

Ключевые слова: природные очаги чумы на территории Российской Федерации, гибридная сборка бактериальных геномов, алгоритм TruSeq, Oxford Nanopore, IonTorrent.

Корреспондирующий автор: Краснов Ярослав Михайлович, e-mail: rusrapi@microbe.ru.

Для цитирования: Федоров А.В., Краснов Я.М., Нарышкина Е.А., Соседова Е.А., Катышев А.Д., Ерошенко Г.А., Куклева Л.М., Осина Н.А., Кутырев В.В. Гибридная сборка полных геномов штаммов *Yersinia pestis*. Проблемы особо опасных инфекций. 2024; 1:176–181. DOI: 10.21055/0370-1069-2024-1-176-181
Поступила 13.10.2023. Отправлена на доработку 30.10.2023. Принята к публ. 06.12.2023.

A.V. Fedorov, Ya.M. Krasnov, E.A. Naryshkina, E.A. Sosodova, A.D. Katyshev, G.A. Eroshenko,
L.M. Kukleva, N.A. Osina, V.V. Kutyrev

Hybrid Assembly of Complete Genomes of *Yersinia pestis* Strains

Russian Research Anti-Plague Institute “Microbe”, Saratov, Russian Federation

Abstract. The aim of the study was to assemble full-length nucleotide sequences of the chromosome and plasmids for 13 *Yersinia pestis* strains from 11 natural plague foci located in the Russian Federation, using data from two sequencing technologies. **Materials and methods.** *Y. pestis* strains were grown on Hottinger’s agar (pH 7.2) at 37 °C. DNA was isolated using phenol-chloroform extraction. For the MinIon genetic analyzer (Oxford Nanopore), DNA fragments were prepared by ligation according to a modified protocol. For the Ion S5 genetic analyzer (IonTorrent), sample preparation was carried out according to the standard protocol for obtaining a library with a DNA fragment size of 400 nucleotide pairs (bp). The resulting single reads were filtered by average quality Q30 for IonTorrent and Q7 for Oxford Nanopore. **Results and discussion.** DNA fragments containing 50 000 or more nucleotide pairs were prepared for subsequent sequencing using nanopore sequencing technology (Oxford Nanopore). The TruSeq algorithm was applied for hybrid assembly of the genome of *Y. pestis* strains and correction of errors arising during this process, allowing the obtainment of full-length nucleotide sequences of the chromosome and plasmids for each genome of the strain. The nucleotide sequences of the chromosomes of 13 *Y. pestis* strains from 11 natural foci of plague located in the Russian Federation have been deposited in the international genetic database NCBI GenBank. It has been established that to assemble full-length genomes of *Y. pestis* strains, a significant number of reads with a size of 50 000 bp or more is required, and the use of the TruSeq algorithm allows one to generate a more accurate assembly of complete bacterial genomes.

Key words: natural foci of plague in the Russian Federation, hybrid assembly of bacterial genomes, TruSeq algorithm, Oxford Nanopore, IonTorrent.

Conflict of interest: The authors declare no conflict of interest.

Funding: The authors declare no additional financial support for this study.

Corresponding author: Yaroslav M. Krasnov, e-mail: rusrapi@microbe.ru.

Citation: Fedorov A.V., Krasnov Ya.M., Naryshkina E.A., Sosodova E.A., Katyshev A.D., Eroshenko G.A., Kukleva L.M., Osina N.A., Kutyrev V.V. Hybrid Assembly of Complete Genomes of *Yersinia pestis* Strains. *Problemy Osobo Opasnykh Infektsii* [Problems of Particularly Dangerous Infections]. 2024; 1:176–181. (In Russian). DOI: 10.21055/0370-1069-2024-1-176-181

Received 13.10.2023. Revised 30.10.2023. Accepted 06.12.2023.

Fedorov A.V., ORCID: <https://orcid.org/0000-0001-7190-4427>
Krasnov Ya.M., ORCID: <https://orcid.org/0000-0002-4909-2394>
Naryshkina E.A., ORCID: <https://orcid.org/0000-0002-9190-099X>
Katyshev A.D., ORCID: <https://orcid.org/0000-0002-8260-4670>

Eroshenko G.A., ORCID: <https://orcid.org/0000-0001-5403-989X>
Kukleva L.M., ORCID: <https://orcid.org/0000-0003-2438-8364>
Osina N.A., ORCID: <https://orcid.org/0000-0003-0954-5683>
Kutyrev V.V., ORCID: <https://orcid.org/0000-0003-3788-3452>

В отличие от небольших геномов вирусов, сборка целого генома бактерий остается сложной задачей ввиду присутствия в геноме множества идентичных протяженных последовательностей. Большинство наиболее распространенных технологий секвенирования имеют значительные ограничения на максимально возможную длину определения фрагментов ДНК, а именно до 600 пар нуклеотидов (п.н.) [1]. Результатом сборки генома бактерий из единичных прочтений не более 600 п.н. является так называемый «черновой геном». В таком геноме нуклеотидная последовательность хромосомы представлена в виде набора из многих десятков/сотен ее крупных фрагментов (контигов), порядок которых в самой структуре хромосомы неизвестен. В ряде случаев сборка «чернового генома» вполне достаточна для решения задач, связанных с анализом особенностей исследуемых бактериальных геномов, однако она не отражает полной структуры генома и не позволяет обнаружить крупные перестановки. Решение этой проблемы возможно только при наличии достаточного количества единичных прочтений, превышающих своей протяженностью любую из повторяющихся областей нуклеотидной последовательности генома. Из современных технологий секвенирования только две из них способны определять фрагменты ДНК размером в десятки тысяч нуклеотидов без их предварительной фрагментации. Особенно перспективной в этом направлении выглядит технология секвенирования ДНК через нанопоры (Oxford Nanopore), которая, в отличие от остальных технологий, не имеет принципиального ограничения по протяженности секвенируемых фрагментов ДНК [2]. Однако ввиду невысокой точности секвенирования с помощью технологии нанопор требуется еще дополнительное секвенирование ДНК исследуемого бактериального генома с применением более точной технологии. Использование данных секвенирования одного образца, полученных с помощью технологий Oxford Nanopore и, например, Illumina или IonTorrent, позволяет провести гибридную сборку, результатом которой будет целый бактериальный геном с минимальным количеством ошибок.

Цель исследования – сборка полноразмерных нуклеотидных последовательностей хромосомы и плазмид для 13 штаммов *Yersinia pestis* из 11 природных очагов чумы, находящихся на территории Российской Федерации, используя данные технологий секвенирования Oxford Nanopore и IonTorrent.

Материалы и методы

Штаммы *Y. pestis* выращивали на агаре Хоттингера (рН 7,2) при 37 °С. Выделение ДНК проводили методом фенол-хлороформной экстракции. Для подготовки проб к секвенированию по технологии Oxford Nanopore на входе брали по 5–10 мкг ДНК от каждого исследуемого штамма. Подготовку

ДНК-фрагментов проводили методом лигирования по модифицированному протоколу. Модификация заключалась в увеличении времени инкубации до 1–1,5 часа на каждой из стадий подготовки ДНК (доставка концов фрагментов, лигирование баркодов и адаптеров). Помимо этого, для очистки образцов использовалось минимальное количество магнитных частиц (AMPure XP, Beckman Coulter, Inc.) – в соотношении объемов образца и магнитных частиц 1:0,4, что способствует удалению коротких фрагментов ДНК. Готовые библиотеки секвенировали на генетическом анализаторе MinIon с использованием прочной ячейки R9.4 и набора для секвенирования SQK-LSK109. Полученные единичные прочтения отфильтровывались по среднему качеству в процессе разделения по баркодам и перевода данных из формата fast5 в формат fastq с помощью программы Guppy 4.0. Для подготовки проб к секвенированию по технологии IonTorrent использовали стандартный протокол получения библиотеки с размером фрагментов ДНК на 400 п.н., рекомендованный производителем данного оборудования. Секвенирование готовых библиотек проводили на генетическом анализаторе Ion S5 с использованием набора Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef и полупроводниковых чипов Ion 530™. Для сборки полных последовательностей исследуемых геномов использовался алгоритм TruSeq [3–6].

Результаты и обсуждение

В результате проделанной работы нами подготовлены библиотеки фрагментов ДНК полного генома каждого из исследуемых штаммов *Y. pestis*. На первом этапе отработана методика подготовки образцов ДНК, позволяющая сохранить значительную часть фрагментов нуклеотидной последовательности в 50000 п.н. и более, для секвенирования по технологии Oxford Nanopore. Наличие фрагментов ДНК в 50000 п.н. и более в количестве, достаточном для многократного покрытия последовательности хромосомы штаммов *Y. pestis*, оказалось необходимым для сборки данных хромосом в виде целостных структур.

Изучение подходов к гибридной сборке полного бактериального генома различными алгоритмами показало эффективность программного пакета TruSeq. Алгоритм TruSeq описан в работе R.R. Wick *et al.* [3], где также приведены результаты ряда тестов по оценке различных подходов к сборке с точки зрения точности. Для оценки качества сборки разными методами авторами использовались как *in silico* сгенерированные наборы ридов из референсных бактериальных геномов, так и реальные результаты секвенирования. Лучший результат в обоих случаях показал подход, предполагающий сборку генома из длинных прочтений Oxford Nanopore с помощью TruSeq и дальнейшее исправление мелких ошибок алгоритмами Medaka (с помощью длинных

прочтений) и Pilon (с помощью коротких прочтений Illumina). С помощью алгоритма TruSeqer нами проведена гибридная сборка генома штаммов *Y. pestis* из данных секвенирования по технологиям Oxford Nanopore и IonTorrent, и в результате получены полные нуклеотидные последовательности хромосомы и плазмид для каждого штамма. Полученные нуклеотидные последовательности хромосом геномов 13 штаммов *Y. pestis* из 11 природных очагов чумы, находящихся на территории Российской Федерации (таблица) депонированы в международную генетическую базу данных NCBI GenBank.

В отличие от алгоритма TruSeqer, стандартный подход к гибридной сборке, использующийся в сборщике Unicycler, предполагает, что сначала из коротких и более точных прочтений будут собраны контиги, а после этого с помощью более длинных прочтений они будут соединены в полную последовательность для каждого репликона. Подход, используемый в алгоритме TruSeqer, наоборот, предполагает сборку полной последовательности из длинных прочтений и дальнейшее исправление в ней мелких ошибок с помощью коротких прочтений. Сборка геномов с помощью программного обеспечения TruSeqer состоит из нескольких шагов и требует вмешательства пользователя. На первом шаге для каждого генома из прочтений, полученных с помощью технологии секвенирования Oxford Nanopore, генерируются несколько случайных относительно независимых наборов ридов, дающих желаемую глубину покрытия генома. Затем эти выборки по отдельности собираются различными алгоритмами сборки длинных прочтений. В большинстве случаев для каждого генома мы генерировали 12 случайных выборок ридов с глубиной покрытия генома примерно равной 50. Выборки делились на 3 группы по 4 выборки и собирались отдельно сборщиками Flye [7], Raven [8, 9] и Miniasm [10, 11] соответственно. Использование выборок из исходных данных и нескольких вариантов алгоритма сборки необходимо, чтобы получить максимально независимые друг от друга сборки, что поможет на следующих шагах избежать типичных ошибок, возникающих при сборке из длинных ридов. Полученные сборки кластеризуются алгоритмом Mash [12, 13], после чего отсеиваются те последовательности, которые не собрались полностью или слишком удалены от остальных вариантов сборки (рис. 1). Оставшиеся последовательности группируются во множественные выравнивания, соответствующие каждому репликону. Спорные области в выравниваниях разрешаются с помощью соответствующих конкретному репликону длинных прочтений Oxford Nanopore. Далее из множественных выравниваний генерируется консенсусная последовательность. На заключительном шаге производится устранение мелких ошибок сборки в полученной консенсусной последовательности. Для этого с помощью программы Medaka [14] на собранную последовательность

сначала выравниваются длинные прочтения, а после этого короткие прочтения IonTorrent картируются алгоритмом Pilon [15], исправляя оставшиеся ошибки. Таким образом, сборка геномов проводилась только из длинных прочтений, что позволяет наиболее точно установить структуру генома, а короткие прочтения использовались только для исправления мелких ошибок сборки.

Чтобы провести сравнение двух подходов к сборке, геномы исследуемых 13 штаммов были также собраны алгоритмом Unicycler [16, 17]. Для оценки качества полученных сборок использовались алгоритмы BUSCO [18, 19] и Ideel [20]. Алгоритм BUSCO использует базы протеин-кодирующих ортогологических генов для их поиска в сборке и оценки их целостности.

Мы проанализировали 26 получившихся сборок с помощью BUSCO, используя базу *enterobacterales odb10*. Для последовательностей, собранных с помощью Unicycler и TruSeqer, получены средние значения BUSCO-целостности 97,9 и 99,5 % соответственно, что подтверждает более высокую точность сборки при использовании TruSeqer.

Также учитывая, что обе применяемые нами технологии секвенирования (Oxford Nanopore и Ion Torrent) часто совершают ошибки в гомополимерных участках ДНК, из-за чего происходит сдвиг рамки считывания, можно предположить, что после сборки мы столкнемся с рядом генов, продукт которых в виде аминокислотной последовательности будет укорочен из-за преждевременного появления стоп-кодона. Чтобы оценить количество таких белков и с этой точки зрения сравнить два подхода к сборке между собой, мы использовали скрипт Ideel. Этот алгоритм сначала предсказывает продукты в заданной геномной последовательности, а затем ищет ближайший гомологичный белок в предварительно загруженной локальной базе белков UniProt TrEMBL. Длина каждого белка из базы данных сравнивается с длиной предсказанных белков из сборки генома, после чего сравнительная информация по всем продуктам генов в собранной генетической последовательности визуализируется (рис. 2).

На рис. 2 представлены результаты обработки скриптом Ideel двух гибридных сборок генома штамма *Y. pestis* M1974, полученных из одних и тех же ридов с помощью алгоритмов Unicycler и TruSeqer. На графике представлено распределение отношения длин продуктов, кодируемых собранной последовательностью, к длинам гомологичных последовательностей белков из базы UniProt. Столбцы, находящиеся левее значения 1 на оси X, отражают число генов, белковый продукт которых короче, чем у ближайшего гена в базе UniProt TrEMBL. Число генов, для которых значение данного отношения меньше 0,9, в случае сборки Unicycler составляет 262, а в случае сборки TruSeqer – 127, что говорит о более высокой точности алгоритма TruSeqer в отношении ошибок такого типа.

Штаммы *Yersinia pestis* из 11 природных очагов чумы на территории России, полные геномы которых секвенированы и собраны в данной работе

***Yersinia pestis* strains from 11 natural plague foci in Russia, the complete genomes of which were sequenced and assembled within the framework of this work**

№ п/п No.	Очаг, подви́д, биовар Focus, subspecies, biovar	Номер штамма Strain number	Источник выделения, год Source of isolation, year	Номер регистрации в базе NCBI GenBank Access number in the NCBI GenBank
1	2	3	4	5
1	Центрально-Кавказский высокогорный (01), основной подви́д, средневековый биовар Central Caucasian high mountain (01), main subspecies, medieval biovar	C-781 [37 КБ] C-781 [37 KB]	Горный суслик (<i>Citellus musticus</i>), 2001 г. Mountain gopher (<i>Citellus musticus</i>), 2001	NZ_CP064117.1
2	Центрально-Кавказский высокогорный (01), основной подви́д, средневековый биовар Central Caucasian high mountain (01), main subspecies, medieval biovar	C-783 [45 КБ] C-783 [45 KB]	Блохи <i>Citellus tesquorum</i> с горного суслика, 2001 г. Flea <i>Citellus tesquorum</i> collected from mountain gopher, 2001	NZ_CP064118.1
3	Терско-Сунженский низкогорный (02), основной подви́д, средневековый биовар Tersko-Sunzha low mountain (02), main subspecies, medieval biovar	14-Д 14-D	Блохи <i>Ceratophyllus tesquorum</i> из нор малого суслика (<i>Citellus pygmaeus</i>), 1970 г. Flea <i>Ceratophyllus tesquorum</i> from the burrows of the small gopher (<i>Citellus pygmaeus</i>), 1970	CP063303.2
4	Дагестанский равнинно-предгорный (03), основной подви́д, средневековый биовар Dagestan plain-piedmont (03), main subspecies, medieval biovar	C-792 [283-Дар.] C-792 [283-Dag.]	Блохи <i>Neopsylla setosa</i> со входов нор малого суслика (<i>Cit. pygmaeus</i>), 2003 г. Flea <i>Neopsylla setosa</i> from the entrances of the burrows of the small gopher (<i>Cit. pygmaeus</i>), 2003	CP064119.2
5	Прикаспийский Северо-Западный (14), основной подви́д, средневековый биовар Precaspian North-Western (14), main subspecies, medieval biovar	КМ 567 [949 Элист. ПЧС] КМ 567 [949 Elist. PChS]	9 молодых малых сусликов (<i>Cit. pygmaeus</i>), 1972 г. 9 young small gophers (<i>Cit. pygmaeus</i>), 1972	NZ_CP064120.1
6	Волго-Уральский степной (15), основной подви́д, средневековый биовар Volga-Ural steppe (15), main subspecies, medieval biovar	M-1482 [330 Ур.] M-1482 [330 Ur.]	5 малых сусликов (<i>Cit. pygmaeus</i>), 1989 г. 5 small gophers (<i>Cit. pygmaeus</i>), 1989	NZ_CP064121.1
7	Волго-Уральский песчаный (16), основной подви́д, средневековый биовар Volga-Ural sandy (16), main subspecies, medieval biovar	M-1770 [769 Аст.] M-1770 [769 Ast.]	6 полуденных песчанок (<i>Meriones meridianus</i>), 2002 г. 6 midday gerbils (<i>Meriones meridianus</i>), 2002	CP064122.2
8	Горно-Алтайский высокогорный (36), основной подви́д, античный биовар Gorno-Altai high mountain (36), main subspecies, antique biovar	M2029	Алтайский сурок (остатки стола хищных птиц), 2018 г. Altai marmot (remains of a meal of birds of prey), 2018	NZ_CP064123.1
9	Горно-Алтайский высокогорный (36), центральноазиатский подви́д, алтайский биовар Gorno-Altai high mountain (36), Central Asian subspecies, altai biovar	M-1974 [3595-Алт.] M-1974 [3595-Alt.]	5 монгольских пищух (<i>Ochotona pallasii</i>), 2012 г. 5 Mongolian pikas (<i>Ochotona pallasii</i>), 2012	NZ_CP064124.1
10	Тувинский горный (37), основной подви́д, античный биовар Tuva mountain (37), main subspecies, antique biovar	M2085	Суслик длиннохвостый (<i>Citellus undulatus</i>), 2015 г. Long-tailed gopher (<i>Citellus undulatus</i>), 2015	CP064125.2
11	Забайкальский степной (38), основной подви́д, античный биовар Transbaikal steppe (38), main subspecies, antique biovar	I-1252 [928] I-1252 [928]	Даурский суслик (<i>Citellus dauricus</i>), 1966 г. Daurian gopher (<i>Citellus dauricus</i>), 1966	NZ_CP064126.1
12	Восточно-Кавказский высокогорный (39), кавказский подви́д East Caucasian high mountain (39), Caucasian subspecies, microtus biovar	C-830	Блохи <i>Ctenophthalmus intermedius</i> из гнезд полевки обыкновенной (<i>Microtus arvalis</i>), 2012 г. Flea <i>Ctenophthalmus intermedius</i> from the nests of the common vole (<i>Microtus arvalis</i>), 2012	NZ_CP064127.1
13	Прикаспийский песчаный (43), основной подви́д, средневековый биовар Caspian sandy (43), main subspecies, medieval biovar	M2086	Блохи <i>Cer. tesquorum</i> , 2015 г. Flea <i>Cer. tesquorum</i> , 2015	NZ_CP064128.1

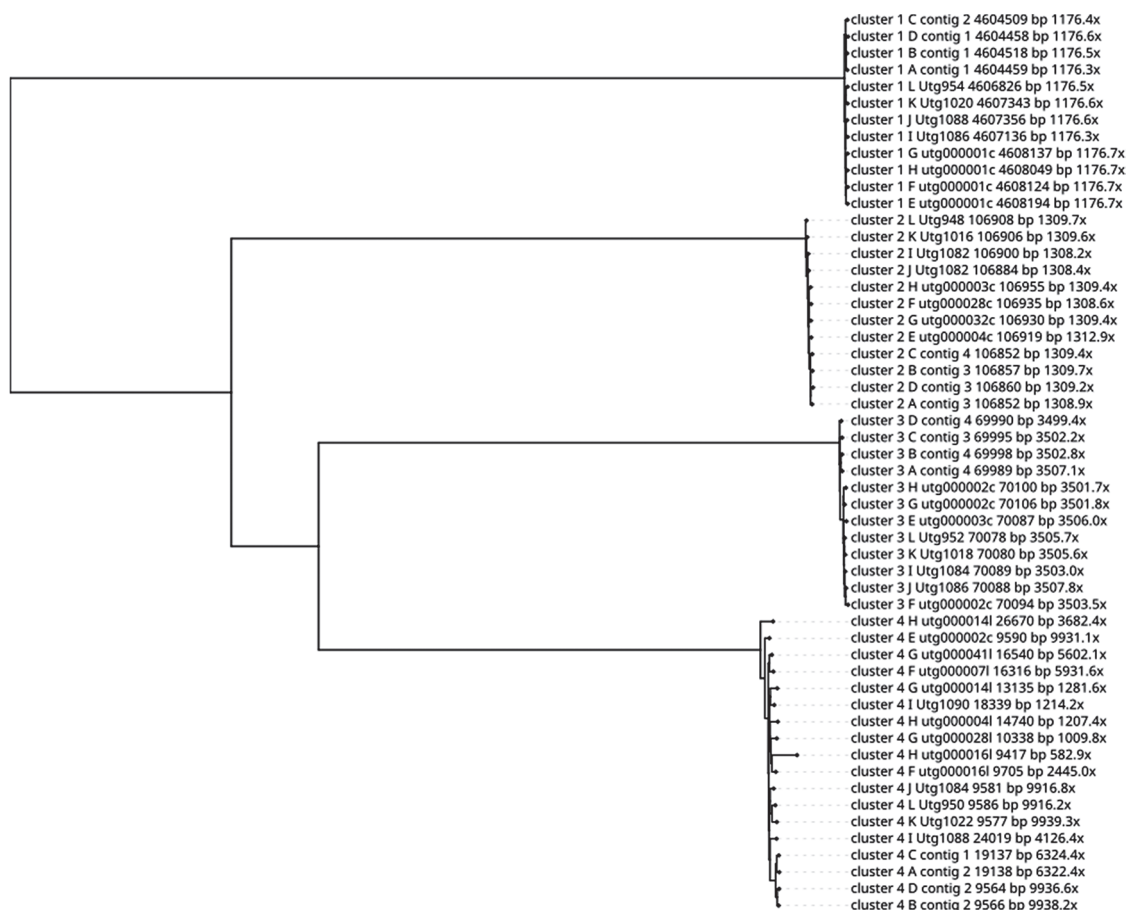


Рис. 1. Дендрограмма, отражающая результаты кластеризации сборок генома штамма *Y. pestis* KM567. Сформировано четыре кластера, состоящих из вариантов сборки трех плазмид и одной хромосомы из разных выборок ридов

Fig. 1. Dendrogram reflecting the results of clustering of genome assemblies of the *Y. pestis* strain KM567. Four clusters were formed, consisting of assembly variants of three plasmids and one chromosome from different selection of reads

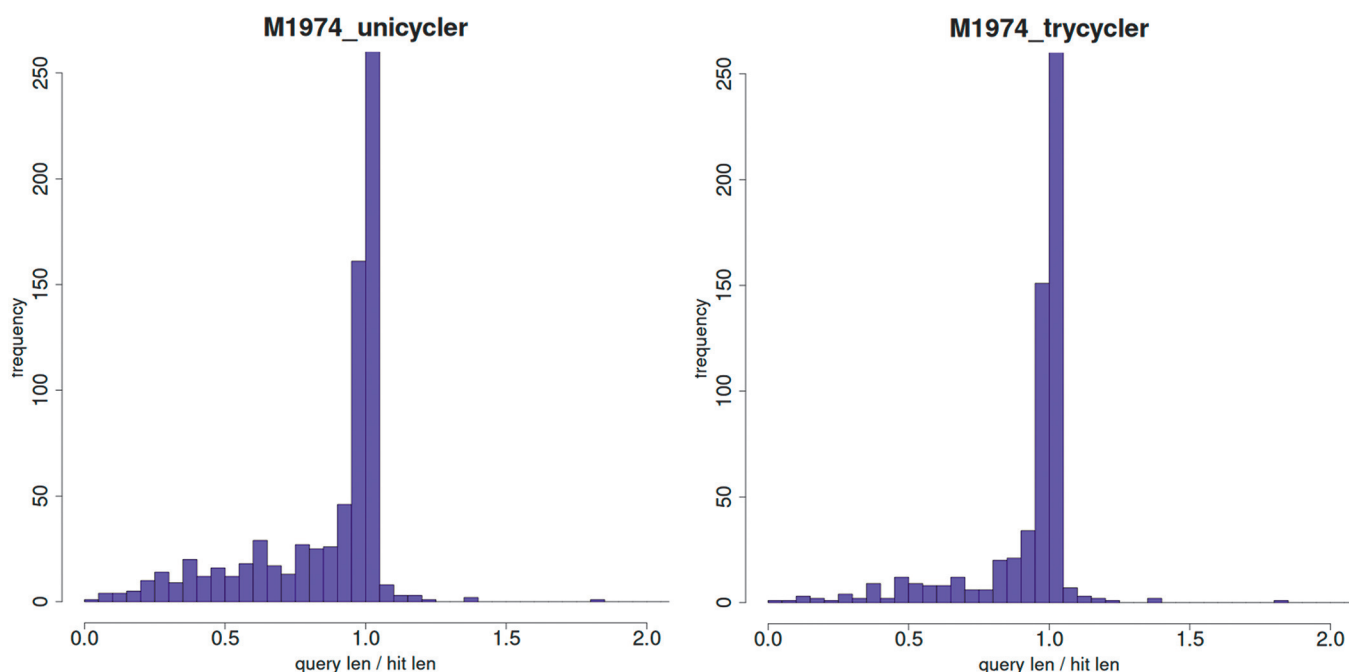


Рис. 2. Сравнение результатов сборки Unicycler и Tricycler с помощью Ideel на примере штамма *Y. pestis* M1974

Fig. 2. Comparison of Unicycler and Tricycler assembly results using Ideel by the example of *Y. pestis* M1974 strain

Полученные нами результаты подтверждают преимущества алгоритма TruSeq, однако алгоритм Unicycler проще в практическом использовании и с успехом продолжает использоваться для гибридных сборок бактериальных геномов [21–23].

Проведенная работа позволила впервые получить целые нуклеотидные последовательности хромосомы и плазмид для 13 геномов штаммов *Y. pestis* из 11 природных очагов чумы, находящихся на территории Российской Федерации.

Конфликт интересов. Авторы подтверждают отсутствие конфликта финансовых/нефинансовых интересов, связанных с написанием статьи.

Финансирование. Авторы заявляют об отсутствии дополнительного финансирования при проведении данного исследования.

References / Список литературы

1. Hu T., Chitnis N., Monos D., Dinh A. Next-generation sequencing technologies: An overview. *Hum. Immunol.* 2021; 82(11):801–11. DOI: 10.1016/j.humimm.2021.02.012.
2. Oxford Nanopore Technologies. (Cited 16 Nov 2023). [Internet]. Available from: <https://nanoporetech.com/applications>.
3. Wick R.R., Judd L.M., Cerdeira L.T., Hawkey J., Méric G., Vezina B., Wyres K.L., Holt K.E. TruSeq: consensus long-read assemblies for bacterial genomes. *Genome Biol.* 2021; 22(1):266. DOI: 10.1186/s13059-021-02483-z.
4. Wick R.R., Judd L.M., Holt K.E. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput. Biol.* 2023; 19(3):e1010905. DOI: 10.1371/journal.pcbi.1010905.
5. Krøvel A.V., Hetland M.A.K., Bernhoff E., Bjørheim A.S., Soma M.A., Löhr I.H. Long-read sequencing for reliably calling the *mompS* allele in *Legionella pneumophila* sequence-based typing. *Front. Cell. Infect. Microbiol.* 2023; 13:1176182. DOI: 10.3389/fcimb.2023.1176182.
6. Martino J.A., Fernandez F.D., Pozzi E.A., Alberione E., Bainotti C., Marquez N., Tolocka P.A., Salines N., Gomez D., Donaire G., Conci L., Alemandri V.M. First report of *Xanthomonas prunicola* causing bacterial leaf streaks on wheat in Argentina. *Plant Dis.* 2022. DOI: 10.1094/PDIS-04-22-0886-PDN.
7. Kolmogorov M., Yuan J., Lin Y., Pevzner P.A. Assembly of long error-prone reads using repeat graphs. *Nat. Biotechnol.* 2019; 37(5):540–6. DOI: 10.1038/s41587-019-0072-8.
8. Vaser R., Sikić M. Raven: a *de novo* genome assembler for long reads. *bioRxiv* 2021. DOI: 10.1101/2020.08.07.242461.
9. Chen Z., Erickson D.L., Meng J. Benchmarking Long-Read Assemblers for genomic analyses of bacterial pathogens using Oxford Nanopore sequencing. *Int. J. Mol. Sci.* 2020; 21(23):9161. DOI: 10.3390/ijms21239161.
10. Li H. Minimap and minimap: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics.* 2016; 32(14):2103–10. DOI: 10.1093/bioinformatics/btw152.
11. Jung H., Jeon M.S., Hodgett M., Waterhouse P., Eyun S.I. Comparative evaluation of genome assemblers from long-read sequencing for plants and crops. *J. Agric. Food Chem.* 2020; 68(29):7670–7. DOI: 10.1021/acs.jafc.0c01647.

12. Ondov B.D., Treangen T.J., Melsted P., Mallonee A.B., Bergman N.H., Koren S., Phillippy A.M. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016; 17(1):132. DOI: 10.1186/s13059-016-0997-x.
13. Criscuolo A. On the transformation of MinHash-based uncorrected distances into proper evolutionary distances for phylogenetic inference. *F1000Res.* 2020; 9:1309. DOI: 10.12688/f1000research.26930.1.
14. Sequence correction provided by ONT Research. (Cited 24 Oct 2023). [Internet]. Available from: <https://github.com/nanoporetech/medaka>.
15. Walker B.J., Abeel T., Shea T., Priest M., Abouelliel A., Sakthikumar S., Cuomo C.A., Zeng Q., Wortman J., Young S.K., Earl A.M. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014; 9(11):e112963. DOI: 10.1371/journal.pone.0112963.
16. Wick R.R., Judd L.M., Gorrie C.L., Holt K.E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 2017; 13(6):e1005595. DOI: 10.1371/journal.pcbi.1005595.
17. Farrow J.M. 3rd, Pesci E.C., Slade D.J. Genome sequences for two *Acinetobacter baumannii* strains obtained using the unicycler hybrid assembly pipeline. *Microbiol. Resour. Announc.* 2021; 10(10):e00017-21. DOI: 10.1128/MRA.00017-21.
18. Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015; 31(19):3210–2. DOI: 10.1093/bioinformatics/btv351.
19. Manni M., Berkeley M.R., Seppey M., Zdobnov E.M. BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* 2021; 1(12):e323. DOI: 10.1002/cpz1.323.
20. Indels are not ideal – quick test for interrupted ORFs in bacterial/microbial genomes. (Cited 08 Jul 2023). [Internet]. Available from: <https://github.com/mw55309/ideel>.
21. Peng Y., Cai X., Li M., Deng L., Wang Y., Qiu Y., Zhao L., Xiao Y., Xu L., Hou Q. The first completed genome of species *Prevotella bivia*, assembled from a clinically derived strain PLW0727. *J. Glob. Antimicrob. Resist.* 2023; 35:268–70. DOI: 10.1016/j.jgar.2023.10.009.
22. El-Sabeh A., Mlesnita A.M., Munteanu I.T., Honceriu I., Kallabi F., Boiangiu R.S., Mihasan M. Characterisation of the *Paenarthrobacter nicotinovorans* ATCC 49919 genome and identification of several strains harbouring a highly syntenic *nic*-genes cluster. *BMC Genomics.* 2023; 24(1):536. DOI: 10.1186/s12864-023-09644-3.
23. Ishida-Kuroki K., Hisatsune J., Segawa T., Sugawara Y., Masuda K., Tadera K., Kashiwayama S., Yokozaki M., Le M.N., Kawada-Matsuo M., Ohge H., Komatsuzawa H., Sugai M. Complete genome sequence of *cfr*(B)-carrying *Enterococcus raffinosus* isolated from bile in a patient in Japan. *J. Glob. Antimicrob. Resist.* 2023; 34:43–5. DOI: 10.1016/j.jgar.2023.06.004.

Authors:

Fedorov A.V., Krasnov Ya.M., Naryshkina E.A., Sosedova E.A., Katyshev A.D., Eroshenko G.A., Kukleva L.M., Osina N.A., Kuttyrev V.V. Russian Research Anti-Plague Institute “Microbe”. 46, Universitetskaya St., Saratov, 410005, Russian Federation. E-mail: rusrapi@microbe.ru.

Об авторах:

Федоров А.В., Краснов Я.М., Нарышкина Е.А., Соседова Е.А., Катышев А.Д., Ерошенко Г.А., Куклева Л.М., Осина Н.А., Кутырев В.В. Российский научно-исследовательский противочумный институт «Микроб». Российская Федерация, 410005, Саратов, ул. Университетская, 46. E-mail: rusrapi@microbe.ru.