

УДК 578.832.1-579.22

НОВЫЕ НЕГАТИВНО-ПОЛЯРНЫЕ ГЕНЫ В РНК-ГЕНОМЕ КОРОНАВИРУСОВ

© 2021 г. Член-корреспондент РАН О. П. Жирнов^{1,2,*}, С. В. Поярков¹

Поступило 29.09.2020 г.
После доработки 07.10.2020 г.
Принято к публикации 09.10.2020 г.

Семейство коронавирусов образует липидсодержащие оболочечные вирусы, имеющие одноцепочечный РНК геном, который у разных вирусов кодирует 25–30 белков по механизму позитивно-полярной стратегии. Дополнительно в геноме коронавирусов обнаружены протяженные открытые рамки трансляции (ОРТ; гены), которые закодированы в негативно-полярной ориентации. Размер негативно-полярных генов колеблется в диапазоне 150–450 нк, что соответствует полипептидам негативно-полярных генов (negative gene protein, NGP) с м.м. $5-30 \times 10^3$ Да. Коронавирусы имеют заметные различия в количестве выявленных негативных генов. Наличие в геноме коронавирусов негативно-полярных генов дает основание рассматривать это семейство как вирусы с амбисенс стратегией генома.

Ключевые слова: коронавирусы, геном, амбисенс гены, полярность генома

DOI: 10.31857/S268673892101025X

Коронавирусы вызывают повышенный интерес в связи с угрожающим развитием пандемии у людей, вызванной штаммом SARS-CoV2 [1, 2]. Вирусы этого семейства имеют наружную липидную оболочку (оболочечные вирусы) и одноцепочечную геномную РНК, которая имеет классическую позитивную полярность и кодирует у разных вирусов 25–30 специфических белков [3]. Экспрессия вирусных белков происходит в две фазы. В первой фазе транслируется 5'-концевая часть генома с образованием 19 неструктурных белков, ns1–ns19 (см. рис. 1а). Во второй фазе транслируется 3'-концевая часть генома посредством промежуточной репликации и транскрипции генома с образованием индивидуальных мРНК, кодирующих структурные белки N, M, S, E, HE и ряд неструктурных вспомогательных белков (см. рис. 1а).

При анализе *in silico* геномной РНК коронавирусов мы обнаружили протяженные открытые рамки трансляции (ОРТ), которые начинались с кодона AUG и заканчивались классическими терминирующими кодонами UAA или UAG (табл. 1; рис. 1б). Анализ негативно-полярных зон предшествующих позиции AUG выявленных генов, в частности, наибольшего гена NGP4 (the negative gene protein 4) (нк 6137–6489) в геноме вируса SARS-CoV2, выполненный с помощью про-

граммы компьютерного предсказания рибосома-связывающих элементов, показал наличие участков, имеющих выраженную вторичную структуру с множеством шпилек и согласно критериям программы [HTTP://bioinfo.net.in/IRESPred](http://bioinfo.net.in/IRESPred) [5], обладающих высокой энергетической стабильностью (свободная энергия 85 ккал/ммоль) и структурными свойствами сайта входа рибосом, так называемого IRES (Internal Ribosome Entry Site) (рис. 1в). Такая структура 5'-прилежащей к AUG зоны может обеспечить узнавание мРНК рибосомами и последующую трансляцию белка [5]. Более того, в этой IRES-подобной зоне выявлялись два дополнительных AUG и три альтернативных CUG инициаторных кодонов в фазе трансляции +1 и/или +2, что также могло облегчать распознавание и экспрессию этого гена рибосомами по механизму сканирования [4].

Протяженность обнаруженных генов (ОРТ) варьировала в диапазоне 150–450 нуклеотидов (нк), что могло обеспечить синтез полипептидов с м.м. в 5–30 кДа. Сравнение геномов различных представителей семейства коронавирусов показало значительное разнообразие как по количеству таких негативно-полярных генов, так и по паттерну их локализации в вирусном геноме (см. табл. 1; рис. 1б). Так, у вирусов Pangolin-CoV и SARS-CoV2, которые, согласно современным представлениям, являются ближайшими родственниками (т.е. поколениями одного предшественника), обнаруживалось 29 и 21 негативных генов, соответственно, при отсутствии совпадения их позиций в геноме (рис. 1б). Напротив, при

¹ АНО Русско-немецкая академия медикосоциальных и биотехнологических наук, Москва, Россия

² Институт вирусологии им. Д.И. Ивановского ФГБУ НЦЭМ им. Н.Ф. Гамалеи, МЗ РФ, Москва, Россия

*e-mail: zhirnov@inbox.ru

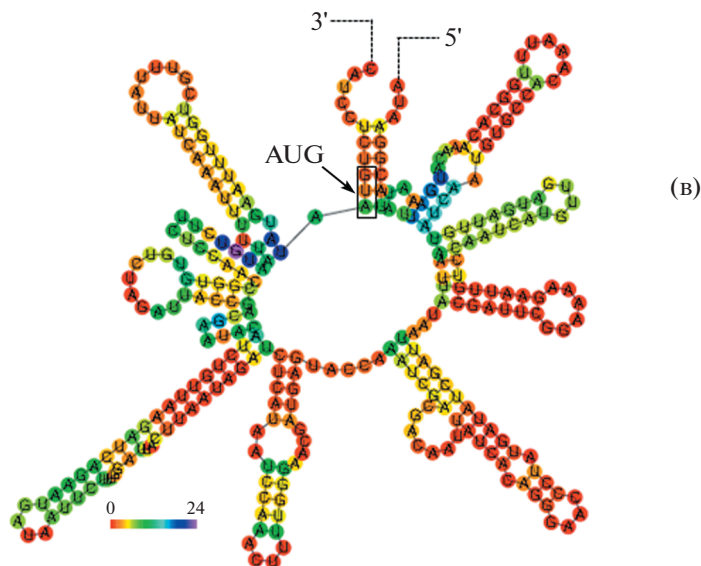
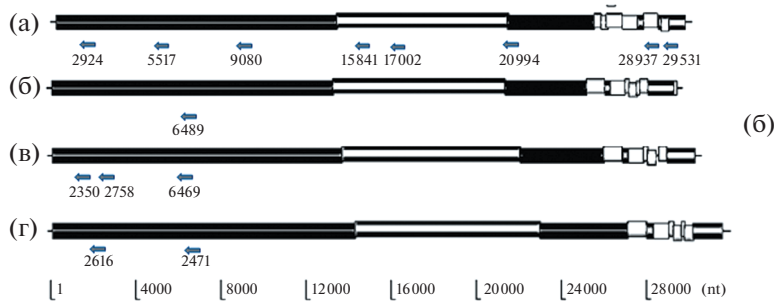
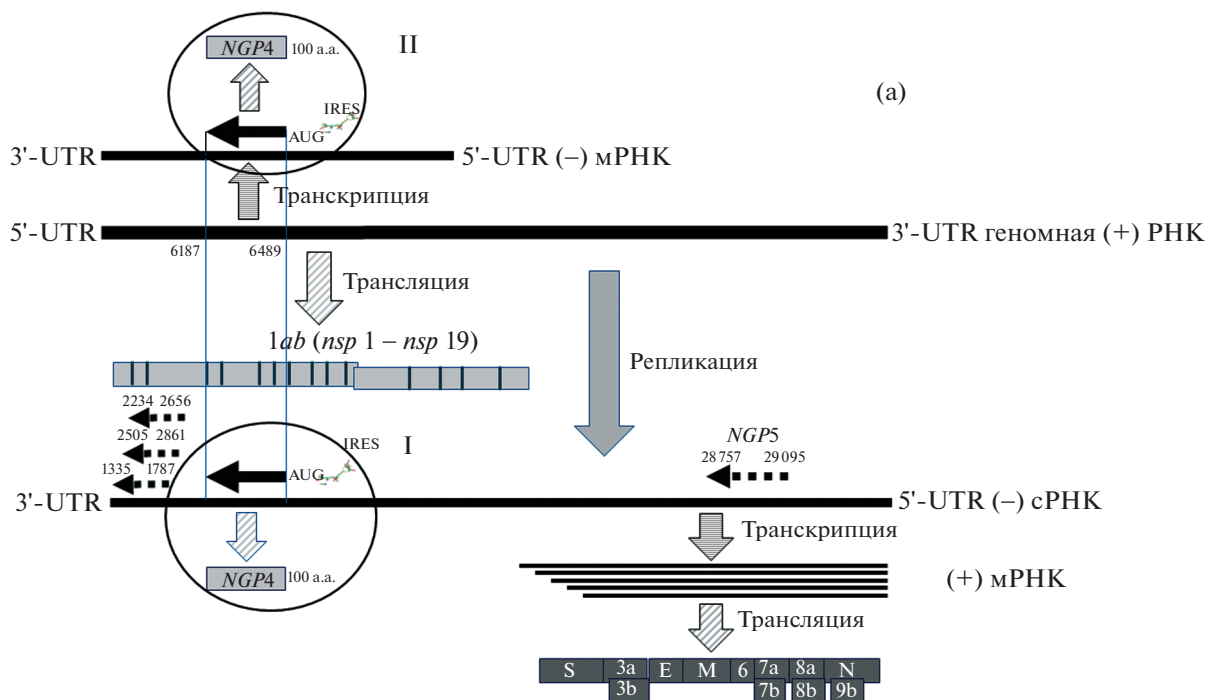


Таблица 1. Количественная характеристика выявления негативно-полярных генов (ОПТ) в геномах корона-вирусов

Вирусы ⁽¹⁾	Количество малых негативно-полярных генов с ATG/CUG кодонами 150–300 нк ⁽²⁾	Количество больших негативно-полярных генов с AUG/CUG кодонами ≥ 300 нк	Диапазон мол. м. для больших полипептидов (NGP) (кДа 10 ⁻³) ⁽³⁾
Альфа-коронавирусы корона-вирус человека HCoV-229E	29/29	1/5	12.37–14.41
Бета-коронавирусы			
SARS-CoV	34/35	0/2	11.47–14.98
MERS-CoV	32/23	8/3	11.07–18.55
SARS-CoV2	21/26	1/4	10.85–17.18
Pangolin-CoV	29/17	3/4	10.8–19.9
BAT-CoV	2/1	17/29	10.95–19.27
HKU1	15/13	1/2	13.1–21.8
Гамма-коронавирусы Avian infectious bronchitis virus	20/8	6/3	12.73–26.46
Дельта-коронавирусы Porcine coronavirus HKU15	26/29	5/3	11.19–17.35

(1) Для каждого вируса показано количество негативно-полярных генов (ОПТ – открытых рамок трансляции) и кодируемых ими полипептидов (negative genes proteins; NGP). Анализ выполнен на основании первичной структуры генов указанных вирусов: HCoV-229 (ac.n. NC_002645.1), MERS (NC_019843.3), SARS-CoV1 (NC_004718.3), SARS-CoV2 (MT635445.1), Pangolin-CoV (MT040335.1), вирус инфекционного бронхита птиц (NC_001451.1), коронавирус свиней HKU-15 (NC_039208.1), коронавирус летучей мыши Bat/RATG13 (MN996532.1), коронавирус человека HKU1 (NC_006577.2).

(2) Показано количество открытых рамок трансляции в негативной полярности генома, которые начинаются с классического инициаторного AUG (левая цифра) и через косую линию с альтернативного CUG (правая цифра) кодонов и заканчиваются терминирующим кодоном (UAA, UAG, UGA) [4].

(3) Приведены диапазоны величин молекулярных масс полипептидов из группы негативных генов протяженностью ≥300 нк, рассчитанные на основании первичной последовательности гена.

сравнении вирусов BAT-CoV и SARS-CoV2, относящихся к одному роду бета-коронавирусов, имеют близкое число классических AUG-содержащих негативных генов 17 и 21, соответственно, которые к тому же имеют сходную локализацию в геноме. Таким образом, присутствие и сходная локализация данных генов в геноме у вирусов человека и летучих мышей подтверждают генетическую и эволюционную близость данных вирусов.

Напротив, выявление 29 AUG-негативных генов в геноме вируса Pangolin-CoV (табл. 1) может говорить о том, что вопреки современным пред-

ставлениям вирус от панголинов является более дальним родственником SARS-CoV2, чем вирус летучих мышей.

Дополнительные протяженные ОПТ выявлялись в геноме коронавирусов, если в качестве стартового кодона использовались альтернативный инициаторный кодон CUG (см. табл. 1). Так же, как и для ОПТ с классическим AUG, ОПТ альтернативного типа имели IRES-подобные структуры и могли обеспечивать синтез протяженных полипептидов с молекулярной массой в диапазоне 5–30 кДа. Наличие дополнительных негатив-

Рис. 1. Схема репликации РНК-генома и локализация негативно-полярных генов в геноме коронавирусов. (а) Показана схема геномной РНК вируса и рамки трансляции вирусных генов на модели коронавирусов (ac.n. MT890462.1; SARS-CoV-2/human/RUS/20200417_10/2020). Вертикальное смещение рамки трансляции в вирусном гене отражает сдвиг фазы трансляции. UTR – концевые, не транслируемые участки РНК. Позитивно-полярные неструктурные гены (nsр1–nsр19, 3а, 4с, 7аb, 8а, 9b) и структурные (N, S, М, Е, HE) показаны рамками, закрашенными серым цветом. Негативно-полярные гены и белки обозначены NGP (negative genes proteins) и показаны широкими стрелками, заштрихованными мелкой сеткой. Белки: NGP1 (нк позиции 1335–1787), NGP2 (2505–2861), NGP3 (2234–2656), NGP5 (28757–29095) начинаются с доминантного альтернативного CUG кодона (пунктирные стрелки); белок NGP4 (нк 6187–6489; зона «позитивно-полярного» белка nsр3 (Mрго)) начинается с классического AUG кодона (сплошная стрелка). (б) Приведены схемы локализации больших негативно-полярных генов (размером ≥ 300 нк) 4 коронавирусов: (а) – MERS-CoV; (б) – SARS-CoV2; (в) – Pangolin-CoV; (г) – Bat SARS-like CoV RatGT13. Схемы составлены на основе анализа нуклеотидной последовательности в базе данных Genbank: коронавирус человека MERS (NC_019843.3), SARS-CoV (NC_004718.3), SARS-CoV2 (MT635445.1), Pangolin-CoV (MT040335.1), Bat/RATG13 (MN996532.1). Цифры обозначают нуклеотид начала гена, отсчет с 5'-конца вирусного генома. (в) Схема вторичной структуры участка перед геном NGP4 (позиции 6187–6489 нк от 5'-конца геномной (+)РНК) получены с помощью программы [HTTP://bioinfo.net.in/IRESPred](http://bioinfo.net.in/IRESPred) [5], стрелкой показан инициаторный кодон AUG.

но-полярных генов альтернативного типа в геноме коронавируса может значительно увеличить его генетическую емкость.

Результаты настоящего сообщения показывают наличие протяженных рамок трансляции (генов) в геноме коронавируса, особенность которых состоит в том, что эти гены локализованы в негативной ориентации. При этом геном коронавируса в настоящее время считается позитивно-полярным, так как все известные гены коронавируса (около 25 генов неструктурных белков и 5 главных генов для структурных белков – E, M, S, N, HE) закодированы в геномной РНК в позитивной ориентации и имеют соответствующую стратегию экспрессии генома в инфицированных клетках-мишенях (рис. 1а). Наличие новых негативно-полярных генов подразумевает существование 2 механизмов их экспрессии и возможного синтеза соответствующих мРНК и последующей трансляции белков двумя возможными способами: либо непосредственной трансляции репликативной копии геномной минус РНК (путь I), либо транскрипции геномной (+)РНК с образованием индивидуальных мРНК “негативной полярности” для их последующей трансляции с образованием специфических полипептидов (путь II) (рис. 1а, выделены окружностью). Интересно отметить, что в геноме другого семейства – вирусов гриппа, относящихся к семейству ортомиксовирусов, характеризующихся негативной полярной стратегией геномной РНК, похожим образом обнаружены (ambisense) позитивно-полярные гены, кодируемые на вирусном негативно-полярном геноме [6–10].

Функция и значение обнаруженных новых амбиполярных вирусных генов пока не раскрыты. В случае с вирусами гриппа существует предположение о том, что выявленные новые амбисенс гены могут иметь значение в регуляции иммунного ответа на вирусные белки и/или в регуляции стабильности вирусных белков в зараженных клетках через систему деубиквитинирования белков [11–14]. Для понимания возможного функционального значения выявленных новых амбиполярных генов следует также иметь в виду две особенности, присущие данным генам. Первое – эволюционная стабильность существования амбиполярных генов у вирусов в течение длительного времени указывает на их биологическую детерминированность [11]. Второе – кодирование на одном участке молекулы РНК генов с противоположной полярностью в формате, так называемой стопки (genes stacking), позволяет значительно увеличить генетическую емкость вирусного генома и открывает для вируса новые возможности для изменчивости, повышения приспособляемости к хозяину и биологической эволюции в природе [11].

Наличие множественных амбисенс генов открывает реальную возможность кодирования мультивирионной популяции, состоящей из вирионов различного структурного типа, когда с одного генома могут синтезироваться более одного типа вирусных частиц, имеющих идентичный геном, но различный состав структурных белков. При этом часть вирионов (возможно инфекционных) может оставаться как бы невидимой (принцип “обратной стороны луны”). Более того, такой запрограммированный вирусным геномом мультивирионный профиль популяции вируса может иметь клеточную или тканевую зависимость, при которой каждый из типов вирусных частиц будет иметь автономную репликацию и репродукцию и доминировать в том или ином хозяине (органе или ткани). Данный пока гипотетический феномен репликации на одном геноме мультивирусных частиц может иметь важное значение в клеточно- или органо-зависимом патогенезе вирусного заболевания и может создать новые платформы для разработки методов лечения и вакцинопрофилактики.

Обнаруженные новые негативно-полярные гены в геноме коронавируса имеют специфическую для вирусного штамма локализацию и количественный состав в геноме (рис. 1б). Таким образом, паттерн негативно-полярных генов в геноме вирусного штамма может служить его молекулярным портретом (molecular signature) и использоваться в диагностике и изучении вирусных связей и биологической эволюции семейства коронавируса.

Наличие потенциальных негативно-полярных генов в геноме коронавируса ставит вопрос о классификации данного семейства. В случае обнаружения в зараженных клетках или инфицированных организмах белковых продуктов, экспрессированных на матрице «негативных» генов, дает основание для отнесения семейства коронавируса к амбисенс вирусам с биполярной стратегией генома.

В настоящее время к таким амбисенс вирусам отнесены вирусы 4 родов: флебо-, тоспо-, арена- и тениовирусов [15]. Амбисенс гены, локализованные в геноме в формате стопки, обнаружены у вирусов гриппа, у которых так же, как и у коронавируса пока не идентифицирована прямая экспрессия данных генов, но имеются косвенные признаки такой экспрессии при естественной вирусной инфекции *in vivo* [12, 13]. Изучение механизмов возможной экспрессии генетической информации указанных новых генов, а также выяснение роли и значения обнаруженных генов и/или их белковых продуктов в репликации вируса может послужить основой создания нового типа вакцин и средств противовирусной химиотерапии коронавирусной инфекции.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Источники финансирования отсутствуют.

БЛАГОДАРНОСТИ

Авторы выражают искреннюю признательность академику РАН Георгию Павловичу Георгиеву и Академику РАН Дмитрию Константиновичу Львову за полезные советы и поддержку данного направления исследований. Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Wu F, Zhao S, Yu B. et al. A new coronavirus associated with human respiratory disease in China // *Nature* 579, 265–269 (2020). <https://doi.org/10.1038/s41586-020-2008-3>
2. Guo Y.R., Cao Q.D., Hong Z.S., Tan Y.Y., Chen S.D., Jin H.J., Tan K.S., Wang D.Y., Yan Y. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status // *Mil Med Res.* 2020 Mar 13; 7 (1): 11. <https://doi.org/10.1186/s40779-020-00240-032169119>
3. Fehr R.A., Perlman S. Coronaviruses: an overview of their replication and pathogenesis // *Methods Mol. Biol.* 2015; 1282: 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1
4. Kearsle M.G., Wilusz J.E. Non-AUG translation: a new start for protein synthesis in eukaryotes // *Genes Dev.* 2017 Sep 1; 31 (17): 1717–1731. <https://doi.org/10.1101/gad.305250.117>
5. Kolekar P., Pataskar A., Kulkarni-Kale U., Pal J., Kulkarni A. IRESPred: Web server for Prediction of Cellular and Viral Internal Ribosome Entry Site (IRES) // *Sci Rep.* 2016 Jun 6; 6: 27436. <https://doi.org/10.1038/srep2743627264539>
6. Palese Baez M., Zazra J.J., Elliott R.M., Young J.F., Palese P. Nucleotide sequence of the influenza A/duck/Alberta/60/76 virus NS RNA: conservation of the NS1/NS2 overlapping gene structure in a divergent influenza virus RNA segment // *Virology.* 1981 Aug; 113 (1): 397–402.
7. Zhirnov O.P., Poyarkov S.V., Vorob'eva I.V., Safonova O.A., Malyshev N.A., Klenk H.D. Segment NS of influenza A virus contains an additional gene NSP in positive-sense orientation // *Dokl Biochem Biophys.* 2007 May–Jun; 414: 127–33.
8. Zhirnov O.P., Vorobjeva I.V., Saphonova O.A., Poyarkov S.V., Ovcharenko A.V., Anhlan D., Malyshev N.A. Structural and evolutionary characteristics of HA, NA, NS and M genes of clinical influenza A/H3N2 viruses passaged in human and canine cells // *J Clin Virol.* 2009 Aug; 45 (4): 322–33. <https://doi.org/10.1016/j.jcv.2009.05.03019546028>
9. Zhirnov O.P., Akulich K.A., Lipatova A.V., Usachev E.V. Negative-sense virion RNA of segment 8 (NS) of influenza a virus is able to translate in vitro a new viral protein // *Dokl Biochem Biophys.* 2017 Mar; 473 (1): 122–127. 2017 May 17. <https://doi.org/10.1134/S160767291702009028510127>
10. Gong Y.N., Chen G.W., Chen C.J., Kuo R.L., Shih S.R. Computational analysis and mapping of novel open reading frames in influenza A viruses // *PLoS One.* 2014 Dec 15; 9 (12): e115016. <https://doi.org/10.1371/journal.pone.0115016>
11. Zhirnov O.P. Unique Bipolar Gene Architecture in the RNA Genome of Influenza A Virus // *Biochemistry (Mosc).* 2020 Mar; 85 (3): 387–392. <https://doi.org/10.1134/S000629792003014132564743>
12. Hickman H.D., Mays J.W., Gibbs J., Kosik I., Magadan J.G., Takeda K., Yewdell J.W. Influenza A virus negative strand RNA is translated for CD8+ T cell immunosurveillance // *J. Immunol.* 2018; 201: 1222–1228. <https://doi.org/10.4049/jimmunol.1800586>
13. Zhirnov O.P., Konakova T.E., Anhlan D., Ludwig S., Isaeva E.I. Cellular immune response in infected mice to NSP protein encoded by the negative strand NS RNA of influenza A virus // *MIR J* 2019; 6 (1), 28–36. <https://doi.org/10.18527/2500-2236-2019-6-1-28-36>
14. Zhong W., Reche P.A., Lai C.C., Reinhold B., Reinherz E.L. Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire // *J Biol Chem.* 2003 Nov 14; 278 (46): 45135–44. Epub 2003 Sep 5. PubMed
15. Nguyen M., Haenni A.L. Expression strategies of ambisense viruses // *Virus Res.* 2003 Jun; 93 (2): 141–50. Review.

NOVEL NEGATIVE SENSE GENES IN THE RNA GENOME OF CORONAVIRUSES

Corresponding Member of RAS O. P. Zhirnov^{a,b,#} and S. V. Poyarkov^a

^a The Russian-German Academy of Medico-Social and Biotechnological Sciences, Moscow, Russian Federation

^b The D.I. Ivanovsky Institute of Virology, The N.F. Gamaleya National Research Center of Epidemiology and Microbiology, Moscow, Russian Federation

[#]e-mail: zhirnov@inbox.ru

The coronavirus family consists of lipid-containing envelope viruses that have a single-stranded RNA genome that encodes 25–0 proteins in different viruses by the mechanism of positive-polar strategy. In addition, extended open reading translation frames (ORF; genes) locating in a negative-sense orientation were found in the genomes of coronaviruses. The size of negative sense genes varies in the range of 150–450 nt corresponding to negative genes polypeptides (NGP) with m. m. 5–30 kDa. Coronaviruses have marked differences from virus to virus in the number of negative genes detected. These negative sense genes in coronavirus genome make it possible to consider this family as viruses developing an ambisense genome strategy.

Keywords: coronaviruses, genome, ambisense genes, genome polarity