Evolution of Synonymous Codon Usage in the Mitogenomes of Certain Species of Bilaterian Lineage with Special Reference to Chaetognatha

Sudeesh Karumathil[1], Vijaya R. Dirisala[2], Uthpala Srinadh[2], Valaboju Nikhil[2], N. Satya Sampath Kumar[2] and Rahul R. Nair[1]

[1]Aushmath Biosciences, Administrative office, Devaraj Corner, Vadavalli Post, Coimbatore, Tamil Nadu, India. [2]Department of Biotechnology, Vignan's University (Vignan's Foundation for Science, Technology and Research University), Guntur, Andhra Pradesh, India.

ABSTRACT: Chaetognatha is a minor phylum, comprising transparent marine invertebrates varying in size from 0.5 to 12 cm. The exact phylogenetic position of Chaetognatha in Metazoa has not been deciphered as some embryological characteristics place chaetognaths among deuterostomes and some morphological characteristics place these among protostomes. In this study, the major factors that drive synonymous codon usage bias (SCUB) in the mitogenomes of representative species of Chaetognatha and chosen species of other closely related phyla were analyzed. Spearman's rank correlation analyses of nucleotide contents suggested that mutational pressure and selection were acting in all examined mitogenomes but with varying intensities. The quantification of SCUB using effective number of codons vs. GC composition at the third codon position (GC_3) plot suggested that mutational pressure due to GC compositional constraints might be one of the major influencing forces driving the SCUB in all chaetognaths except *Sagitta enflata*. However, neutrality plots revealed no significant correlation between GC_3 and cumulative GC content at first and second codon positions (GC_{12}) in all other species, except in *Daphnia pulex*. The parity rule 2 bias plot showed that significant compositional differences existed between C and G, as well as between A and T, contents in most of the protein-coding genes (PCGs) and, comparatively, A and T contents were used more proportionally than C and G contents in all chosen mitogenomes. Chi-square analysis revealed the presence of putative optimal codons in all species, except in *S. enflata*. The correspondence analysis identified that mutational pressure and selection act on the mitogenomes of the selected chaetognaths and other phyla with varying intensities. The cluster analysis based on relative synonymous codon usage (RSCU) values revealed that RSCU variations in the PCGs of mitogenomes of chaetognaths are more comparable with those of protostomes. Apart from mutational pressure and selection, certain unknown selective forces might be acting on the PCGs in the analyzed mitogenomes as the phenomenon of SCUB could not be explained by mutational pressure, by selection, or by both.

KEYWORDS: Chaetognatha, evolution, synonymous codon usage, mitogenomes, bilaterian lineage

Introduction

Most of the amino acids are encoded by more than one codon in the protein-coding genes (PCGs), and those codons are termed as synonymous codons. Synonymous codons have been observed to be used at unequal frequencies in various genomes. Synonymous codon usage (SCU) is species specific.[1] In certain genomes, a subset of codons is used most frequently to encode a given amino acid and this bias in codon usage exists in a wide variety of organisms, from prokaryotes to unicellular and multicellular eukaryotes.[1] Mutational pressure and natural selection have been identified as the two major evolutionary forces that contribute to bias in SCU (SCUB).[2–5] SCU preferences can be attributed to a variety of factors that vary from one species to another.[6,7] In unicellular prokaryotic organisms such as *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, and *Dictyostelium discoideum*, both selectional and mutational pressures dictate SCU,[8–10] whereas in multicellular eukaryotes such as *Drosophila melanogaster* and *Caenorhabditis elegans*, translational selection drives codon usage.[11–13] However, mutational pressure due to GC content is the determining factor of codon usage in mammals.[14]

Though mutational pressure and natural selection have been identified as two major evolutionary forces that drive SCUB, in certain eukaryotes, SCUB and usage of optimal codons are often influenced by certain selective forces,[15] which are unknown. Investigation of the pattern of codon usage and the various factors influencing its diversification form the basis for understanding the evolution of genomes and the habitat adaptation of living organisms.[16–18] Nuclear genomes of various organisms have been exclusively analyzed to detect trends

associated with codon usage.[19,20] However, mitochondrial genomes have been less studied so far, and only very few studies have been carried out on mitogenomes.[21,22]

Chaetognatha is a minor phylum, comprising transparent marine invertebrates varying in size from 0.5 to 12 cm.[23,24] They feed on small animals such as copepods and immature fish.[23] Because chaetognaths are very ancient, only little information is available about their evolutionary origin.[25] The exact phylogenetic position of Chaetognatha in Metazoa is not clearly understood as some embryological characteristics place chaetognaths among deuterostomes and some morphological characteristics place them among protostomes.[26] But, some studies have supported the inclusion of chaetognaths among protostomes.[23,24] The inclusion of Chaetognatha at the appropriate position in the bilaterian lineage has been regarded as one of the most important issues in metazoan phylogeny.[27]

Considering these facts, this study has been designed with the aim of analyzing the various trends associated with SCU in the mitogenomes of all sequenced chaetognaths and other closely related phyla. In this study, evolution of SCUB in the representative species of protostomes and deuterostomes has been extensively studied and compared with that of chaetognaths. The study would certainly help to get insight into the evolution of SCUB of chaetognaths and other chosen species belonging to the bilaterian lineage. The main objective of this study was to analyze the major factors that dictate SCUB in chaetognaths as well as representative species chosen from among protostomes and deuterostomes. This comparative study reveals the closeness of chaetognaths to protostomes in terms of SCUB.

Materials and Methods

Sequence data. Complete mitochondrial genomes of five representative species from the minor phylum Chaetognatha, as well as for one representative species each from Hemichordata, Echinodermata, Arthropoda, Annelida, Mollusca, and Brachiopoda, were retrieved from the National Center for Biotechnology Information (NCBI) (Table 1). The complete coding sequences (CDSs) were examined for the presence of proper initiation and termination codons at the beginning and at the end of the sequences in order to ensure integrity. The final data set contains only sequences having more than 300 nucleotides to avoid sampling errors leading to statistical fluctuations, as suggested in a recent study.[22] The number of CDSs chosen for this study is provided in Table 1.

Measures of SCU. *Relative SCU.* Relative SCU (RSCU) for each informative codon was calculated for the normalization of SCU by avoiding the influence of amino acid composition. The RSCU values are considered to be independent of gene length.[22] RSCU values were obtained as the ratio of the observed frequency of a codon to the expected frequency of that particular codon provided all synonymous codons for the same amino acid are used equally.[8] If the RSCU value is greater than 1, it indicates a bias toward that particular codon, and if the RSCU value is close to 0, it indicates lack of bias.[17]

Plot of effective number of codons vs GC₃. *Plot of effective number of codons vs GC_3.* The extent of SCUB is often quantified using the effective number of codons (ENC), which is independent of gene length and is an indicator of the level of gene expression.[28] ENC is independent of constraints of amino acid composition[22] and takes values ranging from 20 to 61.[28] If the bias is extreme, the ENC would be 20, ie, only one codon would be used for coding one amino acid, and the ENC would be 61 if all synonymous codons for a particular amino acid are used equally, ie, there is no bias at all.[28] If the ENC value of a gene is less than 35, bias is regarded to be high, and if the ENC value is above 50, bias is regarded to be low.[29] The overall and local GC compositions (GC content at the three codon positions) were calculated after excluding the stop codons. The expected ENC values were computed using GC_3 (presence of a guanine or a cytosine at the third codon position) under the assumption that no selection exists.[28] In this study, ENC was plotted against GC_3 to obtain a visual display of the influence of major evolutionary forces, such as mutational pressure and natural selection, on SCUB by the scattering of genes along the expected ENC curve.[28] The ENC values and GC_3 values for all PCGs of the selected mitogenomes were calculated and the ENC-vs-GC_3 plot was developed.

Neutrality plot. Cumulative GC content at the first and second codon positions (GC_{12}) and the GC composition at the third codon position (GC_3) were used to derive the neutrality plot to understand the influence of GC composition in the shaping of codon usage.[17] The influence of mutational pressure on SCUB is low, if the slope of the regression line is close to 0 and it is high if the slope is close to 1.[17,30]

Parity rule 2 plot. Parity rule 2 bias plot (PR2 bias plot) was developed in order to detect the biases for A ≈ T and G ≈ C within a DNA strand (Second parity rule of Chargaff).[31] In the PR2 bias plot, only the third codon positions in the four-fold degenerate amino acids were plotted.[31]

Identification of putative optimal codons. The putative optimal codons were identified using chi-square test for a 2×2 matrix having one degree of freedom.[32] Two data sets for a 2×2 matrix were developed by selecting 10% of the genes that were grouped on the left and the right extremes of the first axis of the correspondence analysis (COA).[33,34] Observed frequencies of codons in the two data sets are given in the first row and the total number of synonymous alternatives of that given codon is given in the second row of the 2×2 matrix.[30] Level of significance was measured at $P < 0.05$.

Codon adaptation index. The codon adaptation index (CAI)[35] was calculated to understand the magnitude of bias in the analyzed mitogenomes toward a subset of referred codons in highly expressed genes. The CAI value ranges from 0 to 1 and a value close to 1 indicates higher codon usage bias and level of gene expression.[35]

Correspondence analysis. COA was performed on the RSCU (COA–RSCU) for identifying variations in SCU by avoiding the influence of amino acid composition.[29] COA

Table 1. List of species examined in this study.

SL. NO:	SPECIES NAME	PHYLUM	ACCESSION ID	NCBI TRANSLATION TABLE	NUMBER OF CDS ANALYZED
1	*Asterina pectinifera*	Echinodermata	NC_001627	Echinoderm and Flatworm Mitochondrial Code. **AAA (N), AGA (S), AGG (S), UGA (W).**	11
2	*Balanoglossus carnosus*	Hemichordata	NC_001887	Echinoderm and Flatworm Mitochondrial Code. **AAA (N), AGA (S), AGG (S), UGA (W)**.	11
3	*Daphnia pulex*	Arthropoda	NC_000844	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	11
4	*Katharina tunicata*	Mollusca	NC_001636	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	11
5	*Lumbricus terrestris*	Annelida	NC_001673	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	11
6	*Sagitta nagae*	Chaetognatha	NC_013810	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	10
7	*Sagitta decipiens*	Chaetognatha	NC_013811	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	11
8	*Paraspadella gotoi*	Chaetognatha	NC_006083	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	10
9	*Sagitta enflata*	Chaetognatha	NC_013814	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	10
10	*Spadella cephaloptera*	Chaetognatha	NC_006386	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	10
11	*Terebratulina retusa*	Brachiopoda	NC_000941	Invertebrate Mitochondrial Code. **AGA (S), AGG (S), AUA (M), UGA (W).**	11

Note: The letters S, W, M, and N stand for serine, tryptophan, methionine, and asparagine, respectively.

partitions the variations along certain numbers of orthogonal axes (the number of orthogonal axes depends upon the total number of synonymous codons excluding stop codons).[36] The first axis accounts for the majority of the variation, with each subsequent axis accounting for diminishing amount of variance.[36] COA considers each CDS as a 59-dimensional vector, and each dimension is defined by the RSCU value of a particular informative codon.[37]

Cluster analysis. Cluster analysis on the RSCU values was performed to show the association between codon bias and other factors.[29] In the cluster analysis, a matrix in which the rows and columns correspond to chosen species and pooled RSCU values of synonymous codons, respectively, was generated.[29] Unweighted pair-group average clustering grouped the selected species based on the variations in the RSCU values, and the distances are in Euclidean distance.[29]

Statistical analysis and software implementation. Nonparametric Spearman's rank correlation method was used for all correlation analyses between various codon usage indexes and other important parameters as it makes no assumptions about the probability relation between the two variables.[30,36] PAST software version 2.12 was used for Spearman's rank correlation analysis and cluster analysis.[38] ENC, aromaticity, and Grand Average of Hydropathy (GRAVY) scores were estimated using CodonW (online version: http://mobyle. pasteur.fr/cgi-bin/portal.py?#forms:codonw).[39] Overall and local nucleotide compositions were calculated using MEGA

5.2.2.[40] RSCU values were calculated using DAMBE version 5.3.31.[41] CAI values were computed using ACUA 1.0[42]

Results

Compositional properties influence codon usage in the analyzed mitogenomes. Complex correlations were noticed between overall and silent base compositions in the examined mitogenomes (Table 2). Significant positive correlations were observed between T and GC_3, C and T_3 in *Sagitta nagae;* C with A_3 (A at the third codon position), T with G_3 in *Paraspadella gotoi;* and A with G_3, T with GC_3, G with A_3, C with T_3 (T at the third codon position) in *S. enflata.* The positive correlation between heterogeneous nucleotide contents suggest the possible influence of natural selection[43] in dictating codon usage in these species belonging to Chaetognatha. However, in the mitogenomes of *Sagitta decipiens* and *Spadella cephaloptera* of Chaetognatha, homogeneous nucleotide contents were in positive correlation and heterogeneous contents were in negative correlation. This suggests the contribution of mutational pressure due to compositional constraints in driving codon usage in *S. decipiens* and *S. cephaloptera.* Positive correlation existed between G and T_3 in *Asterina pectinifera*, indicating that selection of G and T_3 might be one of the contributing factors that play a crucial role in shaping SCUB in *A. pectinifera*, belonging to Echinodermata. All other correlations suggested the strong influence of compositional constraints favoring mutational pressure in *A. pectinifera*. In the

Table 2. Spearman's rank correlation analysis between overall and silent base contents.

SPECIES	BASE CONTENTS	A₃	T₃	G₃	C₃	GC₃
A. pectinifera	A	0.91**	−0.60*	−0.50	0.43	−0.02
	T	−0.80**	0.80**	0.52	−0.70**	−0.40
	G	−0.70**	0.69*	0.75**	−0.70**	−0.30
	C	0.53	−0.90**	−0.80**	0.95**	0.77**
B. carnosus	A	0.77**	−0.20	−0.10	−0.50	−0.70**
	T	0.14	0.12	−0.10	−0.03	−0.30
	G	−0.40	0.35	0.41	−0.20	0.18
	C	−0.40	−0.10	−0.50	0.95**	0.61
D. pulex	A	0.70**	−0.43	−0.85**	0.65*	−0.10
	T	−0.74**	0.71**	0.62	−0.90**	−0.58
	G	−0.69*	0.44	0.77**	−0.57	0.19
	C	0.79**	−0.75**	−0.66	0.91**	0.52
K. tunicata	A	0.82**	−0.86**	−0.77**	0.76**	−0.31
	T	−0.70**	0.72**	0.57	−0.75**	0.03
	G	−0.80**	0.81**	0.91**	−0.72**	0.61
	C	0.77**	−0.78**	−0.79**	0.90**	−0.25
L. terrestris	A	0.77**	−0.20	−0.10	−0.50	−0.70**
	T	0.14	0.12	−0.10	−0.03	−0.30
	G	−0.40	0.35	0.41	−0.20	0.18
	C	−0.40	−0.10	−0.50	0.95**	0.61*
S. nagae	A	0.81**	−0.51	0.28	−0.64*	−0.59*
	T	−0.39	0.02	−0.04	0.62*	0.71**
	G	0.10	−0.06	0.47	−0.58	−0.03
	C	−0.65*	0.78**	−0.75**	0.45	−0.15
S. decipiens	A	0.86**	−0.85**	−0.29	−0.17	−0.34
	T	−0.60*	0.61*	0.43	−0.21	0.12
	G	−0.01	−0.04	0.19	−0.13	0.11
	C	0.25	−0.29	−0.64*	0.63*	0.08
P. gotoi	A	0.94*	−0.93*	−0.73*	0.59	0.05
	T	−0.75*	0.80*	0.94*	−0.75*	−0.095
	G	−0.43	0.32	0.11	−0.27	−0.03
	C	0.69*	−0.73*	−0.97*	0.71*	0.04
S. enflata	A	0.86**	−0.91**	0.81**	−0.83**	−0.73**
	T	−0.61*	0.82**	−0.59	0.46	0.64*
	G	0.82**	−0.82**	0.76**	−0.78**	−0.66**
	C	−0.88**	0.78**	−0.79**	0.90**	0.90**
S. cephaloptera	A	0.79**	−0.91**	−0.15	0.23	0.12
	T	−0.55	0.83**	0.15	−0.36	−0.41
	G	−0.76**	0.39	0.52	−0.19	0.54
	C	0.28	−0.39	−0.37	0.40	0.24
T. retusa	A	0.66*	−0.07	−0.31	−0.37	−0.62
	T	−0.51	0.82**	0.31	−0.35	−0.23
	G	0.04	−0.33	−0.01	0.08	0.24
	C	−0.09	−0.31	−0.2	0.49	0.46

Notes: *P ≤ 0.05, **P ≤ 0.01.

representative species of Hemichordata, ie, *Balanoglossus carnosus,* possible influence of mutational pressure due to compositional constraints in codon usage was noticed as no positive correlation between heterogeneous base contents and no negative correlation between homogeneous contents were observed. In *D. pulex,* belonging to Arthropoda, positive correlations between heterogeneous contents, ie, A with C_3, T with C_3, and C with A_3 revealed the likely influence of selection pressure in shaping codon usage. Similarly, positive correlations existed between A and C_3, C and A_3, and G and T_3 in *Katharina tunicata,* belonging to Mollusca, suggesting the contribution of selection toward the shaping of codon usage, whereas in *Lumbricus terrestris,* belonging to Annelida and *Terebratulina retusa* belonging to Brachiopoda, no positive correlation between heterogeneous base contents and no negative correlation between homogeneous contents were observed, indicating the likely role of mutational pressure in framing the SCUB in these organisms.

Analysis of RSCU. Overall and strand-specific codon usages were analyzed (Tables 3 and 4). Significant strand-specific codon usage bias was not observed in any of the genomes (Table 4). In *A. pectinifera* (Echinodermata) and *B. carnosus* (Hemichordata), A- and C-ending codons were found to be used most frequently, whereas T-ending codons were used most frequently in *D. pulex* (Arthropoda) and *K. tunicata* (Mollusca). Most of the preferred codons in *L. terrestris* (Annelida) were found to be A/T–ending ones, and A/C-ending codons were used most often in *T. retusa* (Brachiopoda). In all representative species of Chaetognatha, A/T-ending codons were used often, and among these, T-ending codons were used more frequently than A-ending codons. The putative optimal codons in *S. enflata* could not be identified, whereas in all other species, the presence of optimal codons was noticed (Table 5). The frequency of putative optimal codons was found to vary from one genome to another, namely, 23 codons in *A. pectinifera*, 12 codons in *B. carnosus*, 8 codons in *D. pulex*, 18 codons in *K. tunicata*, 1 codon in *L. terrestris*, 9 codons in *P. gotoi*, 5 codons in *S. nagae*, 10 codons in *S. cephaloptera,* and 8 codons in *T. retusa*. In *A. pectinifiera* (Echinodermata), among the identified putative optimal codons, T-ending (34.78%), A-ending (17.39%), G-ending (13.04%), and C-ending (30.43%) codons were found. In *B. carnosus*, T-ending (33.33%), C-ending (33.33%), A-ending (16.66%), and G-ending (16.66%) codons were detected. In *D. pulex*, T-ending (37.5%), C-ending (37.5%), and A-ending (25%) codons, and in *K. tunicata,* T-ending (22.22%), A-ending (38.88%), G-ending (22.22%), and C-ending (16.66%) codons were used. *T. retusa* had A-ending (12.5%), T-ending (37.5%), G-ending (12.5%), and C-ending (37.5%) codons. Among chaetognaths, the following proportions of codons were found: T-ending (11.11%), A-ending (44.4%), C-ending (33.33%), and G-ending (11.11%) codons in *P. gotoi*; A-ending (40.0%), T-ending (20.0%), C-ending (20.0%), and G-ending (20.0%) codons in *S. cephaloptera*; C-ending (34.78%), A-ending (39.13%), G-ending (13.04%), and T-ending (13.04%) codons in *S. enflata*; A-ending (40.0%), T-ending (20.0%), G-ending (20.0%), and C-ending (20.0%) codons in *S. nagae*.

Quantification of SCUB. In *S. enflata,* a representative species of the minor phylum Chaetognatha, the majority of the PCGs were found to be grouped considerably below the expected GC_3 curve (Fig. 1). When PCGs are grouped considerably far below the expected GC_3 curve, it indicates the possible influence of selection.[28] However, in other species belonging to Chaetognatha, such as *S. nagae, S. decipiens,* and *S. cephaloptera,* the majority of PCGs were grouped on or just below the expected GC_3 curve, indicating the apparent influence of GC compositional constraints in driving SCUB. In contrast, one gene (*NADH*6) was located considerably below the expected GC_3 curve in *P. gotoi,* indicating the likely influence of selection on this gene. Interestingly, in *T. retusa* (Brachiopoda), all PCGs were located considerably below the expected GC_3 curve, indicating the apparently low influence of GC_3 compositional constraints on codon usage. It was clearly noted that the majority of PCGs were scattered far below the expected GC_3 curve in the cases of *A. pectinifera* (Echinodermata), *B. carnosus* (Hemichordata), *D. pulex* (Arthropoda), and *L. terrestris* (Annelida). This revealed the likely influence of selection or other unknown selective forces in framing the SCUB. To further confirm the influence of GC compositional constraints on codon usage bias, a neutrality plot was developed (Fig. 2). It revealed a positive correlation between GC at the silent positions and cumulative GC composition at the first and second positions, ie, GC_{12} in *D. pulex* (Arthropoda). No such correlations were found in representative species of other selected major and minor phyla. This might suggest the likely influence of GC_3 compositional constraints as an essential contributing factor that shapes SCUB in *D. pulex* (Arthropoda). ENC values of the chosen species of Chaetognatha varied from 37.17 to 50.49, with a mean and standard deviation (SD) of 45.28 and 5.13, respectively. The GC_3 values ranged from 26.35 to 35.17, with a mean and SD of 26.35 and 8.27, respectively. This confirmed the heterogeneity in codon usage among the chosen species belonging to Chaetognatha. In other phyla, GC_3 values varied from 23.68 (*K. tunicata*) to 41.79 (*T. retusa*), with a mean and SD of 34.88 and 6.04, respectively, whereas ENC values varied from 40.94 to 50.88, with a mean and SD of 46.29 and 3.99, respectively. These results suggest the heterogeneity in codon usage in other representative species of the selected phyla. Analyses of GC contents revealed that the average GC_1, GC_2, and GC_3 contents were low in all the analyzed mitogenomes. In addition, mean GC_3 was found to be significantly lower than GC_1 and GC_2. Selection pressure acting against mutational pressure might be the factor behind the narrow distribution of GC contents.

To analyze whether the SCUB is restricted in extremely biased PCGs, the relationship between purines (A and G) and pyrimidines (C and T) in four-codon amino acid families was analyzed using the PR2 bias plot (Fig. 3). Complex

Table 3. Overall relative synonymous codon usage in the mitogenomes of analyzed genomes.

AA	CODONS	A. PECTINIFERA	B. CARNOSUS	D. PULEX	K. TUNICATA	L. TERRESTRIS	T. RETUSA	P. GOTOI	S. DECIPIENS	S. ENFLATA	S. NAGAE	S. CEPHALOPTERA
Ala	GCT	1.013	0.693	**1.399**	**1.725**	**1.457**	0.908	**2.123**	**2.460**	**1.500**	**2.225**	**1.494**
	GCG	0.356	0.231	0.484	0.360	0.223	0.170	0.092	0.214	0.750	0.308	0.582
	GCC	1.298	**1.775**	1.220	0.588	1.145	**1.972**	0.400	0.449	1.025	0.781	0.633
	GCA	**1.333**	1.301	0.897	1.327	1.175	0.950	1.385	0.877	0.725	0.686	1.291
Cys	TGT	0.750	0.690	**1.394**	**1.250**	0.889	0.563	**1.880**	**1.434**	**1.556**	**1.405**	**1.511**
	TGC	**1.250**	**1.310**	0.606	0.750	**1.111**	**1.438**	0.120	0.566	0.444	0.595	0.489
Asp	GAT	0.618	0.423	**1.014**	**1.478**	**1.068**	0.627	**1.516**	**1.692**	**1.400**	**1.522**	**1.077**
	GAC	**1.382**	**1.577**	0.986	0.522	0.932	**1.373**	0.484	0.308	0.600	0.478	0.923
Glu	GAG	0.518	0.642	0.872	0.537	0.841	0.338	0.265	0.906	0.571	0.875	0.921
	GAA	**1.482**	**1.358**	1.128	**1.463**	**1.159**	**1.662**	**1.735**	**1.094**	**1.429**	**1.125**	**1.079**
Phe	TTT	**1.293**	0.776	**1.483**	**1.694**	**1.233**	0.930	**1.772**	**1.645**	**1.432**	**1.457**	**1.560**
	TTC	0.707	1.224	0.517	0.306	0.767	1.070	0.228	0.355	0.568	0.543	0.440
Gly	GGT	0.634	0.564	0.479	0.790	0.818	0.219	1.449	**2.129**	1.108	**1.519**	0.843
	GGG	0.846	0.946	**2.172**	**1.445**	**1.348**	1.005	0.522	0.836	**1.258**	0.833	1.166
	GGC	0.564	1.228	0.494	0.538	0.751	0.590	0.251	0.338	0.563	0.241	0.520
	GGA	**1.956**	**1.261**	0.854	1.227	1.083	**2.186**	**1.778**	0.697	1.070	1.407	**1.471**
His	CAC	**1.439**	**1.529**	0.953	0.481	**1.226**	**1.284**	0.377	0.459	0.866	0.820	0.882
	CAT	0.561	0.471	**1.047**	**1.519**	0.774	0.716	**1.623**	**1.541**	**1.134**	**1.180**	**1.118**
Ile	ATT	1.308	0.793	**1.321**	**1.604**	**1.299**	0.967	**1.838**	**1.733**	1.310	**1.561**	**1.582**
	ATA	**1.385**	0.951	(Met) 1.284	(Met) 1.583	(Met) 1.504	(Met) 1.654	(Met) 1.636	(Met) 1.546	(Met) 0.690	(Met) 1.403	(Met) 1.560
	ATC	0.692	**1.207**	0.679	0.396	0.701	**1.033**	0.162	0.267	**1.515**	0.439	0.418
Lys	AAA	(Asn) 1.351	(Asn) 0.000	1.071	1.674	**1.506**	**1.613**	**1.753**	**1.060**	0.485	**1.302**	**1.694**
	AAG	0.649	**2.000**	0.929	0.326	0.494	0.387	0.247	0.940	0.600	0.698	0.306
Leu	CTA	**1.628**	1.036	1.113	**1.731**	**1.895**	1.522	1.361	1.183	1.283	**1.753**	1.394
	CTC	0.723	**1.436**	0.795	0.615	0.787	1.059	0.218	0.323	0.333	0.351	0.887
	CTG	0.330	0.421	0.453	0.135	0.332	0.369	0.136	0.473	**1.783**	0.701	0.579
	CTT	1.319	1.107	**1.639**	1.519	0.986	1.051	**2.286**	**2.022**	1.226	1.196	1.140
	TTA	1.395	1.271	1.468	1.668	**1.669**	**1.789**	1.663	1.542	0.774	1.521	**1.466**
	TTG	0.605	0.729	0.532	0.332	0.331	0.211	0.337	0.458	0.705	0.479	0.534
Met	ATG	0.615	1.049	0.716	0.417	0.496	0.346	0.364	0.454	1.295	0.597	0.440
Asn	AAC	**1.143**	**1.661**	0.696	0.642	1.008	**1.178**	0.303	0.340	0.617	0.632	0.907
	AAT	0.857	0.339	1.304	1.358	0.992	0.822	**1.697**	**1.660**	**1.383**	**1.368**	**1.093**

AA	Codon											
Pro	CCT	1.349	0.970	**1.662**	**2.394**	**1.225**	0.653	**1.846**	**2.161**	**1.920**	**2.262**	**1.285**
	CCG	0.186	0.198	0.310	0.094	0.370	0.265	0.154	0.323	0.448	0.230	0.496
	CCC	**1.419**	**1.743**	1.380	0.472	1.225	**2.122**	0.462	0.323	0.992	0.754	1.022
	CCA	1.047	1.089	0.648	1.039	1.179	0.959	1.538	1.194	0.640	0.754	1.197
Gln	CAG	0.450	0.419	0.899	0.438	0.286	0.289	0.372	0.766	0.683	**1.024**	0.737
	CAA	**1.550**	**1.581**	**1.101**	**1.563**	**1.714**	**1.711**	**1.628**	**1.234**	**1.317**	0.976	**1.263**
Arg	CGA	**2.535**	**2.190**	**1.614**	**1.825**	**2.000**	**3.213**	**2.367**	1.167	1.156	**1.511**	**1.615**
	CGC	0.282	0.810	0.421	0.491	0.452	0.328	0.163	0.250	1.067	0.356	0.769
	CGG	0.732	0.429	0.491	0.912	0.581	0.262	0.490	0.667	0.533	0.889	1.231
	CGT	0.451	0.571	1.474	0.772	0.968	0.197	0.980	**1.917**	**1.244**	1.244	0.385
	AGC	0.468	0.898	0.456	0.511	0.284	0.452	0.085	0.509	0.402	0.150	0.101
	AGA	(Ser) 1.128	(Ser) 0.502	(Ser) 1.538	(Ser) 2.000	(Ser) 1.420	(Ser) 1.563	(Ser) 2.723	(Ser) 1.139	(Ser) 1.515	(Ser) 1.555	(Ser) 1.691
Ser	TCA	1.340	1.927	1.153	1.106	**2.083**	1.296	1.390	1.348	0.663	1.480	1.615
	TCC	1.745	**2.323**	0.625	0.511	1.633	**2.447**	0.312	0.539	1.349	0.853	0.732
	TCG	0.489	0.660	0.553	0.213	0.189	0.165	0.113	0.330	0.592	0.351	0.429
	TCT	**2.000**	1.215	**2.523**	**2.340**	1.728	1.645	**2.525**	**2.037**	**1.751**	**2.157**	**2.120**
	AGG	(Ser) 0.362	(Ser) 0.000	(Ser) 0.024	(Ser) 0.553	(Ser) 0.284	(Ser) 0.329	(Ser) 0.397	(Ser) 0.719	(Ser) 0.947	(Ser) 0.577	(Ser) 0.934
	AGT	0.468	0.475	1.129	0.766	0.379	0.103	0.454	1.378	0.781	0.878	0.379
Thr	ACA	1.076	1.192	0.894	1.609	**1.438**	1.168	**1.855**	0.908	1.033	1.375	1.418
	ACC	**1.546**	**1.585**	1.140	0.425	1.140	**1.856**	0.261	0.596	1.100	0.650	0.759
	ACG	0.235	0.302	0.291	0.179	0.182	0.112	0.058	0.312	0.467	0.350	0.304
	ACT	1.143	0.921	**1.676**	**1.788**	1.240	0.864	1.826	**2.184**	**1.400**	**1.625**	**1.519**
Val	GTC	0.622	**1.531**	0.612	0.262	0.796	0.921	0.133	0.427	0.711	0.267	0.418
	GTG	0.639	0.536	0.731	0.541	0.718	0.251	0.442	0.444	0.959	0.681	0.803
	GTT	**1.412**	1.167	**1.373**	**1.705**	0.854	1.173	**1.746**	**1.983**	1.091	1.319	1.071
	GTA	1.328	0.766	1.284	1.492	**1.631**	**1.654**	1.680	1.145	**1.240**	**1.733**	**1.707**
Trp	TGG	0.479	0.724	0.825	0.673	0.505	0.163	0.267	0.835	**1.154**	0.519	**1.091**
	TGA	(Trp) **1.521**	(Trp) **1.276**	(Trp) **1.175**	(Trp) **1.327**	(Trp) **1.495**	(Trp) **1.837**	(Trp) **1.733**	(Trp) **1.165**	(Trp) 0.846	(Trp) **1.481**	(Trp) 0.909
Tyr	TAC	**1.119**	**1.439**	**1.232**	0.672	**1.026**	**1.304**	0.397	0.270	0.607	0.738	0.762
	TAT	0.881	0.561	0.768	**1.328**	0.974	0.696	**1.603**	**1.730**	**1.393**	**1.262**	**1.238**

Notes: Figures in bold indicate preferred codon for the corresponding amino acid. Shaded rows represent change in genetic code (changes in translation table).

Table 4. Strand-specific relative synonymous codon usage values.

CODON	A. PECTINIFERA		B. CARNOSUS		D. PULEX		K. TUNICATA		L. TERRESTRIS		P. GOTOI		S. DECIPIENS		S. ENFLATA		S. NAGAE		S. CEPHALOPTERA		T. RETUSA	
	J	N	J	N	J	N	J	N	J	N	J	N	J	N	J	N	J	N	J	N	J	N
TGA	2.717	2.000	2.635	2.000	2.581	0.938	2.455	1.826	3.000	–	2.684	2.800	2.514	2.250	2.516	2.478	2.464	2.615	1.500	2.455	2.750	–
TAG	0.057	0.333	0.081	1.000	0.140	0.000	0.273	0.400	0.000	–	0.158	0.000	0.162	0.150	0.194	0.000	0.107	0.231	1.000	0.136	0.000	–
TAA	0.226	0.667	0.284	0.000	0.279	2.000	0.273	1.600	0.000	–	0.158	0.200	0.324	0.600	0.290	0.522	0.429	0.154	0.500	0.409	0.250	–
GCT	0.688	2.326	0.692	0.400	1.314	1.722	2.455	1.203	1.457	–	1.879	2.421	2.602	2.286	1.937	0.862	2.145	2.280	1.200	1.829	1.043	–
GCG	0.188	0.558	0.176	2.000	0.514	0.608	0.727	0.098	0.223	–	0.182	0.000	0.194	0.238	0.253	1.477	0.464	0.200	0.800	0.343	0.155	–
GCC	1.594	0.465	1.811	0.800	1.171	1.063	0.136	0.911	1.145	–	0.424	0.421	0.466	0.429	1.642	0.123	0.696	0.840	0.800	0.629	1.952	–
GCA	1.531	0.651	1.321	0.800	1.000	0.608	0.682	1.789	1.175	–	1.515	1.158	0.738	1.048	0.168	1.538	0.696	0.680	1.200	1.200	0.850	–
TGT	0.421	1.200	0.593	2.000	1.286	1.385	1.778	0.762	0.889	–	1.778	2.000	1.448	1.417	1.385	1.714	1.474	1.333	1.333	1.533	0.696	–
TGC	1.579	0.800	1.407	0.000	0.714	0.615	0.222	1.238	1.111	–	0.222	0.000	0.552	0.583	0.615	0.286	0.526	0.667	0.667	0.467	1.304	–
GAT	0.375	1.800	0.377	2.000	0.791	1.200	1.714	1.111	1.068	–	1.448	1.857	1.722	1.655	1.415	1.368	1.556	1.500	0.286	1.417	0.750	–
GAC	1.625	0.200	1.623	0.000	1.209	0.800	0.286	0.889	0.932	–	0.552	0.143	0.278	0.345	0.585	0.632	0.444	0.500	1.714	0.583	1.250	–
GAG	0.186	1.524	0.649	0.500	0.755	1.636	0.792	0.176	0.841	–	0.303	0.200	0.750	1.167	0.400	0.757	0.765	1.000	0.421	1.238	0.444	–
GAA	1.814	0.476	1.351	1.500	1.245	0.364	1.208	1.824	1.159	–	1.697	1.800	1.250	0.833	1.600	1.243	1.235	1.000	1.579	0.762	1.556	–
TTT	1.185	1.758	0.729	1.571	1.478	1.583	1.879	1.525	1.233	–	1.648	1.851	1.698	1.590	1.305	1.635	1.528	1.420	1.564	1.579	1.006	–
TTC	0.815	0.242	1.271	0.429	0.522	0.417	0.121	0.475	0.767	–	0.352	0.149	0.302	0.410	0.695	0.365	0.472	0.580	0.436	0.421	0.994	–
GGT	0.431	1.000	0.523	0.889	0.433	0.610	0.851	0.701	0.818	–	1.091	1.681	2.198	2.044	1.867	0.370	1.455	1.547	0.353	1.059	0.229	–
GGG	0.431	1.706	0.860	1.630	2.242	2.271	1.872	0.825	1.348	–	0.646	0.580	0.685	1.022	0.876	1.630	0.788	0.853	1.294	0.941	1.257	–
GGC	0.824	0.353	1.327	0.444	0.561	0.407	0.340	0.825	0.751	–	0.283	0.290	0.360	0.311	0.914	0.222	0.242	0.240	0.118	0.706	0.800	–
GGA	2.314	0.941	1.290	1.037	0.764	0.712	0.936	1.649	1.083	–	1.980	1.449	0.757	0.622	0.343	1.778	1.515	1.360	2.235	1.294	1.714	–
CAC	1.538	0.571	1.529	0.000	0.966	0.842	0.279	0.722	1.226	–	0.444	0.000	0.541	0.333	1.163	0.333	0.545	0.974	1.200	0.846	1.256	–
CAT	0.462	1.429	0.471	0.000	1.034	1.158	1.721	1.278	0.774	–	1.556	2.000	1.459	1.667	0.837	1.667	1.455	1.026	0.800	1.154	0.744	–
ATT	0.928	1.364	0.931	0.857	1.479	1.556	2.047	1.464	1.186	–	1.547	1.745	1.789	1.754	1.304	0.950	1.324	1.339	1.400	1.402	0.962	–
ATA	1.538	1.364	0.634	1.714	0.691	0.444	0.770	0.536	1.173	–	1.280	1.170	0.972	0.934	0.848	1.733	1.368	1.246	1.143	1.225	1.123	–
ATC	0.534	0.273	1.436	0.429	0.829	0.923	0.182	1.841	0.641	–	0.173	0.085	0.239	0.311	0.848	0.317	0.309	0.415	0.457	0.373	0.915	–
AAA	1.474	1.029	0.000	0.000	1.174	1.077	1.310	0.159	1.506	–	1.833	1.692	1.190	0.927	1.609	1.434	1.238	1.364	1.647	1.674	1.521	–
AAG	0.526	0.971	2.000	2.000	0.826	1.097	0.690	1.905	0.494	–	0.167	0.308	0.810	1.073	0.391	0.566	0.762	0.636	0.353	0.326	0.479	–
CTA	1.659	0.681	1.059	0.000	1.140	0.710	1.000	0.738	1.895	–	1.429	1.212	1.087	1.277	0.306	1.314	1.723	1.767	1.662	1.143	1.496	–
CTC	0.853	0.170	1.463	0.533	0.740	0.484	0.100	0.071	0.787	–	0.333	0.000	0.261	0.383	1.576	0.571	0.123	0.465	0.492	0.898	1.089	–
CTG	0.248	1.021	0.375	2.133	0.420	1.710	0.400	1.286	0.332	–	0.143	0.242	0.522	0.426	0.141	0.800	0.800	0.651	0.862	0.571	0.397	–
CTT	1.240	2.128	1.103	1.333	1.700	1.444	2.500	1.963	0.986	–	2.095	2.545	2.130	1.915	1.976	1.314	1.354	1.116	0.985	1.388	1.018	–
TTA	1.797	1.019	1.452	0.167	1.467	0.556	1.455	0.037	1.669	–	1.733	1.578	1.636	1.467	1.440	1.097	1.393	1.595	1.585	1.368	1.750	–
TTG	0.203	0.981	0.548	1.833	0.533	0.814	0.545	0.120	0.331	–	0.267	0.422	0.364	0.533	0.560	0.903	0.607	0.405	0.415	0.632	0.250	–
ATG	1.000	1.000	1.000	1.000	1.000	1.186	1.000	1.880	1.000	–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	–
AAC	1.333	0.235	1.658	0.000	0.776	0.650	0.189	0.868	1.008	–	0.473	0.054	0.327	0.353	0.840	0.364	0.588	0.656	0.933	0.809	1.086	–

AAT	0.667	1.765	0.342	0.000	1.224	1.350	1.811	1.132	0.992	–	1.527	1.946	1.673	1.160	1.636	1.647	1.412	1.344	1.067	1.191	0.914	–
CCA	1.087	0.800	1.117	0.000	0.500	2.000	1.636	1.636	1.179	–	1.587	1.514	1.189	0.205	1.362	1.200	1.156	0.519	1.103	1.220	1.027	–
CCC	1.670	0.400	1.766	0.800	1.636	0.286	0.262	0.000	1.225	–	0.762	0.000	0.324	1.282	0.511	0.320	0.800	0.727	1.517	0.475	2.081	–
CCT	1.126	2.133	0.975	0.800	1.500	0.857	3.213	0.667	1.225	–	1.651	2.054	2.162	2.359	1.191	2.160	1.689	2.597	0.828	1.966	0.622	–
CCG	0.117	0.667	0.142	2.400	0.364	0.857	0.197	1.697	0.370	–	0.000	0.432	0.324	0.154	0.936	0.320	0.356	0.156	0.552	0.339	0.270	–
CAA	1.739	0.875	1.581	0.000	1.125	0.952	1.286	0.222	1.714	–	1.579	1.846	1.333	1.455	1.158	1.130	1.053	0.909	1.714	0.909	1.679	–
CAG	0.261	1.125	0.419	0.000	0.875	1.048	0.714	1.778	0.286	–	0.421	0.154	0.667	0.545	0.842	0.870	0.947	1.091	0.286	1.091	0.321	–
AGA	1.778	1.077	2.000	2.000	2.000	1.333	1.228	2.333	1.667	–	1.828	1.619	1.200	1.000	1.261	1.250	1.308	1.525	1.692	1.085	1.627	–
AGG	0.222	0.923	0.000	0.000	0.000	0.190	0.772	1.000	0.333	–	0.172	0.381	0.800	1.000	0.739	0.750	0.692	0.475	0.308	0.915	0.373	–
CGA	2.421	1.333	2.272	0.000	1.744	0.762	1.455	0.333	2.000	–	2.727	2.000	1.379	0.444	2.222	0.842	1.400	1.600	1.600	1.714	3.351	–
CGC	0.316	0.267	0.84	1.000	0.513	1.714	0.121	0.333	0.452	–	0.000	0.400	0.276	1.778	0.000	0.211	0.000	0.640	0.400	0.381	0.216	–
CGG	0.947	1.067	0.444	0.000	0.308	0.408	1.333	0.643	0.581	–	0.364	0.400	0.552	0.148	1.111	0.842	1.200	0.640	0.800	1.524	0.324	–
CGT	0.316	1.333	0.444	3.000	1.436	1.376	1.091	2.372	0.968	–	0.909	1.200	1.793	1.63	0.667	2.105	1.400	1.120	1.200	0.381	0.108	–
AGC	1.385	0.267	1.395	0.889	0.579	0.815	0.471	1.688	0.857	–	0.222	0.667	0.733	0.485	1.059	0.364	0.333	0.261	1.000	0.308	1.700	–
AGT	0.615	1.733	0.605	1.111	1.421	0.510	1.529	0.844	1.143	–	1.778	1.333	1.267	1.515	0.941	1.636	1.667	1.739	1.000	1.692	0.300	–
TCA	1.203	0.200	1.275	1.000	0.918	0.561	0.465	0.080	1.479	–	1.655	1.200	1.351	0.356	1.306	1.176	1.356	1.164	1.576	1.333	0.919	–
TCC	1.464	0.267	1.537	0.000	0.59	2.955	0.140	1.809	1.160	–	0.207	0.267	0.541	0.481	0.571	0.471	0.746	0.687	0.848	0.485	1.799	–
TCG	0.288	0.667	0.419	1.000	0.426	0.051	0.372	0.161	0.134	–	0.069	0.200	0.432	0.237	1.388	0.176	0.339	0.269	0.364	0.283	0.115	–
TCT	1.046	2.867	0.769	2.000	2.066	1.325	3.023	0.402	1.227	–	2.069	2.333	1.676	1.926	0.735	2.176	1.559	1.881	1.212	1.899	1.167	–
ACC	1.654	0.667	1.588	1.333	1.171	0.706	0.000	1.916	1.140	–	0.267	0.316	0.564	1.623	0.392	0.635	0.839	0.531	1.067	0.677	1.884	–
ACA	1.160	0.500	1.206	0.000	0.976	0.941	1.000	0.639	1.438	–	2.187	1.053	0.974	0.406	1.882	0.825	1.484	1.306	1.733	1.108	1.163	–
ACG	0.123	0.333	0.260	2.667	0.228	0.471	0.200	0.168	0.182	–	0.053	0.105	0.308	0.174	0.863	0.317	0.452	0.286	0.267	0.369	0.140	–
ACT	1.062	2.500	0.947	0.000	1.626	1.882	2.800	1.277	1.240	–	1.493	2.526	2.154	1.797	0.863	2.222	1.226	1.878	0.933	1.846	0.814	–
GTT	0.844	2.020	1.061	1.857	1.129	0.467	2.139	0.440	0.854	–	1.165	2.338	2.081	1.485	0.828	1.874	1.412	1.262	1.000	1.069	1.323	–
GTG	0.444	1.069	0.354	1.714	0.663	1.000	0.889	0.040	0.718	–	0.354	0.369	0.455	0.536	1.241	0.432	0.667	0.690	0.727	1.000	0.315	–
GTC	1.022	0.158	1.746	0.143	0.712	1.700	0.139	1.080	0.796	–	0.152	0.123	0.325	1.155	0.414	0.541	0.078	0.381	0.455	0.414	1.039	–
GTA	1.689	0.752	0.84	0.286	1.497	0.833	0.833	2.440	1.631	–	2.329	1.169	1.138	0.825	1.517	1.153	1.843	1.667	1.818	1.517	1.323	–
TGG	1.000	1.000	1.000	0.000	1.000	1.063	1.000	0.174	1.000	–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	–
TAC	1.298	0.529	1.520	0.286	1.254	1.000	0.536	0.788	1.026	–	0.533	0.275	0.299	0.780	0.444	0.247	0.634	0.790	0.800	0.786	1.373	–
TAT	0.702	1.471	0.480	1.714	0.746	1.000	1.464	1.212	0.974	–	1.467	1.725	1.701	1.220	1.556	1.753	1.366	1.210	1.200	1.214	0.627	–

Table 5. Identified putative optimal codons in the mitogenomes of chosen species.

A. PECTINIFERA	B. CARNOSUS	D. PULEX	K. TUNICATE	L. TERRESTRIS	P. GOTOI	S. DECIPIENS	S. ENFLATA	S. NAGAE	S. CEPHALOPTERA	T. RETUSA
GCT	TTT	GAT	TTT	TTA	TTA	–	GCC	GGC	GAG	GGC
GAT	TTC	GAC	GGA	–	AAC	–	GCA	ATA	GAA	AAC
GAC	AGT	GAG	ATA	–	CCC	–	CTC	ATG	GCT	AAT
GAG	CGA	AAG	ATG	–	AGA	–	CTT	TTA	GCA	CAA
GAA	CGT	CCT	CTA	–	GTT	–	TTA	AGT	GAA	CAG
TTT	ACC	TCC	TTG	–	ATG	–	AGA	–	GGT	CGT
TTC	ACG	TAC	CCA	–	AAC	–	AGC	–	GGA	ACC
CAC	GTG	TAT	CCC	–	ACA	–	TCT	–	CTG	GTT
CAT	GTC	–	CCT	–	GTA	–	CGC	–	TCC	–
AAC	GTA	–	CAA	–	–	–	ACA	–	CGC	–
AAT	TAC	–	CAG	–	–	–	GTT	–	–	–
CAA	TAT	–	TCA	–	–	–	GTA	–	–	–
CAG	–	–	TCC	–	–	–	GGC	–	–	–
AGG	–	–	TCT	–	–	–	GGA	–	–	–
AGC	–	–	ACC	–	–	–	CCA	–	–	–
TCA	–	–	GTT	–	–	–	CAA	–	–	–
CGA	–	–	GTG	–	–	–	CAG	–	–	–
ACC	–	–	GTA	–	–	–	AGG	–	–	–
CGT	–	–	–	–	–	–	TCC	–	–	–
ACT	–	–	–	–	–	–	CGA	–	–	–
GTC	–	–	–	–	–	–	ACC	–	–	–
TAC	–	–	–	–	–	–	ACG	–	–	–
TAT	–	–	–	–	–	–	GTC	–	–	–

Figure 1. ENC vs. GC$_3$ plots for the quantification of SCUB. The SCUB of genes lying on or close to the expected GC$_3$ is dictated by GC compositional constraints.

associations between purines and pyrimidines were observed in Chaetognatha and the other examined genomes. The analysis of PR2 bias plot revealed that A and T contents were used more proportionally than G and C contents. In addition, significant compositional differences between C and G and between A and T contents were found in most of the PCGs.

Factors influencing codon usage bias. COA–RSCU was carried out to identify the major factors influencing the SCUB (Fig. 4). Correlation between codon usage indexes and COA axes was analyzed using Spearman's rank correlation method (Table 6). Axis 1 and Axis 2 accounted for the majority of the total variations. Thus, Axis 1 could be considered to be a single major explanatory axis as it accounted for more than 40% variations in all chosen species. In *A. pectinifera* (Echinodermata), Axis 1 was in significant negative correlation with A$_3$, C$_3$, and GC$_3$ and was in significant positive correlation with T$_3$ and G$_3$. Thus, silent base contents might play a critical role in shaping the SCUB in the mitogenome of *A. pectinifera*. In *B. carnosus* (Hemichordata), Axis 1 was in significant negative correlation with CAI and
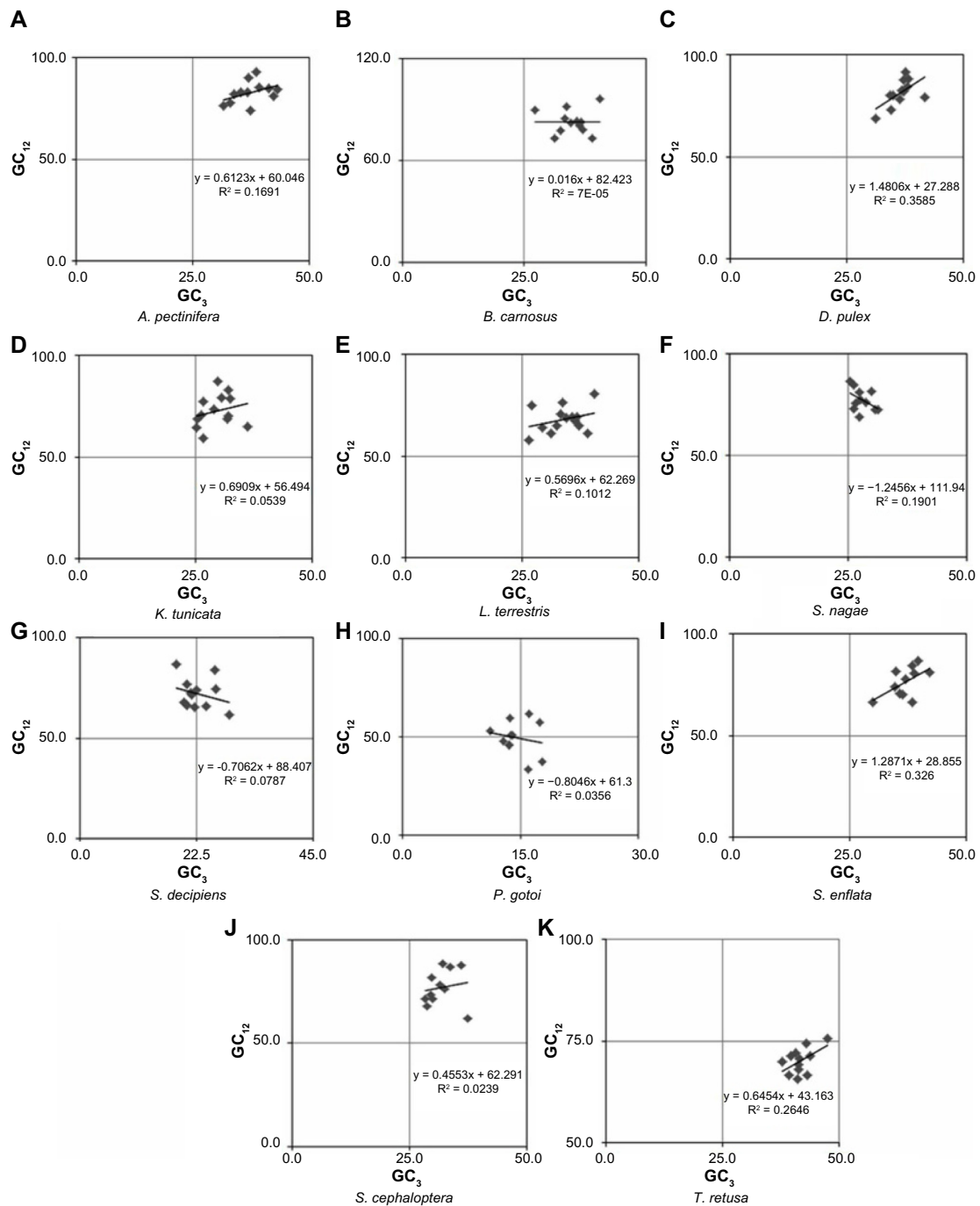
Figure 2. Neutrality plot. If correlation exists between GC_3 and GC_{12}, GC compositional constraints may have profound influence in shaping SCUB. If no correlation exists between GC_3 and GC_{12}, selection may have significant role in framing SCUB.

was in positive correlation with ENC and GRAVY score. This suggested that the level of gene expression might play a crucial role in driving SCUB in *B. carnosus*. In *D. pulex* (Arthropoda), A_3 and C_3 were found to be in significant positive and negative correlation with Axis 1, respectively, suggesting that A and C contents at the third codon position might be a contributing factor in framing SCUB in *D. pulex*. All silent base contents were found to be in significant correlation with Axis 1 in *K. tunicata*, indicating the apparent role of nucleotide compositional constraints in shaping SCUB. No correlations

were observed between any of the codon usage indexes and Axis 1 in *L. terrestris* (Annelida), *S. nagae, S. cephaloptera* (Chaetognatha), and *T. retusa* (Brachiopoda), suggesting the influence of some other unknown selective forces in shaping variations in SCU in the mitogenomes of these organisms. In *S. decipiens* (Chaetognatha), Axis 1 was negatively correlated with A_3 and positively correlated with T_3 and aromaticity, suggesting their role in CUB. In *S. enflata*, Axis 1 was significantly correlated with all silent base contents, indicating the likely role of mutational pressure due to compositional
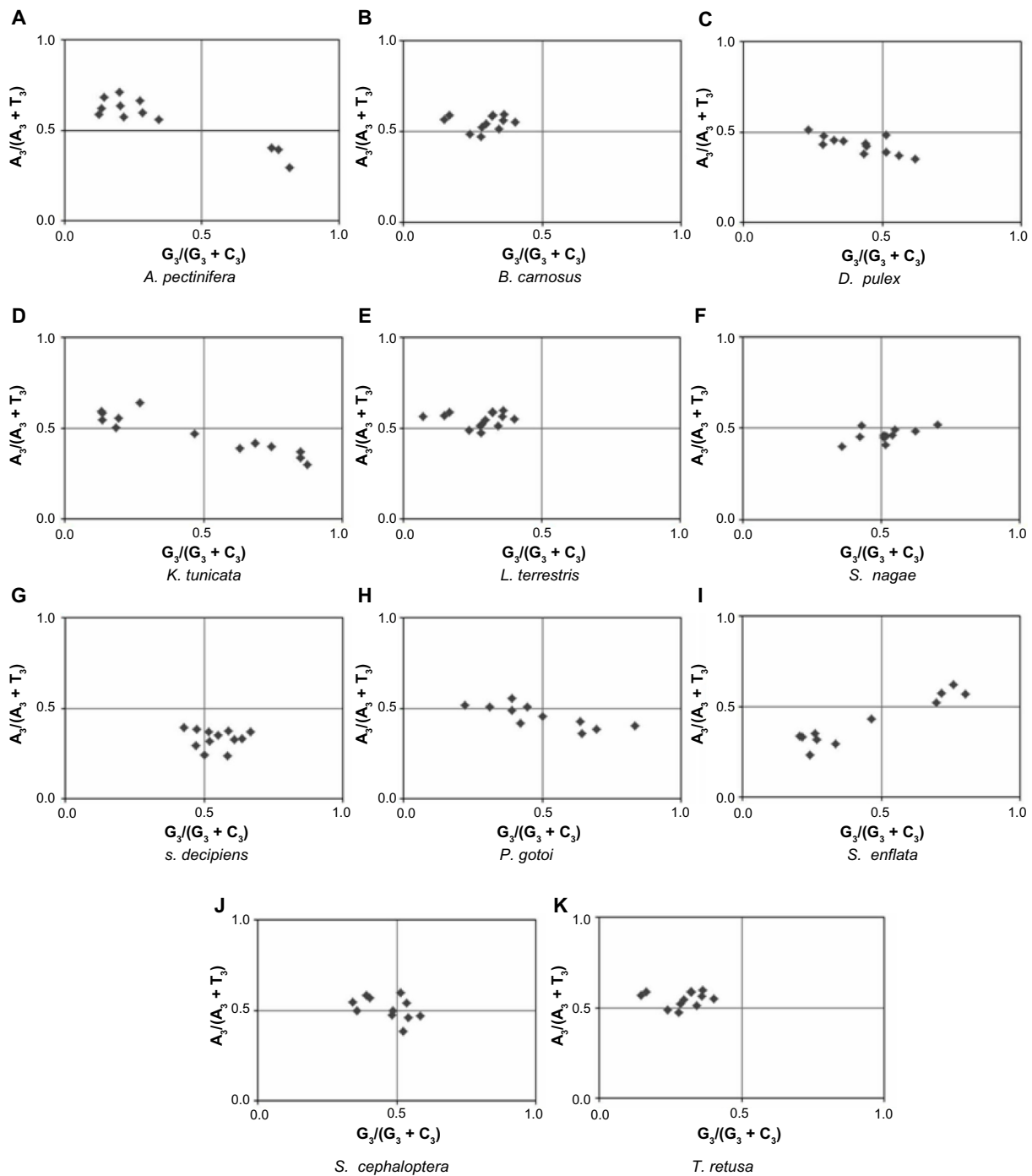
Figure 3. Parity rule 2 bias plot. This plot reveals the proportionate usage of A and T contents in comparison with that of G and C contents.

constraints. Interestingly, Axis 1 was in significant negative correlation with ENC in *P. gotoi* (Chaetognatha). This suggested that the level of gene expression might influence codon usage in the mitogenomes of *P. gotoi*.

Cluster analysis yielded two clusters, ie, one upper cluster (major) and one lower cluster (minor), based on the variation in RSCU values (Fig. 5). The upper cluster included all chosen species from Chaetognatha, Mollusca, and Arthropoda, and the lower minor cluster included the species chosen from Annelida, Echinodermata, Brachiopoda, and Hemichordata. This grouping suggested that variations of RSCU values in

the PCGs of mitogenomes in the chaetognaths were more similar to those of the chosen species of protostomes than those of deuterostomes. The results suggest that mutational pressure, natural selection, and some other unknown selective forces were acting at varying intensities (species specific) in the examined mitogenomes (Fig. 6). In general, terms such as strong, moderate, and weak are used in Figure 6 to express the idea that in any selected mitogenome, the intensity of mutational pressure and selection pressure vary significantly. In this study, this factor was assessed by using the following strategy.
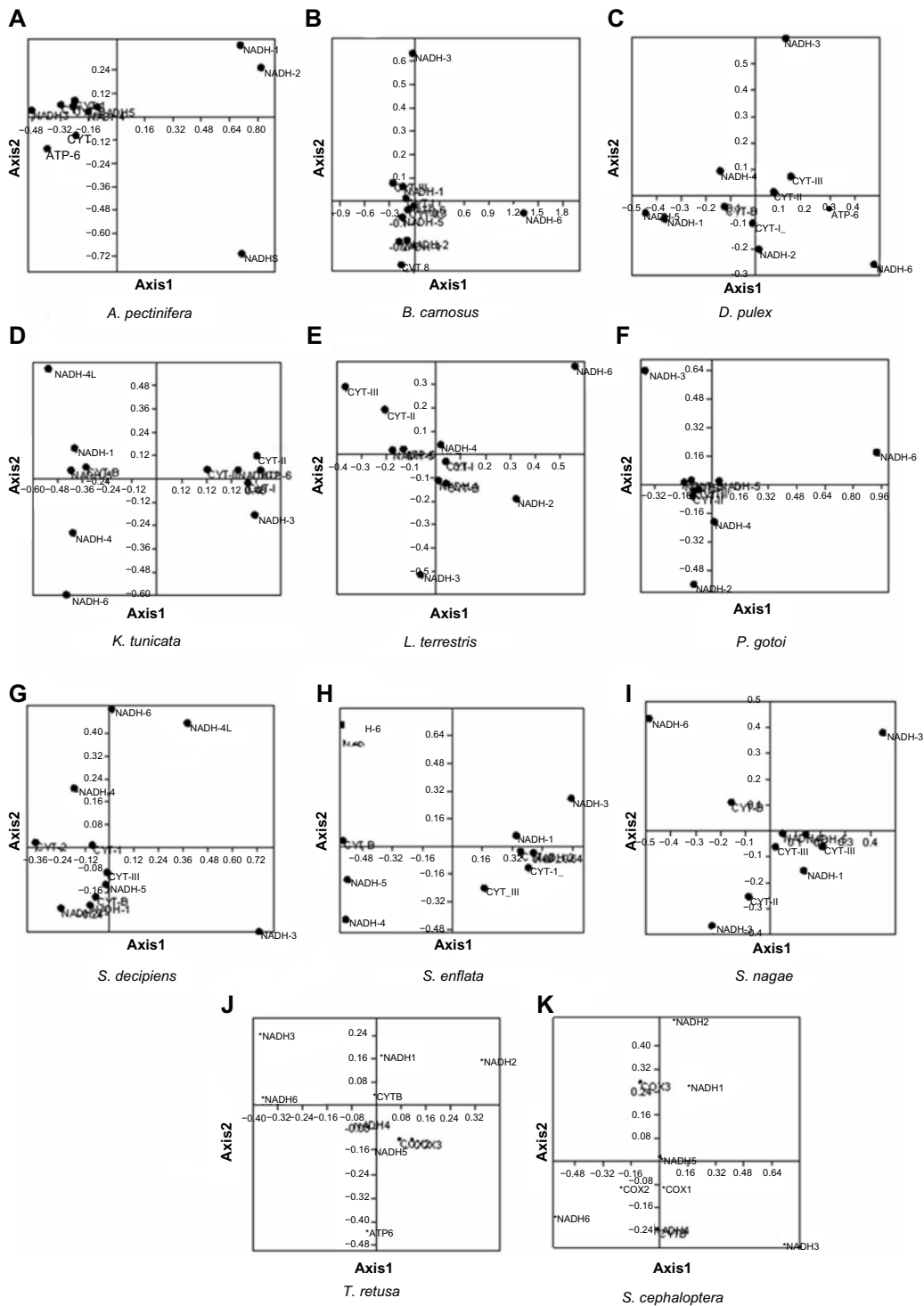
Figure 4. Correspondence analysis. Axis 1 accounts for the majority of variations. COA–RSCU plot reveals various factors that are responsible for the variations in RSCU values.

Among the following five conditions,

- no significant correlation between GC_3 and GC_{12};
- significant correlation between Axis 1 COA, with the indexes showing the level of gene expression;
- positive correlation between heterogeneous nucleotide contents;

- negative correlation between homogeneous nucleotide contents; and
- significant correlation between Axis 1 COA and any of the silent base contents,

if any three conditions are met, then it is considered that strong selection pressure acts on such genomes. If two

Table 6. Spearman's rank correlation analysis between various codon usage indexes and major axes of COA.

INDICES	A. PECTINIFERA		B. CARNOSUS		D. PULEX		K. TUNICATA		L. TERRESTRIS		S. NAGAE		S. DECIPIENS		P. GOTOI		S. ENFLATA		S. CEPHALOPTERA		T. RETUSA	
	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2	AXIS 1	AXIS 2
A_3	−0.64*	−0.16	−0.22	−0.64*	0.59	−0.23	−0.90**	0.20	0.05	−0.16	0.36	−0.45	−0.70*	0.05	−0.28	−0.45	0.75**	−0.05	0.03	0.60*	0.52	0.15
T_3	0.82**	0.39	0.34	0.27	−0.50	0.37	−0.87**	−0.18	−0.24	−0.22	−0.15	0.15	0.74**	−0.11	0.33	0.31	0.74**	0.17	−0.29	−0.54	−0.10	−0.26
G_3	0.75**	0.39	0.07	−0.23	−0.63	−0.29	0.76**	0.11	−0.09	−0.64*	0.13	0.79**	0.30	−0.44	0.47	0.41	0.74**	−0.32	0.56	0.35	−0.18	−0.32
C_3	−0.88**	−0.49	−0.02	0.62*	−0.70*	0.02	−0.78**	−0.26	0.07	0.51	0.24	0.79**	0.17	0.64*	−0.49	−0.21	0.74**	0.12	−0.46	−0.04	−0.43	0.14
GC_3	−0.65*	−0.42	0.02	0.64*	0.18	−0.23	0.26	0.24	0.45	−0.12	0.48	−0.33	0.48	0.17	0.16	−0.11	0.70*	−0.06	0.08	−0.19	−0.53	−0.02
ENC	0.40	−0.48	0.57	−0.11	−0.11	0.32	0.09	−0.42	−0.36	−0.50	0.36	−0.69*	0.01	0.23	−0.81**	−0.72*	−0.26	−0.62	0.38	−0.04	−0.21	0.12
CAI	−0.20	0.14	−0.83**	−0.40	−0.37	−0.17	0.45	0.20	−0.43	−0.19	0.18	−0.20	−0.10	−0.45	0.11	0.01	−0.38	0.11	−0.23	−0.10	0.19	−0.25
Aromaticity	−0.39	−0.10	−0.25	0.27	−0.42	0.48	0.15	0.11	0.05	0.16	−0.15	0.66*	0.74**	−0.18	0.10	0.49	0.35	0.15	−0.29	−0.68*	−0.47	0.44
Gravy	0.32	−0.20	0.58	−0.07	−0.05	−0.35	−0.23	−0.35	0.49	−0.18	−0.05	0.66*	0.48	0.14	0.06	−0.34	0.32	0.44	0.04	−0.18	−0.07	−0.12

Notes: *$P \leq 0.05$; **$P \leq 0.01$.

conditions are met, then the term "moderate" is used to express the intensity of selection, and if only one or no condition is met, the term "weak" is used. The intensity of mutational pressure was also assessed by using the above strategy, albeit vice versa. If both mutational and selection pressures are strong, then the influence of the unknown selective forces on SCUB was reckoned as weak.

Discussion

Natural selection and mutational pressure are regarded as the two major forces behind the species-specific SCU.[17,36] Analysis of codon usage revealed that influence of selection or mutational pressure was not absolute in representative species of Chaetognatha. In *P. gotoi*, one gene, ie, *NADH6* (ENC = 30.92), was found to lie considerably below the expected GC_3 curve, indicating the possible influence of selection in shaping codon usage of that particular gene as selection might be dominant in highly biased genes (ENC < 35). All other genes were lying on or close to the GC_3 curve. However, in other species of Chaetognatha, ie, *S. nagae* and *S. cephaloptera,* Axis 1 was not in correlation with any codon usage indexes (silent base contents and indexes indicating level of gene expression) and the majority of genes were lying far below the GC_3 curve, indicating the apparent influence of some unknown selective forces other than selection and mutational pressures in driving SCUB in their mitogenomes. However, significant correlation of Axis 1 with silent base contents and grouping of genes on or just below the GC_3 curve indicated the role of mutational pressure due to compositional constraints in driving SCUB in *S. decipiens*. In the mitogenomes of *S. enflata*, though the majority of the genes were spread considerably below the expected GC_3 curve, Axis 1 was found to be in significant correlation with all silent base contents, indicating the possible influence of compositional constraints. In addition, positive correlation between heterogeneous base contents was observed in *S. enflata*. This suggested the likely role of some unknown selective forces in driving SCUB in *S. enflata*.

We confirmed that trends associated with codon usage in the representative species of the minor phylum Chaetognatha were not uniform and absolute, and these findings were supported by the genome theory that codon usage is species specific.[44] In addition, we suggest that SCU variation in the mitogenomes of chaetognaths was not completely dictated by GC_3 composition and selection pressure. No correlation between GC_{12} and GC_3 in the mitogenomes of selected chaetognaths indicated that mutational pressure due to GC compositional constraints was apparently less influential in framing the SCUB in the mitogenomes of these species. Our analyses showed the presence of certain other unknown selective forces in shaping codon usage across PCGs in the mitogenomes of species belonging to Chaetognatha.

In general, the unknown selective forces act at the mRNA level.[15] The mRNA sequences contain both information (coding sequences) required for synthesizing protein
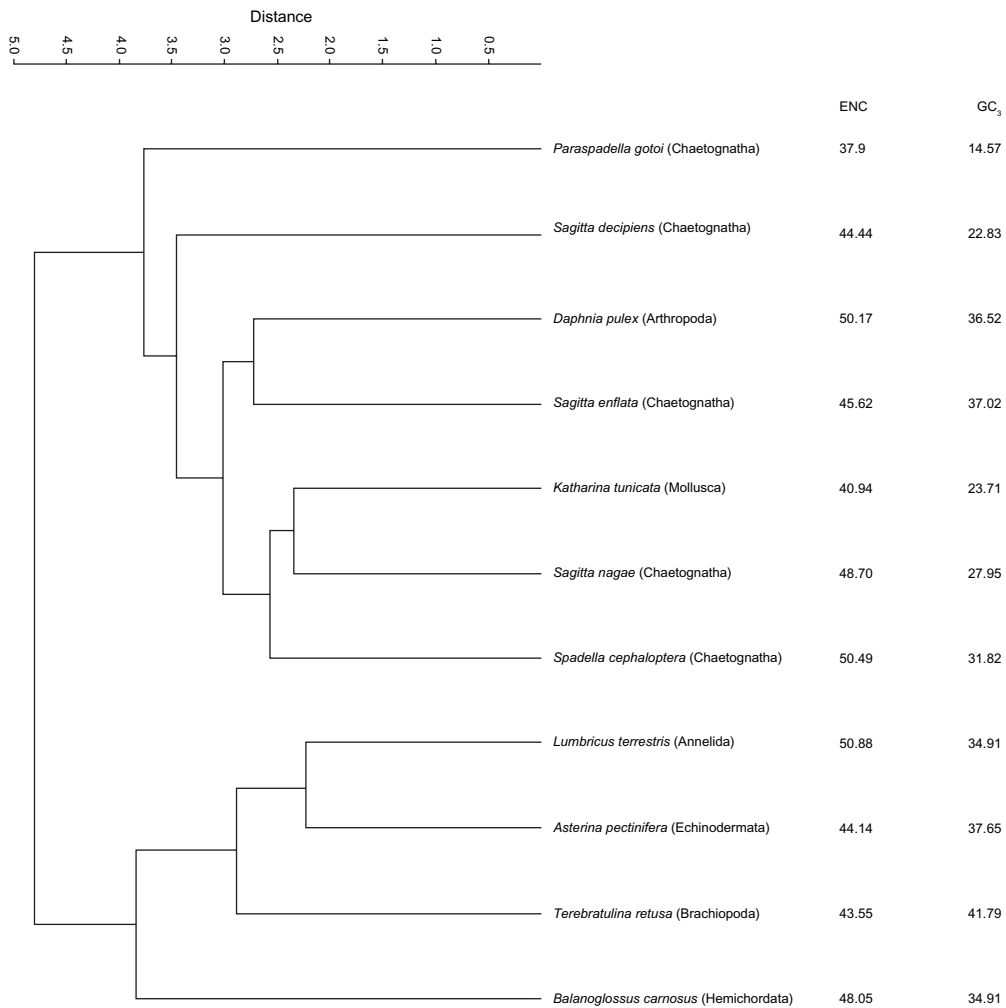
Distance

	ENC	GC₃
Paraspadella gotoi (Chaetognatha)	37.9	14.57
Sagitta decipiens (Chaetognatha)	44.44	22.83
Daphnia pulex (Arthropoda)	50.17	36.52
Sagitta enflata (Chaetognatha)	45.62	37.02
Katharina tunicata (Mollusca)	40.94	23.71
Sagitta nagae (Chaetognatha)	48.70	27.95
Spadella cephaloptera (Chaetognatha)	50.49	31.82
Lumbricus terrestris (Annelida)	50.88	34.91
Asterina pectinifera (Echinodermata)	44.14	37.65
Terebratulina retusa (Brachiopoda)	43.55	41.79
Balanoglossus carnosus (Hemichordata)	48.05	34.91

Figure 5. Cluster analysis. All examined species were grouped into two clusters based on the variations in RSCU values. All chaetognaths were grouped with species belonging to protostomes.

Major forces

		Mutational pressure	Selection pressure	Unknown selective forces
Brachiopoda	*T. retusa*	Moderate	Moderate	Strong
Hemichordata	*B. carnosus*	Moderate	Strong	Moderate
Echinodermata	*A. pectinifiera*	Strong	Moderate	Moderate
	S. nagae	Moderate	Strong	Moderate
	S. enflata	Moderate	Moderate	Strong
Chaetognatha	*S. decipiens*	Moderate	Weak	Strong
	P. gotoi	Weak	Strong	Moderate
	S. cephaloptera	Moderate	Moderate	Strong
Arthropoda	*D. pulex*	Strong	Weak	Moderate
Annelida	*L. terrestris*	Strong	Moderate	Moderate
Mollusca	*K. tunicata*	Strong	Moderate	Moderate

Figure 6. Identified major forces driving SCUB in mitogenomes of chaetognaths and other closely related phyla. The intensity of these forces vary and terms such as strong, moderate, and weak are used to express it.

as well as information (motifs) required for regulating the expression.[15] It is common knowledge that most of the motifs are mRNA secondary structures. Though the functional significance of these mRNA secondary structures are not fully understood,[15] a few studies have reported that these structures interfere with the process of translation[44,45] and the stability of such structures is higher than expected.[46–48] If highly stable mRNA secondary structures interfere with the process of translation, it is obvious that selective forces will act against such stable secondary structures.[15] But, no foolproof method has been developed for predicting the structure of mRNA molecules as the structures predicted using thermodynamic and comparative methods possess several drawbacks.[15] These methods predict the structures that are best approximations of the mRNA secondary structures present inside the cell and, hence, in general, it is appropriate to use the term "unknown selective forces" to express "seemingly profound influences" that are not explainable on the basis of natural selection and mutational pressure.[15] Note that even extremely biased genes use nonoptimal codons, as usage of optimal codons vary according to the influence of alternative selective forces, depending upon the strength and direction of translational selection.[15] Because the accuracy of the mRNA structures predicted using bioinformatic tools is still questionable, it has been suggested that there may exist selective forces[15] that are yet to be known.

In *A. pectinifera* (Echinodermata; deuterostome) and *B. carnosus* (Hemichordata; deuterostome), compositional constraints favoring mutational pressure (as revealed by high correlation between Axis 1 and silent base contents) and selection for translational accuracy (revealed by the significant correlation between Axis 1 and indexes of level of gene expression, namely, ENC and CAI, respectively) were found to play the major roles in shaping SCUB. In the mitogenome of *T. retusa* (Brachiopoda; deuterostome), apart from mutational pressure, selection might be one of the crucial forces as almost all genes were lying considerably below the expected GC_3 curve and no correlation was observed between Axis 1 and the silent base contents, as well as between GC_3 and GC_{12}. In *K. tunicata* (Mollusca; protostome), *D. pulex* (Arthropoda; protostomes), and *L. terrestris* (Annelida; protostome), the influence of both mutational pressure and selection was noticed. Though mutational pressure due to compositional constraints and selection have been found to influence the SCUB of representative species chosen from protostomes and deuterostomes, the presence of unknown selective forces cannot be neglected as the entire phenomenon of SCUB in these organisms could not be explained either by mutational pressure or by selection favoring translational accuracy.

Cluster analysis based on RSCU values revealed that RSCU variations in the PCGs of the mitogenomes of chaetognaths are more comparable with those in protostomes as all chosen species of Chaetognatha, Mollusca, and Arthropoda formed one major cluster (upper cluster).

The putative optimal codons identified in the mitogenomes of the chosen species of Chaetognatha and those of other phyla could be correlated with the gene expression levels.[30] The presence of putative optimal codons could be suggested as a marker for genes having unknown expression levels.[30] Because the frequency of optimal codons was found to vary in all representative species of Chaetognatha, optimal codon usage was not considered to be conserved among these species.[49] Putative optimal codons are of paramount significance in enhancing heterologous gene expression and, moreover, the a software named Visual Gene Developer improves the expression levels of a synthetic gene by adopting a codon optimization strategy.[50,51] Hence, the identification of putative optimal codons in chaetognaths may help in studies relating to transcriptional control of gene expression as chaetognaths are regarded as good model systems for comparative genomics to study the evolution of animal genomes.[52]

This study was able to identify the trends associated with SCU in the representative species of the minor phylum Chaetognatha and other closely related phyla, such as Hemichordata, Echinodermata, Brachiopoda (deuterostomes), Mollusca, Annelida, and Arthropoda (protostomes). From the results, we understand that some forces other than mutational and selection pressures act on the PCGs in the mitogenomes of all chosen species. The high degree of complexity in the process of speciation of chaetognaths[53] might have invited certain other forces that can interfere with the SCUB. We conclude that the identification of these unknown selective forces would certainly help get insight into the evolution of mitogenomes in chaetognaths.

Author Contributions

Abbreviations

SCUB: synonymous codon usage bias
ENC: effective number of codons
PCG: protein-coding genes
GC_3: GC composition at third codon position
GC_{12}: cumulative GC contents at first and second positions
RSCU: relative synonymous codon usage
COA: correspondence analysis
CAI: codon adaptation index
PR2: parity rule 2.

REFERENCES

1. Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet*. 2008; 42:287–99.
2. Shields DC, Sharp PM. Synonymous codon usage in *Bacillus subtilis* reflects both traditional selection and mutational biases. *Nucleic Acids Res*. 1987;15:8023–40.

3. Shields DC, Sharp PM, Higgins DG, Wright F. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol Biol Evol*. 1988;5:704–16.

4. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics*. 1991;129:897–907.

5. Duret L. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev*. 2002;12:640–9.

6. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985;2:13–34.

7. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. Codon usage between genomes is constrained by genome-wide mutational processes. *Proc Natl Acad Sci*. 2004;101:3480–5.

8. Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res*. 1986; 14:5125–43.

9. Sharp PM, Li WH. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. *Nucleic Acids Res*. 1986;14:7737–49.

10. Sharp PM, Devine KM. Codon usage and gene expression level in *Dictyostelium discoideum*: highly expressed genes do 'prefer' optimal codons. *Nucleic Acids Res*. 1989;17:5029–39.

11. Powell JR, Moriyama EN. Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci U S A*. 1997;94:7784–90.

12. Akashi H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics*. 1994;136:927–35.

13. Stenico M, Lloyd AT, Sharp PM. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res*. 1994;22:2437–46.

14. Maruyama T, Gojobori T, Aota SI, Ikemura T. Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res*. 1986;14:151–3.

15. Stoletzki N. Conflicting selection pressures on synonymous codon use in yeast suggest selection on mRNA secondary structures. *BMC Evol Biol*. 2008;8:224.

16. Botzman M, Margalit H. Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol*. 2011;12:R109.

17. Nair RR, Nandhini MB, Sethuraman T, Doss G. Mutational pressure dictates synonymous codon usage in freshwater unicellular α – cyanobacterial descendant *Paulinella chromatophora* and β – cyanobacterium *Synechococcus elongatus* PCC6301. *Springerplus*. 2013;2:492.

18. Kannaujiya VK, Rastogi RP, Sinha RP. GC constituents and relative codon expressed amino acid composition in cyanobacterial phycobiliproteins. *Gene*. 2014;546(2):162–71.

19. Moriyama EN, Hartl DL. Codon usage bias and base composition of nuclear genes in *Drosophila*. *Genetics*. 1993;134:847–58.

20. Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, *and Arabidopsis*. *Proc Natl Acad Sci U S A*. 1999;96(8):4482–7.

21. Selva KC, Nair RR, Sivaramakrishnan KG, et al. Influence of certain forces on evolution of synonymous codon usage bias in certain species of three basal orders of aquatic insects. *Mitochondr DNA*. 2012;23:447–60.

22. Chen H, Sun S, Norenburg JL, Sundberg P. Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). *PLoS One*. 2014;9:e85631.

23. Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad Sci U S A*. 2004;101:10639–43.

24. Papillon D, Perez Y, Caubit X, Le-Parco Y. Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. *Mol Biol Evol*. 2004;21(11):2122–9.

25. Faure E, Casanova JP. Comparison of chaetognath mitochondrial genomes and phylogenetical implications. *Mitochondrion*. 2006;6:258–62.

26. Nielsen C. Animal Evolution: Interrelationships of the Living Phyla. UK: Oxford University Press; 2012:402.

27. Giribet G, Distel DL, Polz M, Sterrer W, Wheller WC. Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. *Syst Biol*. 2000;49:539–62.

28. Wright F. The "effective number of codons" used in a gene. *Gene*. 1990;87:23–9.

29. RoyChoudhury S, Mukherjee DA. Detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res*. 2010;148:31–43.

30. Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol Biol Rep*. 2009;36:2039–46.

31. Sueoka N. Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A=T and G=C. *J Mol Evol*. 1999; 49:49–62.

32. McInerney JO. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. *Proc Natl Acad Sci U S A*. 1998;95:10698–703.

33. Nair RR, Raveendran NT, Dirisala VR, et al. Mutational pressure drives evolution of synonymous codon usage in genetically distinct *Oenothera* plastomes. *Iran J Biotech*. 2014;12(4):e1156.

34. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Struct Dyn*. 2004;21:527–36.

35. Sharp PM, Li WH. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucl Acid Res*. 1987;15:1281–95.

36. Sablok G, Nayak KC, Vazquez F, Tatarinova TV. Synonymous codon usage, GC_3, and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Mol Biotechnol*. 2011;49:116–28.

37. Zhang J, Wang M, Liu WQ, et al. Analysis of codon usage and nucleotide composition bias in polioviruses. *Virol J*. 2011;8:146.

38. Hammer Q, Harper DAT, Ryan PD. PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron*. 2001;4:1–9.

39. Peden JF. *Analysis of Codon Usage* [Ph. D. thesis]. Nottingham: University of Nottingham; 1999.

40. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.

41. Xia X. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol*. 2013;30:1720–8.

42. Umashankar V, Arunkumar V, Dorairaj S. ACUA: a software tool for automated codon usage analysis. *Bioinformation*. 2007;2:62–3.

43. Liu YS, Zhou JH, Chen HT, et al. The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. *Infect Genet Evol*. 2011;11:1168–73.

44. Doma MK, Parker R. Endonucleotic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*. 2006;440:561–4.

45. Liebhaber SA, Cash F, Eshleman SS. Translation inhibition by an mRNA coding region secondary structure is determined by its proximity to the AUG initiation codon. *J Mol Biol*. 1992;226:609–21.

46. Seffens W, Digby D. mRNAs have greater negative folding free energies than shuffled or codon choice randomized sequences. *Nucleic Acids Res*. 1999; 27:1578–84.

47. Katz L, Burge CB. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res*. 2003;13:2042–51.

48. Chamary JV, Hurst LD. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol*. 2005;6:R75.

49. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*. 1980;8:49–62.

50. Jung SK, McDonald K. Visual gene developer: a fully programmable bioinformatics software for synthetic gene optimization. *BMC Bioinformatics*. 2011;12:340.

51. Wang X, Wu L, Zhou P, et al. Analysis of synonymous codon usage patterns in the genus *Rhizobium*. *World J Microbiol Biotechnol*. 2013;29:2015–24.

52. Marletaz F, Gilles A, Caubit X, et al. Chaetognath transcriptome reveals ancestral and unique features among bilaterians. *Genome Biol*. 2008;9:R94.

53. Jennings RM, Bucklin A, Pierrot-Bults A. Barcoding of arrow worms (Phylum Chaetognatha) from three oceans: genetic diversity and evolution within an enigmatic phylum. *PLoS One*. 2010;5:e9949.