Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI:10.26479/2019.0501.50**

COMPARATIVE MITOCHONDRIAL GENOME ANALYSIS OF HYPOCREALEAN FUNGI WITH DIFFERENT LIFESTYLES

Prajna Muthabathula[1, 2*], Uma Devi Koduru[1], Suneetha Saragadam[1]

1.Department of Botany, Andhra University, Visakhapatnam, Andhra Pradesh, India.

2. Department of Botany, Adikavi Nannaya University, Rajamahendravaram, Andhra Pradesh, India.

ABSTRACT: Mitochondria play a key role in energy production in all eukaryotic systems and these are autonomous organelles containing their own genetic material in the form of mt genome. With the advent of DNA sequencing, mt genomes were sequenced and with the help of computational biology tools sequenced genomes can be annotated and analyzed in various ways to bring out the aspects like genetic structure, gene compositions, rearrangements, codon usage bias (CUB), intron analysis, synteny, phylogeny etc. The present study compares Hypocrealean fungal mt genomes. A variety of comparative analyses were performed using various bioinformatics tools to explore the details of mt genomes. Mt genomes being smaller with a limited number of key genes, the comparative analysis showed highly conserved nature among the organisms with similar lifestyle and they were grouped together when a phylogenetic tree is constructed. All the mt genomes encoded 14 conserved core protein coding genes with a slight varying number of tRNA genes.

KEYWORDS: mt genome, Hypocreale, bioinformatics, synteny, phylogeny.

Corresponding Author: Prajna Muthabathula*

1. Department of Botany, Andhra University, Visakhapatnam, Andhra Pradesh, India.
2. Department of Botany, Adikavi Nannaya University, Rajamahendravaram, Andhra Pradesh, India.

Email Address:prajna.muthabathula@gmail.com

1.INTRODUCTION

In eukaryotic cells, mitochondria play essential roles, primary function in respiration and energy production using electron transport coupled with oxidative phosphorylation to generate ATP [1]. The

mitochondria have a genome originating from an endosymbiotic α-proteobacterial ancestor whose genetic function is well conserved [2]. Mitochondrial genome (mt DNA) being small in size has conserved gene content [3]. Over the past few years, many mitochondrial genomes were sequenced and data is available through databases. A comparative analysis of the nuclear genomes of Entomopathogenic fungi (EPFs) with fungi with plant pathogenic (PPFs) habit has been done [4]. A number of studies provided insights into fungal mitochondrial (mt) genomes till to date [5, 6, 7, 8, 9, 10, 11]. Genome sequencing has become an important tool in understanding the organism to its genetic level, with the emerging strategies of sequencing. So many economically important organisms are being sequenced now-a-days. When compared to nuclear gnomes, organellar genomes have a quite smaller size and have been studied intensively. Hypocrealean fungi belong to the group Ascomycota which contains the sac fungi. The order Hypocreales comprises of fungi with diversified lifestyles viz., saprobes, pathogens (Plant pathogens, Entomopathogens, Nematopathogens etc.) and were reported to be a crucial part in forest succession [12]. PPFs pose a threat to the economically important crops. EPFs are responsible for significant natural control of insect populations [13, 14] with insect pests in crop fields decimated due to their epizootics. Mimicking this natural phenomenon, EPFs are being used in the biological control of insect pests in agriculture. All commercial biopesticide formulations are made from Hypocrealean EPFs. Nematopathogenic fungi (NPFs) play a vital role in natural biocontrol of nematodes. With the advent of genomics the nuclear and mitochondrial genomes of these fungi have been sequenced and analyzed. However, to our knowledge there has not been a comparative analysis on EPFs, PPFs and NPFmt genomes providing a picture on the mt genome analysis. Mt genome comparison paves way to identify the similarities and dissimilarities (if any) among the compared genomes that can provide insights into the evolutionary aspects. In the present study mitochondrial genomes of 7 fungi belonging to the order Hypocreales are investigated, which include EPFs (*Beauveria bassiana, Metarhizium anisopliae* and *Cordyceps militaris*), PPFs (*Fusarium graminearum, F. oxysporum* and *Verticillium dahliae*) and a NPF (*Hirsutella minnesotensis*). Here, therefore we set our investigation on comparing aspects like synteny (gene order), codon usage (codon bias), intron analysis, *trn* analysis.

2. MATERIALS AND METHODS

1. mt genome Sequence retrieval

Mitochondrial genomes of the three EPFs: *Beauveria bassiana, Metarhizium anisopliae* and *Cordyceps militaris*; three PPFs: *Fusarium graminearum, F. oxysporum* and *Verticillium dahliae* and a NPF: *Hirsutella minnesotensis* were downloaded from NCBI. The mt genome sequences were retrieved from NCBI in (.gbk, .faa, .fasta) formats (Accession numbers NC_010652, NC_008066,

NC_022834, HG970331, NC_017930, NC_008248 and NC_027660 respectively) (http://www.ncbi.nlm.nih.gov/genome/). The selected organisms' genomes were sequenced and these include both asexually and sexually reproducing fungi with different genome sizes.

2. Examination of General features of mt genomes

The general features of the genome such as size, GC content, gene content were examined. The GC content and Nucleotide composition (NC) was calculated using CAIcal tool (http://genomes.urv.cat/CAIcal/) [15].

3. tRNAand Codon Usage Bias (CUB) analysis

tRNA count in mt genomes is less than that of the nuclear tRNA count. In order to examine if the tRNA genes cater entirely to the protein synthesis of the mt protein coding genes, codon usage analysis was done with CodonW (http://codonw.sourceforge.net/) [16]. For this analysis, the coding sequences of all the seven mitochondrial genomes were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/genome/). All coding sequences of each corresponding organism are concatenated using concatenate option in codonW and are given as input data for calculation of overall Relative Synonymous Codon Usage (RSCU).

4. Intron analysis

Presence and enumeration of types of introns in the mt genes were identified using RNAweasel [17, 18] (http://megasun.bch.umontreal.ca/cgibin/RNAweasel/RNAweaselInterface.pl) and Gene Structure Display Server GSDS2.0 tool (http://gsds.cbi.pku.edu.cn/index.php) [19].

5. Synteny analysis

The synteny analysis of the genomes was done by Progressive mauve software (http://darlinglab.org/mauve/mauve.html) [20].

6. Phylogenetic analysis

The phylogenetic analysis of the Hypocrealean fungi was carried out by concatenation of the protein coding genes, by using Maximum Likelihood (ML) approach using MEGA version 6 [21] using default parameters.

3. RESULTS AND DISCUSSION

1. Gene content and genome organization

The comparative analysis of the fungal genomes shows that the fungi are very divergent [22] and are very dynamic in nature [23]. The mt genomes of seven fungi belonging to the order Hypocreales show their genome size ranging from 24.67kb (*M. anisopliae*) to 95.68kb (*F. graminearum*) and their GC content ranging from 26.8% (*C. militaris*) to 31.8% (*F. graminearum*) (Table 1). The mt genomes of fungi tend to be AT rich [24], which is evident from our study as well. The mitochondrial genomes encode a set of conserved genes including cytochrome c oxidase subunits (*cox1, cox2 &cox3*),

apocytochrome b (*cob*), three ATP synthase subunits (*atp6, atp8* & *atp9*), seven NADH dehydrogenase subunits (*nad1, nad2, nad3, nad4, nad4L, nad5* & *nad6*), the small and large ribosomal subunits (*rns* & *rnl*), *rps3* and tRNAs ranging from 24 to 28 (Table 1&2).

Table 1: General features of mitochondrial genomes of Hypocrealean fungi with different habit

S.No	Fungus	Gen bank Acc	Length (kb)	% GC	Genes				% of coding
					Protein	rRNA	tRNA	Total	
Entomopathogens (EPFs)									
1	***Beauveria bassiana***	NC_010652.2	29.96	27.2	15	2	25	42	48.27
2	***Metarhizium anisopliae***	NC_008068	24.67	28.4	15	2	24	41	59.01
3	***Cordyceps militaris***	NC_022834.1	33.28	26.8	15	2	26	43	43.10
Plant pathogens (PPFs)									
4	***Fusarium oxysporum***	NC_017930.1	34.48	31	15	2	25	42	44.66
5	***Fusarium graminearum***	NC_009493.1	95.68	31.8	50	2	28	46*	55.97
6	***Verticillium dahlia***	NC_008248.1	27.18	27.3	15	2	25	42	53.66
Nematophagous (NPF)									
7	***Hirsutella minnesotensis***	NC_027660.1	52.25	28.4	30	2	25	56*	57.34

***Having ORFs coding hypothetical proteins and endonuclease coding regions**

The fungi of the order Hypocreales with different lifestyles show varying genome sizes (Table 1). EPFs *B. bassiana, C. militaris* and *M. anisopliae* show a little variation among their genome size i.e., 29.96kb, 33.28kb and 24.67kb respectively. Comparatively there is a remarkable difference in the genome size between *C. militaris* and *M. anisopliae*. *B. bassiana* being a close relative of *C. militaris* shows a moderate genome size between the three EPFs. The three PPFs *F. graminearum*, *F. oxysporum* and *V. dahliae* also show considerable variations in their genome sizes i.e., 95.68kb, 34.48kb and 27.18kb respectively. The NPF *H. minnesotensis* showing a genome size of 52.25kb. The genome size is highly varying with the plant pathogen *F. graminearum*, the size difference is due to the presence of more number of introns in *cox, cob* and *nad1,2* genes (Table 1 and 3). There is no considerable variation in the GC content among the compared mt genomes (Table 1), EPFs *B. bassiana, C. militaris* and *M. anisopliae* showing 27.2%, 26.8% and 28.4% respectively. Though the genome size of *C. militaris*is larger than the other two EPFs (*B. bassiana* and *M. anisopliae*), it is showing lower GC content than *M. anisopliae* which is showing higher GC content despite its small genome size. *B. bassiana* showing a moderate GC content like its genome size in comparison with *C. militaris* and *M. anisopliae*. A moderate level of GC content exhibited by the PPFs and NPF, *F. graminearum* (31.8%), *F. oxysporum* (31%) and *V. dahliae* (27.3%) and *H. minnesotensis* (28.4%). Despite larger genomes, *F. graminearum* and *H. minnesotensis* exhibit GC content similar to other fungi. The overall Nucleotide composition (NC) showed that the mt genome is AT rich. Coding proportion of *M. anisopliae* (59.01%) is the highest and *C. militaris* (43.10%) is the lowest of all the

compared fungi (Table 1).

2. Synteny

Genes in the mt genomes are encoded on only one strand i.e., plus strand (Figures 1 and 2). Genes were encoded on a single strand which is typical in mt genomes of Ascomycota [10]. The Mauve alignments between the three EPFs *B. bassaiana, C. militaris* and *M. anisopliae* (Figure 1), and between *B. bassiana* and the PPFs *F. graminearum, F. oxysporum, V. dahlia* and NPF *H. minnesotensis* (Figure 2), visualize locally collinear blocks (LCBs), representing homologous regions of sequences that do not contain major rearrangements [25]. Four LCBs are seen in figure 1, gene clusters in each LCB are highly conserved, three LCBs contained all the protein and rRNA coding genes following conserved order and content within each LCB. The syntenic structure of the three EPFs is similar. Figure 2 also shows four LCBs in all the organisms except *F. oxysporum* which is lacking the smallest LCB. Three LCBs in figure 2 contained all protein and rRNA coding genes, gene clusters in between the organisms (*B. bassiana*, PPFs and NPF) is not highly conserved as in case of EPFs (Figure 1). Gene order of protein coding genes in the studied Hypocreales is *rnl-rps3-nad2-nad3-atp9-cox2-nad4L-nad5-cob-cox1-nad1-nad4-atp8-atp6-rns-cox3-nad6*, this synteny is being deviated by *V. dahliae* showing a translocation of the *nad1-nad4-atp8-atp6-rns-cox3-nad6* region in between *nad3* and *atp9* (Table 2 and Figure 2). Highly conserved gene order is identified with a very little variation among the studied organisms. Gene orders tend to be conserved, especially within major phyla although they can be variable between them [26]. Syntenic *trn* clusters in between EPFs are conserved, in LCB1 a cluster (TEM) is conserved in between *C. militaris, M. anisopliae* and *F. oxysporum* while EM found as a cluster in *B. bassiana* and another cluster (MLAFKLQHM) is conserved in between *B. bassiana*, *M. anisopliae* and *F. oxysporum,* this cluster separated into MLAFK, LH and M in *C. militaris* and *V. dahliae* showing (TEMMLAFKLQHM) as a single cluster.

Table 2: Gene order of the mitochondrial protein coding genes of Hypocrealean fungi with different habit

Life style	Organism	Protein coding genes																
EPFs	*Beauveria bassiana*	rnl	rps3	ND2	ND3	atp9	COXII	nad4L	ND5	cob	COXI	ND1	nad4	atp8	atp6	rns	COXIII	ND6
	Metarhizium anisopliae	rnl	rps3	nad2	nad3	atp9	cox2	nad4L	nad5	cob	cox1	nad1	nad4	atp8	atp6	rns	cox3	nad6
	Cordyceps militaris	rnl	rps3	nad2	nad3	atp9	cox2	nad4L	nad5	cob	cox1	nad1	nad4	atp8	atp6	rns	cox3	nad6
PPFs	*Fusarium oxysporum*	rnl	rps3	nad2	nad3	atp9	cox2	nad4L	nad5	cob	cox1	nad1	nad4	atp8	atp6	rns	cox3	nad6
	Fusarium graminearum	rnl	rps3	nad2	nad3	atp9	cox2	nad4L	nad5	cob	cox1	nad1	nad4	atp8	atp6	rns	cox3	nad6
	Verticillium dahliae	rnl	rps3	nad2	nad3	nad1	nad4	atp8	atp6	rns	cox3	nad6	atp9	cox2	nad4L	nad5	cob	cox1
NPF	*Hirsutella minnesotensis*	rnl	rps3	nad2	nad3	atp9	cox2	nad4L	nad5	cob	cox1	nad1	nad4	atp8	atp6	rns	cox3	nad6

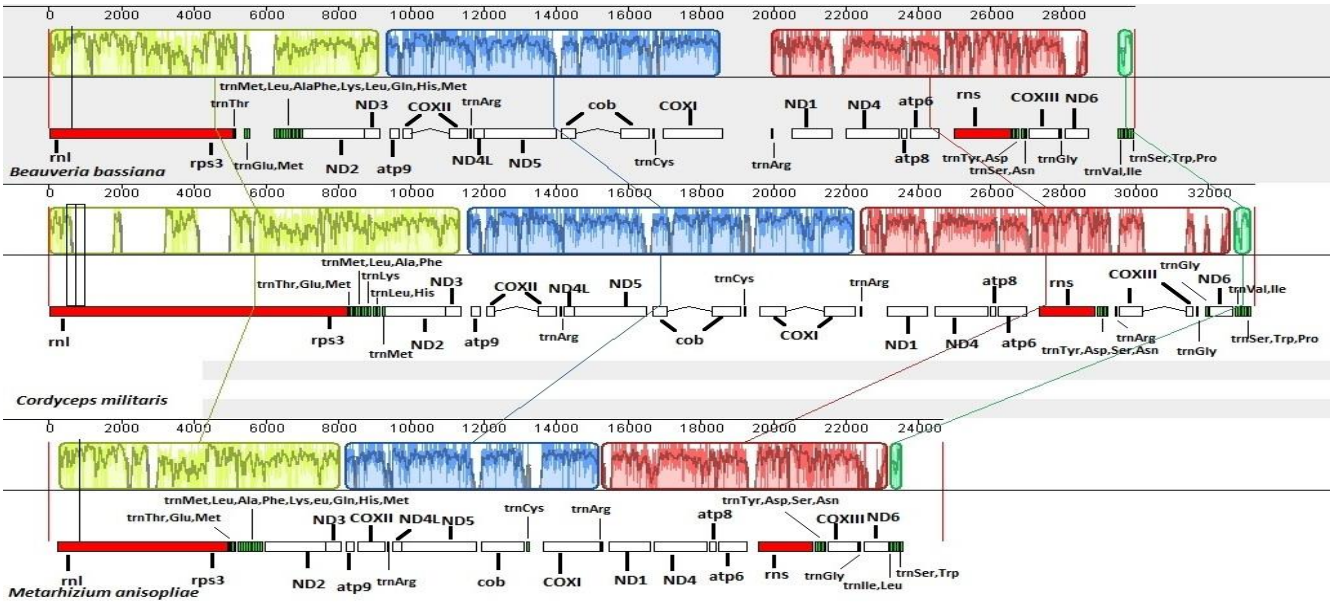

Figure 1: Alignments between the mt genomes of *Beauveria bassiana, Cordyceps militaris* and *Metarhizium anisopliae*. The ProgressiveMauve algorithm was used to align the mt DNAs of *Beauveria bassiana, Cordyceps militaris and Metarhizium anisopliae.* Corresponding colour boxes are locally collinear blocks (LCBs). A sequence identity similarity profile is shown in each box. Annotations are shown above and below the LCBs. White boxes are protein coding genes, Red boxes are rRNA genes and Green boxes are tRNA genes.

In LCB3 a cluster (YDSN) is conserved in *C. militaris* and *M. anisopliae*, while YD translocated next to D in *B. bassiana* and YD rearranged as DY in *F. graminearum*, and in *V. dahliae* YN formed a cluster with translocation of DS in between G and V which visible in LCB4. In LCB4, two clusters (VI; SWP) are conserved in *B. bassiana* and *C. militaris* while *M. anisopliae* showing SW as a cluster as P is missing completely (Figure 1 and 2). Conserved protein coding or core genes, tRNA clusters and conserved syntenic regions were identified which aid in understanding genomic architecture and evolutionary history of genomes.

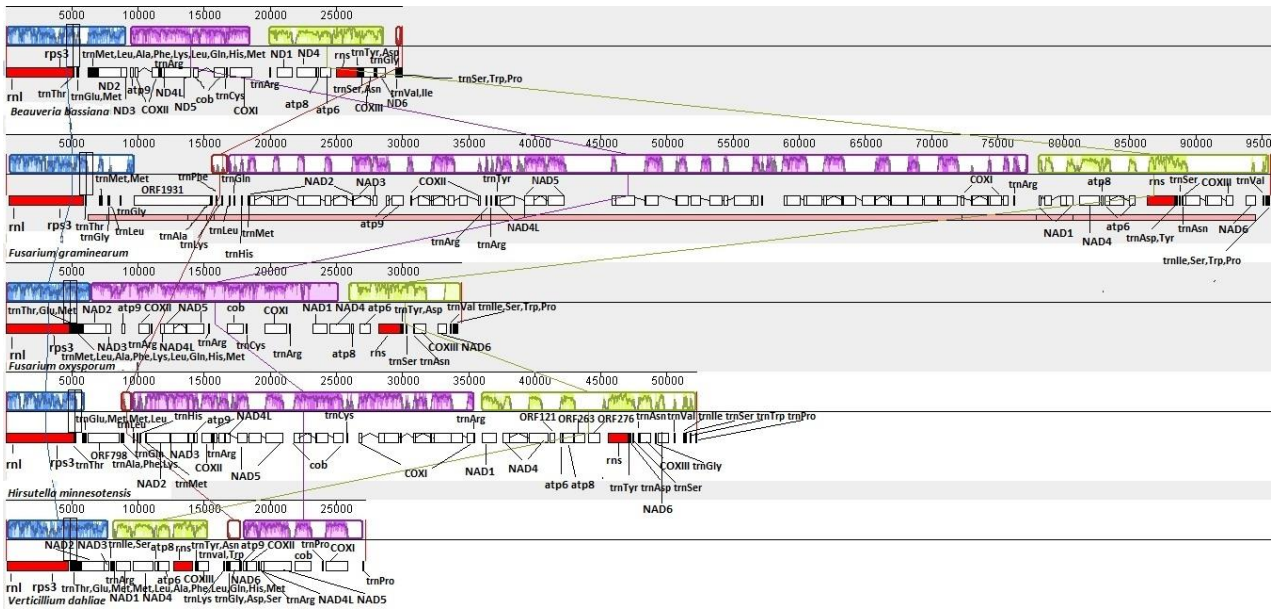

Figure 2: Alignments between the mt genomes of *B.bassiana, F.graminearum, F.oxysporum, H.minnesotensis* and *V.dahliae.* The ProgressiveMauve algorithm was used to align the mt DNAs of *B.bassiana, F.graminearum, F.oxysporum, H. minnesotensis and V.dahliae*. Corresponding colour boxes are locally collinear blocks (LCBs). A sequence identity similarity profile is shown in each box. Annotations are shown above and below the LCBs. White boxes are protein coding genes, Red boxes are rRNA genes, Green boxes are tRNA genes and large pink box is repeat region.

3. Intron comparison

A comparison of mt genomes in Sordariomycetes suggests that intergenic regions or intronic regions contribute the size of the most mt genomes [11]. Fungi mostly have group I introns [27], whose number is variable [10]; here in our study also different groupI introns were observed whose number was varied. The present data support for intronic variability among Hypocreales. Whole mtDNA analysis and protein coding genes analysis revealed the presence of group I introns in the fungi (Table 3). All the fungi are showing intronic ORF coding for a putative ribosomal protein *rps3*. The introns number is ranging from 1 to 35. The lowest numbers of introns are possessed by *M. anisopliae* and *V. dahliae* having a single intron which is an intronic ORF encoding *rps3* protein. The next lowest number of introns is shown by *B. bassiana*, two introns ID in *cob* and IB complete in *cox2*; *F. oxysporum* showing two introns ID in *cob* and IB complete in *nad5*. *C. militaris* showing

eight introns ID in *cob*, IB complete in *cox1* and *cox2*, I derived A in *cox3* and unlike others, three additional introns (IB complete, two IC1s) are shown in *rnl*. The NPF *H. minnesotensis* is showing eleven introns, two introns in *cob* (IA, ID), five introns in *cox1* (IB complete-4, IderivedB2), IC2 in *nad3* and *nad4*, one intron in *nad4L* (I derived A) and unlike others, IA 5'partial in *rnl*. *F. graminearum* showing 35 introns, one intron in *atp6* (IB complete), one intron in *atp9* (IA), five introns in *cob* (IB complete, IA, ID, IC1-2), twelve introns in *cox1* (IB complete-7, IB extra insertion-2, IB 3' partial, I derivedB1, ID), three introns in *cox2* (IB complete, IC1, IC2), three introns in *cox3* (IB complete, I derivedA, IC2),

Table 3: Details of introns in the mitochondrial genes of Hypocrealean fungi with different habit

Life style	**EPF**			**PPF**			**NP**
Organism / Gene	*Beauveria bassiana*	*Metarhizium anisoplieae*	*Cordyceps militaris*	*Fusarium oxysporum*	*Fusarium graminearum*	*Verticillium dahliae*	*Hirsutella minnesotensis*
atp6	-	-	-	-	1(IB complete)	-	-
atp8	-	-	-	-	-	-	-
atp9	-	-	-	-	1(IA)	-	-
cob	1(ID)	-	1(ID)	1(ID)	5 (IB complete, IA, ID, IC1-2)	-	2(IA, ID)
coxI	-	-	1(IB complete)	-	12 (IB complete-7, IB extra insertion-2, IB 3' partial, I derivedB1, ID)	-	5(IB complete-4, IderivedB2)
coxII	1(IB complete)	-	1(IB complete)	-	3(IB complete, IC1, IC2)	-	-
coxIII	-	-	1(I; derived,A)	-	3(IB complete, I derivedA, IC2)	-	-
nad1	-	-	-	-	2(IB complete, I derived)	-	-
nad2	-	-	-	-	4(IA,IC2-3)	-	-
nad3	-	-	-	-	1(IC2)	-	1(IC2)
nad4	-	-	-	-	-	-	1(IC2)
nad4L	-	-	-	-	1(IC1)	-	1(IderivedA)
nad5	-	-	-	1(IB complete)	1(IB complete)	-	-
nad6	-	-	-	-	-	-	-
rps3	-	-	-	-	-	-	-
rnl	1(IA)	1(IA)	4(IA, IB complete, IC1-2)	1(IA)	1(IA)	1(IA)	1(IA 5' partial)
Total no. of introns	3	1	8	3	35	1	11

two introns in *nad1* (IB complete, I derived), four introns in *nad2* (IA,IC2-3), one intron in *nad4L*(IC1), and one intron in *nad5*(IB complete). IA intron is found in all studied organisms in the *rnl* coding gene. The next commonly found intron is ID in the *cob* gene which is found in all except *M. anisopliae* and *V. dahliae*. Though the entomopathogens *B. bassiana* and *C.militaris* are closely related they are having varied intron number. The presence of exceptions in the number of introns

is an interesting feature which provides information for understanding fungal evolution [9]. *F. graminearum* and *H. minnesotensis* are with larger genomes where larger genome sizes were due to the presence of homing endonuclease regions/ genes (HEGs) (viz., LAGLIDAG and GIY) which are considered to be mobile invasive elements common in fungi and plants [28]. HEGs drive the mobility and persistence of their own coding sequences [29, 30]. The variability among HEGs can be exploited to develop phylogenetic markers which may be useful for differentiating various species and also being employed in biotechnology, medicine and agriculture by engineering for targeting specific cleavage sites in the genome [31].

4. tRNAs

The total number of tRNAs encoded by mitochondrial genomes of present study range from 25-28 (Table 1 and 4). All 20 amino acids coding tRNA are present only in 3 organisms (*B. bassiana, F. graminearum* and *H. minnesotensis*), while the others lack one or two amino acid(s) coding tRNA(s). *C. militaris* lacks Glutamine (Q) coding tRNA, *M. anisopliae* lacks Valine (V) and Proline (P) coding tRNAs and *V. dahliae* lacks Cysteine(C) coding tRNA (Table 4). Most mt proteins are specified by nuclear genes which are synthesized in the cytosol and imported to the organelle [32]. tRNA gene order is also a highly conserved phenomenon, all the mt genomes of present study followed a conserved gene order of tRNA with a very slight variation, forming tRNA clusters.

Figure 3: Gene order of the conserved *trn* gene clusters in the mt genomes of the order Hypocreales. Flanking are the protein coding or the rRNA genes / or the ORFs represented by black boxes and gaps. Gene order is based on Genbank sequences: *B. bassiana* (NC_010652.2), *M. anisopliae* (NC_008068), *C. militaris* (NC_022834.1), *F. oxysporum* (NC_017930.1), *F. graminearum* (NC_009493.1), *V. dahliae* (NC_008248.1), and *H. minnesotensis* (NC_027660.1)

trn genes are maintained in clusters while some genes are scattered as single genes (Figure 3). Apart from scattered single genes rest of the genes are clustered into groups of 2, 3, 4, 5, 7, 11, 12 and 14 clusters. Clustering of *trn* genes is suggested to be a unique characteristic of all Pezizomycotina [6]. Out of these grouped clusters the groups containing 12, 4 and 5 genes are highly conserved. First highly conserved cluster containing 12 genes is located.between *rnl* and *nad2* genes in *B. bassiana, M. anisopliae* and *F. oxysporum*, *H. minnesotensis* also maintaining a 12 gene cluster with an ORF

after fifth *trn* gene. 14 genes cluster is in *F. graminearum* where it is having an ORF after seventh *trn* gene. 11 gene clusters are in *C. militaris,* which is missing Q (Table 4) in its 12 gene cluster and in *V. dahliae,* which is missing K. Second highly conserved cluster (YDSN) is located between *rns* and *cox3* genes in almost all organisms except *V. dahliae* and *C. militaris* showing YSDNR, addition of R gene to the YSDN cluster. Third highly conserved cluster (VISWP) is adjacent to *nad6* gene and also maintained by all organisms except *V. dahliae* and *M. anisopliae* shows a cluster with VISW with a missing P *trn* gene (Table 4). YDSN and VISWP clusters are conserved clusters in the order Hypocreales which is evident from a study by Lin et al. 2015 [11]. *V. dahliae* showing a four gene cluster KGDS in between *cox3* and *nad6* genes. In all organisms scattered single genes are *trnRs, trnG* and *trnC*. Variation in gene order is due to tRNA location change which is a rare event [33]. Change in the location of *trnG* is reported in previous studies [6, 8, 11, 34]. *B. bassiana* and *M. anisopliae* are showing highly conserved *trn* gene order and clustering. All the organisms are conserved in maintaining the *trn* gene order and clusters except *V. dahliae*, there is a lot of rearrangements in the order of*trn* genes (Figure 3). tRNA rearrangements were observed in pezizomycotina[11]. tRNAs display editing, excision and integration capabilities, that allow them to change location within the genome and participate in Horizontal gene transfer (HGT) events [35]. Rearrangement events may provide insight into the phylogeny and evolution of the fungi [11].

Table 4: The order of tRNA genes (referred by the letter of aminoacid for which they carry an anticodon) in the mt genomes of Hypocrealean fungi with different habit.

Habit	Organism	tRNA genes	Total tRNA genes
EPF	*Beauveria bassiana*	T E M M L A F K L Q H M R C R Y D S N G V I S W P	25
	Metarhizium anisopliae	T E M M L A F K L Q H M R C R Y D S N G L I S W	24
	Cordyceps militaris	T E M M L A F K L H M R C R Y D S N R G G V I S W P	26
PPF	*Fusarium oxysporum*	T E M M L A F K L Q H M R R C R Y D S N V I S W P	25
	Fusarium graminearum	T E M M G L G A F K L Q H M R R Y C R Y D S N V I S W P	28
	Verticillium dahliae	T E M M L A F L Q H M I S R Y N K G D S V W R P P	25
NPF	*Hirsutella minnesotensis*	T E M M L A F K L Q H M R C R Y D S N G V I S W P	24

5. Codon bias

In general 61 codons of 64, code for 20 amino acids where the remaining 3 codons are stop codons (UAA, UAG and UGA). But this universal concept varies with mitochondrial genetic code where 'UGA' being a universal stop codon here codes for 'Tryptophan' (W), making the genetic code nearly universal. Mitochondria often use a different genetic code different from standard genetic code [36, 3].

Table 5: Codon usage table

AA		Bb		Man		Cm		F.gram		F.oxy		V.dah		H.min	
	Codon	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU	N	RSCU
Ala(A)	GCU*	152	2.24	151	2.28	151	2.29	386	2.09	142	2.06	186	2.71	238	2.17
	GCC	19	0.28	25	0.38	20	0.3	81	0.44	27	0.39	21	0.31	38	0.35
	GCA	95	**1.4**	86	**1.3**	84	**1.27**	202	**1.09**	89	**1.29**	64	**0.93**	137	**1.25**
	GCG	5	0.07	3	0.05	9	0.14	70	0.38	18	0.26	4	0.06	26	0.24
Cys(C	UGU*	26	1.58	36	1.67	22	1.63	149	1.57	25	1.67	25	2	115	1.63
	UGC	7	**0.42**	7	**0.33**	5	**0.37**	41	**0.43**	5	**0.33**	0	0	26	**0.37**
Asp(D)	GAU*	106	1.8	90	1.48	110	1.91	573	1.62	105	1.62	95	1.58	261	1.69
	GAC	12	**0.2**	32	**0.52**	5	**0.09**	135	**0.38**	25	**0.38**	25	**0.42**	48	**0.31**
Glu(E)	GAA*	109	**1.79**	119	**1.84**	100	**1.75**	609	**1.52**	124	**1.75**	122	**1.83**	306	**1.65**
	GAG	13	0.21	10	0.16	14	0.25	191	0.48	18	0.25	11	0.17	65	0.35
Phe(F)	UUU*	260	1.38	190	0.99	260	1.39	833	1.56	258	1.32	268	1.35	526	1.59
	UUC	116	**0.62**	192	1.01*	115	**0.61**	233	**0.44**	132	**0.68**	129	**0.65**	137	**0.41**
Gly(G)	GGU*	174	2.21	139	1.76	169	2.15	441	1.9	156	1.93	151	1.91	267	2.15
	GGC	2	0.03	8	0.1	1	0.01	50	0.22	4	0.05	0	0	21	0.17
	GGA*	120	**1.52**	148	**1.87**	123	**1.57**	321	**1.38**	138	1.7	142	**1.79**	166	**1.34**
	GGG	19	0.24	21	0.27	21	0.27	117	0.5	26	0.32	24	0.3	42	0.34
His(H)	CAU*	87	1.81	44	0.96	81	1.74	297	1.6	67	1.4	70	1.59	153	1.64
	CAC	9	**0.19**	48	**1.04***	12	**0.26**	74	**0.4**	29	**0.6**	18	**0.41**	34	**0.36**
Ile(I)	AUU	177	0.88	131	0.71	196	0.99	619	1.12*	157	0.78	210	1.14*	386	1.06*
	AUC	34	**0.17**	55	**0.3**	24	**0.12**	146	**0.26**	41	**0.2**	29	**0.16**	71	**0.2**
	AUA*	390	1.95	365	1.99	376	1.89	899	1.62	406	2.02	316	1.71	632	1.74
Lys(K)	AAA*	159	**1.86**	186	**1.83**	161	**1.86**	1132	**1.64**	205	**1.86**	169	**1.92**	573	**1.7**
	AAG	12	0.14	17	0.17	12	0.14	248	0.36	15	0.14	7	0.08	102	0.3
Leu(L)	UUA*	619	**5.37**	618	**5.27**	617	**5.36**	1233	**3.72**	569	**4.75**	597	**5.1**	823	**4.17**
	UUG	19	0.16	20	0.17	29	0.25	152	0.46	21	0.18	12	0.1	93	0.47
	CUU	29	0.25	24	0.2	23	0.2	240	0.72	60	0.5	62	0.53	122	0.62
	CUC	2	0.02	2	0.02	0	0	40	0.12	2	0.02	1	0.01	14	0.07
	CUA	21	**0.18**	38	**0.32**	21	**0.18**	266	**0.8**	57	**0.48**	31	**0.26**	114	**0.58**
	CUG	2	0.02	2	0.02	1	0.01	56	0.17	9	0.08	0	0	18	0.09
Met(M	AUG	123	**1**	124	**1**	124	**1**	324	**1**	135	**1**	114	**1**	218	**1**
Asn(N)	AAU*	225	1.7	169	1.37	222	1.71	961	1.59	250	1.64	227	1.68	580	1.66
	AAC	39	**0.3**	77	**0.63**	38	**0.29**	246	**0.41**	55	**0.36**	43	**0.32**	118	**0.34**
Pro(P)	CCU*	101	2.66	98	2.67	99	2.55	331	2.36	97	2.49	123	3	174	2.53
	CCC	2	0.05	7	0.19	11	0.28	47	0.34	10	0.26	4	0.1	18	0.26
	CCA	48	**1.26***	40	1.09*	41	**1.06***	146	**1.04***	42	**1.08***	34	**0.83**	66	**0.96**
	CCG	1	0.03	2	0.05	4	0.1	36	0.26	7	0.18	3	0.07	17	0.25
Gln(Q)	CAA*	97	**1.96**	90	**1.91**	94	**1.86**	376	**1.65**	99	**1.9**	90	**1.94**	208	**1.74**
	CAG	2	0.04	4	0.09	7	0.14	80	0.35	5	0.1	3	0.06	31	0.26
Arg(R)	CGU	15	**0.82**	14	**0.78**	15	**0.87**	117	**0.99**	15	**0.78**	15	**0.82**	40	**0.71**
	CGC	1	0.05	1	0.06	0	0	17	0.14	0	0	0	0	10	0.18
	CGA	0	0	0	0	0	0	30	0.25	0	0	2	0.11	29	0.51

	CGG	0	0	3	0.17	0	0	15	0.13	1	0.05	1	0.05	8	0.14
	AGA*	92	**5.02**	81	**4.5**	83	**4.83**	457	**3.87**	97	**5.02**	90	**4.91**	200	**3.54**
	AGG	2	0.11	9	0.5	5	**0.29**	72	0.61	3	0.16	2	0.11	52	0.92
Ser(S)	UCU*	98	1.41	97	**1.48**	97	1.4	494	1.95	135	1.87	159	2.27	270	1.92
	UCC	4	0.06	7	0.11	7	0.1	80	0.32	15	0.21	1	0.01	39	0.28
	UCA*	121	**1.74**	117	**1.79**	122	**1.76**	337	**1.33**	104	**1.44**	74	**1.05**	168	**1.2**
	UCG	5	0.07	3	0.05	8	0.12	55	0.22	4	0.06	1	0.01	18	0.13
	AGU*	169	2.43	151	2.31	164	2.36	439	1.73	148	2.05	168	2.39	285	2.03
	AGC	20	**0.29**	18	**0.27**	19	**0.27**	116	**0.46**	28	**0.39**	18	**0.26**	63	**0.45**
Thr(T)	ACU*	112	1.68	75	1.2	113	1.69	450	1.94	129	1.97	124	1.95	244	1.89
	ACC	2	0.03	4	0.06	3	0.04	85	0.37	5	0.08	1	0.02	33	0.26
	ACA*	151	**2.26**	165	**2.65**	149	**2.23**	338	**1.46**	125	**1.91**	127	**1.99**	223	**1.73**
	ACG	2	0.03	5	0.08	2	0.03	56	0.24	3	0.05	3	0.05	17	0.13
Val(V)	GUU*	73	1.09	114	1.31	76	1.14	326	1.33	120	1.41	137	1.73	225	1.63
	GUC	4	0.06	4	0.05	2	0.03	54	0.22	5	0.06	0	0	19	0.14
	GUA*	177	**2.63**	203	**2.33**	170	**2.55**	493	**2.01**	173	**2.04**	168	**2.13**	264	**1.92**
	GUG	15	0.22	27	0.31	19	0.28	108	0.44	42	0.49	11	0.14	43	0.31
Trp(W)	UGA*	66	**1.97**	65	**1.83**	65	**1.97**	178	**1.61**	70	**1.94**	68	**2**	112	**1.66**
	UGG	1	0.03	6	0.17	1	0.03	43	0.39	2	0.06	0	0	23	0.34
Tyr(Y)	UAU*	218	1.79	180	1.43	220	1.8	760	1.6	211	1.6	198	1.6	449	1.63
	UAC	25	**0.21**	72	**0.57**	24	**0.2**	191	**0.4**	52	**0.4**	50	**0.4**	101	**0.37**
TER	UAA*	14	**1.87**	15	**2**	15	**2**	96	**1.21**	13	**1.63**	14	**1.87**	40	**1.21**
	UAG	1	0.13	0	0	0	0	63	0.79	3	0.38	1	0.13	26	0.79

***Codons /RSCU values showing high RSCU values >1; RSCU values in boldrepresent the codons for which corresponding tRNA present in mt genome and RSCU values underlined are the codons for which wobbling is possible.**

So, in case of mitochondria 62 codons code for 20 amino acids and 2 are stop codons. The genes coding for the proteins of mt genomes all started with translation initiation codon AUG, whereas the preferred stop codon is UAA, with UAG as an alternative stop codon. The most frequently used codons are UUA (RSCU value 3.5 to 5.37), AGA (3.5 to 5.02) and CCU (2.36 to 2.67) that code for Leucine (Leu), Arginine (Arg) and Proline (Pro) respectively, the corresponding tRNAs for UUA and AGA are present in all the corresponding mt genomes but CCU corresponding tRNA is absent in all mt genomes though it is showing high RSCU value (>2 to 3) (Table 5). Bb and Cm are showing identity in their codon composition in relation to RSCU values with only one difference i.e., trnQ is totally absent in Cm, whereas Man showing equal similarity and dissimilarity with both *B. bassiana* and *C. militaris*. All the 3 PPs and the NP are showing more similarity with little RSCU variations within them as well as with the EPFs (Table 5). Genome hypothesis of Grantham postulates that each species uses certain synonymous codons (codons that code for the same amino acid) systematically in coding sequences [37]. Codon usage bias may be influenced by many factors like selection, mutations, effective population size, random drift, biased gene conversion, mRNA

secondary structure, translational initiation and evolutionary history [38, 3].

6. Phylogenetic analysis

The phylogenetic relationships of the 7 Hypocrealean fungi were inferred using *Neurospora crassa* as outgroup. Concatenated phylogenetic trees were constructed using 14 conserved protein-coding genes (*cox1, cox2* and *cox3; cob; atp6, atp8* and *atp9; nad1, nad2, nad3, nad4, nad4L, nad5* and *nad6)* using Maximum Likelihood (ML) approach (Figure 4). Phylogenetic analysis inferred that organisms of same habitat are grouped together, because of the conserved nature of the genes of mt (mentioned in the results in *synteny*). The concatenated ML tree shows that the EPFs were grouped together in adjacent clades while the *B. bassiana* and *C. militaris* are highly closely related sharing the same clade, while the PPFs grouped in adjacent clades and the *H. minnesotensis* forms a clade in between the EPFs and PPFs. Phylogenetic analysis showed a similar topology to the trees that are constructed using β-tubulin, small and large subunit rRNA sequences by Neelapu et al. 2009 [39]. We used protein coding genes devoid of rRNA coding genes and we're able to view the similar topology of the resulted phylogenetic trees with the phylogenetic tree produced from the analysis of amino acid sequences for Ascomycota mitochondrial genomes [18].

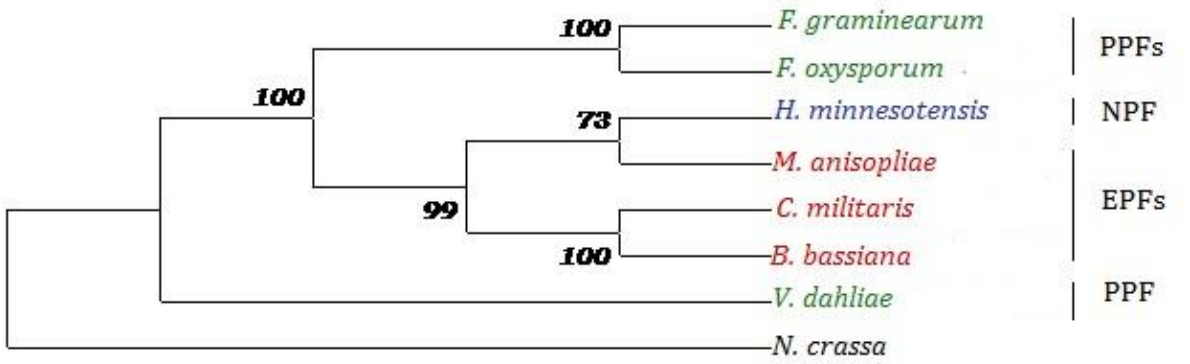

Figure 4: Phylogenetic tree constructed using 14 conserved protein coding genes (concatenated) (*cox1-3; cob; atp6, 8* and *9; nad1-4, nad4L, nad5* and *6*) of seven hypocrealean fungi using ML approach. Number at nodes indicate bootstrap value.

This may suggest that mt protein coding genes may be used as markers for identification and phylogenetic analyses just like rRNA genes of mitochondria. Mitochondrial markers also have been considered due to their favourable features,like their high copy number, the possibility of an easier and cheaper recovering of their sequences and the paucity of repetitive regions which could produce misleading results owing to the comparison of non-orthologous sequences pairs [40].

4. CONCLUSION

Comparative analysis of mt genomes in the present study revealed the conserved nature of mt genomes. All the core protein coding genes were conserved in showing synteny (conserved gene order) with seldom seen rearrangements. EPFs' has shown the high level of synteny when compared

to others. IA introns were common in all organisms. RSCU analysis revealed the most of the tRNAs were coded by the mt genome but certain tRNAs showing high RSCU value were missing indicating that such tRNAs probably being catered by the nuclear genome. Phylogenetic analysis inferred that the mt core genes are helpful in grouping the related organisms together meaning these core genes may be developed as phylogenetic markers on par with the nuclear gene markers. Further in depth analyses may dwell more into the evolutionary aspects of these organisms.

ACKNOWLEDGEMENT

We are thankful to the Departmental facilities of Andhra University.

CONFLICT OF INTEREST

Authors have no any conflict of interest.

REFERENCES

1. Saraste M. Oxidative phosphorylation at the fin de siecle. Science. 1999 Mar 5;283(5407):1488-93.
2. Burger G, Gray MW, Lang BF. Mitochondrial genomes: anything goes. Trends in genetics. 2003 Dec 1;19(12):709-16.
3. Kamatani T, Yamamoto T. Comparison of codon usage and tRNAs in mitochondrial genomes of Candida species. Biosystems. 2007 Sep 1;90(2):362-70.
4. Xiao G, Ying SH, Zheng P, Wang ZL, Zhang S, Xie XQ, Shang Y, Leger RJ, Zhao GP, Wang C, Feng MG. Genomic perspectives on the evolution of fungal entomopathogenicity in Beauveria bassiana. Scientific reports. 2012 Jul 2;2:483.
5. Pfeifer TA, Hegedus DD, Khachatourians GG. The mitochondrial genome of the entomopathogenic fungus Beauveria bassiana: analysis of the ribosomal RNA region. Canadian journal of microbiology. 1993 Jan 1;39(1):25-31.
6. Ghikas DV, Kouvelis VN, Typas MA. The complete mitochondrial genome of the entomopathogenic fungus Metarhizium anisopliae var. anisopliae: gene order and trn gene clusters reveal a common evolutionary course for all Sordariomycetes, while intergenic regions show variation. Archives of microbiology. 2006 Jun 1;185(5):393.
7. Pantou MP, Kouvelis VN, Typas MA. The complete mitochondrial genome of the vascular wilt fungus Verticillium dahliae: a novel gene order for Verticillium and a diagnostic tool for species identification. Current genetics. 2006 Aug 1;50(2):125-36.
8. Pantou MP, Kouvelis VN, Typas MA. The complete mitochondrial genome of Fusarium oxysporum: insights into fungal mitochondrial evolution. Gene. 2008 Aug 1;419(1):7-15.
9. Ghikas DV, Kouvelis VN, Typas MA. Phylogenetic and biogeographic implications inferred by mitochondrial intergenic region analyses and ITS1-5.8 S-ITS2 of the entomopathogenic fungi Beauveria bassiana and B. brongniartii. BMC microbiology. 2010 Dec;10(1):174.

10. Aguileta G, De Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, Gabaldón T. High variability of mitochondrial gene order among fungi. Genome biology and evolution. 2014 Feb 1;6(2):451-65.
11. Lin R, Liu C, Shen B, Bai M, Ling J, Chen G, Mao Z, Cheng X, Xie B. Analysis of the complete mitochondrial genome of Pochonia chlamydosporia suggests a close relationship to the invertebrate-pathogenic fungi in Hypocreales. BMC microbiology. 2015 Dec;15(1):5.
12. Chaverri P, Vilchez B. Hypocrealean (Hypocreales, Ascomycota) Fungal Diversity in Different Stages of Tropical Forest Succession in Costa Rica 1. Biotropica. 2006 Jul;38(4):531-43.
13. Keller S. Investigations on the effect of herbicides on aphid pathogenic Entomophthoraceae. SER. ENTOMOL. 1986.. 1986.
14. Tanada Y, Kaya HK. Insect pathology. Academic press; 2012 Dec 2.
15. Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. Biology direct. 2008 Dec;3(1):38.
16. Peden J. CodonW 1.4. 2. Distributed by the author, Nottingham. 2005.
17. Gautheret D, Lambert A. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles1. Journal of molecular biology. 2001 Nov 9;313(5):1003-11.
18. Shen XY, Li T, Chen S, Fan L, Gao J, Hou CL. Characterization and phylogenetic analysis of the mitochondrial genome of Shiraia bambusicola reveals special features in the order of Pleosporales. PloS one. 2015 Mar 19;10(3):e0116466.
19. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2014 Dec 10;31(8):1296-7.
20. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one. 2010 Jun 25;5(6):e11147.
21. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular biology and evolution. 2007 May 7;24(8):1596-9.
22. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. Genome research. 2005 Dec 1;15(12):1620-31.
23. Mohanta TK, Bae H. The diversity of fungal genome. Biological procedures online. 2015 Dec;17(1):8.
24. Hudspeth ME. The fungal mitochondrial genome—a broader perspective. Handbook of applied mycology. 1992;4:213-41.
25. Leliaert F, Lopez-Bautista JM. The chloroplast genomes of Bryopsis plumosa and Tydemania expeditiones (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. Bmc Genomics. 2015 Dec;16(1):204.

26. Hoffmann RJ, Boore JL, Brown WM. A novel mitochondrial genome organization for the blue mussel, Mytilusedulis. Genetics. 1992 Jun 1;131(2):397-412.
27. Lang BF, Laforest MJ, Burger G. Mitochondrial introns: a critical view. Trends in Genetics. 2007 Mar 1;23(3):119-25.
28. Belfort M, Roberts RJ. Homing endonucleases: keeping the house in order. Nucleic acids research. 1997 Sep 1;25(17):3379-88.
29. Stoddard BL. Homing endonucleases from mobile group I introns: discovery to genome engineering. Mobile DNA. 2014 Dec;5(1):7.
30. Turmel M, Otis C, Lemieux C. Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. Genome biology and evolution. 2015 Jul 1;7(7):2062-82.
31. Stoddard BL. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. Structure. 2011 Jan 12;19(1):7-15.
32. Herrmann JM. Converting bacteria to organelles: evolution of mitochondrial protein sorting. Trends in microbiology. 2003 Feb 1;11(2):74-9.
33. Cedergren R, Lang BF. Probing fungal mitochondrial evolution with tRNA. Biosystems. 1985 Jan 1;18(3-4):263-7.
34. Xu J, Huang B, Qin C, Li ZZ. Sequence and phylogenetic analysis of Beauveria bassiana with mitochondrial genome. Mycosystema. 2009;28(5):718-23.
35. Tuller T, Girshovich Y, Sella Y, Kreimer A, Freilich S, Kupiec M, Gophna U, Ruppin E. Association between translation efficiency and horizontal gene transfer within microbial communities. Nucleic acids research. 2011 Feb 22;39(11):4743-55.
36. Swire J, Judson OP, Burt A. Mitochondrial genetic codes evolve to match amino acid requirements of proteins. Journal of molecular evolution. 2005 Jan 1;60(1):128-39.
37. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. Nucleic acids research. 1980 Jan 11;8(1):197-.
38. Qin H, Wu WB, Comeron JM, Kreitman M, Li WH. Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. Genetics. 2004 Dec 1;168(4):2245-60.
39. Neelapu NR, Reineke A, Chanchala UM, Koduru UD. Molecular phylogeny of asexual entomopathogenic fungi with special reference to Beauveria bassiana and Nomuraea rileyi. Revistaibero americana de micología. 2009 Jun 1;26(2):129-45.
40. Santamaria M, Vicario S, Pappadà G, Scioscia G, Scazzocchio C, Saccone C. Towards barcode markers in Fungi: an intron map of Ascomycota mitochondria. InBMC bioinformatics 2009 Jun (Vol. 10, No. 6, p. S15). BioMed Central.