



Совершенствование методики SNP-типирования штаммов *Vibrio cholerae* на основе анализа первичных данных полногеномного секвенирования

А.С. Водопьянов[✉], Р.В. Писанов, С.О. Водопьянов, И.П. Олейников

ФКУЗ «Ростовский-на-Дону противочумный институт» Роспотребнадзора, Ростов-на-Дону, Россия

Цель работы — совершенствование метода оценки качества единичных нуклеотидных замен, используемых для SNP-типирования, на основе анализа их распределения в первичных данных полногеномного секвенирования (ридах).

Материалы и методы. В работе использованы данные полногеномного секвенирования 56 штаммов *Vibrio cholerae*, полученные на секвенаторах разных типов. Программное обеспечение разрабатывали на языке программирования Java. Кластерный анализ и построение дендрограммы проведены с использованием авторского программного обеспечения по методу UPGMA.

Результаты и обсуждение. Показана «нестабильность» определения ряда SNP в геноме возбудителя холеры. Разработан метод подбора перечня SNP для филогенетического анализа на основе обработки первичных данных полногеномного секвенирования (ридов). Предложена методика использования «контрольных геномов» при проведении кластерного анализа данных полногеномного секвенирования.

Заключение. Составлен перечень из 3198 «стабильных SNP» для проведения филогенетического анализа. Показана генетическая близость нетоксигенных штаммов, содержащих ген *tcpA* (*ctxAB-tcpA*⁺), и *preCTX*-штаммов *V. cholerae*.

Ключевые слова: холера; генотипирование; секвенирование; филогенетический анализ; единичные нуклеотидные замены.

Источник финансирования. Авторы заявляют об отсутствии финансирования при проведении исследования.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Для цитирования: Водопьянов А.С., Писанов Р.В., Водопьянов С.О., Олейников И.П. Совершенствование методики SNP-типирования штаммов *Vibrio cholerae* на основе анализа первичных данных полногеномного секвенирования. *Журнал микробиологии, эпидемиологии и иммунобиологии.* 2020; 97(6): 587–593.

DOI: <https://doi.org/10.36233/0372-9311-2020-97-6-9>

Поступила 10.07.2019
Принята в печать 15.09.2020

Improvement of the technique of SNP-typing of *Vibrio cholerae* strains on the basis of the analysis of the primary data of whole genome sequencing

Alexey S. Vodopianov[✉], R.V. Pisanov, Sergey O. Vodopianov, Igor P. Oleynikov

Rostov-on-Don Research Institute for Plague Control, Russia, 344019, Rostov-on-Don, Russia

Aim. To improve the method of the quality assessment of single nucleotide polymorphisms, which are used for SNP-typing, based on the analysis of their distribution in the primary data of whole genome sequencing (reads).

Materials and methods. Data of the whole genome sequencing of 56 *Vibrio cholerae* strains obtained using different types of sequencers were used. The software was developed using Java programming language. Cluster analysis and construction of the dendrogram were performed with the author's software using the UPGMA method.

Results and discussion. The «instability» of detection the number of SNP in the genome of cholera causative agent was shown. The method of selection of the SNP list for phylogenetic analysis based on the analysis of the primary data of whole genome sequencing (reads), has been developed. The method of using «control genomes» for cluster analysis of whole genome sequencing data has been proposed.

Conclusion. The list of 3198 «stable SNP» for phylogenetic analysis has been composed. Genetic affinity between the non-toxicogenic strains that contain the *tcpA* gene (*ctxAB-tcpA*⁺) and *preCTX*-strains of *V. cholerae* was shown.

Keywords: *cholera; genotyping; sequence; phylogenetic analysis; single nucleotide polymorphisms.*

Acknowledgments. The study had no sponsorship.

Conflict of interest. The authors declare no apparent or potential conflicts of interest related to the publication of this article.

For citation: Vodopianov A.S., Pisanov R.V., Vodopianov S.O., Oleynikov I.P. Improvement of the technique of SNP-typing of *Vibrio cholerae* strains on the basis of the analysis of the primary data of whole genome sequencing. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, èpidemiologii i immunobiologii*. 2020; 97(6): 587–593. (In Russ.).
 DOI: <https://doi.org/10.36233/0372-9311-2020-97-6-9>

Received 10 July 2019

Accepted 15 September 2020

Введение

Методы генотипирования на основе анализа распределения единичных нуклеотидных замен (single nucleotide polymorphism, SNP) возбудителей опасных инфекционных болезней, в том числе холеры, широко используются в научных исследованиях и эпидемиологическом анализе, поскольку позволяют оценить филогенетические связи между разными штаммами, установить их возможное происхождение, а также источники и пути распространения инфекции. Исследователи используют различные наборы SNP-маркеров, что может приводить к расхождению результатов даже при изучении одного набора штаммов. Например, ранее анализ данных полногеномного секвенирования (whole genome sequencing, WGS) позволил выявить, что клинические штаммы *Vibrio cholerae* O1, выделенные в 2010 г. в Москве (завоз из Индии), относятся к штаммам «гаитянской группы» [1], однако при анализе с использованием другого набора SNP они попали в группу «непальских штаммов», дистанцированную от вызвавших вспышку на острове Гаити [2].

В ряде случаев набор SNP-маркеров строится непосредственно для анализируемого набора штаммов. Например, именно такой алгоритм используется в программе kSNP 3.0 [3], что было использовано для генотипирования штаммов *V. cholerae*, выделенных в Демократической Республике Конго [4].

Это подчеркивает актуальность работ по совершенствованию и стандартизации подходов к анализу полногеномных сиквенсов [5] и, соответственно, отбору SNP-маркеров, используемых для генотипирования.

Как правило, результаты WGS представлены «ридами» — последовательностями ДНК размером 100–300 п.о., которые многократно дублируются. В дальнейшем с помощью специализированных программ проводится их сборка в контиги, размеры которых могут достигать сотен тысяч нуклеотидов. При этом существенно сокращается общий объем: так, если набор ридов составляет около 600–900 Мб, то суммарный объем контигов, как правило, не превышает 7 Мб. Это приводит к тому, что большинство исследователей предпочитают работать именно с контигами, а не с ридами.

Однако такой подход имеет существенный недостаток, заключающийся в обработке ошибок. Важно отметить, что если в ридах присутствуют замены нуклеотидов, то программа-сборщик в качестве «правильного» выбирает самый часто встречающийся вариант, вне зависимости от того, встречается он в 99% или в 50% случаев. Таким образом, при работе с контигами может теряться весьма важная информация о достоверности выявляемых SNP.

В связи с этим **цель** работы состояла в разработке метода оценки качества SNP, используемых для SNP-типирования, на основе анализа их распределения в первичных данных WGS (ридах).

Материалы и методы

В работе использованы данные WGS 56 штаммов разных серогрупп, полученные на платформе «MiSeq Illumina» в лаборатории диагностики особо опасных инфекций ФКУЗ «Ростовский-на-Дону противочумный институт», и геномы, включенные в локальную базу геномов, секвенирование которых проведено другими группами исследователей (табл. 1).

Программное обеспечение разрабатывали на языке программирования Java. Кластерный анализ и построение дендрограммы проводили с использованием обеспечения по методу UPGMA. Для построения дендрограммы использовали программу MEGA 5 [6].

Результаты и обсуждение

Для анализа был использован набор SNP-маркеров, выявленный ранее при анализе данных WGS штаммов *V. cholerae* [7]. Для этого нами было разработано программное обеспечение, позволяющее оценивать встречаемость каждого SNP непосредственно в наборе ридов. Для каждого изучаемого SNP была посчитана встречаемость в наборе данных первичного секвенирования (ридах) каждого нуклеотида. Примеры результатов анализа нескольких SNP-маркеров представлены в табл. 2, из которой можно видеть, что SNP в позиции 316 в гене *VC0275* штамма *V. cholerae* 81 представлена именно тиминотом (Т), который обнаружен в 79 ридах, в то время как наличие ридов с аденином (А) и

Таблица 1. Геномы штаммов, использованные в работе
Table 1. Genomes of the strains used in the study

Штамм Strain	Место выделения Place of isolation	Год выделения Year of isolation	Штамм Strain	Место выделения Place of isolation	Год выделения Year of isolation
<i>V. cholerae</i> O1 CTX⁺VPI-I⁺			<i>V. cholerae</i> O1 CTX⁺VPI-I⁺		
16228	Республика Дагестан, Россия Republic of Dagestan, Russia	1994	18785	Республика Саха, Россия Republic of Sakha, Russia	2003
17290	Республика Дагестан, Россия Republic of Dagestan, Russia	1994	2687*	Республика Калмыкия, Россия Republic of Kalmykia, Russia	2015
17296	Республика Дагестан, Россия Republic of Dagestan, Russia	1994	94	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2018
1786	Гаити Haiti	2010	M1501*	Республика Калмыкия, Россия Republic of Kalmykia, Russia	2011
18329	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2001	M1518*	Республика Калмыкия, Россия Republic of Kalmykia, Russia	2012
18369	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2007	M1524*	Республика Калмыкия, Россия Republic of Kalmykia, Russia	2013
301	Таганрог, Россия Taganrog, Russia	2001	МАК97*	Индонезия Indonesia	1937
31*	Мариуполь, Украина Mariupol, Ukraine	2011	P-18778*	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2005
3265_80*	Москва, Россия Moscow, Russia	2014	P18899-D*	Тверь, Россия Tver, Russia	2006
3569-08*	США USA	2008	RND18899*	Тверь, Россия Tver, Russia	2006
39*	Мариуполь, Украина Mariupol, Ukraine	2011	<i>V. cholerae</i> O1 CTX⁺VPI-I⁺		
43*	Мариуполь, Украина Mariupol, Ukraine	1994	18984	Краснодарский край, Россия Krasnodar Kray, Russia	2007
56*	Мариуполь, Украина Mariupol, Ukraine	1994	19178	Иркутская область, Россия Irkutsk region, Russia	2015
81-S*	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2014	19308	Астраханская область, Россия Astrakhan region, Russia	2012
81c	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2014	20000	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2016
CIRS101*	Бангладеш Bangladesh	2002	433*	Сочи, Россия Sochi, Russia	2015
E506*	США USA	1974	434	Сочи, Россия Sochi, Russia	2015
HC-23A1*	Гаити Haiti	2010	CP1037(10)*	Мексика Mexico	2003
МАК676*	Индонезия Indonesia	1937	Env-390*	Гаити Haiti	2012
MJ-1236*	Бангладеш Bangladesh	1994	M1522*	Республика Калмыкия, Россия Republic of Kalmykia, Russia	2014
MS6*	Тайланд Thailand	2008	<i>V. cholerae</i> O1 preCTX⁺		
P13762	Ташкент, Узбекистан Tashkent, Uzbekistan	1988	pre_13767	Узбекистан Uzbekistan	1998
RND6878*	Москва, Россия Moscow, Russia	2012	pre_15500	Крым, Украина Crimea, Ukraine	1991
VC51*	Индия India	1992	pre_18963	Ростов-на-Дону, Россия Rostov-on-Don, Russia	2007
<i>V. cholerae</i> O139 CTX⁺VPI-I⁺			pre_9961	Москва, Россия Moscow, Russia	1977
4260B*	Бангладеш Bangladesh	1993	<i>V. cholerae</i> nonO1/nonO139		
E306*	Китай China	2013	16002	Узбекистан Uzbekistan	1971
MO10*	Бангладеш Bangladesh	1992	CP1117*	США USA	2010
VC4370*	Малайзия Malaysia	2008	V51*	США USA	1987
			V52*	Судан Sudan	1968
			nag16150	Узбекистан Uzbekistan	1987

Примечание. *Геномы получены из базы данных NCBI.
Note. *Genomes obtained from the NCBI database.

Таблица 2. Оценка встречаемости единичных нуклеотидных замен в наборе ридов штаммов *V. cholerae* 81 и *V. cholerae* HC-72 (фрагмент). Указано количество ридов, в которых встречается тот или иной нуклеотид**Table 2.** Estimation of the occurrence of single nucleotide polymorphisms in the read set of strains *V. cholerae* 81 and *V. cholerae* HC-72 (fragment). The number of rows in which a particular nucleotide occurs is indicated

N	Ген Gene	Позиция Position	<i>V. cholerae</i> 81				<i>V. cholerae</i> HC-72			
			A	T	G	C	A	T	G	C
1	VC0275	316	1	79	0	2	0	126	0	0
2	VC0289	564	36	0	0	0	134	1	0	0
3	VC0321	334	0	0	0	207	0	0	0	443
4	VC0345	1047	0	0	1	116	0	0	0	176
5	VC0362	833	148	0	154	1	268	1	241	0
6	VC0362	944	0	133	1	133	0	226	0	206

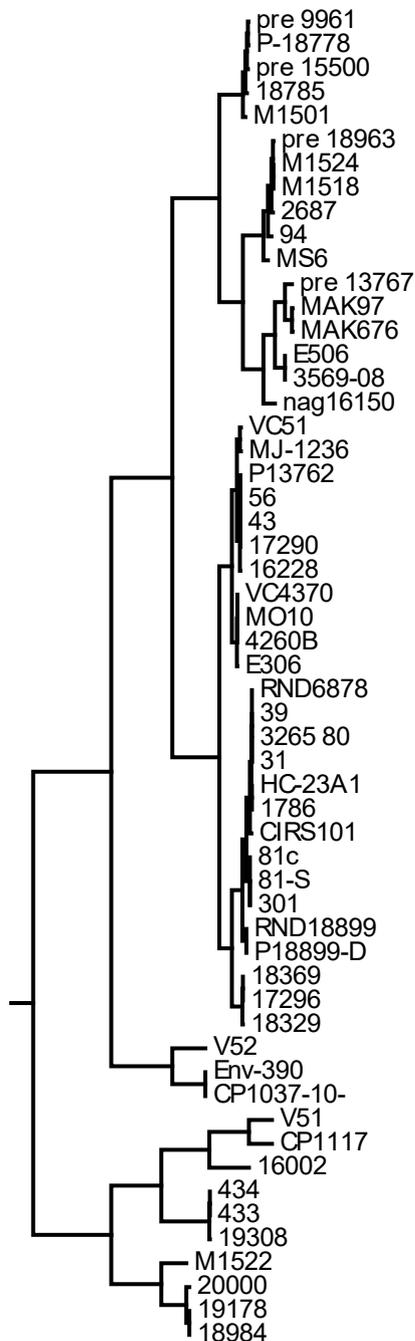
цитозином (2 рида) является ошибкой секвенирования. Аналогичные результаты получены и для генов *VC0289*, *VC0321* и *VC0345*.

В противовес этому для гена *VC0362* почти половина ридов у штамма *V. cholerae* 81 в позиции 833 содержала аденин (148 ридов), а половина (154 рида) — гуанин. Такое распределение не позволяет рассматривать это как ошибку секвенирования, тем более что аналогичное распределение наблюдается у штамма *V. cholerae* HC-72, WGS которого проведено другой группой исследователей. Причиной такой «нестабильности» нуклеотидов может являться существование в геноме нескольких копий гена или сходных нуклеотидных мотивов в разных генах, содержащих замены, однако это требует дальнейшего изучения. Тем не менее, вне зависимости от причины, при сборке контигов выбор «итонового» нуклеотида может быть практически случайным и зависеть от незначительного преобладания ридов с тем или иным нуклеотидом. Ярким примером может служить вышеописанный SNP в гене *VC0362*: так, у штамма *V. cholerae* 81 преобладают риды, содержащие в позиции 833 гуанин, а у *V. cholerae* HC-72 — аденин. Нам представляется нецелесообразным использовать при анализе подобных «нестабильные» SNP, в связи с чем описанный ранее перечень [7] был сокращен с 3683 до 3198 SNP путем удаления SNP, в которых наиболее часто встречаемый вариант обнаруживался менее чем в 70% случаев. Также очевидно, что выбор схемы SNP-типирования должен осуществляться, исходя из анализа первичных данных секвенирования (ридов), а не результата сборки (контигов).

Довольно информативным методом кластерного анализа является UPGMA — метод невзвешенных парных групп со средним арифметическим [8–11]. Сравнительный анализ показывает конкордантность результатов UPGMA и других методов анализа, в частности, методов минимального остовного дерева и максимальной бережливости [12, 13].

На основе кластерного анализа по методу UPGMA с использованием предлагаемого набора SNP нами была построена дендрограмма, отражающая генетическую близость между геномами 56 штаммов *V. cholerae* различного происхождения (**рисунок**). Ранее мы предложили использовать в качестве контрольных образцов при проведении биоинформационного анализа штаммы, геномы которых были секвенированы различными группами авторов, либо штаммы, генетическая близость которых не вызывает сомнений. При этом критерием корректности проведения анализа является попадание «контрольных геномов» в один кластер [7]. В данной работе в качестве таких контролей мы использовали штамм *V. cholerae* O1 81, геном которого секвенирован дважды: на платформе MiSeq (ФКУЗ «Ростовский-на-Дону противочумный институт») и платформе IonTorrent (ФКУЗ РосНИПЧИ «Микроб»). Как видно из дендрограммы, оба контрольных штамма попадают в одну ветку, что свидетельствует о корректности проведения кластерного анализа. Аналогичное распределение получено и для двух других пар контрольных геномов: штамма *V. cholerae* O1 18899 и его изогенного варианта, лишённого гена холерного токсина, и двух относящихся к одному клональному комплексу штаммов *V. cholerae* O1 № 433 и № 434, выделенных из воды реки Агура в 2015 г. (секвенированных в РосНИПЧИ «Микроб» и ФКУЗ «Ростовский-на-Дону противочумный институт» соответственно на разных платформах).

Использование нового набора SNP-маркеров позволило более достоверно установить родственные связи между токсигенными штаммами O1 (биовара El Tor) и O139 серогрупп, содержащих разные наборы маркеров эпидемического потенциала: аллели генов *ctxB* (*ctxB3*, *ctxB1* или *ctxB7*), *tcpA* (*eltor* или *CIRS*), *rtxA* (*rtxA1*, *rtxA4* или *rtxA4a*), структура острова пандемичности *VSP11* (интактный или содержащий протяженную делецию — *VSP11Δ*). Все



Дендрограмма, построенная на основе SNP-типирования.

A dendrogram based on SNP typing.

они сгруппировались в один большой кластер из двух основных ветвей. Наиболее опасные штаммы с генотипом *ctxB7 tcpA^{CIRS} rtxA4a VSPIIΔ* и близким ему *ctxB1 tcpA^{CIRS} rtxA4 VSPIIΔ* попали в одну ветвь с гаитянскими штаммами и штаммом *CIRS101* из Бангладеш; отдельную группу в этой ветви образовали штаммы *ctxB1 tcpA^{eltor} rtxA1 VSPII* из Ростова-на-Дону, Казани и Дагестана. В то же время другие дагестанские штаммы с таким же генотипом оказались в составе другой ветви, куда отдельной

группой вошли и 4 штамма серогруппы O139, что согласуется с более ранними данными на основе INDEL-типирования [1].

Особый интерес представляют нетоксигенные штаммы, лишенные генов *ctxAB* и не склонные к эпидемическому распространению. Вместе с тем они могут вызывать спорадические случаи и локальные вспышки холероподобной диареи за счет экспрессии ряда детерминант факторов патогенности. Часть из них содержит остров патогенности *VPI-I*, в состав которого входит ген *tcpA*, а некоторые из последних — профаг *preCTX*, их вирулентность показана в опытах *in vivo* [14]. Их происхождение, пути эволюции, а также возможности приобретения генов холерного токсина изучены крайне недостаточно.

По результатам SNP-типирования с использованием другого набора маркеров сотрудниками РосНИПЧИ «Микроб» [15] показано, что штаммы с генотипом *ctxAB⁻tcpA⁺VSPI⁺VSPII⁺* близкородственны токсигенным штаммам. На этом основании авторами было предложено считать их потенциально эпидемически опасными, т.е. способными приобрести профаг *CTX* путем «традиционной» и TSP-зависимой трансдукции и восстановить эпидемический потенциал. В то же время штаммы с генотипом *ctxAB⁻tcpA⁺VSPI⁻VSPII⁻* даже в случае приобретения ими *CTX* эпидемически опасными стать не могут, поскольку на дендрограмме они образовали отдельный удаленный от токсигенных штаммов кластер. Третий отдаленный от обоих указанных кластер был образован штаммами *ctxAB⁻tcpA⁻VSPI⁻VSPII⁻*. Эти данные позволили оценить сходство и различия между штаммами с разными наборами генетических детерминант и проследить пути их эволюции.

Вместе с тем в анализ была включена большая, но все же ограниченная выборка штаммов, и некоторые генотипы оказались за ее пределами. Мы включили в исследование ряд других нетоксигенных штаммов и установили, что они образуют большее число кластеров на дендрограмме и иногда группируются с токсигенными. Например, штаммы *ctxAB⁻tcpA⁺VSPI⁻VSPII⁻*, выделенные из водоемов в Ростове-на-Дону и Калмыкии, попали в один кластер со штаммом *ctxAB⁺tcpA⁺VSPI⁺VSPII⁻V. cholerae* O1 MS6 из Тайланда [16]. При этом один из штаммов, выделенных в Ростовской области, (18963) содержал *preCTX*. Два других штамма *preCTX⁺tcpA⁺VSPI⁻VSPII⁻* сгруппировались со штаммами *preCTX⁻tcpA⁺VSPI⁻VSPII⁻* — представителями клонального комплекса, вызвавшими вспышку заболеваний в Каменском районе Ростовской области в 2005 г. [17].

Еще более гетерогенной группой возбудителей являются штаммы *V. cholerae* nonO1/nonO139, в том числе имеющие гены холерного токсина. Например, именно такими штаммами была вызвана продолжительная вспышка в Узбекистане в 1987–1990 гг.

[18]. Настоящее исследование показало, что узбекский штамм *ctxAB⁺tcpA⁺VSPF-VSPII* (16150) попадает в общий гетерогенный кластер, содержащий *preCTX⁺* и *ctxAB⁺* штаммы *V. cholerae* O1. Ряд других токсигенных штаммов *V. cholerae* nonO1/nonO139 (V51, CP1117, 16002) образовали отдаленный от него кластер и оказались ближе к нетоксигенным вариантам *V. cholerae* O1 серогруппы. Штамм V52 *ctxAB⁺tcpA⁺VSPF-VSPII* серогруппы O37 — один из возбудителей крупной эпидемической вспышки в Судане в 1968 г. — оказался близок атоксигенному штамму, выделенному на Гаити в 2012 г.

Очевидно, пути эволюции нетоксигенных штаммов крайне многообразны и требуют отдельного, более детального изучения, чему будет способствовать дальнейшее использование разработанного нами усовершенствованного метода SNP-типирования.

Заключение

В ходе выполнения настоящей работы разработан метод подбора перечня SNP для проведения филогенетического анализа на основе анализа первичных данных WGS (ридов). Составлен перечень из 3198 «стабильных SNP» для проведения филогенетического анализа. Предложена и апробирована методика использования «контрольных геномов» при проведении кластерного анализа данных WGS. Показана генетическая близость между нетоксигенными штаммами, содержащими ген *tcpA* (*ctxAB⁻tcpA⁺*) и *preCTX*-штаммами *V. cholerae*.

ЛИТЕРАТУРА

1. Водопоьянов А.С., Водопоьянов С.О., Олейников И.П., Мишанькин Б.Н. INDEL-типирование штаммов *Vibrio cholerae*. *Эпидемиология и инфекционные болезни*. 2017; 22(4): 195-200.
DOI: <http://doi.org/10.18821/1560-9529-2017-22-4-195-200>
2. Kuleshov K.V., Vodop'yanov S.O., Dedkov V.G., Markelov M.L., Deviatkin A.A., Kruglikov V.D., et al. Travel-associated *Vibrio cholerae* O1 El Tor, Russia. *Emerg. Infect. Dis.* 2016; (11): 2006-8.
DOI: <https://dx.doi.org/10.3201/eid2211.151727>.
3. Gardner S.N., Slezak T., Hall B.G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genomes. *Bioinformatics*. 2015; 31: 2877-8.
DOI: <https://dx.doi.org/10.1093/bioinformatics/btv271>.
4. Ireng L.M., Ambrose J., Mitangala P.N., Bearzatto B., Kabangwa R.K.S., Durant J.F., Gala J.L. Genomic analysis of pathogenic isolates of *Vibrio cholerae* from eastern Democratic Republic of the Congo (2014–2017). *PLoS Negl. Trop. Dis.* 2020; 14(4): e0007642.
DOI: <https://dx.doi.org/10.1371/journal.pntd.0007642>.
5. Миронова Л.В., Балахонов С.В. Полногеномный анализ однонуклеотидных полиморфизмов в изучении молекулярной эпидемиологии холеры и эволюционной истории возбудителя. *Эпидемиология и вакцинопрофилактика*. 2014; 4(77):10-8.
6. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maxi-

- mum parsimony methods. *Mol. Biol. Evol.* 2011; 28: 2731-9.
DOI: <https://dx.doi.org/10.1093/molbev/msr121>.
7. Водопоьянов А.С., Писанов Р.В., Водопоьянов С.О., Мишанькин Б.Н., Олейников И.П., Кругликов В.Д., Титова С.В. Молекулярная эпидемиология *Vibrio cholerae* — разработка алгоритма анализа данных полногеномного секвенирования. *Эпидемиология и инфекционные болезни*. 2016; 21(3): 146-52.
8. García D.F., Astudillo M. MIRU-VNTR genotyping of *Mycobacterium tuberculosis* in a population of patients in Cali, Colombia, 2013–2015. *Biomedica*. 2019; 39(s1): 71-85.
DOI: <https://dx.doi.org/10.7705/biomedica.v39i2.3924>.
9. Shastri A.A., Ahuja K., Ratnaparkhe M.B., Shah A., Gagrani A., Lal A. Vector quantized spectral clustering applied to whole genome sequences of plants. *Evol. Bioinform. Online*. 2019; 15: 1176934319836997.
DOI: <https://dx.doi.org/10.1177/1176934319836997>.
10. Singh R.B., Mahenderakar M.D., Jugran A.K., Singh R.K., Srivastava R.K. Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. *Gene*. 2020; 753: 144800.
DOI: <https://dx.doi.org/10.1016/j.gene.2020.144800>.
11. Subramanian S., Ramasamy U., Chen D. VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs. *PeerJ*. 2019; 7:e8213.
DOI: <https://dx.doi.org/10.7717/peerj.8213>.
12. Chatterjee S., Rudra S.K., Azmi S.A., Bandyopadhyay R. Phylogenetic study based on 28S rRNA gene sequencing of *Wuchereria bancrofti* isolated from the filaria endemic areas of Bankura district, West Bengal, India. *J. Parasit. Dis.* 2017; 41(4): 981-6.
DOI: <https://dx.doi.org/10.1007/s12639-017-0922-6>.
13. Yokoyama E., Hirai S., Ishige T., Murakami S. Application of whole genome sequence data in analyzing the molecular epidemiology of Shiga toxin-producing *Escherichia coli* O157:H7/H. *Int. J. Food Microbiol.* 2018; 264: 39-45.
DOI: <https://dx.doi.org/10.1016/j.ijfoodmicro.2017.10.019>.
14. Монахова Е.В., Миронова А.В., Алексеева Л.П., Мазрухо А.Б. Вирулентность холерных вибрионов, содержащих *pre-CTXφ*: генотипическая и фенотипическая характеристика. *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2008; (4): 27-32.
15. Смирнова Н.И., Кульшань Т.А., Баранихина Е.Ю., Краснов Я.М., Агафонов Д.А., Кутырев В.В. Структура генома и происхождение нетоксигенных штаммов *Vibrio cholerae* биовара Эль Тор с различной эпидемиологической значимостью. *Генетика*. 2016; 52(9): 1029-41.
16. Okada K., Roobthaisong A., Swaddiwudhipong W., Hamada S., Chantaroj S. *Vibrio cholerae* O1 isolate with novel genetic background, Thailand–Myanmar. *Emerg. Infect. Dis.* 2013; 19(6): 1015-7.
DOI: <https://dx.doi.org/10.3201/eid1906.120345>
17. Онищенко Г.Г., Ломов Ю.М., Москвитина Э.А., Подосинникова Л.С., Водяницкая С.Ю., Прометной В.И. и др. Холера, обусловленная *Vibrio cholerae* O1 *ctxAB*–*tcpA*+. *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2007; (1): 23-9.
18. Онищенко Г.Г., Водопоьянов С.О., Ломов Ю.М., Мишанькин Б.Н., Сучков И.Ю., Черепихина И.Я. и др. Холерные вибрионы серогрупп неO1, выделенные в Узбекистане в 1987–1990 гг.: ретроспективный VNTR-анализ. *Эпидемиология и инфекционные болезни*. 2003; (6): 25-9.

REFERENCES

1. Vodop'yanov A.S., Vodop'yanov S.O., Oleynikov I.P., Mishan'kin B.N. INDEL-genotyping of *Vibrio cholerae* strains. *Epidemiologiya i infektsionnye bolezni*. 2017; 22(4): 195-200.
DOI: <http://doi.org/10.18821/1560-9529-2017-22-4-195-200> (in Russian)

2. Kuleshov K.V., Vodop'yanov S.O., Dedkov V.G., Markelov M.L., Deviatkin A.A., Kruglikov V.D., et al. Travel-associated *Vibrio cholerae* O1 El Tor, Russia. *Emerg. Infect. Dis.* 2016; (11): 2006-8. DOI: <https://dx.doi.org/10.3201/eid2211.151727>.
3. Gardner S.N., Slezak T., Hall B.G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genomes. *Bioinformatics.* 2015; 31: 2877-8. DOI: <https://dx.doi.org/10.1093/bioinformatics/btv271>.
4. Ireng L.M., Ambroise J., Mitangala P.N., Bearzatto B., Kambangwa R.K.S., Durant J.F., Gala J.L. Genomic analysis of pathogenic isolates of *Vibrio cholerae* from eastern Democratic Republic of the Congo (2014–2017). *PLoS Negl. Trop. Dis.* 2020; 14(4): e0007642. DOI: <https://dx.doi.org/10.1371/journal.pntd.0007642>.
5. Mironova L.V., Balakhonov S.V. Whole-genome analysis of single-nucleotide polymorphisms in study of cholera modular epidemiology and agent evolutionary history. *Epidemiologiya i vaktinoprofilaktika.* 2014; 4(77): 10-8. (in Russian)
6. Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 2011; 28: 2731-9. DOI: <https://dx.doi.org/10.1093/molbev/msr121>.
7. Vodop'yanov A.S., Pisanov R.V., Vodop'yanov S.O., Mishankin B.N., Oleynikov I.P., Kruglikov V.D., Titova S.V. Molecular epidemiology of *Vibrio cholerae* — development of the algorithm for data analysis of whole genome sequencing. *Epidemiologiya i infektionnyye bolezni.* 2016; 21(3): 146-52. (in Russian)
8. García D.F., Astudillo M. MIRU-VNTR genotyping of *Mycobacterium tuberculosis* in a population of patients in Cali, Colombia, 2013–2015. *Biomedica.* 2019; 39(s1): 71-85. DOI: <https://dx.doi.org/10.7705/biomedica.v39i2.3924>.
9. Shastri A.A., Ahuja K., Ratnaparkhe M.B., Shah A., Gagrani A., Lal A. Vector quantized spectral clustering applied to whole genome sequences of plants. *Evol. Bioinform. Online.* 2019; 15: 1176934319836997. DOI: <https://dx.doi.org/10.1177/1176934319836997>.
10. Singh R.B., Mahenderakar M.D., Jugran A.K., Singh R.K., Srivastava R.K. Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. *Gene.* 2020; 753: 144800. DOI: <https://dx.doi.org/10.1016/j.gene.2020.144800>.
11. Subramanian S., Ramasamy U., Chen D. VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs. *PeerJ.* 2019; 7 :e8213. DOI: <https://dx.doi.org/10.7717/peerj.8213>.
12. Chatterjee S., Rudra S.K., Azmi S.A., Bandyopadhyay R. Phylogenetic study based on 28S rRNA gene sequencing of *Wuchereria bancrofti* isolated from the filaria endemic areas of Bankura district, West Bengal, India. *J. Parasit. Dis.* 2017; 41(4): 981-6. DOI: <https://dx.doi.org/10.1007/s12639-017-0922-6>.
13. Yokoyama E., Hirai S., Ishige T., Murakami S. Application of whole genome sequence data in analyzing the molecular epidemiology of Shiga toxin-producing *Escherichia coli* O157:H7/H. *Int. J. Food Microbiol.* 2018; 264: 39-45. DOI: <https://dx.doi.org/10.1016/j.ijfoodmicro.2017.10.019>.
14. Monakhova E.V., Mironova A.V., Aickscova L.P., Mazrukh A.B. Virulence of pre-ctx9-carrying *Vibrio cholerae*: genotypic and pheno-typic characteristics. *Zhurnal mikrobiologii, epidemiologii i immunobiologii.* 2008; (4): 27-32. (in Russian)
15. Smirnova N.I., Kul'shan' T.A., Baranikhina E.Y., Krasnov Y.M., Agafonov D.A., Kuttyrev V.V. Genome structure and origin of nontoxic strains of *Vibrio cholerae* of El Tor biovar with different epidemiological significance. *Russian Journal of Genetics.* 2016; 52(9): 914-25.
16. Okada K., Roobthaisong A., Swaddiwudhipong W., Hamada S., Chantaroj S. *Vibrio cholerae* O1 isolate with novel genetic background, Thailand–Myanmar. *Emerg. Infect. Dis.* 2013; 19(6): 1015-7. DOI: <https://dx.doi.org/10.3201/eid1906.120345>
17. Onishchenko G.G., Lomov Yu.M., Moskvitina E.A., L S. Podosinnikova, Vodyanitskaya S.Yu., Prometnoy V.I., et al. Cholera caused by *Vibrio cholerae* O1 ctxAB– tcpA+. *Zhurnal mikrobiologii, epidemiologii i immunobiologii.* 2007; (1): 23-9. (in Russian)
18. Onishchenko G.G., Mishankin B.N., Vodopyanov A.S., Lomov Yu.M., Voronezhskaya L.G., Suchkov I.Yu., et al. Serogroups non-O1 *Vibrio cholerae* isolated in Uzbekistan in 1987–1990: a retrospective VNTR analysis. *Epidemiologiya i infektionnyye bolezni.* 2003; (6): 25-9. (in Russian)

Информация об авторах:

Водопьянов Алексей Сергеевич[✉] — к.м.н., с.н.с. группы вирусологии ФКУЗ «Ростовский-на-Дону противочумный институт», 344002, Ростов-на-Дону, Россия.
ORCID ID: <https://orcid.org/0000-0002-9056-3231>.
E-mail: alexvod@gmail.com

Писанов Руслан Вячеславович — к.б.н., зав. лаб. диагностики особо опасных инфекций ФКУЗ «Ростовский-на-Дону противочумный институт», 344002, Ростов-на-Дону, Россия.
ORCID ID: <http://orcid.org/0000-0002-7178-8021>.

Водопьянов Сергей Олегович — д.м.н., зав. лаб. биохимии микробов ФКУЗ «Ростовский-на-Дону противочумный институт», 344002, Ростов-на-Дону, Россия.
ORCID ID: <https://orcid.org/0000-0003-4336-0439>.

Олейников Игорь Павлович — н.с. лаб. биохимии микробов ФКУЗ «Ростовский-на-Дону противочумный институт», 344002, Ростов-на-Дону, Россия.
ORCID ID: <https://orcid.org/0000-0002-2390-9773>.

Участие авторов: все авторы сделали эквивалентный вклад в подготовку публикации.

Information about the authors:

Alexey S. Vodop'yanov[✉] — PhD (Med.), senior researcher, Virology group, Rostov-on-Don Plague Control Research Institute, 344002, Rostov-on-Don, Russia.
ORCID ID: <https://orcid.org/0000-0002-9056-3231>.
E-mail: alexvod@gmail.com

Ruslan V. Pisanov — PhD (Biol.), Head, Department of epidemiology of especially targeted infections, Rostov-on-Don Research Institute for Plague Control, 344019, Rostov-on-Don, Russia.
ORCID ID: <http://orcid.org/0000-0002-7178-8021>.

Sergey O. Vodop'yanov — D. Sci. (Med.), leading researcher, Head, Department of microbial chemistry, Rostov-on-Don Research Institute for Plague Control, 344019, Rostov-on-Don, Russia.
ORCID ID: <https://orcid.org/0000-0003-4336-0439>.

Igor P. Oleynikov — researcher, Department of microbial chemistry, Rostov-on-Don Research Institute for Plague Control, 344019, Rostov-on-Don, Russia.
ORCID ID: <https://orcid.org/0000-0002-2390-9773>.

Contribution: the authors contributed equally to this article.