Highly Conserved Elements and Chromosome Structure Evolution in Mitochondrial Genomes in Ciliates

Roman A. Gershgorin [1], Konstantin Yu. Gorbunov [1], Oleg A. Zverkov [1,*], Lev I. Rubanov [1], Alexandr V. Seliverstov [1] and Vassily A. Lyubetsky [1,2]

[1]   Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoy Karetny per. 19, build.1, Moscow 127051, Russia; gershr@mail.ru (R.A.G.); gorbunov@iitp.ru (K.Y.G.); rubanov@iitp.ru (L.I.R.); slvstv@iitp.ru (A.V.S.); lyubetsk@iitp.ru (V.A.L.)

[2]   Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Leninskiye Gory 1, Main Building, Moscow 119991, Russia

*   Correspondence: zverkov@iitp.ru; Tel.: +7-495-694-3338

Abstract: Recent phylogenetic analyses are incorporating ultraconserved elements (UCEs) and highly conserved elements (HCEs). Models of evolution of the genome structure and HCEs initially faced considerable algorithmic challenges, which gave rise to (often unnatural) constraints on these models, even for conceptually simple tasks such as the calculation of distance between two structures or the identification of UCEs. In our recent works, these constraints have been addressed with fast and efficient solutions with no constraints on the underlying models. These approaches have led us to an unexpected result: for some organelles and taxa, the genome structure and HCE set, despite themselves containing relatively little information, still adequately resolve the evolution of species. We also used the HCE identification to search for promoters and regulatory elements that characterize the functional evolution of the genome.

1. Introduction

ATP and other compounds are synthesized in mitochondria [1]. Generally, many eukaryotes living under anaerobic conditions either lack mitochondria [2], or contain mitochondrial remnants such as hydrogenosomes or mitosomes. For example, *Nyctotherus ovalis*, anaerobic, lives in the hindgut of the cockroaches *Periplaneta americana* and *Blaberus* sp. [3]; its mitochondria generate hydrogen [4]. The role of mitochondria varies between different organisms, and is reflected in the size of the mitochondrial genomes. Parasitic apicomplexans have extremely small mitochondrial genomes coding for only three proteins and short rRNA fragments [5].

The ciliates (Ciliophora) include parasitic *Ichthyophthirius multifiliis* which causes death in many freshwater fish species reared in aquaria, fish farms, and aquacultures [6]. Mitochondria of ciliates can serve as targets for therapeutic intervention in parasitic diseases, and analysis of the structure and evolution of their genomes as well as the regulation of their gene expression can be of practical importance, in particular in veterinary medicine, e.g., for organization and veterinary care in fish hatcheries.

We analyzed the mitochondrial genomes in Ciliophora. The phylum Apicomplexa as well as the phylum Ciliophora belong to the superphylum Alveolata. We considered genera that belong to three classes: Armophorea (*Nyctotherus*), Oligohymenophorea (*Ichthyophthirius*, *Paramecium*,

and *Tetrahymena*), and Spirotrichea (*Moneuplotes*, *Oxytricha*). Twelve complete mitochondrial genomes considered here are listed at the beginning of Materials and Methods. Oligohymenophorea and Spirotrichea substantially differ [7]. Armophorea includes anaerobes but is closer to Spirotrichea than to Oligohymenophorea [4]. Many ciliates are free-living organisms. For example, *Moneuplotes minuta* cells can be collected in the Mediterranean Sea near Corsica [7]. Both *Moneuplotes minuta* and *Oxytricha trifallax* can be cultured in inorganic salt medium with *Chlamydomonas reinhardtii* or *Klebsiella* spp. as food sources. On the contrary, the ciliate *Ichthyophthirius multifiliis* is a pathogen of freshwater fish occurring in both temperate and tropical regions throughout the world [8]. It is less tolerant of salt than fish. Both *Tetrahymena* and *Paramecia* are free-living ciliate protozoa. *Tetrahymena* are common in freshwater ponds. *Paramecia* are widespread in freshwater, brackish, and marine environments and are often very abundant in stagnant basins and ponds. The endosymbionts of *Paramecium aurelia* are Gram-negative bacteria. Most of the endosymbionts produce toxins which kill sensitive strains of *Paramecia* [9].

The mitochondria considered here code for tens of proteins [4,7–15]. The functions of some of them remain unclear, and they relatively rapidly accumulate substitutions [16]. The mitochondrial chromosome is circular in *Ichthyophthirius* and linear in other species considered here [15,17]. In the mitochondria of *Tetrahymena*, *Moneuplotes*, and *Oxytricha*, most genes are transcribed in opposite directions from the middle of the linear chromosome. In contrast, most genes are transcribed in the same direction in the mitochondria of *Paramecium* and *Nyctotherus*.

The considered mitochondrial genomes are very compact. Genes form long operons with short non-coding regions; the coding regions can overlap in some cases. The order of genes differs between the classes considered, which makes the analysis of evolution of the chromosome structure alone a nontrivial task. The class Oligohymenophorea features relatively long non-coding regions upstream of the gene encoding apocytochrome b.

The goal of this report was confined to the application of the algorithm for the identification of highly conserved elements (HCEs) as well as the algorithm of chromosome structure evolution presented in [18,19] to the mitochondrial data of taxonomically distant species. In addition, the study of the statement formulated in the next paragraph was initiated as well as the analysis of the identified HCEs.

Traditionally, studies of species evolution to a large extent relied on the comparative analysis of genomic regions coding for rRNAs and proteins apart from the analysis of morphological characters. Later, analyses made use of regulatory elements and the structure of the genome as a whole. More recently, phylogenetic analyses start to incorporate ultraconserved elements (UCEs) and highly conserved elements (HCEs). Models of evolution of the genome structure and HCE initially faced considerable algorithmic challenges, which gave rise to (often unnatural) constraints on these models, even for conceptually simple tasks such as the calculation of distance between two structures or the identification of UCEs. These constraints are now being addressed with fast and efficient solutions with no constraints on the underlying models [18,19]. These approaches have led us to an unexpected result: at least for some organelles and taxa, the genome structure and HCE set, despite themselves containing relatively little information, still adequately resolve the evolution of species.

We also used the HCE identification to search for promoters and regulatory elements that characterize the functional evolution of the genome.

2. Results and Discussion

2.1. Highly Conserved Elements in Mitochondrial Genome in Ciliates (Ciliophora)

Our computer program based on the original algorithm [18] was used to identify highly conserved DNA elements referred to as HCEs. As a result, 393 HCEs have been identified and assigned unique numbers (see Table S1). Figure 1 demonstrates the tree generated by RAxML [20] from a matrix with 12 rows and 393 columns showing the presence or absence in each mitochondrial genome of each HCE. Notice that this popular program deals with a matrix of ones and zeros, which distinguishes it from, e.g., PhyloBayes and the neighbor-joining method. RAxML implements the maximum

likelihood method (ML). This tree is in good but naturally imprecise agreement with the species tree based on GenBank taxonomy. In particular, *Moneuplotes crassus* more commonly shared HCEs with *Oxytricha trifallax* than with *Moneuplotes minuta*, while *Paramecium aurelia* notably differed from *P. caudatum* by the HCE pattern.
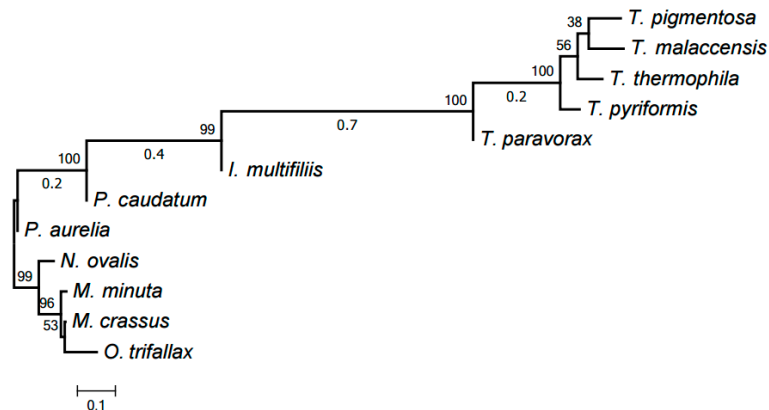
Figure 1. The tree of mitochondrial evolution generated using 393 HCEs identified by our algorithm. The tree was generated by the RAxML program based on a matrix with 12 rows and 393 columns, with the matrix cells containing 1 or 0 to indicate the presence or absence of a given HCE in the mitochondrial genome of a given species, respectively.

Five HCEs have been found in Oligohymenophorea (assigned numbers 138, 234, 287, 290, and 315), neither overlapping with gene coding regions nor corresponding to RNA species described in Rfam. Four out of five of these HCEs have been found only within the *Tetrahymena* genus. The identified HCEs are described in Table S1. Table 1 exemplifies six HCEs found in Oligohymenophorea.

Table 1. Six highly conserved elements (HCEs) represented in the class Oligohymenophorea.

Species	1st Position	Sequence Fragments
		HCE 287
T. malaccensis	2984	AATTTAAATACTTGCATTAAGACTAATCGTGG
T. pigmentosa	2988	AATTTAAATACTTGCATTAAGACTAATCGTGG
T. pyriformis	2988	AATTTAAAAGCTTGCATTAATACTAATCTTGG
T. thermophila	2943	AATTTAAACACTTGCATTAAAACTAATCTTGG
		HCE 299
T. malaccensis	10523	GACACAC**CATA**TGAATTTAAATCATTAATAATTCAA
T. pigmentosa	10558	GATAAAC**CATA**TGAATTTAAATTATTACTAATTAAA
T. pyriformis	10589	GATAGAC**CATA**AGAATTTAAGTCATTATTTATTCAA
T. thermophila	10500	GATAGAC**CATA**TGAATTTAAATCATTATTAATTCAA
		HCE 290
T. malaccensis	4810	ATAAAATAAGTTCTAAAAATG**TGTA**TTAATTCCTTAAACATTTA
T. paravorax	5270	ATAAAATAAGTTCTTAATATA**TGTA**TAAATTCTTTAAACATTTA
T. pigmentosa	4811	ATAAAATATGTTCTAAAAATA**TGTA**TTAATTCTTTAAACATTTA
T. pyriformis	4839	ATAAAATAAGTTCTAAAAATA**TGTA**TCAATTCTTTAAACATTTA
		HCE 234
T. malaccensis	4788	TTTTTTTAAATATCTAAAAGTAATAAAATAAGTTCTAAA
T. paravorax	5248	TTTTTTAAATATCTAAATGTTATAAAATAAGTTCTTAA
T. pigmentosa	4789	TTTTTTAAAATATCTAAAAGTTATAAAAATATGTTCTAAA
T. pyriformis	4817	TTTTTTGATATATCTAAAAGTGATAAAATAAGTTCTAAA
T. thermophila	4756	TTTTTTTTAAATATCTAAAAGTAATAAAATAAGTTCTAAA
		HCE 138
I. multifiliis	1364	TTTAGGTGCAGCTAT
I. multifiliis	47702	TATAGCTGCACCTAAAAAAAAAAAA
T. malaccensis	27009	AATAGCCGCACCTAAAAGAAAAAAATCTA
T. paravorax	26884	AATAGCTGCTCCAAAAAGAAAAAAATCAA
T. pigmentosa	26364	AATAGCCGCACCTAAAAGAAAAAAATCCA
T. pyriformis	26770	AATGGCCGCACCTAAAAGAAAAAAATCAA
T. thermophila	27061	AATAGCCGCACCTAAAAGAAAAAAATCTA
		HCE 315
T. malaccensis	26891	ATAA**CGTA**TTTACAATAAAAAAAATAAT
T. pigmentosa	26211	TCAA**CGTA**TTTACAATAAAATAATAAA
T. pyriformis	26678	TTAA**CGAA**TTTACAATAAAAAAATAAA
T. thermophila	26921	TTAA**CGTA**TCTACAATAAAAAAATAAA

HCE 287 has been found only in four *Tetrahymena* species. It is located upstream of the rRNA large subunit (on the complementary strand). It can be involved in the regulation of transcription or in post-transcriptional modifications of rRNA.

HCE 299. The mitochondrial *nad2* and *nad7* genes have opposite orientations and close positions in Oligohymenophorea; each of them starts a long operon. The alignment of Nad2 amino acid sequences annotated in GenBank demonstrates that there are nearly no conserved positions at the N terminus. Conversely, the *nad7* genes are highly conserved, and their 5′ ends overlap with HCE 151 in *Ichthyophthirius multifiliis*, *Tetrahymena malaccensis*, *T. paravorax*, *T. pigmentosa*, *T. pyriformis*, and *T. thermophile*.

This suggests that the *nad2* gene overlaps the promoter upstream of *nad7*. HCE 299, containing a potential promoter, has been found within the *nad2* coding regions in *Tetrahymena malaccensis*, *T. pigmentosa*, *T. pyriformis*, and *T. thermophile*. The CATA sequence (boldfaced in Table 1) corresponds to the YRTA consensus of promoters in plant mitochondria [21].

HCE 234 has been found in all *Tetrahymena* species between the *ymf76* and *ymf66* genes (both on the complementary strand). In four species, *T. malaccensis, T. paravorax, T. pigmentosa*, and *T. pyriformis*, HCE 234 is neighbored by HCE 290. The TGTA sequence (boldfaced in Table 1) corresponds to the YRTA consensus of promoters in plant mitochondria [21]. Analysis of potential promoters within HCE 290 and HCE 299 exposes a conserved motif, YRTAnnAATTY. However, the genes around HCE 290 are on the complementary strand.

HCEs 138 and 315. The *Tetrahymena* spp. *cob* gene coding for apocytochrome b is in an opposite orientation to the *ymf77* gene. The *Tetrahymena pyriformis* genome annotation indicates a PAL2 element between these genes close to *ymf77*, which is similar to the parasitic PAL2-1 element from the mitochondria of *Neurospora* and *Podospora*, a senescence factor in these fungi [22]. A conserved motif has been found in this intergenic region closer to the *cob* gene. It corresponds to HCE 138 found in a wide range of species including *Ichthyophthirius multifiliis* (two regions, both between pairs of the gene encoding the large subunit ribosomal RNA), *Tetrahymena malaccensis, T. paravorax, T. pigmentosa*, *T. pyriformis*, and *T. thermophila*. Different localization of HCE 138 in *Ichthyophthirius multifiliis* and *Tetrahymena* spp. confirms that this element is associated with the mobile element rather than with the gene.

The same genome region harbors HCE 315, which was found only in four *Tetrahymena* species. Three out of four of these elements contain the CGTA sequence corresponding to the YRTA consensus of promoters in plant mitochondria [21]. This can be a promoter of the operon starting with the *cob* gene. However, a nucleotide was substituted in this site in *T. pyriformis*.

HCE 315 has not been found in other Oligohymenophorea, which suggests the presence of another promoter upstream of the *cob* gene in them. Indeed, a potential promoter with a different sequence has been identified in *Ichthyophthirius multifiliis* and *Paramecium* spp.

Figure 2 shows the alignment of the 5′-leader sequences upstream of the *cob* gene in *Ichthyophthirius multifiliis* and *Paramecium* spp. The conserved region with low similarity to plant mitochondrial promoters is marked in grey; however, this region contains no YRTA site typical for such promoters [21]. The *cob* gene in these species is surrounded with other genes in the same DNA strand; however, the 5′-intergenic region of *cob* is relatively long.

I. multifiliis ATAATTGTATTAGAGAAACTAACGACAACAATATAATTTTTTATAAAATTACTTCTAATAATTATATATTTAAAATTTT
P. caudatum TTTAAATTTTTAAAGAAATAAAAATTATATTGTATTTTTTATACAATAAAATACCTAAACAGGTATTGAATTATTCG
*P. aurelia* CACACATTTTTGAAGAGAAAAAGCTACCTCTTTTAGAGGTGGCTATCGCTCTGCTCGGCAGCAGCGTAGATTTAG

Figure 2. Alignment of 5′-leader sequences upstream of the cob gene.

2.2. Clustering of Proteins Encoded in Mitochondria in Ciliates

We used our algorithm [23] to divide the proteins encoded in mitochondria into clusters, presumable protein families. The obtained protein families are available at [24] as a database, which can be searched by protein phylogenetic profile. It should be noted that different clustering methods are also discussed in [25].

Thus, 550 proteins from 12 mitochondria were divided into 63 non-single-element (nontrivial) clusters and 109 single-element clusters (singletons). Most singletons are represented by proteins from *Oxytricha trifallax* and *Nyctotherus ovalis*.

Only one cluster including NADH dehydrogenase subunit 9 (Nad9) proteins contains paralogs. Specifically, two *Tetrahymena* species, *T. malaccensis* and *T. thermophila*, include very similar pairs of proteins YP_740744.1 (Nad9_1) and YP_740745.1 as well as (Nad9_2) NP_149392.1 (Nad9_1) and NP_149393.1 (Nad9_2), emerging from a recent duplication, presumably in their nearest common ancestor. Indeed, these species form a clade in two evolutionary trees discussed below, while they essentially form a polytomous group in the HCE-based tree (Figure 1). However, this conclusion can be refined. The proteins in each of the two pairs differ by a single position (specific for each pair), while the four proteins composing these pairs differ by 18 positions. Hence, it is more reasonable to propose independent duplications in these two species. The evolution of these paralogs was reconstructed by generating the tree of the Nad9 cluster using the PhyloBayes program (Figure 3), in particular demonstrating that each paralog is nearly equidistant from other proteins of the family. PhyloBayes implements commonly used Bayesian inference.
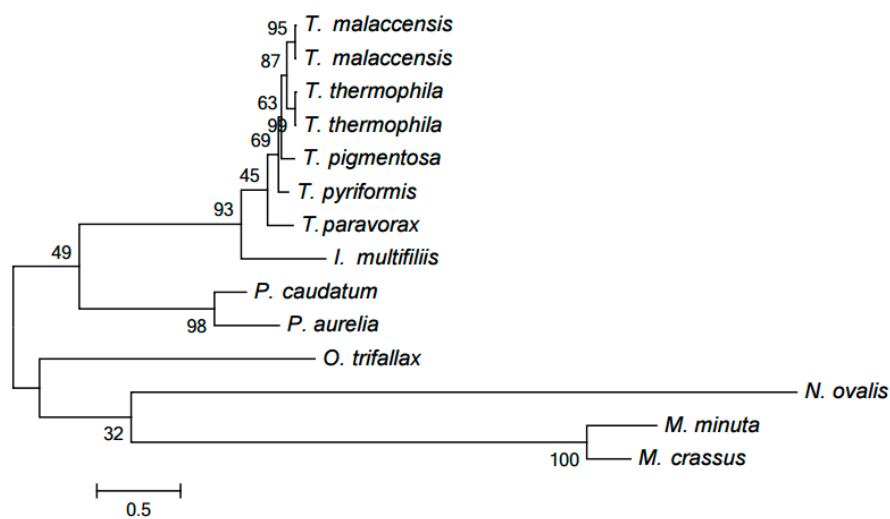
Figure 3. Tree of NADH dehydrogenase subunit 9 (Nad9) family according to our clustering. The tree was generated by PhyloBayes.

The size distribution of the clusters is shown in Figure 4; the number of proteins in each species in clusters and singletons is given in Table 2.
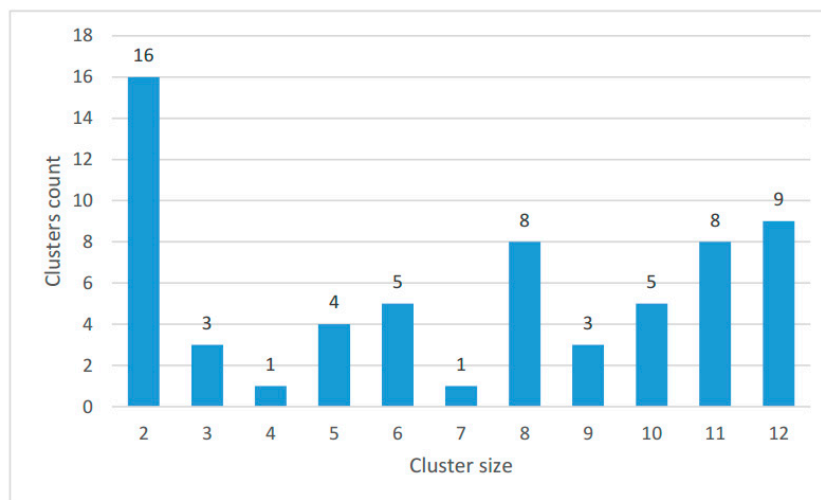
Figure 4. Size distribution of the clusters. The bar height shows the number of clusters including proteins from the number of species indicated on the abscissa.

Table 2. Distribution of proteins in clusters and singletons. Three columns on the right specify the numbers of proteins encoded in the mitochondrion, nontrivial clusters, and singletons for each species.

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

Finally, all clusters (39 in total) representing at least six species were selected. An alignment was generated for each of them using MUSCLE as described below in Materials and Methods. The trimAl program was then used to remove low-informative alignment columns. The alignments were concatenated into a single one with a total length of 8701 amino acids and the missing data ratio of 26%. RAxML was used to generate an evolutionary tree for the mitochondria of the species considered from this concatenated alignment; the tree was in a good agreement with the generally accepted taxonomy. Exactly the same tree has been generated by the PhyloBayes program (Figure 5). The tree has maximum support at all nodes (100% bootstrap values for RAxML and posterior probability of 1 for PhyloBayes). This is a common practice in tree building from protein data.
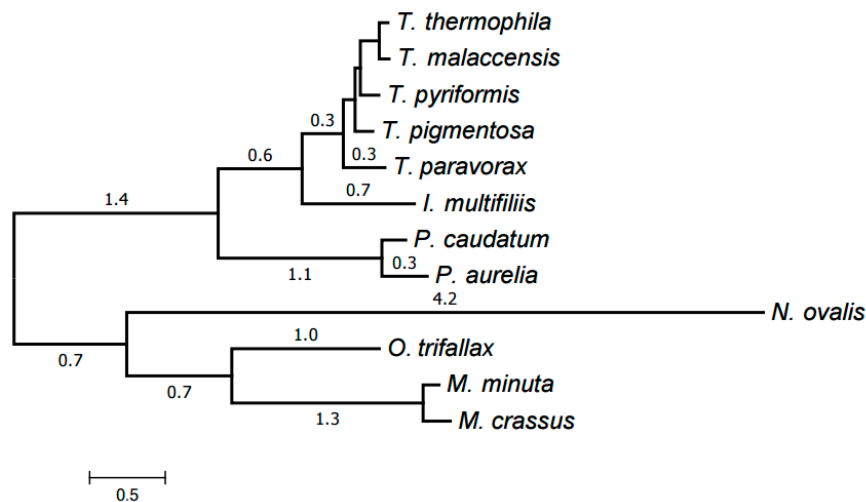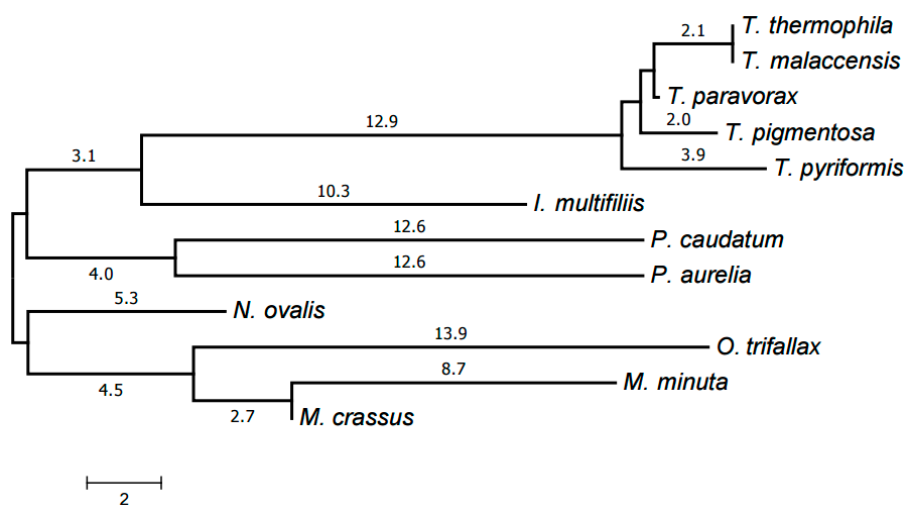
Figure 5. Evolutionary tree of mitochondria generated by PhyloBayes using the identified protein families. All nodes have the maximum support values.

2.3. Evolution of Mitochondrial Chromosome Structure in Ciliates

The evolutionary tree of mitochondrial chromosome structures was generated from the distances between them, calculated using the chromosome structure model proposed in [19] and the program available at [26].

The resulting tree is shown in Figure 6. Each genus forms a clade in it. The Armophorea, Oligohymenophorea, and Spirotrichea classes also form clades. The close position of Armophorea and Spirotrichea on the tree is consistent with published data [4]. Thus, there is a largely good agreement between the HCE-based tree (Figure 1), the tree of proteins (Figure 5), the tree of chromosome structures (Figure 6), and the generally accepted taxonomy. Minor differences between the trees shown in Figures 1, 4 and 5 can be attributed to the small size of the mitochondrial genomes and the corresponding relative scarcity of data.

Figure 6. Evolutionary tree of mitochondrial chromosome structures. The tree was generated by the neighbor-joining method using distances between chromosome structures calculated as described in [19].

2.4. Reconstruction of Mitochondrial Chromosome Structure in Ciliates

The results of the reconstruction of the mitochondrial chromosome structures in ciliates at the internal nodes of the tree generated by the method described in [19] are presented in Table S2.

The left column of the table lists the tree leaf designated as (*l*) or terminal (according to Figure 6) leaves descending from the considered internal node. The middle column contains the order of genes on the chromosome at the corresponding node; *L* and *C* indicate linear and circular chromosomes, respectively; the asterisk preceding the gene name indicates that it is encoded in the complementary strand. The second chromosome (if any) in the species corresponding to the node begins a new line. The chromosomes at the leaves served as initial data for our algorithm. The right column lists evolutionary events that occurred at the edge incident to the considered node.

3. Materials and Methods

Complete mitochondrial genomes were extracted from GenBank for the following species: *Ichthyophthirius multifiliis* (NC_015981), *Paramecium aurelia* (NC_001324), *P. caudatum* (NC_014262), *Tetrahymena malaccensis* (NC_008337), *T. paravorax* (NC_008338), *T. pigmentosa* (NC_008339), *T. pyriformis* (NC_000862), and *T. thermophila* (NC_003029). The same source was used to extract four incomplete mitochondrial genomes of *Moneuplotes minuta* (GQ903130), *M. crassus* (GQ903131), *Nyctotherus ovalis* (GU057832), and *Oxytricha trifallax* (*Sterkiella histriomuscorum*; JN383843).

The search for HCEs was performed using the algorithm based on the dense subgraph identification described elsewhere [18]. A similar method relying on pseudo-boolean programming is discussed in [27]. The following parameters were used: key length, 8; minimum word length, 24; maximum cost of difference between words, 3.1; minimum overlap length of merged words, 20; number of consecutive deletions, 0; deletion cost, 2.1; maximum key repeat count, 1000; maximum word compaction ratio, 2.2; minimum number of different words in a word and a key, 4 and 3, respectively.

The HCE-based tree in Figure 1 was generated using the RAxML program [20] from a matrix with 12 rows and 393 columns with the cells containing 1 or 0 to indicate the presence or absence of a given HCE in the mitochondrial genome of a given species, respectively. Maximum likelihood search followed by rapid bootstrapping was performed in RAxML v. 8.2.4 with the binary substitution model and maximum likelihood estimate for the base frequencies; number of bootstrap replicates was limited to 300 using the frequency-based criterion.

The distances between chromosome structures as well as the reconstruction of chromosome rearrangements were obtained by the methods described elsewhere [19,28,29]. The default operation costs were used, specifically: the linear variant and double-cut-and-paste, 1.2; sesqui-cut-and-paste, 1.1; *a*-edge insertion and *b*-edge deletion, 1; *b*-edge insertion or *a*-edge deletion, 0.9; deletion of special *a*-edges, 2.0; deletion of special *b*-edges, 2.5. The unrooted tree shown in Figure 6 was generated from the distances between chromosome structures using the neighbor-joining method [30].

Proteins were clustered using the method described and tested in [23,31–34]. BLAST hit threshold $E = 0.001$ and the most relaxed values for additional parameters: $L = 0$, $H = 1$, and very high *p* were used for clustering.

Protein alignment was performed by the MUSCLE program v. 3.8.31 [35] with default settings. Then, the trimAl program v. 1.2 [36] was used to remove positions with more than 50% gaps or with the similarity below 0.001. RAxML [20] and PhyloBayes v. 4.1 [37–39] with the MtZoa mitochondrial model [40] were used for tree generation. In the case of PhyloBayes, four chains ran in parallel for more than a thousand cycles. Upon convergence of likelihood values, alpha parameter, and tree length of the four chains, the discrepancy of bipartition frequencies between all chains was equal to zero (as shown by bpcomp utility in the PhyloBayes package). The first hundred cycles of each chain were discarded as burn-in and the majority rule consensus tree containing the posterior probabilities was calculated from the remaining trees of all chains. Both algorithms yielded the same tree. Trees with the same topology were also generated using more general CAT + GTR + Γ model in PhyloBayes and GTR + Γ model in RAxML.

Potential binding sites of transcription factors and promoters were identified using the method described elsewhere [41,42]. This method was used previously to identify binding sites of transcription factors in algal plastids [23,33,34].

GenBank gene annotations overlapping with HCEs were additionally checked against the Rfam database v. 12.1 [43].

4. Conclusions

At least for some organelles and taxa, the genome structure and HCE set, despite themselves containing relatively little information, still adequately describe the evolution of species. Indeed, the trees of HCEs, proteins, chromosome structures, and species are in agreement for the considered material. HCEs were found in mitochondrial genomes of ciliates (Ciliophora). Families of proteins encoded in mitochondria as well as the evolution of the chromosome structure were described in ciliate species. The data obtained were used to propose a method of combined application of our original methods to describe HCEs, protein families, and chromosome structures and eventually their evolution.

Author Contributions: Roman A. Gershgorin and Konstantin Yu. Gorbunov analyzed rearrangements of chromosome structures. Oleg A. Zverkov performed protein clustering and generated trees of HCEs and proteins. Lev I. Rubanov identified HCEs and inquired their functions in the databases. Alexandr V. Seliverstov analyzed the regulation of gene expression; Alexandr V. Seliverstov and Vassily A. Lyubetsky conducted the study and were major contributors in writing the manuscript. All authors discussed, read, and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Garmash, E.V. Mitochondrial respiration of the photosynthesizing cell. *Russ. J. Plant Physiol.* **2016**, *63*, 13–25. [CrossRef]
2. Karnkowska, A.; Vacek, V.; Zubáčová, Z.; Treitli, S.C.; Petrželková, R.; Eme, L.; Novák, L.; Žárský, V.; Barlow, L.D.; Herman, E.K.; et al. A eukaryote without a mitochondrial organelle. *Curr. Biol.* **2016**, *26*, 1274–1284. [CrossRef] [PubMed]
3. Van Hoek, A.H.; van Alen, T.A.; Sprakel, V.S.; Hackstein, J.H.; Vogels, G.D. Evolution of anaerobic ciliates from the gastrointestinal tract: Phylogenetic analysis of the ribosomal repeat from *Nyctotherus ovalis* and its relatives. *Mol. Biol. Evol.* **1998**, *15*, 1195–1206. [CrossRef] [PubMed]
4. 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.; et al. The organellar genome and metabolic potential of the hydrogen-producing mitochondrion of *Nyctotherus ovalis*. *Mol. Biol. Evol.* **2011**, *28*, 2379–2391. [CrossRef] [PubMed]
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**, *13*, 898–905. [PubMed]
6. Vanyatinsky, V.F.; Mirsoeva, L.M.; Poddubnaya, A.V. *Bolezni Ryb [Fish Diseases]*; Musselius, V.A., Ed.; Food Processing Industry: Moscow, Russia, 1979. (In Russian)
7. 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 Genom.* **2009**, *10*, 514. [CrossRef] [PubMed]
8. Matthews, R.A. *Ichthyophthirius multifiliis* Fouquet and ichthyophthiriosis in freshwater teleosts. *Adv. Parasitol.* **2005**, *59*, 159–241. [PubMed]
9. Preer, J.R., Jr.; Preer, L.B.; Jurand, A. Kappa and other endosymbionts in *Paramecium aurelia*. *Bacteriol. Rev.* **1974**, *38*, 113–163. [PubMed]
10. 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 Genom.* **2011**, *12*, 272. [CrossRef] [PubMed]
11. 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**, *31*, 1673–1682. [CrossRef] [PubMed]

12. 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**, *297*, 365–380. [CrossRef] [PubMed]

13. Cummings, D.J. Mitochondrial genomes of the ciliates. *Int. Rev. Cytol.* **1992**, *141*, 1–64. [PubMed]

14. Edqvist, J.; Burger, G.; Gray, M.W. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J. Mol. Biol.* **2000**, *297*, 381–393. [CrossRef] [PubMed]

15. 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.; et al. The *Oxytricha trifallax* mitochondrial genome. *Genome Biol. Evol.* **2012**, *4*, 136–154. [CrossRef] [PubMed]

16. 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**, *2*, E650. [CrossRef] [PubMed]

17. 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. USA* **1981**, *78*, 7341–7345. [CrossRef] [PubMed]

18. 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 Bioinform.* **2016**, *17*, 385. [CrossRef] [PubMed]

19. Lyubetsky, V.A.; Gershgorin, R.A.; Seliverstov, A.V.; Gorbunov, K.Y. Algorithms for reconstruction of chromosomal structures. *BMC Bioinform.* **2016**, *17*, 40. [CrossRef] [PubMed]

20. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]

21. Kühn, K.; Bohne, A.-V.; Liere, K.; Weihe, A.; Börner, T. *Arabidopsis* phage-type RNA polymerases: Accurate in vitro transcription of organellar genes. *Plant Cell* **2007**, *19*, 959–971. [CrossRef] [PubMed]

22. Villarreal, L.P. *Origin of Group Identity: Viruses, Addiction and Cooperation*; Springer: Berlin/Heidelberg, Germany, 2008.

23. Lyubetsky, V.A.; Seliverstov, A.V.; Zverkov, O.A. Transcription regulation of plastid genes involved in sulfate transport in Viridiplantae. *BioMed Res. Int.* **2013**, *2013*, 413450. [CrossRef] [PubMed]

24. The Ciliophora Mitochondria-Encoded Protein Clusters. Available online: http://lab6.iitp.ru/mpc/cilio/ (accessed on 25 November 2016).

25. Kim, S.; Kwack, K.B. A Fast Comparison Algorithm to Measure the Accuracy of Ortholog Clusters. *Curr. Bioinf.* **2016**, *11*, 324–329. [CrossRef]

26. The ChromoGGL Programs. Available online: http://lab6.iitp.ru/en/chromoggl/ (accessed on 25 November 2016).

27. Seliverstov, A.V. Monomials in quadratic forms. *J. Appl. Ind. Math.* **2013**, *7*, 431–434. [CrossRef]

28. Gorbunov, K.Y.; Gershgorin, R.A.; Lyubetsky, V.A. Rearrangement and inference of chromosome structures. *Mol. Biol.* **2015**, *49*, 327–338. [CrossRef]

29. Gorbunov, K.Y.; Lyubetsky, V.A. A linear algorithm of the shortest transformation of graphs under different operation costs. *Inf. Process.* **2016**, *16*, 223–236. (In Russian)

30. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [PubMed]

31. Lyubetsky, V.A.; Seliverstov, A.V.; Zverkov, O.A. Elaboration of the homologous plastid-encoded protein families that separate paralogs in Magnoliophytes. *Math. Biol. Bioinform.* **2013**, *8*, 225–233. (In Russian) [CrossRef]

32. Zverkov, O.A.; Seliverstov, A.V.; Lyubetsky, V.A. Plastid-encoded protein families specific for narrow taxonomic groups of algae and protozoa. *Mol. Biol.* **2012**, *46*, 717–726. [CrossRef]

33. 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 Res. Int.* **2015**, *2015*, 510598. [CrossRef] [PubMed]

34. Zverkov, O.A.; Seliverstov, A.V.; Lyubetsky, V.A. Regulation of expression and evolution of genes in plastids of rhodophytic branch. *Life* **2016**, *6*, 7. [CrossRef] [PubMed]

35. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2014**, *32*, 1792–1797. [CrossRef] [PubMed]

36. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef] [PubMed]

37. Lartillot, N.; Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **2004**, *21*, 1095–1109. [CrossRef] [PubMed]

38. Lartillot, N.; Philippe, H. Computing Bayes factors using thermodynamic integration. *Syst. Biol.* **2006**, *55*, 195–207. [CrossRef] [PubMed]

39. Lartillot, N.; Brinkmann, H.; Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **2007**, *7* (Suppl. 1), S4. [CrossRef] [PubMed]

40. Rota-Stabelli, O.; Yang, Z.; Telford, M.J. MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies. *Mol. Phylogenet. Evol.* **2009**, *52*, 268–272. [CrossRef] [PubMed]

41. Seliverstov, A.V.; Lysenko, E.A.; Lyubetsky, V.A. Rapid evolution of promoters for the plastome gene *ndhF* in flowering plants. *Russ. J. Plant Physiol.* **2009**, *56*, 838–845. [CrossRef]

42. Lyubetsky, V.A.; Rubanov, L.I.; Seliverstov, A.V. Lack of conservation of bacterial type promoters in plastids of Streptophyta. *Biol. Direct.* **2010**, *5*, 34. [CrossRef] [PubMed]

43. 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.; et al. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* **2015**, *43*, D130–D137. [CrossRef] [PubMed]