

Реконструкция эволюции митохондрий растений и простейших, предсказание регуляции экспрессии митохондриальных генов на основе разработки нового метода поиска высоко консервативных участков ДНК.

(На примере эволюции хромосомной структуры митохондриального генома простейших типа Ciliophora)

Предварительные результаты к заявке на проект

Аннотация

Предпосылки

Митохондрии простейших типа Ciliophora могут служить мишенями для терапевтического воздействия на паразитов, изучение структуры и эволюции их геномов, регуляции экспрессии их генов имеет разнообразное практическое значение, например, для организации и ветеринарии нерестово-выростных хозяйств.

Результаты

Геномы митохондрий простейших типа Ciliophora изучаются на основе описания высоко консервативных элементов ДНК, семейств белков, хромосомных структур и сравнения различных эволюционных деревьев, построенных на этой основе.

Заключение

На указанном материале сравниваются оригинальные методы авторов для описания высоко консервативных элементов, семейств белков и хромосомных структур. Показано, что методы дают согласованные и адекватные результаты.

Предпосылки

Работа посвящена описанию геномов митохондрий у простейших типа Ciliophora. К ним относится, например, *Ichthyophthirius multifiliis* – паразит, вызывающий гибель многих пресноводных рыб, в том числе – разводимых в аквариумах, прудовых, промышленных, нерестово-выростных хозяйствах и на рыбозаводах [1]. Митохондрии могут служить мишенями для терапевтического воздействия на паразитов, изучение структуры и эволюции их геномов, регуляции экспрессии их генов имеет разнообразное практическое значение, в том числе в области ветеринарии.

Нами рассмотрены роды, принадлежащие трём классам: Armophorea (*Nyctotherus*), Oligohymenophorea (*Ichthyophthirius*, *Paramecium* и *Tetrahymena*) и Spirotrichea (*Moneuplotes*, *Oxytricha*). Классы Oligohymenophorea и Spirotrichea значительно различаются между собой [2]. Класс Armophorea состоит из анаэробов, но ближе к классу Spirotrichea, чем к классу Oligohymenophorea [3].

Обычно в митохондриях происходит окисление жирных кислот и синтез некоторых соединений [4]. Многие эукариоты, живущие в анаэробных условиях, лишены митохондрий, однако митохондрии у анаэроба *Nyctotherus ovalis* продуцируют водород [3]. Роль митохондрий у разных организмов различна, что отражается и на размерах их геномов. Экстремально маленькие геномы у митохондрий паразитов из типа Apicomplexa, в них кодируются лишь три белка и короткие фрагменты рРНК [5]. Этот тип наряду с Ciliophora принадлежит надтипу Alveolata.

В рассматриваемых нами митохондриях кодируются десятки белков [6, 7, 8, 9, 2, 3, 10, 11]. Роль некоторых из них не известна, но они относительно быстро накапливают замены [12]. Митохондриальная хромосома кольцевая у *Ichthyophthirius* и линейная у других рассмотренных нами видов [13, 11]. При этом в митохондриях родов *Tetrahymena*, *Moneuplotes* и *Oxytricha* большинство генов транскрибируются в двух направлениях, начиная из середины линейной хромосомы. Напротив, в митохондриях родов *Paramecium* и *Nyctotherus* большинство генов транскрибируется в одном направлении.

Рассмотренные геномы митохондрий очень компактны. Гены образуют длинные опероны с маленькими некодирующими промежутками; иногда кодирующие области перекрываются. Порядок генов у рассмотренных видов различается внутри классов, что делает нетривиальным изучение эволюции даже только хромосомной структуры. Общим для класса Oligohymenophorea является относительно длинный некодирующий участок в 5'-лидерной области гена, кодирующего белок аросутоchrome b.

Результаты и обсуждение

Высоко консервативные элементы у простейших типа Ciliophora

Предложенным нами алгоритмом и соответствующей программой [14] выполнен поиск высоко консервативных элементов ДНК, которые далее сокращаются как ВКЭ. В результате найдено 393 ВКЭ. На рис. 1 показано дерево, построенное программой RAxML [15] по двенадцати строкам, которые показывают представленность каждого из ВКЭ в митондрии каждого вида. Дерево хорошо согласуется с деревом видов, но, естественно, не совпадает с ним. В частности, *Moneuplotes crassus* чаще имеет общий ВКЭ с *Oxytricha trifallax*, чем с *Moneuplotes minuta*, а *Paramecium aurelia* по представленности различных ВКЭ заметно отличается от *P. caudatum*.

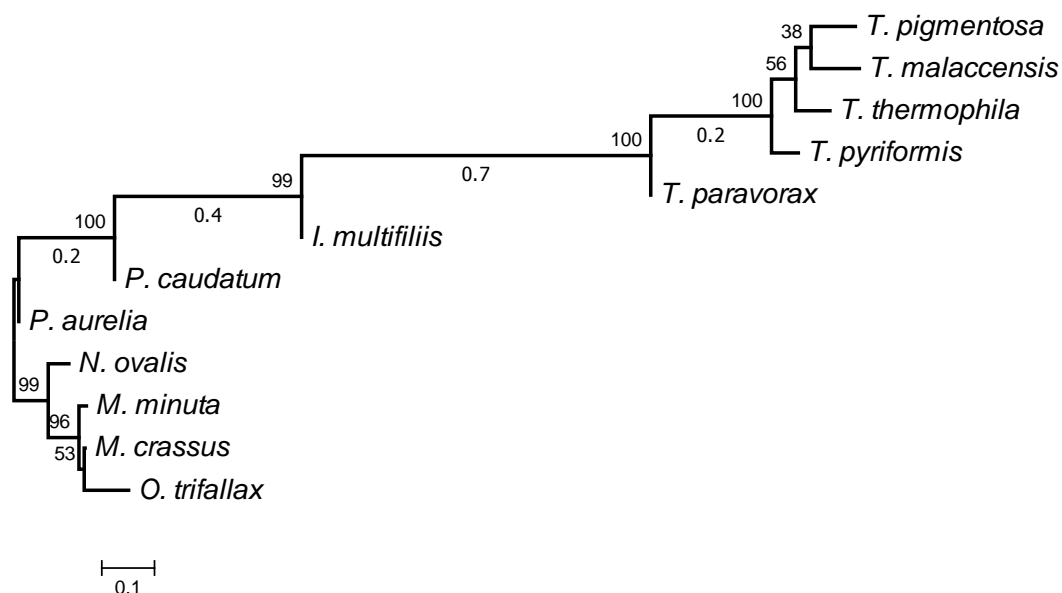


Рис. 1. Дерево эволюции митохондрий, построенное по 393 ВКЭ, найденным нашим алгоритмом. Дерево построено программой RAxML на основе матрицы из 12 строк и 393 столбцов, элемент которой 1 или 0 указывает на присутствие или соответственно отсутствие у митондрии каждого вида представителя каждого ВКЭ.

ВКЭ 287 найден лишь у четырёх видов рода *Tetrahymena*. Он расположен перед геном большой субъединицы рРНК (на комплементарной цепи). Возможно, он участвует в регуляции транскрипции или посттранскрипционных модификациях рРНК.

ВКЭ 299. В митохондриях видов из класса Oligohymenophorea гены *nad2* и *nad7* расположены дивергентно на очень маленьком расстоянии друг от друга; каждый из них служит началом длинного оперона. Выравнивание размеченных в аннотации GenBank белков Nad2 говорит о почти полном отсутствии у них консервативных позиций на N-концах. Напротив, гены *nad7* высоко консервативны, их 5'-концы покрыты ВКЭ 151 у *Ichthyophthirius multifiliis*, *Tetrahymena malaccensis*, *Tetrahymena paravorax*, *Tetrahymena pigmentosa*, *Tetrahymena pyriformis* и *Tetrahymena thermophile*.

Поэтому можно предположить, что ген *nad2* перекрывает промотор перед геном *nad7*. У *Tetrahymena malaccensis*, *Tetrahymena pigmentosa*, *Tetrahymena pyriformis* и *Tetrahymena thermophile* внутри кодирующих областей гена *nad2* обнаружен ВКЭ 299, содержащий потенциальный промотор. Выделенное в его составе полужирным шрифтом слово САТА соответствует консенсусу YRTA промоторов в митохондриях растений [16].

ВКЭ 234 найден у всех видов рода *Tetrahymena* и расположен между генами *ymf76* и *ymf66* (оба гена на комплементарной цепи). У четырёх видов *T. malaccensis*, *T. paravorax*, *T. pigmentosa* и *T. pyriformis* около ВКЭ 234 расположен ВКЭ 290. Выделенное в его составе полужирным шрифтом слово TGTA соответствует консенсусу YRTA промоторов в митохондриях растений [16]. Сравнение потенциальных промоторов в составе ВКЭ 290 и ВКЭ 299 позволяет выделить консервативный мотив YRTA_{nn}AATTY. Однако гены вокруг ВКЭ 290 расположены на комплементарной цепи.

ВКЭ 138 и 315. У *Tetrahymena* spp. ген *cob*, кодирующий аросytochrome b, расположен дивергентно с геном *ymf77*. Согласно аннотации генома митохондрии *Tetrahymena pyriformis* в промежутке между этими генами около гена *ymf77* расположен элемент PAL2, аналогичный паразитическому элементу PAL2-1 из митохондрий грибов *Neurospora* и *Podospora*, который препятствует старению и увеличивает время жизни [17, стр. 203]. В этом межгенном участке, но ближе к гену *cob* найден консервативный мотив. Он соответствует ВКЭ 138, который найден у большего числа видов: у *Ichthyophthirius multifiliis* (два участка, оба между парами соседних генов, кодирующих large subunit ribosomal RNA), у *Tetrahymena malaccensis*, *T. paravorax*, *T. pigmentosa*, *T. pyriformis* и *T. thermophila*. Различие локализации ВКЭ 138 у *Ichthyophthirius multifiliis* и *Tetrahymena* spp. подтверждает, что элемент связан не с геном, а с мобильным элементом.

В той же области генома расположен ВКЭ 315, найденный лишь у четырёх видов рода *Tetrahymena*. В трёх из четырёх случаев он содержит слово CGTA, соответствующее консенсусу YRTA промоторов в митохондриях растений [16]. Возможно, это – промотор оперона, начинающегося с гена *cob*. Однако у *T. pyriformis* в сайте произошла замена одного нуклеотида.

ВКЭ 315 не продолжается на другие виды из класса Oligohymenophorea, что позволяет предположить наличие у них другого промотора перед геном *cob*. Действительно, потенциальный промотор с иным нуклеотидным составом обнаружен у *Ichthyophthirius multifiliis* и *Paramecium* spp.

На рис. 2 показано выравнивание фрагментов 5'-лидерных областей непосредственно перед геном *cob*, кодирующим аросytochrome b у *Ichthyophthirius multifiliis* и *Paramecium* spp. Серым цветом выделен консервативный участок, имеющий слабое сходство с промоторами в митохондриях растений, однако указанный участок не содержит подслово

вида YRTA, типичного для таких промоторов [16]. У этих видов ген *cob* окружён генами, расположенными на той же цепи ДНК, однако длина межгенного промежутка в 5'-лидерной области гена *cob* относительно большая.

```
I. multifiliis  ATAATTGTTAGAGAAAСТААСGАСААСААТААТТТТТТААААТТАСТТСТААТААТТАТАТАТТТАААТТТТ
P. caudatum   ТТТАААТТТТТАААGАААТАААААТТАТАТТГТАТТТТТТАСААТААААТАССТАААСAGGTATTGAATTATTСG
P. aurelia    САСАСАТТТТТGAAGAGААААAGСТАССТТТТТАGAGGTGGCTATCGCTCTGCTCGGCAGCAGCGTAGATTTAG
```

Рис. 2. Выравнивание фрагментов 5'-лидерных областей перед геном *cob*

Кластеризация белков, кодируемых в митохондриях

Нашим алгоритмом [18] множество белков, кодируемых в митохондриях рассмотренных видов, разбито на кластеры, предполагаемые семейства белков. Полученный результат представлен на странице лаборатории по адресу <http://lab6.iitp.ru/mpc/cilio/> в форме базы данных, включающей, в том числе, возможность поиска кластера по филогенетическому профилю белка. Отметим, что различные методы кластеризации обсуждаются также в [19].

Кластеризация 550-ти белков из 12-ти митохондрий представлена 63-мя неоднородными кластерами («нетривиальными»), и 109 одноэлементными кластерами («синглетоны»). Большинство синглетонов представлены белками из *Oxytricha trifallax* и *Nyctotherus ovalis*.

Только один кластер, состоящий из белков NADH dehydrogenase subunit 9 (Nad9), содержит паралоги. А именно, два вида *T. malaccensis* и *T. thermophila* рода *Tetrahymena* содержат очень близкие пары белков YP_740744.1 (Nad9_1) и YP_740745.1 (Nad9_2), NP_149392.1 (Nad9_1) и NP_149393.1 (Nad9_2), которые возникли в результате недавней дупликации, возможно, в ближайшем предке этих видов. Действительно, на деревьях, показанных далее на рис. 5 и 6, эти виды образуют кладу, а на дереве, показанном на рис. 1, в этом месте по существу имеется политомический узел. Однако возможно более тонкое заключение. Белки в каждой паре различаются лишь по одной позиции (своей для каждой пары), в то время как четыре белка, составляющие эти пары, различаются по 18-ти позициям. Поэтому представляется более вероятным, что в каждом из этих двух видов произошли независимые дупликации. Чтобы проследить эволюцию этих паралогов, с помощью программы PhyloBayes построено дерево кластера Nad9, рис. 3, на котором видно, в частности, что каждый из паралогов практически равноудалён от других белков семейства.

Распределение размеров кластеров показано на рис. 4; число белков каждого вида в кластерах и синглетонах приведено в таблице 2.

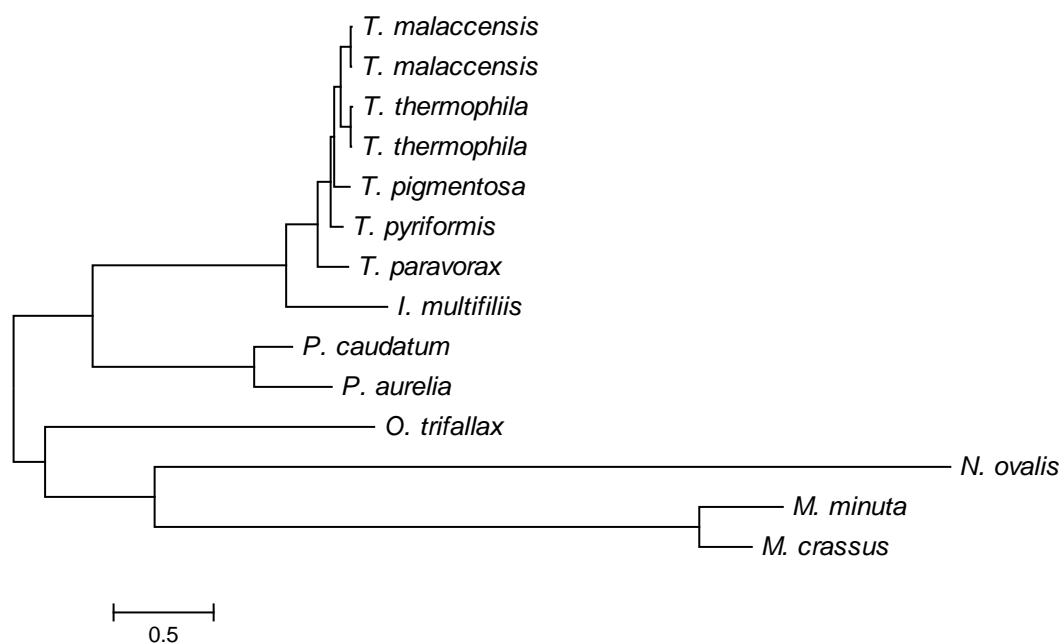


Рис. 3. Дерево белкового семейства NADH dehydrogenase subunit 9 (Nad9) из нашей кластеризации, построенное программой PhyloBayes

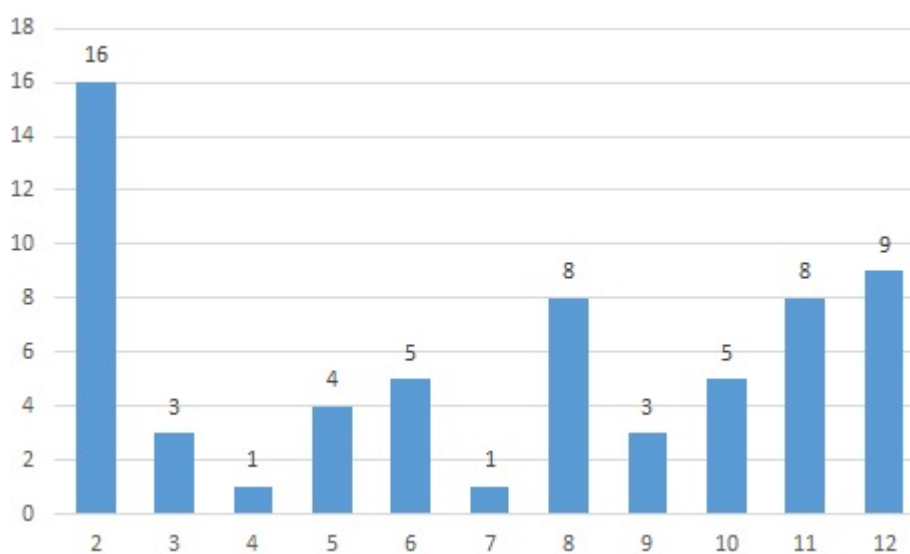


Рис. 4. Распределение размеров кластеров. По оси ординат показано число кластеров, включающих белки из указанного по оси абсцисс числа видов.

Наконец, отобраны все кластеры, представляющие не менее шести видов, таковых 39. По каждому из них, как и выше, с помощью MUSCLE получено выравнивание. К нему применена программа trimAl, удаляющая «малоинформативные» столбцы выравнивания. По выравниваниям получен конкатенат с общей длиной 8701 aa и долей missing data 26%. По конкатенату программой RAxML построено дерево эволюции митохондрий рассмотренных видов, которое хорошо согласуется с принятым таксономическим делением. Точно такое же дерево получено программой PhyloBayes. Последнее показано на рис. 5. Дерево весьма устойчиво, все клады имеют максимальную поддержку (бутстрэп-поддержку 100% в случае RAxML и единичную апостериорную вероятность в случае PhyloBayes).

Таблица 2. Распределение белков по кластерам и синглтонам. В трёх последних столбцах указано число кодируемых в митохондриях белков из соответствующего вида, а также – нетривиальных кластеров и синглтонов, в которых представлен вид.

Locus	Species	Proteins	Clusters	Singletons
NC_015981.1	<i>Ichthyophthirius multifiliis</i>	41	39	2
GQ903131.1	<i>Moneuplotes crassus</i>	29	25	4
GQ903130.1	<i>Moneuplotes minuta</i>	36	30	6
GU057832.1	<i>Nyctotherus ovalis</i>	35	13	22
JN383843.1	<i>Oxytricha trifallax</i>	99	31	68
NC_001324.1	<i>Paramecium aurelia</i>	46	41	5
NC_014262.1	<i>Paramecium caudatum</i>	42	41	1
NC_008337.1	<i>Tetrahymena malaccensis</i>	45	44	0
NC_008338.1	<i>Tetrahymena paravorax</i>	44	43	1
NC_008339.1	<i>Tetrahymena pigmentosa</i>	44	44	0
NC_000862.1	<i>Tetrahymena pyriformis</i>	44	44	0
NC_003029.1	<i>Tetrahymena thermophila</i>	45	44	0

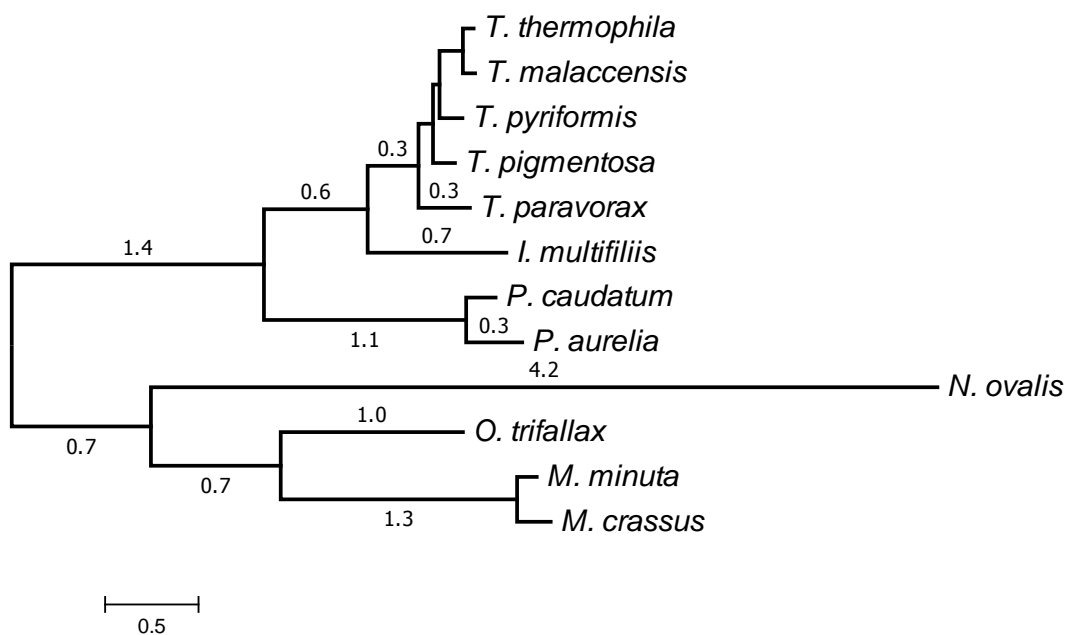


Рис. 5. Дерево эволюции митохондрий, построенное по полученным нами белковым семействам программами PhyloBayes. Все клады имеют максимальную поддержку.

Методы

Из базы данных GenBank получены полные геномы митохондрий видов *Ichthyophthirius multifiliis* (NC_015981), *Paramecium aurelia* (NC_001324), *Paramecium caudatum* (NC_014262), *Tetrahymena malaccensis* (NC_008337), *Tetrahymena paravorax* (NC_008338), *Tetrahymena pigmentosa* (NC_008339), *Tetrahymena pyriformis* (NC_000862), *Tetrahymena thermophila* (NC_003029). Из того же источника получены четыре неполных генома митохондрий *Moneuplotes minuta* (GQ903130), *Moneuplotes crassus* (GQ903131), *Nyctotherus ovalis* (GU057832) и *Oxytricha trifallax* или *Sterkiella histriomuscorum* (JN383843).

Поиск ВКЭ выполнен предварительным алгоритмом, который основан на поиске плотного подграфа и описан в работе [14]. Близкий метод, связанный с псевдобулевым программированием, обсуждается в [21]. В программе использовались следующие значения параметров: длина ключа 8, минимальная длина слова 24, максимальная цена различий между словами 3.1; минимальная длина перекрытия склеиваемых слов 20, число подряд делеций 0, цена делеции 2.1; максимальное число повторов ключа 1000, максимальный коэффициент сжатия слова 2.2; минимальное число разных букв – 4 в слове и 3 в ключе.

Дерево ВКЭ на рис. 1 построено программой RAxML [15] на основе матрицы из 12 строк и 393 столбцов, элемент которой 1 или 0 указывает соответственно на присутствие или отсутствие у данного вида представителя данного ВКЭ. Поиск по максимуму правдоподобия с интенсивным бутстрепом выполнялся программой RAxML v. 8.2.4 с двоичной моделью замен и оценкой частот нуклеотидов на основе максимального правдоподобия. Число бутстрепных повторений в частотном критерии было установлено равным 300.

Расстояния между хромосомными структурами и реконструкция хромосомных перестроек получены методами, описанными в [22, 20, 23]. Используются значения цен операций, которые предполагаются в программе по умолчанию; а именно, линейный вариант и значения: двойная переклейка 1.2, полуторная переклейка 1.1, добавление *a*-ребра и удаление *b*-ребра 1, добавление *b*-ребра или удаление *a*-ребра 0.9, удаление *a*-особых вершин 2.0, удаление *b*-особых вершин 2.5. По расстояниям между хромосомными структурами методом neighbor-joining [24] построено бескорневое дерево, показанное на рис. 2.

Кластеризация белков выполнена методом, описанным и апробированным в [25, 26, 27, 28, 18]. В кластеризации использовалась программа BLAST с порогом сходства $E=0.001$ и максимально ослабленные значения прочих параметров: $L=0$, $H=1$, $p=\infty$.

Выравнивание белков выполнялось программой MUSCLE v. 3.8.31 [29] с параметрами по умолчанию. Затем из каждого выравнивания с помощью программы trimAl v. 1.2 [30] удалялись позиции, содержащие более 50% гэпов или имеющие сходство менее 0.001. Для построения дерева использовались программы RAxML [15] и PhyloBayes v. 4.1 [31, 32, 33] с митохондриальной моделью MtZoa [34]. В случае PhyloBayes четыре независимые цепи запускались более чем на тысячу циклов. По достижении сходимости значения правдоподобия, параметра альфа и длины дерева в четырех цепях, величина рассогласования частот между всеми цепями при разбиении пополам обратилась в ноль (как показала утилита brcomp из пакета PhyloBayes). Первые 100 циклов каждой цепи использовались для раскрутки, а консенсусное дерево с апостериорными вероятностями вычислялось по остальным деревьям всех цепей. Оба алгоритма привели к одинаковому дереву. К дереву с такой же топологией привело и применение более общей модели CAT+GTR+Г в PhyloBayes и GTR+Г в RAxML.

Поиск потенциальных сайтов связывания транскрипционных факторов и промоторов выполнен методом, описанным в работах [35, 36]. Этот же метод применён для поиска сайтов связывания транскрипционных факторов в пластидах водорослей [25, 26, 27].

Полученные из базы данных GenBank аннотации генов, перекрывающих ВКЭ, согласованы с содержанием базы данных Rfam v. 12.1 [37].

Литература

1. Ванятинский В.Ф., Мирзоева Л.М., Поддубная А.В. *Болезни рыб* / Под редакцией В.А. Мусселиус. М.: Пищевая промышленность, 1979.
2. de Graaf R.M., van Alen T.A., Dutilh B.E., Kuiper J.W.P., van Zoggel H.J.A.A., Huynh M.B., Görtz H.-D., Huynen M.A., Hackstein J.H.P. The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*. *BMC Genomics*. 2009. V. 10, no. 514. DOI: 10.1186/1471-2164-10-514
3. de Graaf R.M., Ricard G., van Alen T.A., Duarte I., Dutilh B.E., Burgtorf C., Kuiper J.W., van der Staay G.W., Tielens A.G., Huynen M.A., Hackstein J.H. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Mol. Biol. Evol.* 2011. V. 28, no. 8. P. 2379–2391. DOI: 10.1093/molbev/msr059
4. Гармаш Е.В. Митохондриальное дыхание фотосинтезирующей клетки. *Физиология растений*. 2016. Т. 63, № 1. С. 17–30. DOI: 10.1134/S1021443715060072
5. Kairo A., Fairlamb A.H., Gobright E., Nene V. A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. *EMBO J.* 1994. V. 13, no. 4. P. 898–905.
6. Barth D., Berendonk T.U. The mitochondrial genome sequence of the ciliate *Paramecium caudatum* reveals a shift in nucleotide composition and codon usage within the genus *Paramecium*. *BMC Genomics*. 2011. V. 12, no. 272. DOI: 10.1186/1471-2164-12-272
7. Brunk C.F., Lee L.C., Tran A.B., Li J. Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucleic Acids Res.* 2003. V. 31, no. 6. P. 1673–1682. DOI: 10.1093/nar/gkg270
8. Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N., Lang B.F., Gray M.W. Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *J. Mol. Biol.* 2000. V. 297, no. 2. P. 365–380. DOI: 10.1006/jmbi.2000.3529
9. Cummings D.J. Mitochondrial genomes of the ciliates. *Int. Rev. Cytol.* 1992. V. 141. P. 1–64.
10. Edqvist J., Burger G., Gray M.W. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J. Mol. Biol.* 2000. V. 297, no. 2. P. 381–393. DOI: 10.1006/jmbi.2000.3530
11. Swart E.C., Nowacki M., Shum J., Stiles H., Higgins B.P., Doak T.G., Schotanus K., Magrini V.J., Minx P., Mardis E.R., Landweber L.F. The *Oxytricha trifallax* mitochondrial genome. *Genome Biol Evol.* 2012. V. 4, no. 2. P. 136–154. DOI: 10.1093/gbe/evr136
12. Moradian M.M., Beglaryan D., Skozylas J.M., Kerikorian V. Complete mitochondrial genome sequence of three tetrahymena species reveals mutation hot spots and accelerated nonsynonymous substitutions in Ymf genes. *PLoS ONE*. 2007. V. 2, no. 7. E650. DOI: 10.1371/journal.pone.0000650

13. Pritchard A.E., Cummings D.J. Replication of linear mitochondrial DNA from *Paramecium*: sequence and structure of the initiation-end crosslink. *Proc. Natl. Acad. Sci. U.S.A.* 1981. V. 78, no. 12. P. 7341–7345.
14. Rubanov L.I., Seliverstov A.V., Zverkov O.A., Lyubetsky V.A. A method for identification of highly conserved elements and evolutionary analysis of superphylum Alveolata. *BMC Bioinformatics*. 2016. V. 17, no. 385, 16 pages. DOI: 10.1186/s12859-016-1257-5
15. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014. V. 30, no. 9. P. 1312–1313. DOI: 10.1093/bioinformatics/btu033
16. Kühn K. Analysis of components of the mitochondrial transcription machinery in *Arabidopsis thaliana*. Dissertation. Humboldt-Universität zu Berlin, 2006.
17. Villarreal L.P. *Origin of Group Identity: Viruses, Addiction and Cooperation*. Springer Science & Business Media, 2008. 614 p.
18. Любецкий В.А., Селиверстов А.В., Зверков О.А. Построение разделяющих паралоги семейств гомологичных белков, кодируемых в пластидах цветковых растений. *Математическая биология и биоинформатика*. 2013. Т. 8, № 1. С. 225–233.
19. Sunshin Kim, KyuBum Kwack. A Fast Comparison Algorithm to Measure the Accuracy of Ortholog Clusters. *Current Bioinformatics*. 2016. V. 11, no. 3. P. 324–329. DOI: 10.2174/1574893611666160322233309
20. Lyubetsky V.A., Gershgorin R.A., Seliverstov A.V., Gorbunov K.Yu. Algorithms for reconstruction of chromosomal structures. *BMC Bioinformatics*. 2016. V. 17, no. 40, 23 pages. DOI: 10.1186/s12859-016-0878-z
21. Seliverstov A.V. Monomials in quadratic forms. *Journal of Applied and Industrial Mathematics*. 2013. V. 7, no. 3. P. 431–434. DOI: 10.1134/S1990478913030162
22. Горбунов К.Ю., Гершгорин Р.А., Любецкий В.А. Перестройка и реконструкция хромосомных структур. *Молекулярная биология*. 2015. Т. 49, № 3. С. 372–383. DOI: 10.7868/S0026898415030076
23. Горбунов К.Ю., Любецкий В.А. Линейный алгоритм кратчайшей перестройки графов при разных ценах операций. *Информационные процессы*. 2016. Т. 16, № 2. С. 223–236.
24. Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987. V. 4, no. 4. P. 406–425.
25. Lyubetsky V.A., Seliverstov A.V., Zverkov O.A. Transcription regulation of plastid genes involved in sulfate transport in Viridiplantae. *BioMed Research International*. 2013. V. 2013, Article ID 413450, 6 pages. DOI: 10.1155/2013/413450
26. Zverkov O.A., Seliverstov A.V., Lyubetsky V.A. A Database of plastid protein families from red algae and Apicomplexa and expression regulation of the *moeB* gene. *BioMed Research International*. 2015. V. 2015, Article ID 510598, 5 pages. DOI: 10.1155/2015/510598
27. Zverkov O.A., Seliverstov A.V., Lyubetsky V.A. Regulation of expression and evolution of genes in plastids of rhodophytic branch. *Life*. 2016. V. 6, no. 7, 13 pages. DOI: 10.3390/life6010007
28. Зверков О.А., Селиверстов А.В., Любецкий В.А. Белковые семейства, специфичные для пластовов небольших таксономических групп водорослей и простейших. *Молекулярная биология*. 2012. Т. 46, № 5. С. 799–809. DOI: 10.1134/S0026893312050123
29. Edgar R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2014. V. 32. P. 1792–1797. DOI: 10.1093/nar/gkh340

30. Capella-Gutierrez S.; Silla-Martinez J.M.; Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009. V. 25, P. 1972–1973. DOI: 10.1093/bioinformatics/btp348
31. Lartillot N., Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*. 2004. V. 21. no. 6. P. 1095–1109. DOI: 10.1093/molbev/msh112
32. Lartillot N., Philippe H. Computing Bayes factors using thermodynamic integration. *Systematic Biology*. 2006. V. 55, no. 2. P. 195–207. DOI: 10.1080/10635150500433722
33. Lartillot N., Brinkmann H., Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evolutionary Biology*. 2007. V. 7 (Suppl 1.):S4. DOI: 10.1186/1471-2148-7-S1-S4.
34. Rota-Stabelli O., Yang Z., Telford M.J. MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies. *Molecular Phylogenetics and Evolution*. 2009. V. 52. P. 268–272. DOI:10.1016/j.ympev.2009.01.011
35. Селиверстов А.В., Лысенко Е.А., Любецкий В.А. Быстрая эволюция промоторов пластомных генов *ndhF* у цветковых растений. *Физиология растений*. 2009. Т. 56, № 6. С. 926–934. DOI: 10.1134/S1021443709060144
36. Lyubetsky V.A., Rubanov L.I., Seliverstov A.V. Lack of conservation of bacterial type promoters in plastids of Streptophyta. *Biology Direct*. 2010. V. 5, no. 34. DOI: 10.1186/1745-6150-5-34
37. Nawrocki E.P., Burge S.W., Bateman A., Daub J., Eberhardt R.Y., Eddy S.R., Floden E.W., Gardner P.P., Jones T.A., Tate J., Finn R.D. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*. 2015. V. 43, no. D1. P. D130–D137. DOI: 10.1093/nar/gku1063

Таблица 3. Реконструкции митохондриальных хромосомных структур инфузорий. В корне и двух его сыновьях – по две хромосомы (линейная и циклическая). В других внутренних вершинах – по одной линейной хромосоме.

лист или крайние листья – потомки вершины	Хромосомная структура	События на ребре, входящем в вершину
<i>T. thermophila</i> (l)	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_2 *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	нет событий
<i>T. malaccensis</i> (l)	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_2 *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	нет событий
<i>T. thermophila</i> – <i>T. malaccensis</i>	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_2 *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	1 дупликация
<i>T. paravorax</i> (l)	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	нет событий

<i>T. paravorax</i> – <i>T. malaccensis</i>	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	нет событий
<i>T. pigmentosa</i> (l)	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 trnH *nad3 *ymf72 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	2 инверсии
<i>T. paravorax</i> – <i>T. pigmentosa</i>	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	нет событий
<i>T. pyriformis</i> (l)	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 rpl6 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	1 замена гена
<i>T. paravorax</i> – <i>T. pyriformis</i>	*trnY *rnl_a_1 *trnL_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf74 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf69 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *ymf72 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 ymf71 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 trnL_2 rnl_a_2 trnM (L)	6 возникновений гена
<i>I. multifiliis</i> (l)	*trnY *rnl_a_1 *rnl_b_1 *ymf66 *ymf57 *ymf76 *rps13 *rps3 *rps19 *rpl2 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *atp9 *ymf63 *ymf65 *ymf59 *rpl16 *yejR *ymf61 *nad3 *nad4L *nad9_1 *ymf77 *nad1_b cob nad5 cox2 rns_a rns_b ymf67 trnW ymf68 cox1 nad1_a ymf62 rpl14 ymf70 nad4 ymf73 trnE rnl_b_2 rnl_a_2 trnY_2 (C)	2 потери гена, 1 замена гена, 2 транспозиции, 1 зацикливание хромосомы
<i>I. multifiliis</i> – <i>T. Pyriformis</i>	*trnY *rnl_a_1 *rnl_b_1 *ymf57 *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *nad10 *rps12 *nad2 nad7 rps14 ymf60 *ymf64 *ymf75 *trnF *nad1_b *atp9 *ymf63 *ymf65 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *nad4L *nad9_1 *ymf77 cob nad5 cox2 rns_a rns_b ymf56 ymf67 trnW ymf68 cox1 nad1_a ymf62 rpl14 trnE ymf70 nad4 ymf73 rnl_b_2 rnl_a_2 trnM (L)	1 вставка участка, 4 возникновения гена, 1 замена гена, 1 трансверсия, 1 транспозиция, 1 вставка циклической хромосомы в линейную
<i>P. aurelia</i> (l)	*trnY *rnl_b *trnM *rnl_a *ymf66_1 *ymf66 *ymf57 *nad4 *ymf80 *rpl14 *ymf62 *nad1_a *cox1 *ymf68 *trnW *ymf67_b *ymf67_a *ymf56 *rns_b *rns_a *cox2 *nad5 *cob *ymf81 *ymf85 *rps13 *rps3 *rpl2 *ymf84 *nad10 *rps12 *nad2_a nad7 rps14 ymf79 ymf60 *ymf64 *ymf86 *ymf83 *atp9 *ymf63 *nad1_b *trnF *ymf65 *ymf65_1 *ymf78 *ymf59 *rpl16 *ymf82 *yejR *ymf61 *nad3 *nad4L *nad9_1 (L)	2 возникновения гена 3 дубликации, 2 потери гена, 1 вставка участка
<i>P. caudatum</i> (l)	*trnY *rnl_a_1 *rnl_b_1 *ymf66 *ymf57 *nad4 *ymf80 *rpl14 *nad6 *nad1_a *cox1 *ymf68 *trnW *ymf67_a *ymf56 *rns_b *rns_a *cox2 *nad5 *cob *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf84 *nad10 *rps12 *nad2_a nad7 rps14 ymf79 rpl6 *ymf64 *ymf83 *atp9 *ymf63 *ymf87 *nad1_b *trnF *ymf65 *ymf78 *ymf59 *rpl16 *yejR *ymf61 *nad3 *nad4L *nad9_1 (L)	2 замены гена, 1 потеря гена, 1 возникновение гена
<i>P. aurelia</i> – <i>P. caudatum</i>	*trnY *rnl_a_1 *rnl_b_1 *trnM *ymf66 *ymf57 *nad4 *ymf80 *rpl14 *ymf62 *nad1_a *cox1 *ymf68 *trnW *ymf67_a *ymf56 *rns_b *rns_a *cox2 *nad5 *cob *ymf76 *rps13 *rps3 *rps19 *rpl2 *ymf84 *nad10 *rps12 *nad2_a nad7 rps14 ymf79 ymf60 *ymf64 *ymf83 *atp9 *ymf63 *nad1_b *trnF *ymf65 *ymf78 *ymf59 *rpl16 *yejR *ymf61 *nad3 *nad4L *nad9_1 (L)	1 потеря гена, 5 возникновений гена, 1 замена гена, 1 инверсия, 1 транспозиция, 1 трансверсия, 1 вставка циклической хромосомы в линейную
<i>P. aurelia</i> – <i>T. pyriformis</i>	*trnM *ymf66 *ymf76 *rps13 *rps3 *rps19 *rpl2 *nad10 *rps12 *nad2_a nad7 rps14 ymf60 *ymf64 *trnF *ymf65 *ymf59 *rpl16 *yejR *ymf61 *trnH *nad3 *nad4L *nad9_1 cob nad5 cox2 rns_a rns_b ymf56 trnW ymf68 cox1 nad1_a ymf62 rpl14 trnE nad4 ymf57 rnl_b_1 rnl_a_1 trnY (L) nad1_b ymf63 atp9 (C)	1 возникновение гена, 1 вставка участка, 2 инверсии, 4 транспозиции, 2 трансверсии
<i>M. minuta</i> (l)	*nad5 *ccmF *cytb *trnM *rnl *cox2 *rpl14 *cox1 *nad4L *rps3_b *trnW *rns *trnY trnF nad9_1 nad2_a rpl16 nad4 rps12 nad10 rpl2 rps4 nad7 nad1_b trnG atp9 nad3 trnH nad1_a (L)	нет событий

<i>M. crassus</i> (I)	*nad5 *ccmF *cytb *trnM *rnl *cox2 *rpl14 *cox1 *nad4L *rps3_b *rns *trnY trnF nad9_1 nad2_a rpl16 nad4 rps12 nad10 rpl2 rps4 nad7 (L)	1 удаление участка, 1 потеря гена
<i>M. minuta</i> – <i>M. crassus</i>	*nad5 *ccmF *cytb *trnM *rnl *cox2 *rpl14 *cox1 *nad4L *rps3_b *trnW *rns *trnY trnF nad9_1 nad2_a rpl16 nad4 rps12 nad10 rpl2 rps4 nad7 nad1_b trnG atp9 nad3 trnH nad1_a (L)	3 потери гена, 1 удаление участка, 3 возникновения гена, 1 инверсия
<i>O. trifallax</i> (I)	*rps2_sin *nad5 *nad5_iii_sin *nad5_ii_sin *nad5_i_sin *ccmF *nad1_a *trnH *cob *nad3 *rpl6_ii_sin *rpl6_i_sin *rps7 *rps3_a_sin *atp9 *rps8_sin2 *trnG *nad1_b *rps14 *nad7 *rps4 *rps13 *rps19 *rpl2 *nad10 *rps12 *nad4_i_sin *rpl16 *nad2_a_sin *nad2_b *trnL *rps10 *trnE *nad9 *nad9_i_sin *trnF trnY trnW rps3_b nad4L cox1 nad6_sin rpl14 cox2 trnM (L)	4 возникновения гена, 3 вставки участка, 3 замены гена, 2 дубликации
<i>M. minuta</i> – <i>O. trifallax</i>	*nad5 *ccmF *nad1_a *trnH *cob *nad3 *atp9 *trnG *nad1_b *rps14 *nad7 *rps4 *rps13 *rps19 *rpl2 *nad10 *rps12 *nad4 *rpl16 *nad2_a *trnE *nad9_1 *trnF trnY trnW rps3_b nad4L cox1 rpl14 cox2 trnM (L)	4 возникновения гена, 1 инверсия, 1 транспозиция, 2 трансверсии, 1 вставка циклической хромосомы в линейную
<i>N. ovalis</i> (I)	nad1_b nad1_a nad3 nad9_1 nad4 nad10 nad4L rpl14 nad2_a nad5 rpl2a_sin rpl2 nad7 rps14_sin rps8_sin1 rps4_sin rpl6 rps12 (L)	5 потерь гена, 1 вставка участка, 1 дубликация, 1 замена гена, 4 транспозиции, 1 вставка циклической хромосомы в линейную
<i>M. minuta</i> – <i>N. ovalis</i>	nad1_b atp9 rpl6 rps12 nad10 rpl2 rps19 rps13 nad7 rps14 *trnW *trnY trnF nad9_1 trnE nad2_a nad5 nad4L cox1 rpl14 cox2 trnM nad4 (L) trnH nad1_a nad3 cob(C)	4 удаления участка, 3 потери гена, 1 инверсия, 3 транспозиции, 2 трансверсии, 1 вставка циклической хромосомы в линейную, 1 вырезание циклической хромосомы из линейной
The tree root	*cob *nad1_a *yejR *ymf61 *trnH *nad3 *nad4L *rpl14 *ymf62 *cox1 *ymf68 *trnW *ymf56 *rns_b *rns_a *ymf66 *rps13 *rps3 *rps19 *rpl2 *nad10 *rps12 nad9_1 trnE nad2_a nad5 cox2 trnM rpl16 ymf59 ymf65 nad7 rps14 nad4 ymf57 ymf60 *ymf64 *trnF trnY (L) nad1_b ymf63 atp9 (C)	–