Порядок митохондриальных генов как дополнительный маркер в филогенетических исследованиях насекомых

Ф.С. Шарко¹, А.В. Недолужко², С.М. Расторгуев², С.В. Цыганкова², Е.С. Булыгина², А.А. Полилов³, Е.Б. Прохорчук^{1, 3}, К.Г. Скрябин^{1, 2, 3}

За миллионы лет эволюции геномы современных насекомых накопили значительное количество мутаций. Геномные различия некоторых представителей класса Insecta, принадлежащих одному семейству или отряду, настолько велики, что могут завести в тупик при проведении филогенетических исследований и требуют использования нетрадиционных методов анализа. Известно, что молекулярная эволюция идет не только путем единичных нуклеотидных замен, но и включает в себя более крупные геномные перестройки, такие как изменение порядка генов. Гены митохондриальной ДНК (мтДНК) достаточно часто используются в качестве маркера для филогенетических исследований у многих организмов, в том числе членистоногих, поскольку мтДНК многокопийна, наследуется по материнской линии, не подвержена рекомбинации и достаточно быстро (относительно ядерного генома) накапливает мутации. К настоящему времени в общедоступных базах данных собрано большое количество полных нуклеотидных последовательностей митогеномов (тысячи организмов), однако их филогенетический анализ имеет свои сложности, особенно для представителей класса Haceкомые (Insecta), чья эволюция занимает значительный отрезок геологического времени. Целью данного исследования была оценка возможности использования новых методических приемов филогенетического анализа насекомых. Сравниваются два способа филогенетического анализа. Первый метод использует изменчивость нуклеотидной последовательности мтДНК, второй – порядок генов в полных митохондриальных геномах как дополнительный маркер. Показано, что порядок генов может быть применен в качестве дополнительного маркера при филогенетических исследованиях представителей отряда Hymenoptera. Разработана программа mitoSpider, с помощью которой выявлены 63 последовательности митогеномов насекомых, где количества генов отличаются от стандартного (для высокоорганизованных Metazoa).

Ключевые слова: порядок генов; митохондриальная ДНК; насекомые, Hymenoptera; перепончатокрылые; mitoSpider; CYTB; филогения.

КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ:

Шарко Ф.С., Недолужко А.В., Расторгуев С.М., Цыганкова С.В., Булыгина Е.С., Полилов А.А., Прохорчук Е.Б., Скрябин К.Г. Порядок митохондриальных генов как дополнительный маркер в филогенетических исследованиях насекомых. Вавиловский журнал генетики и селекции. 2017;21(3):368-373. DOI 10.18699/VJ17.254

HOW TO CITE THIS ARTICLE:

Sharko F.S., Nedoluzhko A.V., Rastorguev S.M., Tsygankova S.V., Boulygina E.S., Polilov A.A., Prokhortchouk E.B., Skryabin K.G. The mitochondrial gene order and CYTB gene evolution in insects. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2017;21(3):368-373. DOI 10.18699/VJ17.254

Received 24.10.2016 Accepted for publication 04.02.2017

© AUTHORS, 2017

The mitochondrial gene order and CYTB gene evolution in insects

F.S. Sharko¹, A.V. Nedoluzhko², S.M. Rastorguev², S.V. Tsygankova², E.S. Boulygina², A.A. Polilov³, E.B. Prokhortchouk^{1, 3}, K.G. Skryabin^{1, 2, 3}

¹ Institute of Bioengineering, Federal Research Center "Fundamentals of Biotechnology" RAS, Moscow, Russia

Over millions of years of evolution, the genomes of modern insects have accumulated a significant number of mutations, which often can lead up a blind alley when carrying out phylogenetic research. Genomic differences between some representatives belonging to the same family or group are often so great that they demand using nonconventional methods of the phylogenetic analysis. It is known that molecular evolution goes by the way of not only single nucleotide substitutions, but also by larger genomic reorganizations, such as insertion or deletion of large genome fragments, and even changing the order of genes. Mitochondrial DNA genes (mtDNA) are quite often used as markers for phylogenetic research into many organisms including arthropods, because mtDNA is multicopied, is inherited maternally, does not undergo recombination and accumulates mutations quickly enough (relative to the nuclear genome). To date, a large number of full nucleotide sequences of mitogenomes (thousands of organisms) has been deposited in public databases; however, their phylogenetic analysis has obstacles, especially for representatives of the insects (Insecta), whose evolution takes a considerable part of geological time. In this work we describe the application and a comparison of two ways of the phylogenetic analysis for different groups of insects. The first method uses the variability of the nucleotide sequence of mtDNA, and the second one analyses the order of genes in full mitochondrial genomes of insects that can be used as an additional marker in phylogenetic research into representatives of the order Hymenoptera.

Key words: gene order; mitochondrial DNA; Insects; Hymenoptera; mitoSpider; CYTB; phylogeny.

¹ Институт биоинженерии, Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук,

² Федеральное государственное бюджетное учреждение «Национальный исследовательский центр «Курчатовский институт», Москва, Россия

³ Биологический факультет, Московский государственный университет им. М.В. Ломоносова, Москва, Россия

² National Research Centre "Kurchatov Institute", Moscow, Russia

³ Lomonosov Moscow State University, Faculty of Biology, Moscow, Russia

оявление и развитие методов современного секвенирования ДНК (next-generation sequencing – NGS) способствовало значительному прогрессу в эволюционной биологии, включая проведение многочисленных работ в области геномики, транскриптомики или эпигеномики. К настоящему времени применение NGS-технологий стало рутинной процедурой, которую используют многие исследовательские группы, в том числе в России (Rastorguev et al., 2013; Nedoluzhko et al., 2016; Sokolov et al., 2016). Успех этих методов привел к экспоненциальному росту количества отсеквенированных и аннотированных геномов (ядерных и митохондриальных), которые депонированы в различные общедоступные базы данных. Эти последовательности ДНК активно применяются в филогенетических и филогеномных исследованиях.

Гены митохондриальной ДНК (мтДНК) считаются признанными молекулярными маркерами и задействованы в различных эволюционных исследованиях, посвященных систематике и генетическому разнообразию животных и растений. В последнее время для подобных работ стали использоваться полные последовательности митохондриальных геномов, которые более информативны, нежели отдельные гены мтДНК, их фрагменты или некодирующие регионы (Li et al., 2012; Mao et al., 2015).

Количество генов в митохондриальном геноме эукариот (включая одноклеточные) варьирует от таксона к таксону: например, число белок-кодирующих генов может меняться от 3 до 67, а гены транспортных РНК (тРНК) и вовсе могут отсутствовать (Adams, Palmer, 2003). Митогеном многоклеточных животных, за исключением представителей типа Knidaria, в большинстве случаев содержит 37 генов: 13 белок-кодирующих генов, 22 гена тРНК и два гена рибосомальной РНК (Castellana et al., 2011; Kayal et al., 2012).

В ранее опубликованных работах, посвященных определению последовательности и порядку генов в мтДНК насекомых (Babbucci et al., 2014; Jiang et al., 2016; Nedoluzhko et al., 2016), высказано предположение, что количество генов и их порядок значительно варьируют внутри как небольших, так и более крупных таксонов высокоорганизованных эукариот (например, членистоногих). Более того, данные о количестве, дупликациях и элиминировании митохондриальных генов могут служить дополнительным источником информации для понимания филогенетических взаимосвязей эволюционно древних животных, позволяя уточнить (в том числе с помощью разнообразных биоинформатических подходов) филогенетическое положение конкретного таксона.

Настоящее исследование направлено на оценку возможностей использования новых методических приемов филогенетического анализа у насекомых. Для анализа опубликованных в базе данных NCBI (Национальный центр биотехнологической информации США) митохондриальных геномов насекомых применялись data mining технологии. Нами разработана программа mitoSpider, которая позволяет оценить количество генов в опубликованных последовательностях мтДНК.

Найденные последовательности митохондриальных геномов представителей отрядов насекомых Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera и Orthoptera

(наиболее часто представлены в базе NCBI) были использованы для проведения филогенетического анализа, опирающегося на порядок генов. Параллельно для представителей этих же отрядов выполнялся филогенетический анализ на основе изменчивости нуклеотидной последовательности гена цитохрома в (СУТВ) и всех белок-кодирующих генов (CDS) мтДНК. Показано, что филогенетическое дерево, построенное в соответствии с порядком генов в мтДНК у отряда Hymenoptera, в основном совпало с филогенетическим деревом, построенным с использованием нуклеотидной последовательности гена СҮТВ. Поскольку в отрядах Hemiptera, Diptera, Coleoptera, Orthoptera, и Lepidoptera порядок генов в мтДНК фактически неизменен, филогенетическая реконструкция по этому признаку невозможна. Продемонстрировано, что взятый в качестве маркера порядок генов делает топологию филогенетического дерева более точной, приближая ее к топологии, построенной по всем белок-кодирующим генам у Hymenoptera. Кроме того, в статье обсуждается обнаруженное значительное число митохондриальных геномов насекомых, количество генов в которых отличается от стандартного набора (37 генов).

Материалы и методы

Для сбора информации в базе данных NCBI, на языке Perl авторами была написана программа mitoSpider, основанная на EntrezAPI (API ncbi) (исходный код программы доступен по адресу: https://github.com/megaphragma/mitoSpider). Программа позволяет получить статистику по генам и их порядку в митогеномах, также с ее помощью можно получить нуклеотидные последовательности выбранных организмов, которые представлены в открытом доступе. Данный программный продукт использовался для сбора статистики о количестве генов и их порядке в митохондриальных геномах всех насекомых, представленных в базе данных (mitoSpider может применяться для таксонов произвольного ранга).

С помощью mitoSpider все митогеномы насекомых, по названиям которых можно было определить, что они неполные и/или дублирующиеся (разные версии генома одного и того же вида), отфильтровывались на предварительной стадии. В итоге статистику по количеству генов и их порядку собирали для 895 последовательностей мтЛНК насекомых.

Полные митохондриальные геномы насекомых из отрядов Двукрылые (Diptera), Жесткокрылые (Coleoptera), Прямокрылые (Orthoptera), Полужесткокрылые (Hemiptera), Перепончатокрылые (Hymenoptera) и Чешуекрылые (Lepidoptera) были отобраны для построения филогенетических реконструкций на основе порядка генов в их последовательности при помощи программы MLGO (Ни et al., 2014) со стандартными настройками, которая использует метод максимального правдоподобия для статистической модели. Для возможности сравнения порядка генов у разных видов гены были расположены согласно их последовательности в митогеноме, начиная с tRNA-Ile и следующих за ним генов tRNA-Gln и tRNA-Met.

Параллельно в программе MEGA 7.0 и методом максимального правдоподобия (Maximum likelihood) для кластеризации (Kumar et al., 2016) с моделью замен Тамуры—Нея

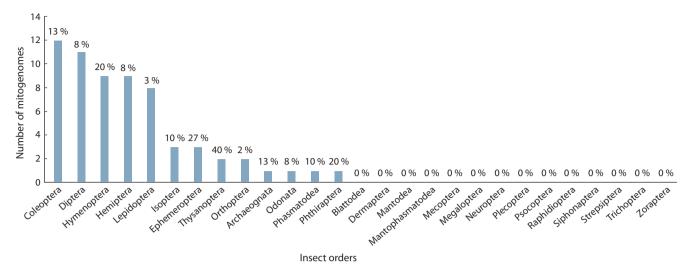


Fig. 1. The number of annotated mitogenomes of Insecta (June 2016) with deviations from the standard numbers of genes and their percentages in the total number of mitogenomes in each order.

проводился филогенетический анализ с использованием нуклеотидной последовательности гена *CYTB* из всех отобранных ранее полных последовательностей мтДНК. Нуклеотидные последовательности предварительно были выровнены с помощью алгоритма ClustalW. Таким же методом реконструировались филогенетические деревья по 13 белок-кодирующим генам для выбранных отрядов насекомых. Объединение информации по последовательности нуклеотидов гена *CYTB* и порядку генов выполняли в программе TNT, используя встроенный скрипт "hybtree. run" (Goloboff et al., 2008).

Топологию филогенетических деревьев, построенных с помощью двух подходов, сравнивали, используя программу TreeKO со стандартными параметрами. В результате проведенного анализа получали значение дистанций (strict distance) между двумя топологиями. Искомая дистанция представляет собой расстояние между двумя деревьями, если отсутствие исключений для эволюционно значимых событий не принимается во внимание. Таким образом, чем ближе дистанция TreeKO к нулю, тем ближе топологии ветвления построенных деревьев, а чем ближе к единице, тем сильнее топологии различаются (Huerta-Cepas et al., 2010; Marcet-Houben, Gabaldón, 2011).

Результаты

Программа mitoSpider позволила обнаружить 895 полных митохондриальных геномов насекомых в базе данных NCBI (июнь 2016 г.) (http://www.treefrog.ru/geneorder/Приложение1.xlsx), причем в 63 из них были выявлены отклонения от стандартного количества генов (больше или меньше 37). На рис. 1 показано распределение «нестандартных» митохондриальных геномов по различным отрядам класса Насекомые и их процентное соотношение относительно общего числа опубликованных митогеномов.

Из общей выборки митогеномов, отсортированных при помощи mitoSpider, были взяты последовательности мтДНК шести отрядов насекомых: Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera и Orthoptera (наиболее часто представлены в базе NCBI). С использованием этих

митогеномов было построено шесть филогенетических деревьев (Доп. материалы) 1 , основанных на порядке генов (метод Maximum likelihood) (рис. 2), и еще шесть филогенетических деревьев (метод Maximum likelihood) – на данных об изменчивости митохондриального гена *CYTB*.

Топологии филогенетических деревьев были проанализированы с помощью программы ТreeKO. Показано, что филогенетические деревья, построенные на основании порядка генов митохондриальных геномов Hymenoptera, в значительной степени совпали с теми, которые были построены с использованием нуклеотидной последовательности гена *CYTB*, в отличие от других отрядов насекомых (Hemiptera, Diptera, Coleoptera, Orthoptera и Lepidoptera), где порядок генов фактически неизменен, поэтому не несет дополнительной информации (см. таблицу).

Более того, добавление данных по порядку генов к нуклеотидным последовательностям *СҮТВ* у Hymenoptera уточняет топологию филогенетического дерева, приближая ее к топологии, построенной по всем белок-кодирующим генам. Значения дистанции TreeKO при сравнении филогении *СҮТВ* против CDS равны 0.232, в то время как при добавлении данных по порядку генов к нуклеотидным последовательностям *СҮТВ* они снижались до 0.195. В то же время порядок генов не может быть использован для филогенетических построений представителей Hemiptera, Diptera, Coleoptera, Orthoptera и Lepidoptera, поскольку он идентичен у большинства опубликованных к настоящему времени митогеномов.

Как и в других исследованиях (Dowton, 1999; Mao et al., 2015; Kim et al., 2016), нами показано, что порядок генов в отряде Нутепортега может различаться даже у представителей одного рода, например у шмелей (*Bombus* sp.) или безжальных пчел (*Melipona* sp.).

Анализ 63 митогеномов насекомых, в которых были выявлены отклонения от стандартного количества генов для Метагоа, показал, что наиболее часто из митохондриального генома насекомых элиминируются гены *trn-lle* и *ATP8*

¹ Дополнительные материалы см. в Приложении 2 по адресу: http://www.bionet.nsc.ru/vogis/download/pict-2017-21/appx6.pdf

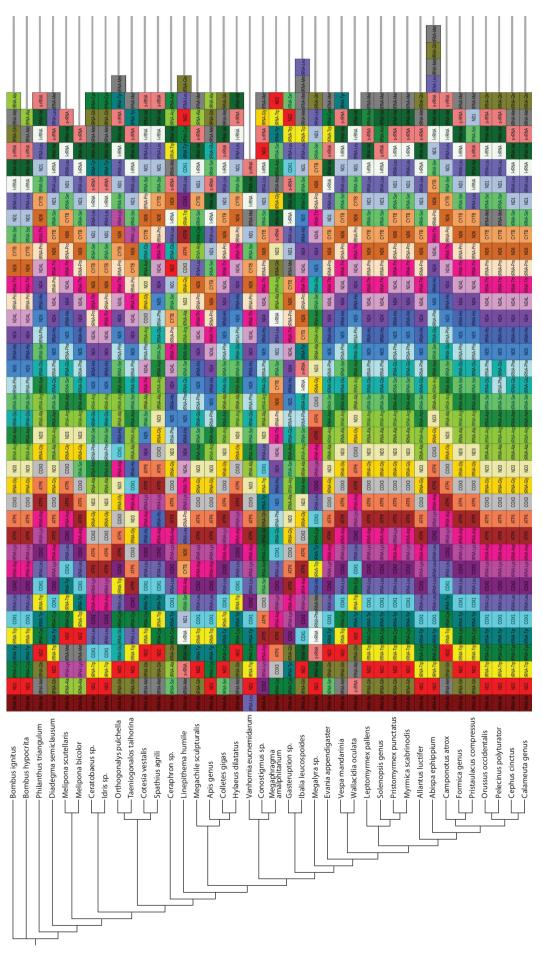


Fig. 2. Phylogenetic tree of Hymenoptera based on the gene order in mitogenomes.

TreeKO distances between two tree topologies based on the mitogenome gene order and sequence variation of the CYTB gene

Insecta orders	The number of complete mitochondrial genomes used in the analysis	TreeKO distances between two tree topologies
Lepidoptera	254	0.565
Orthoptera	93	0.516
Diptera	128	0.455
Coleoptera	83	0.319
Hemiptera	102	0.312
Hymenoptera	37	0.271

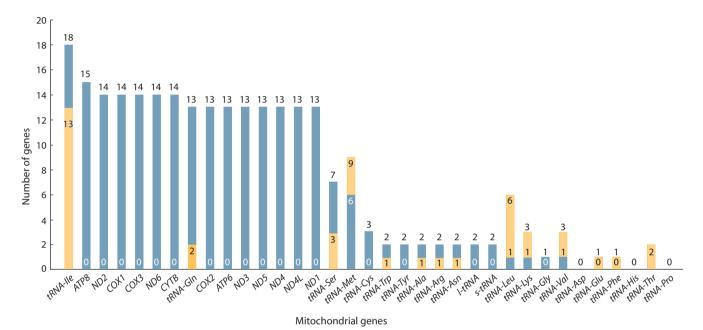


Fig. 3. Mitochondrial genes that are most often missed (marked in black) or duplicated (marked in white).

(не обнаружены в 18 и 15 митогеномах соответственно). Тем не менее мы не выявили каких-либо закономерностей, которые подтвердили бы направленный отбор по этому признаку. По-видимому, события утери некоторых генов в митогеномах насекомых случайны. Более того, нередко даже единичная мутация в гене тРНК может привести к потере его функциональности. Также стоит отметить, что в последовательностях митохондриальных геномов насекомых чаще отсутствуют некоторые белок-кодирующие гены, по сравнению с тРНК и рРНК, к тому же нами не обнаружено ни одного случая дупликации этих генов (рис. 3).

Кроме того, типичная нуклеотидная последовательность размером 69 пар оснований, обнаруженная нами между генами trnQ и trnV в опубликованном ранее митогеноме микроскопического жука Scydosella musawasensis (Nedoluzhko et al., 2016), не формировала характерную шпильку при анализе ее вторичной структуры.

Обсуждение

Маркеры мтДНК активно используются в филогенетических исследованиях насекомых, однако подобные реконструкции нередко противоречат друг другу в зависимости

от метода, выбора объектов и генов. На сегодняшний день для уточнения филогенетических связей между организмами применяется дополнительная информация об исследуемых таксонах. В нашей работе в качестве дополнительного маркера был использован порядок генов в митохондриальном геноме насекомых.

Полученные результаты показывают перспективность использования информации о порядке генов в качестве дополнительного филогенетического маркера для отдельных видов, родов и семейств перепончатокрылых. Наши данные указывают на то, что информация о порядке генов, добавленная к данным о нуклеотидной последовательности гена *СҮТВ*, улучшает топологию филогенетических построений, приближая их к филогенетическим деревьям, построенным с использованием всех 13 белок-кодирующих генов мтДНК.

Ранее наибольшее число перестроек внутри митохондриального генома Hymenoptera было обнаружено у паразитических видов, и первоначально предполагалось, что существует связь между паразитизмом и многочисленными перестройками в геноме (Dowton, Austin, 1999), однако в дальнейшем эта гипотеза была опровергнута.

В настоящее время принято считать, что эти перестройки эволюционно нейтральны (Boore, Brown, 1998). Некоторые примеры перестроек в митохондриальных геномах были обнаружены и в других, более высокоорганизованных таксонах животных, но их эволюционная ценность остается неясной.

Разработанная нами программа mitoSpider позволила обнаружить 63 последовательности митогеномов насекомых, у которых выявлены отклонения от стандартного количества генов (для высокоорганизованных Metazoa).

Нами показано, что наиболее часто из митохондриального генома насекомых элиминируются некоторые белок-кодирующие гены, которые, вероятно, мигрируют в ядерный геном в процессе эволюции (Berg, Kurland, 2000). Мы предполагаем, что в случае *S. musawasensis* имели место замены, которые деактивировали ген *trnI*. Следует особо отметить, что далеко не все полные митохондриальные геномы с нестандартным количеством генов отсеквенированы полностью. По-видимому, столь значительное количество «нестандартных» митогеномов связано с их низкокачественной *de novo* сборкой и секвенированием, а также может объясняться неправильной аннотацией митогеномов при загрузке в базу NCBI, что показано в настоящей работе и в исследованиях наших коллег (Gissi et al., 2008).

Acknowledgments

This work was supported by the Russian Science Foundation, project 14-24-00175.

Conflict of interest

The authors declare no conflict of interest.

References

- Adams K.L., Palmer J.D. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 2003; 29(3):380-395. DOI 10.1016/S1055-7903(03)00194-5.
- Babbucci M., Basso A., Scupola A., Patarnello T., Negrisolo E. Is it an ant or a butterfly? Convergent evolution in the mitochondrial gene order of Hymenoptera and Lepidoptera. Genome Biol. Evol. 2014; 6(12):3326-3343. DOI 10.1093/gbe/evu265.
- Berg O.G., Kurland C.G. Why mitochondrial genes are most often found in nuclei. Mol. Biol. Evol. 2000;17(6):951-961.
- Boore J.L., Brown W.M. Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. 1998;8:668-674. DOI 10.1016/S0959-437X (98)80035-X.
- Castellana S., Vicario S., Saccone C. Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. Genome Biol. Evol. 2011;3:1067-1079. DOI 10.1093/gbe/evr040.
- Dowton M. Relationships among the cyclostome braconid (Hymenoptera, Braconidae) subfamilies inferred from a mitochondrial tRNA

- gene rearrangement. Mol. Phylogenet. Evol. 1999;11:283-287. DOI 10.1006/mpev.1998.0580.
- Dowton M., Austin A.D. Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. Mol. Biol. Evol. 1999; 16:298-309.
- Gissi C., Iannelli F., Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity (Edinb). 2008;101(4):301-320. DOI 10.1038/hdy.2008.62.
- Goloboff P.A., Farris J.S., Nixon K.C. TNT, a free program for phylogenetic analysis. Cladistics. 2008;24:774-786. DOI 10.1111/j.1096-0031.2008. 00217.x.
- Hu F., Lin Y., Tang J. MLGO: phylogeny reconstruction and ancestral inference from gene-order data. BMC Bioinformatics. 2014;15(1): 354. DOI 10.1186/s12859-014-0354-6.
- Huerta-Cepas J., Dopazo J., Gabaldón T. ETE: a python Environment for Tree Exploration. BMC Bioinformatics. 2010;11:24. DOI 10.1186/1471-2105-11-24.
- Jiang P., Li H., Song F., Cai Y., Wang J., Liu J., Cai W. Duplication and remolding of tRNA genes in the mitochondrial genome of *Reduvius tenebrosus* (Hemiptera: Reduviidae). Int. J. Mol. Sci. 2016;17(6): 951. DOI 10.3390/ijms17060951.
- Kayal E., Bentlage B., Collins A.G., Kayal M., Pirro S., Lavrov D.V. Evolution of linear mitochondrial genomes in medusozoan cnidarians. Genome Biol. Evol. 2012;4(1):1-12. DOI 10.1093/gbe/evr123.
- Kim M.J., Hong E.J., Kim I. Complete mitochondrial genome of *Camponotus atrox* (Hymenoptera: Formicidae): a new tRNA arrangement in Hymenoptera. Genome. 2016;59(1):59-74. DOI 10.1139/gen-2015-0080.
- Kumar S., Stecher G., Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 2016;33(7):1870-1874. DOI 10.1093/molbev/msw054.
- Li H., Liu H., Shi A., Stys P., Zhou X., Cai W. The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). PLoS ONE. 2012; 7(1):e29419. DOI 10.1371/journal.pone.0029419.
- Mao M., Gibson T., Dowton M. Higher-level phylogeny of the Hymenoptera inferred from mitochondrial genomes. Mol. Phylogenet. Evol. 2015;84:34-43. DOI 10.1016/j.ympev.2014.12.009.
- Marcet-Houben M., Gabaldón T. TreeKO a duplication aware algorithm for the comparison of phylogenetic trees. Nucl. Acids Res. 2011;39(10):e66.
- Nedoluzhko A., Sharko F., Boulygina E., Tsygankova S., Sokolov A., Mazur A., Polilov A., Prokhortchouk E., Skryabin K. The complete mitochondrial genome of the smallest known free-living insect *Scy-dosella musawasensis*. Mitochondrial DNA. Part B. 2016;1(1):171-172. DOI 10.1080/23802359.2016.1149785.
- Rastorguev S.M., Nedoluzhko A.V., Mazur A.M., Gruzdeva N.M., Volkov A.A., Barmintseva A.E., Mugue N.S., Prokhortchouk E.B. High-throughput SNP-genotyping analysis of the relationships among Ponto-Caspian sturgeon species. Ecol. Evol. 2013;3(8):2612-2618. DOI 10.1002/ece3.659.
- Sokolov A.S., Nedoluzhko A.V., Boulygina E.S., Tsygankova S.V., Sharko F.S., Gruzdeva N.M., Shishlov A.V., Kolpakova A.V., Rezepkin A.D., Skryabin K.G., Prokhortchouk E.B. Six complete mitochondrial genomes from Early Bronze Age humans in the North Caucasus. J. Archaeol. Sci. 2016;73:138-144. DOI 10.1016/j.jas. 2016.07.017.