

ГЕНОМИКА.
ТРАНСКРИПТОМИКА

УДК 577.29

ПОВЫШЕНИЕ РАВНОМЕРНОСТИ ПОКРЫТИЯ
ФРАГМЕНТОВ ГЕНОМА ВИРУСА ГРИППА А
ПРИ ВЫСОКОПРОИЗВОДИТЕЛЬНОМ СЕКВЕНИРОВАНИИ

© 2020 г. Ю. В. Михайлова^{а, *}, А. А. Шеленков^а, Ю. Г. Янушевич^а, Д. А. Шагин^{а, б}

^аЦентральный научно-исследовательский институт эпидемиологии Роспотребнадзора, Москва, 111123 Россия

^бРоссийский национальный исследовательский медицинский университет им. Н.И. Пирогова, Москва, 117997 Россия

*e-mail: mihailova@cmd.su

Поступила в редакцию 24.03.2020 г.

После доработки 22.05.2020 г.

Принята к публикации 25.05.2020 г.

Высокая степень изменчивости вируса гриппа А создает значительную угрозу для общественного здравоохранения, что делает актуальной задачу мониторинга вирусных штаммов и изучения их генетических характеристик. Одно из направлений мониторинга – секвенирование вирусов гриппа А любого подтипа и анализ их полных геномов, что особенно важно при изучении межвидовой адаптации и реассортации вирусов гриппа в природе. Технология высокопроизводительного секвенирования (NGS) значительно расширила возможности эпидемиологического надзора за вирусом гриппа. При проведении NGS полный геном вируса гриппа А амплифицируют с помощью одноступенчатой ОТ-ПЦР, результат которой сильно зависит от типа образца и его качества, что, в свою очередь, отражается на равномерности покрытия фрагментов вирусного генома и на результатах секвенирования в целом. В данной работе мы предлагаем дополнить методику полногеномной амплификации вируса гриппа А последовательными супрессионными ПЦР с целью получения равномерной представленности разных по длине сегментов вируса, что позволяет секвенировать образцы с меньшей глубиной прочтения без снижения качества секвенирования.

Ключевые слова: вирус гриппа А, полногеномное секвенирование, супрессионная ПЦР

DOI: 10.31857/S0026898420060087

ВВЕДЕНИЕ

Вирус гриппа А – один из наиболее распространенных вирусов, имеющий широкий спектр хозяев, и, следовательно, высокий пандемический потенциал. Вирус гриппа А характеризуется значительной вариабельностью из-за способности к антигенной изменчивости в результате мутационных событий и реассортации.

Технологии секвенирования следующего поколения (NGS) значительно расширили объем получаемой информации и возможности аннотации генома. Анализ данных NGS позволяет выявлять генетические вариации множества вирусных частиц гриппа в одном образце. Кроме того, можно выделить всю кодирующую область геномов. Полногеномный анализ вируса гриппа дает возможность получить информацию о реассортантах и мутациях, возникающих во всех генах, раскрывая эволюционные паттерны и информацию о разнообразии вируса внутри хозяина [1].

Геном вируса гриппа состоит из восьми сегментов РНК отрицательной полярности разме-

ром 850–2300 н. (PB2, PB1, PA, NA, NP, NA, MP и NS). У всех штаммов вируса каждый сегмент фланкирован одинаковыми консервативными последовательностями – 12 н. на 3'-конце и 13 н. на 5'-конце [2], на основе которых генерируются универсальные наборы праймеров для проведения обратной транскрипции и амплификации РНК вируса и последующего полногеномного секвенирования [3–6].

Опубликован первый универсальный набор праймеров для полноразмерной амплификации всех вирусов гриппа А, который включает несколько праймеров, специфичных для каждого сегмента [3]. В настоящее время наибольшее распространение получила методика [5], в которой используются два праймера, направленных на универсальные концевые последовательности всех восьми сегментов генома вируса гриппа А. По данному протоколу за один этап в одной пробирке последовательно осуществляется обратная транскрипция РНК вируса гриппа А и его полногеномная амплификация. Этот протокол оптимизирован для рутинного применения [4, 6].

Однако по нашему опыту с помощью этого метода самые крупные геномные сегменты (PB1, PB2 и PA) удается амплифицировать не во всех образцах, а данные NGS свидетельствуют о низком покрытии концевых участков этих фрагментов, что подтверждают и результаты других авторов [7, 8]. Мы дополнили протокол полногеномной амплификации вируса гриппа разработанной ранее методикой регулирования средней длины комплексного ПЦР-продукта [9], что позволило увеличить представленность наиболее высокомолекулярных сегментов в образцах.

Молекулы ДНК, амплифицируемые с использованием этой методики, снабжаются инвертированными повторами. После денатурации двухцепочечных молекул эти повторы способны к внутримолекулярной гибридизации, которая замедляет амплификацию последовательностей, препятствуя отжигу праймеров. Этот эффект ПЦР-супрессии усиливается с увеличением длины повтора и снижением концентрации ПЦР-праймеров. Поскольку внутримолекулярная гибридизация с большей вероятностью происходит на коротких последовательностях ДНК, более длинные получают преимущество при амплификации.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Анализировали РНК вируса гриппа А (H1N1), выделенную из клинического материала (четыре образца из назофарингеальных мазков) и материала, обогащенного путем пассирования на клетках линии MDCK (21 образец, пять из которых относились к подтипу H3N2, остальные – к H1N1) с концентрацией РНК не менее 10 млн. копий/мл. От пациентов получено письменное согласие на использование их клинического материала (назофарингеальные мазки) в научных целях.

Выделение РНК. Вирусную РНК выделяли из надосадочной культуральной жидкости с использованием набора реагентов Qiagen RNeasy mini kit

(“Qiagen”, Германия) в соответствии с рекомендациями производителя.

Праймеры. Для секвенирования на приборах Illumina последовательности должны быть фланкированы определенными адаптерами (табл. 1), поэтому мы удлинители универсальные для всех сегментов вируса праймеры соответствующей последовательностью. При этом в целях достижения эффекта супрессии эта последовательность была одинаковой в прямом и обратном праймерах (Nts-inf1 и Nts-inf2). Последовательности всех использованных праймеров приведены в табл. 1.

Обратная транскрипция. РНК транскрибировали в кДНК с использованием обратной транскриптазы Mint (“Евроген”, Россия) и праймера NTS-inf1. Реакционные смеси, содержащие 4 мкл РНК и 1 мкл 10 мкМ NTS-inf1, нагревали при 70°C в течение 2 мин и охлаждали на льду в течение 5 мин. Затем к смеси добавляли 2 мкл 5× буфера для первой цепи (“Евроген”), 1 мкл 10 мМ dNTP, 1 мкл DTT и 1 мкл обратной транскриптазы Mint. Реакцию проводили при 42°C в течение 60 мин.

Амплификация первой цепи кДНК. Первую цепь кДНК амплифицировали с использованием ДНК-полимеразы Encyclo (“Евроген”) и набора праймеров NTS-inf1/NTS-inf2. Смеси для ПЦР (общий объем 25 мкл) содержали 2.5 мкл 10-кратного буфера, по 0.5 мкл 10 мкМ праймеров, 0.5 мкл 10 мМ dNTP, 5 мкл одноцепочечной кДНК, 0.5 мкл ДНК-полимеразы и 14 мкл воды. Первые пять циклов программы амплификации включали денатурацию (20 с при 94°C), отжиг праймеров (35 с, 42°C) и элонгацию (3 мин, 72°C), в последующих 25 циклах отжиг праймеров проводили при 57°C, конечную достройку – при 72°C в течение 10 мин.

Результаты амплификации образцов проверяли с помощью электрофоретического разделения в 1.2%-ном агарозном геле в TAE-буфере.

Таблица 1. Праймеры для ПЦР-супрессии (5' → 3')

Название	Последовательность праймера
Nts-inf1	<u>A GAT GTG TAT AAG</u> AGA CAG <u>AGC AAA AGC AGG</u>
Nts-inf2	<u>A GAT GTG TAT AAG</u> AGA CAG <u>AGT AGA AAC AAG G</u>
Nts1	<u>A GAT GTG TAT AAG</u> AGA CAG
NtSeq1	TCG TCG GCA GCG TCA <u>GAT GTG</u>
Not1-NtSeq1	<u>AGC GTG GTC GCG GCC GAG GTT</u> CGT CGG CAG CGT TCA <u>GAT GTG</u>
Na21Not1	TGT AGC GTG AAG ACG ACA GAA <u>AGC GTG GTC GCG GCC GAG GT</u>

Примечание. Жирной линией подчеркнуты консервативные фланкирующие последовательности сегментов вируса гриппа А; двойным подчеркиванием показаны одинаковые последовательности разных праймеров.

Обогащение амплифицированной кДНК с помощью ПЦР-супрессии. Для обогащения образцов амплифицированных кДНК вируса гриппа А по длинным сегментам были проведены дополнительные последовательные супрессионные ПЦР (супПЦР). Результаты каждого этапа детектировали с помощью электрофореза в агарозном геле. Первую амплификацию с эффектом супрессии (супПЦР1) осуществляли с помощью праймеров NTS1 в концентрации 0.012 мкМ и NTSeq1 в концентрации 0.12 мкМ. В качестве матрицы использовали тысячекратное разведение амплифицированной кДНК (исходная концентрация 5–15 нг/мкл), амплификацию проводили согласно следующей программе: 95°C, 20 с (95°C, 15 с; 57°C, 25 с; 72°C, 1.5 мин) × 14–17 циклов (амплификатор Eppendorf nexus X2). СупПЦР 2 осуществляли с праймерами NTSeq1 и NotNTSeq1 в концентрации 0.012 и 0.12 мкМ соответственно. Проводили 16–18 циклов амплификации по той же программе, что и для супПЦР1. Финальную супПЦР3 с 23 циклами амплификации проводили аналогично с праймером Na21NotI в концентрации 0.12 мкМ.

Приготовление библиотек для полногеномного секвенирования. Продукты полногеномной амплификации вируса гриппа А очищали с использованием SpeedBead Magnetic Carboxylate Modified Particles (“GE Healthcare”, США). Концентрацию двухцепочечной кДНК в ПЦР-продуктах определяли с использованием флуориметра Qubit и набора реагентов Qubit HS dsDNA kit (“Life Technologies”, США).

Приготовление библиотек для полногеномного секвенирования осуществляли с помощью набора реагентов Nextera Library Prep kit (“Illumina”, США) согласно рекомендациям фирмы-изготовителя. Образцы индексировали с использованием набора Illumina Nextera XT Index kit v2 Set A (“Illumina”).

Секвенирование и анализ данных. Высокопроизводительное секвенирование проводили на секвенаторе Illumina HiSeq 1500 с использованием наборов Illumina HiSeq PE Rapid Cluster Kit v2 и Illumina HiSeq Rapid SBS Kit v2.

Первичную обработку прочтений с целью их фильтрации по качеству и удаления адаптеров секвенирования выполняли с использованием программного обеспечения bcl2fastq 2.20 (“Illumina”), flexbar 2.5 и специально разработанного программного обеспечения (скриптов). Сборку сегментов вируса проводили путем картирования прочтений на геном референсного штамма вируса А (Crie-162, H1N1, идентификаторы Genbank для восьми сегментов KX775364.1-KX775371.1), поскольку сборка с применением программного обеспечения SPAdes [10] и некоторых других программ сборки вирусных геномов (IVA [11], MIRA [12]) позволи-

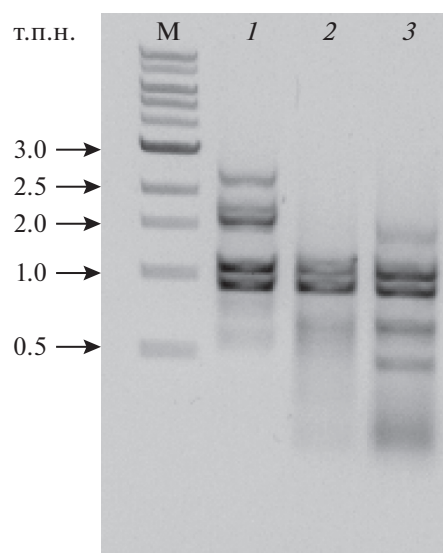


Рис. 1. Амплифицированная кДНК вируса гриппа А (H1N1) из образцов разного качества. М – маркер молекулярных масс (1 kb DNA Ladder, “Евроген”). 1 – Амплифицированная кДНК положительного контрольного образца вируса гриппа А; 2 и 3 – амплифицированная кДНК образцов вируса гриппа А, у которых плохо нарабатывались высокомолекулярные сегменты.

ла получить лишь множество контигов (более 20), которые не соответствовали сегментам вируса, сильно различались по длине и содержали перекрывающиеся последовательности. Сложности сборки геномов РНК-вирусов связаны с большим внутривидовым разнообразием и неравномерной представленностью различных сегментов, а также с наличием повторов, разрешить которые в рамках существующих алгоритмов не всегда удается [11].

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Продукт амплификации первой цепи вируса гриппа А представляет собой семь–восемь видимых фрагментов, соответствующих вирусным сегментам. В части образцов, как и в положительном контроле, после электрофореза в агарозном геле, видны все восемь амплифицированных сегментов (рис. 1, дорожка 1). В то же время во многих образцах (в нашем исследовании в 10 из 25 образцов) высокомолекулярные сегменты размером от 1.5 до 2.3 т.п.н. амплифицируются намного слабее, что подтверждают результаты электрофоретического разделения в агарозном геле (рис. 1, дорожки 2 и 3). Отсутствие видимых полос, соответствующих высокомолекулярным сегментам, может быть обусловлено несколькими причинами: частичной деградацией РНК вируса до или после выделения; низкой концентрацией кДНК – большое число

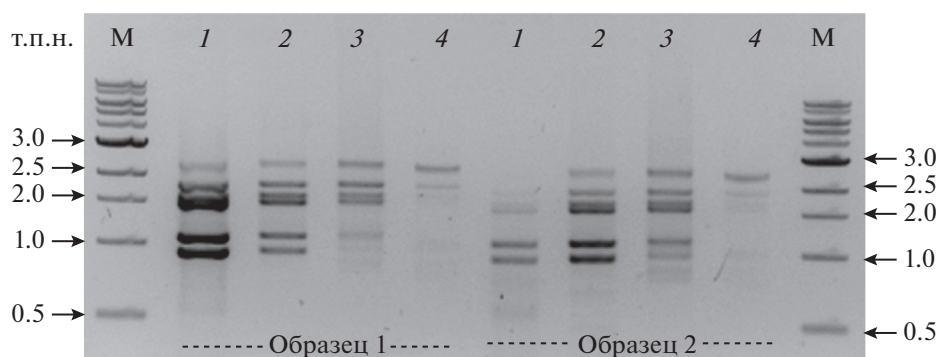


Рис. 2. Применение ПЦР-супрессии к разным по качеству образцам амплифицированной кДНК вируса гриппа А. М – Маркер молекулярных масс (1 kb DNA Ladder, “Евроген”); 1 и 5 – амплифицированная кДНК образцов 1 и 2 соответственно; дорожки 2 и 6 – продукт амплификации супПЦР1 образцов 1 и 2; дорожки 3 и 7 – продукт амплификации супПЦР2 образцов 1 и 2; дорожки 4 и 8 – продукт амплификации финальной супПЦР3 образцов 1 и 2.

циклов ПЦР, приводящее к существенному отставанию высокомолекулярных фрагментов.

Применение ПЦР-супрессии смещает эффективность ПЦР в сторону длинных последовательностей, что обеспечивает более равномерную представленность всех фрагментов генома вируса. Так, на рис. 2 показано, что все фрагменты амплифицированной кДНК вируса гриппа А представлены неравномерно независимо от того, видны они или нет (дорожки 1 и 5), при этом их представленность выравнивается с каждой супПЦР (дорожки 2 и 6, 3 и 7, 4 и 8). Из рис. 2 видно, что увеличение степени супрессии (каждый последующий этап супПЦР) сдвигает представленность фрагментов в сторону более длинных (дорожки 4 и 8).

Эффект ПЦР-супрессии подтвержден и результатами полногеномного секвенирования, в котором мы использовали продукты амплификации первой цепи (рис. 2, дорожки 1 и 5) и продукты второго этапа ПЦР-супрессии (рис. 2, дорожки 3 и 7). Всего для проверки эффекта супрессии проанализированы четыре образца вируса гриппа А до и после супПЦР.

Для каждого образца получено 40–70 тысяч парноконцевых прочтений длиной 251 п.н. В целях корректного сравнения покрытия сегментов каждую пару образцов до и после ПЦР-супрессии нормировали по минимальному количеству прочтений в паре. В качестве примера рассмотрим изменение покрытия генома штамма 304 вируса гриппа А (H1N1). На рис. 3 (до супПЦР) отмечена сильная неравномерность покрытия сегментов генома, особенно PB1, PB2, NA, PA, где перепредставлены преимущественно 3'-концевые области. Повышение покрытия на концах сегментов генома объясняется особенностями приготовления библиотеки для секвенирования; главным образом тем, что продукт ПЦР уже фланкирован необходимыми для NGS адаптерными последовательностями,

и для получения пригодной для секвенирования молекулы достаточно однократного разреза транспозазой. Кроме того, использование этого фермента не обеспечивает полностью равномерного покрытия, поскольку транспозаза имеет предпочтение к сайтам связывания ДНК. Как видно из рис. 3 и табл. 2, после супПЦР среднее нормированное покрытие в прочтениях каждого высокомолекулярного сегмента (PB1, PB2, PA и NA) увеличилось почти в 2 раза, и покрытие стало более равномерным. При этом в случае сегментов меньшего размера (NP, NA, M и NP) количество прочтений либо не изменилось, либо стало меньше (рис. 3, табл. 2), но не повлияло на равномерность покрытия сегментов. Среднее покрытие коротких сегментов осталось на уровне, достаточном для проведения последующих анализов (>1000). Таким образом, результаты полногеномного секвенирования подтверждают выравнивание представленности всех сегментов вируса гриппа А в амплифицированном образце с помощью супрессионной ПЦР. Всего с помощью данной технологии мы амплифицировали и проанализировали 25 образцов вируса гриппа А.

Во всех случаях удалось секвенировать полные геномы вируса гриппа с покрытием, достаточным для понимания полиморфизма геномов и сравнения их между собой при проведении эпидемиологических исследований.

Следует отметить, что в работе [8], по сути, также предложен вариант супПЦР, с помощью которого удалось добиться увеличения покрытия на концевых участках фрагментированных сегментов вируса гриппа А. Указанное среднее покрытие сегментов вируса (4000) во всех проанализированных образцах говорит о том, что на секвенирование каждого из них отводился относительно большой процент от всего запуска (высокая глубина прочтения). Вариант супПЦР, предлагаемый нами, бо-

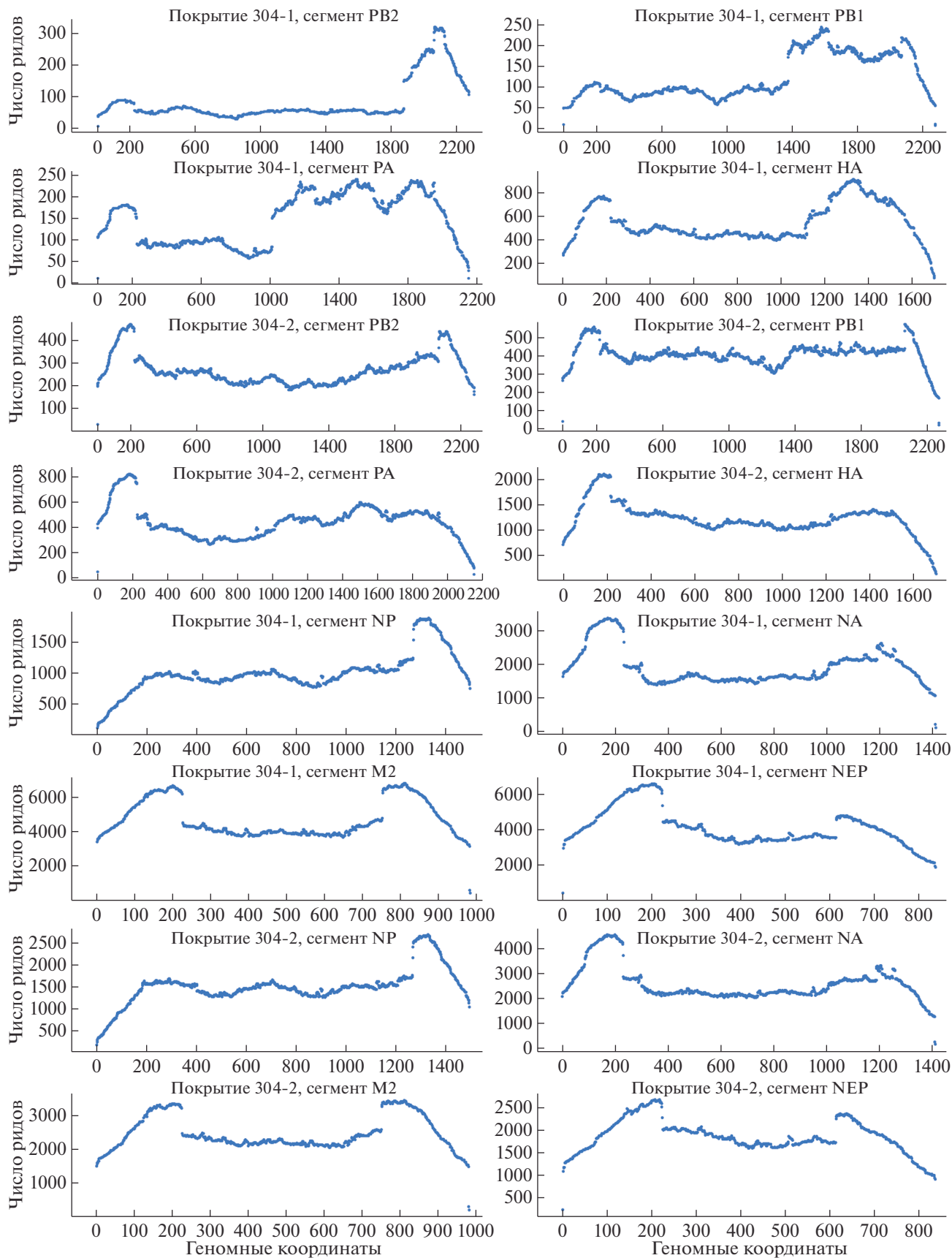


Рис. 3. Покрытие секвенированных с помощью NGS сегментов вируса гриппа А до (304-1) и после ПЦР-супрессии (304-2).

Таблица 2. Среднее покрытие сегментов образцов вируса гриппа А (H1N1) до и после супПЦР

Код образца, название сегмента	PB2	PB1	PA	HA	NP	NA	M2	NEP
Длина сегмента, п.н.	2300	2300	2200	1800	1600	1400	1000	900
304	54	98	167	484	957	1680	4189	3760
304 супПЦР	1200	1928	1584	1624	1485	2291	2277	1813
288	60	22	92	146	2854	2778	5724	5382
288 супПЦР	460	520	390	501	760	820	850	751
314	103	272	115	247	453	318	1938	1298
314 супПЦР	249	631	186	373	450	229	620	288
371	165	168	26	430	1818	1488	6176	3799
371 супПЦР	698	667	549	878	1344	1751	1359	1194

лее жесткий, что позволяет не только получить равномерную представленность разных по длине сегментов вируса, но и секвенировать каждый образец вируса гриппа А с меньшей глубиной прочтения, что значительно снижает стоимость полногеномного секвенирования.

Написание данной работы не требовало специального финансирования.

Все процедуры, выполненные в данной работе, соответствуют этическим стандартам институционального комитета по исследовательской этике и Хельсинкской декларации 1964 года и ее последующим изменениям или сопоставимым нормам этики.

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Poon L.L., Song T., Rosenfeld R., Lin X., Rogers M.B., Zhou B., Sebra R., Halpin R.A., Guan Y., Twaddle A., DePasse J.V., Stockwell T.B., Wentworth D.E., Holmes E.C., Greenbaum B., Peiris J.S., Cowling B.J., Ghedin E. (2016) Quantifying influenza virus diversity and transmission in humans. *Nat. Genet.* **48**, 195–200.
2. Desselberger U., Racaniello V.R., Zazra J.J., Palese P. (1980) 3'-Terminal and 5'-terminal sequences of influenza-A, influenza-B and influenza-C virus-RNA segments are highly conserved and show partial inverted complementarity. *Gene*. **8**, 315–328.
3. Hoffmann E., Stech J., Guan Y., Webster R.G., Perez D.R. (2001) Universal primer set for the full-length amplification of all influenza A viruses. *Arch. Virol.* **146**, 2275–2289.
4. Meinel D.M., Heinzinger S., Eberle U., Ackermann N., Schonberger K., Sing A. (2018) Whole genome sequencing identifies influenza A H3N2 transmission and offers superior resolution to classical typing methods. *Infection*. **46**, 69–76.
5. Zhou B., Donnelly M.E., Scholes D.T., St George K., Hatta M., Kawaoka Y., Wentworth D.E. (2009) Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza A viruses. *J. Virol.* **83**, 10309–10313.
6. Inoue E., Wang X., Osawa Y., Okazaki K. (2010) Full genomic amplification and subtyping of influenza A virus using a single set of universal primers. *Microbiol. Immunol.* **54**, 129–134.
7. McGinnis J., Laplante J., Shudt M., George K.S. (2016) Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. *J. Clin. Virol.* **79**, 44–50.
8. Lee H.K., Lee C.K., Tang J.W., Loh T.P., Koay E.S. (2016) Contamination-controlled high-throughput whole genome sequencing for influenza A viruses using the MiSeq sequencer. *Sci Rep.* **6**, 33318.
9. Shagin D.A., Lukyanov K.A., Vagner L.L., Matz M.V. (1999) Regulation of average length of complex PCR product. *Nucl. Acids Res.* **27**, e23.
10. Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477.
11. Hunt M., Gall A., Ong S.H., Brener J., Ferns B., Goulder P., Nastouli E., Keane J.A., Kellam P., Otto T.D. (2015) IVA: accurate de novo assembly of RNA virus genomes. *Bioinformatics*. **31**, 2374–2376.
12. Chevreaux B., Pfisterer T., Drescher B., Driesel A.J., Muller W.E., Wetter T., Suhai S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–1159.

INCREASING THE UNIFORMITY OF GENOME FRAGMENT COVERAGE FOR HIGH-THROUGHPUT SEQUENCING OF INFLUENZA A VIRUS

Y. V. Mikhaylova^{1,*}, A. A. Shelenkov¹, Y. G. Yanushevich¹, and D. A. Shagin^{1,2}

¹*Central Research Institute of Epidemiology of the Federal Service on Customers, Moscow, 111123 Russia*

²*Pirogov Russian National Research Medical University, Moscow, 117997 Russia*

**e-mail: mihailova@cmd.su*

High variability of the influenza A virus poses a significant threat to public health, thereby monitoring of viral strains and studying their genetic properties represent important task. One part of this monitoring includes sequencing of influenza A viruses of any subtype and analysis of their whole genomes, which is especially important in case of interspecies adaptation and reassortment events of influenza viruses. High-throughput sequencing technologies have significantly extended the capabilities of influenza virus epidemiological surveillance. Preparation stages for NGS of influenza A virus include whole genome amplification using one-step RT-PCR, the results of which vary greatly depending on the sample type and quality, that, in turn, affects the coverage of virus fragments and sequencing results in general. In this work, we propose to supplement the aforementioned technique of whole genome amplification of influenza A virus with sequential suppression PCRs to obtain an even coverage of viral segments having different lengths, which allows to sequence samples with lower read coverage without decreasing sequencing quality.

Keywords: influenza virus A, whole-genome sequencing, suppression PCR