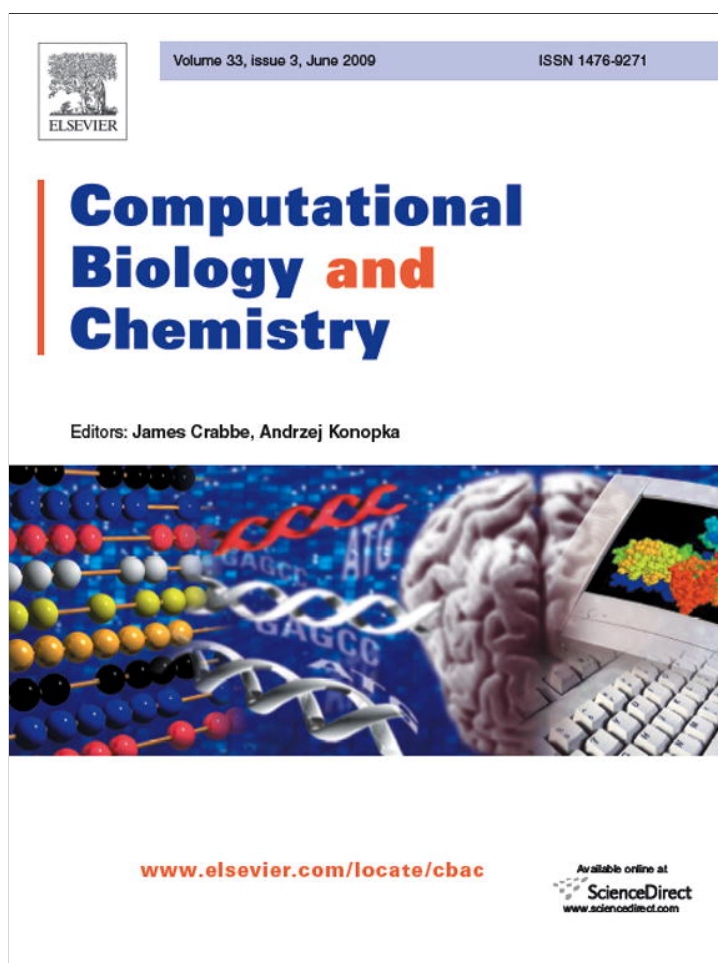


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/combiolchem

Brief Communication

Ambush hypothesis revisited: Evidences for phylogenetic trends

Tiratha Raj Singh^{a,b,*}, Kamal Raj Pardasani^c^a Department of Zoology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel^b Department of Bioinformatics, MANIT, Bhopal 462007, India^c Department of Mathematics, MANIT, Bhopal 462007, India

ARTICLE INFO

Article history:

Received 28 August 2008

Received in revised form 15 April 2009

Accepted 23 April 2009

Keywords:

Ambush hypothesis

Frame-shifting

Hidden stops

Codon usage

Phylogenetic trends

ABSTRACT

Recoding events occur in competition with standard readout of the transcript, and are site-specific. Recoding is the reprogramming of mRNA translation by localized alterations in the standard translational rules. Frame-shifting is one class of recoding and defined as protein translations that start not at the first, but either at the second (+1 frame-shift) or the third (−1 frame-shift) nucleotide of the codon. Coding sequences lack stop codons, but frame-shifted sequences contain many stop codons, termed off-frame stops or hidden stops. These hidden stops terminate frame-shifted translation, potentially decreasing energy, and resource waste on non-functional proteins. Our results support this putative ancient adaptive event for the selection of codons that can be part of hidden stop codons. All taxonomic groups represent positive correlation between codon usage frequencies and contribution of codons to hidden stops in off-frame context. Our analysis on nuclear and mitochondrial genomic data revealed phylogenomic selection of ambush mechanism. Strongest impact of this event was found in viruses and bacteria. It has been suggested that this mechanism has occurred and been utilized in the early stages of evolution.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Recoding is the reprogramming of mRNA translation by localized alterations in the standard translational rules. Recoding events occur in competition with standard readout of the transcript, and are site-specific. The efficiency of recoding at the site is usually influenced by stimulatory signals present on the mRNA (*cis*-elements) and in some cases by protein products or other cellular components (*trans*-elements) (Miller et al., 1997; Baranov et al., 2001). The three classes of recoding are (1) frame-shifting (2) bypassing (Hopping) and (3) codon redefinition (Gesteland and Atkins, 1996; Farabaugh, 1996).

Frame-shifting at a particular site can yield two protein products from one coding sequence or one protein product from two overlapping open reading frames (ORFs). The known cases of frame-shifting where the product is utilized involve shifts of one base either +1 or −1, but shifts of two bases have been demonstrated in artificial systems.

Bypassing (hopping) occurs when a block of nucleotides within a coding sequence is not translated. Translation is suspended temporarily, ribosomes traverse the coding gap and protein synthesis resumes yielding a single protein.

Codon redefinition involves site-specific alteration of codon meaning. The altered meaning of a codon can be the redefinition of an initiation codon or stop codon to specify an amino acid.

Frame-shifts are defined as protein translations that start not at the first, but either at the second (+1 frame-shift) or the third (−1 frame-shift) nucleotide of the codon. Presumably, most frame-shifts would yield non-functional proteins. Therefore frame-shifts lead to waste of energy, resources and activity of the biosynthetic machinery. In addition, some peptides synthesized after frame-shifts are probably cytotoxic.

Coding sequences lack stop codons, but many stop codons appear off-frame. Off-frame stops, i.e. stop codons in +1 and −1 shifted reading frames, are termed hidden stop codons or hidden stops. The 'ambush' hypothesis (Seligmann and Pollock, 2004) on the hidden stop codons and their role in off-frame (incorrect) gene reading suggests that hidden stop codons prevent off-frame gene reading and there is an advantage in using codons that can be part of hidden stop codons. The ambush hypothesis implies that hidden stop codons are sometimes selected and this selection is reflected in sequences. These hidden stops terminate frame-shifted translation, potentially decreasing energy, and resource waste on non-functional proteins.

There are several occurrences of codon reassignments, premature stop codons, and read through stop codons, and several codon usage biases in protein coding sequences at various taxonomic levels (Tuite and McLaughlin, 1982; Lovett et al., 1991; Castresana et al., 1998; Bidou et al., 2004; Brooks et al., 2006; Henry and Sharp, 2007;

* Corresponding author at: Department of Zoology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel. +972 527246022.

E-mail address: tiratharaj@gmail.com (T.R. Singh).

Table 1
The numbers of +1 and –1 hidden stop codons for genetic codes with different stop codon assignments.

Stop codons	+1	Seligmann and Pollock (2004) +1	–1	Seligmann and Pollock (2004) –1	Genetic system
TAG	19	–	19	–	Alternative flatworm mt
TGA	19	–	19	–	Nuclear ciliate, dasycladacean and hexamita
TAA TAG	34	–	22	–	Nuclear euplotid; mt ascidian, echinoderm, flatworm, mold, trematode, yeast, invertebrates
TAA TGA	22	–	22	–	Nuclear Blepharisma, mt chlorophycean
TAA TAG TGA	36	–	25	22	Standard
TAA TCA TGA	25	–	25	–	mt <i>Scenedesmus obliquus</i>
TAA TTA TCA TGA	28	–	44	–	mt Thraustochytrium
AGA AGG TAA TAG	36	42	40	46	mt vertebrates

Columns 2–5 show the number of codons that can contribute to hidden stops in +1 and –1 frame-shift, respectively. The last column indicates the genetic system in which the genetic code is found.

Morton and Wright, 2007). This ambiguous role of stop codons, codon reassignments and its implications during the course of evolution is an interesting research topic.

In this study we have analyzed and evaluated evidences for, and against the ambush hypothesis. Computational and statistical analyses have been performed to resolve the ambiguous role of hidden stops and their evolutionary implications and applications. In particular, here we ask whether there is a statistically significant correlation between average codon usage and their potential contribution to hidden stops? We also considered whether this scenario is unique to certain taxonomic groups, i.e. is there an existence of any phylogenetic trends? Large data sets from several taxonomic groups of organisms such as bacteria, viruses, bacteriophages, invertebrates, vertebrates, and plants have been subject to detailed statistical and computational analyses.

2. Materials and Methods

The genomic data for codon usage has been downloaded from the codon usage database (<http://www.kazusa.or.jp/codon/>; Nakamura et al., 2000). Assignment of hidden stops in frame-shifted context has been calculated according to the respective genetic code (Table 1). Average usage of codons that can be part of off-frame hidden stops, for nuclear, mitochondrial (mt), and chloroplast genomic data, has been calculated according to the respective genetic codes (Table 1). The coefficient of correlation (r) has been calculated between hidden stops contribution to codons and their usage frequencies. Binomial and chi square (χ^2) distributions have been employed with 5% level of significance for statistical support of analyses and robustness of the results.

3. Results and Discussion

3.1. Alternative Genetic Codes, Stop Codons Assignments and Selection Pressure

Discoveries of the deviations of the universal genetic code in prokaryotes, eukaryotic nuclear and mitochondrial genomes have revealed that the genetic code is still evolving despite strong negative forces working against the fixation of mutations that result in codon reassignment (Santos et al., 2004). These rather unexpected discoveries prompted researchers to investigate both the underlying molecular mechanism and the putative evolutionary forces that might drive the diversification of genetic code assignments and codon reassignments.

All known non-standard genetic codes appears to be secondarily derived minor modifications of the canonical code. The arrangement of amino acid assignments to the codons of the standard genetic code appears to be a direct product of natural selection for a system that minimizes the phenotypic impact of genetic error. Known variants of the standard genetic code appear less adaptive,

suggests that different evolutionary factors predominated before and after fixation of the canonical code. It has been suggested that universal genetic code can efficiently carry arbitrary parallel codes in a much better way than the alternative genetic codes (Freeland and Hurst, 1998; Freeland et al., 2000; Giulio, 2005; Itzkovitz and Alon, 2007).

Statistical and biochemical studies of the genetic code have found evidences of non-random patterns in the distribution of codon assignments. These non-random patterns can be either the assignment of amino acids to codons appears to be optimal for minimizing the effect of translational misread errors; or amino acids with less complex chemical structure have more codon assignment tendency towards themselves. Although the code still evolves today, as reflected by its departure from universality in some organisms, but limited and concerns only reassignments of a few codons (Gillis et al., 2001; Knight et al., 2001; Archetti, 2004).

We first considered stop codon assignments in standard and alternative genetic codes. Stop codon assignments for +1 and –1 frame-shifts have been observed in known genetic code variants (Elzanowski et al., 2000). For genetic codes having only one stop codon like alternative flatworm mt code, and ciliate, dasycladacean and hexamita nuclear code, 19 stop codons have been assigned in +1 and –1 frame-shifts each. The number of off-frame stop codons assignment increases in genetic codes with multiple stop codons (Table 1). Codon count for a new genetic code system of *Thraustochytrium* mt has also been included (Table 1). We found some differences with Seligmann and Pollock (2004), in frame-shifted codon assignments in some genetic systems, shown in column 3 and 5 (Table 1).

Several studies have demonstrated translational selection in codon usages. General agreement is that codons are translated at different rates (Kurland, 1991; Fernández et al., 2001). The first indication of non-uniform translation rates was the observation that there are pauses during polypeptide elongation and that these can be identified with short strings of rarely used codons (Randall et al., 1980; Varenne and Lazdunski, 1986). As soon as a significant number of genes had been sequenced, it became an accepted opinion that biased codon usage could regulate the expression levels of individual genes by modulating the rates of polypeptide formation (Chavancy and Garel, 1981; Gouy and Gautier, 1982; Robinson et al., 1984; Henry and Sharp, 2007).

A translational frame-shift event is estimated to occur at a probability of about 3.3×10^{-5} codons (Parker, 1989; Farabaugh and Björk, 1999). Different codons contribute to hidden stops in different ways. Some contribute up to six times and some cannot become part of a stop codon assignment in any frame-shift scenario (+1 or –1). In case of the standard genetic code, AAU (asparagine) and ACU (serine), contribute six times to stop codon assignment, AUU (isoleucine), ACU (threonine), and GAU (aspartic acid) contribute five times to stop codon assignment, whereas 20 codons cannot contribute to stop codon assignment in any frame-shifted conditions. In vertebrate mt genetic code, only GUA (valine) contributes

Table 2

Statistical analyses of different taxonomic groups between codon usage frequencies and contribution of codons to hidden stops in off-frame context, according to the standard, bacterial, and alternative yeast nuclear genetic codes.

Category	Total number of organisms	Organisms with positive correlation	% of data with positive correlation from total number of organisms	Significant positive correlation ($P < 0.05$)	Binomial distribution (P)
Bacteria	4495	2360	52.5	318	0
Virus	3457	2844	82.2	326	0
Bacteriophages	692	573	82.8	85	1.9762E-14
Invertebrates	1724	1139	66	85	0.000134014
Plants	2572	1634	63.5	84	0.36954
Vertebrates	876	322	36.8	2	0.999989183
Primates	186	41	22	0	0.877913
Rodents	122	27	22.1	0	0.749656
Other mammals	344	69	20	1	0.86552

six times to stop codon assignment, 10 codons contribute five times, and no codon falls in the category of one stop codon assignment in frame-shifted conditions. It is believed that the number of peptide bonds produced per unit time is one of the main selection pressures in growing microorganisms. There are observations of reduction in length of frame-shifted peptides, which can also have beneficial effect such as reducing potential toxicity of the non-sense peptides (Parker, 1989; Woese, 1998; Farabaugh and Björk, 1999; Dekel and Alon, 2005; Wagner, 2005).

Each genetic code is supposed to have n -mer sequences, termed 'difficult' sequences, which contain stop codons and thus cannot appear in one of three frames. It has been observed that real genetic code is able to include even the most difficult n -mers, because of special property: 'its stop codons when frame-shifted tend to form abundant codons'. The optimality of the genetic code seems to increase with the length of the n -mers. This is because as the length of the n -mers increases, the fraction of n -mers that include stop codons increases dramatically (Parker, 1989; Farabaugh and Björk, 1999; Seligmann and Pollock, 2004; Itzkovitz and Alon, 2007). One possibility for this is the ability to include parallel codes within protein-coding sequences conferred a selection advantage during the early evolution of genetic code. Alternatively, the genetic code might have been fixed in such a way in early stages of evolution before most parallel codes existed.

3.2. Codon Usage Biases and Compositional Aspects: Selectionist Versus Neutralist View

Kimura's revolutionary proposal "the main cause of evolutionary change at the molecular level – changes in the genetic material itself – is random fixation of selectively neutral or nearly neutral mutants", started a neutralist/selectionist debate, which is still going on (Kimura, 1968, 1983).

Various evolutionary theories like the "codon disappearance" or "codon capture" theory, and "ambiguous intermediate" theory have emerged that could explain how the meaning of a codon could be changed without the extinction of the species. The "codon capture" theory is based on the observed genomic bias against GC content in an AT-rich genome, can drive GC containing codons to extinction. The "codon capture" theory proposes that specific codons, made rare by AT or GC pressure, disappear from the genome entirely. Any mutation to the tRNAs that translate these codons will be allowed, since such mutations will be neutral. If the mutation pressure reverses, causing these codons to reappear, they may now code for a different amino acid. Thus the process of codon reassignment can be 'entirely neutral' (Muto and Osawa, 1987; Osawa and Jukes, 1989; Schultz and Yarus, 1994a,b; Wan and Wootton, 2000).

The "ambiguous intermediate" theory is based on ambiguous decoding of a single codon by both of cognate and mutant tRNAs. Same mechanism might also apply to reassignment of a stop codon

to a sense codon. Such codon reassignments might lead to the synthesis of an aberrant, non-functional polypeptide chain, and a misfolded protein. This theory has been supported through the discovery of some codon-specific ambiguous decoding in *Candida* species. Several occurrences of codon disappearance have been observed. Once a codon has vanished from the genome of an organism, it can potentially re-emerge through genetic drift. Genetic drift is a change in the frequency of an allele in a population occurring by chance and in the absence of any evolutionary selection against that allele (Suzuki et al., 1997; Santos et al., 1999, 2004).

In organisms where large proportion of the genome consists of coding sequences, selective bias acting on a local scale can potentially influence global nucleotide composition. There are two ways of looking at the evolutionary changes of nucleotide compositional aspects; the selectionist and the neutralist point of view. These hypotheses differ in the estimate of the role of selection on base substitution. According to selectionist hypothesis, isochors are the result of positive selection for GC content as an adaptation to the high body temperature in warm blooded vertebrates. Mutational hypothesis is based on directional mutation pressure, and regarded as neutral rather than selective (Bernardi et al., 1985; Bernardi, 2007; Sueoka, 1988; Frank and Lobry, 1999). Ambush hypothesis (Seligmann and Pollock, 2004) and some recent studies (Itzkovitz and Alon, 2007), supported the selectionist view and suggested that there is selection in favor of some combination of codons for which frame-shift mutations will lead to a stop codon.

3.2.1. Correlation Between Hidden Stop Codons and their Usage

We have analyzed the correlation between codon usage frequencies (%) and contribution of codons to hidden stops in off-frame context, for nuclear, mt and chloroplast genomic data, according to the respective genetic codes (Table 1), for various taxonomic groups such as viruses, archaea, bacteria and metazoa. Correlation was strongest among viruses and bacteriophages with more than 80% of the genomic data in these taxonomic groups indicate positive correlation. Among invertebrates, plants, and bacteria correlations were also positive in more than 50% of the genomic data analyzed (Table 2).

Most organisms use a preferred set of codons, and selection acting on codon choice could create local asymmetries between coding and non-coding strands. Biased codon usage has been explained as the result of selection at translational level. It is well known that the bias of codon usage in bacteria is related to the level of expression. It has been suggested that codon usage may be biased toward codons that can form hidden stops. Real genetic code encounters a stop more rapidly on average than 99.3% of the alternative codes. Additionally it has been shown that real code has the capability to abort translation eight codons earlier than the average alternative code. Relative fitness advantage of real code upon alternative codes has been suggested a selection consequence (Gouy and Gautier, 1982;

Table 3
Statistical analyses of different taxonomic groups between codon usage frequencies and contribution of codons to hidden stops in off-frame context, according to the invertebrate, yeast and vertebrate mitochondrial codes, and plant plastid codes.

Category	Total number of organisms	Organisms with positive correlation	% of data with positive correlation from total number of organisms	Significant positive correlation ($P < 0.05$)	Binomial distribution (P)
Invertebrates	1099	1034	94.09	147	0
Plant plastids	260	257	98.85	2	0.999801
Chloroplasts	6470	6450	99.70	458	1.07025E–13
Yeasts	35	35	100	10	6.63817E–07
Vertebrates	4658	2685	57.6	0	1
Primates	186	181	97.3	0	0.999907109
Rodents	606	592	97.7	0	1
Other mammals	1044	981	94	0	1

Sharp and Matassi, 1994; Sueoka, 1995; Rocha et al., 1998; Frank and Lobry, 1999; Seligmann and Pollock, 2004; Itzkovitz and Alon, 2007).

3.2.2. Significance of Correlation and Taxonomic Bias

Ambush hypothesis implies that at the level of genetic codes, ancient adaptive events may have adjusted codon assignments to increase frequencies of codons that can be part of hidden stop codons (Seligmann and Pollock, 2004; Itzkovitz and Alon, 2007). Assuming that hidden stop codons are uncorrelated with the average codon usage frequencies (*null hypothesis*), we expect 0.05 of the organisms tested to show significant correlation between hidden stop codon and codon usage frequencies. Thus, to test whether the *null hypothesis* holds we performed the binomial test.

Our results (Tables 2 and 3) support this putative ancient adaptive event for the selection of codons that can be part of hidden stop codons. All taxonomic groups represent statistically significant positive correlation ($P < 0.05$, one-tailed test) between codon usage frequencies and contribution of codons to hidden stops in off-frame context (Table 2). Further analysis with mitochondrial genomic data also supports this fact as most of the correlations were positive from vertebrates, primates, rodents and other mammals according to vertebrate mitochondrial genetic code (Table 3).

For bacteria, viruses, bacteriophages, and invertebrates we found that indeed a substantial number of genomes display a significant positive correlation between hidden stop codons and codon usage frequencies ($P < 0.05$; binomial test). In contrast, in all other taxonomic groups tested such as vertebrates, primates, rodents, and other mammals, this result was neither obtained for nuclear nor for mitochondrial genomes (Tables 2 and 3). It has been proposed that this putative adaptive event might have occurred in the early stages

of evolution and ancient adaptive events may have adjusted codon assignments towards hidden stops.

3.2.3. Codon Usage and Independence Analysis of Significant Correlation: Phylogenetic Trends

In some recent studies on overlapping genes and related evolutionary aspects, it has been suggested that genomes come under a variety of different mutational and selectional pressures. There are more than one evolutionary mechanisms account for all the variability and biases in genomes. Mutational and selectional constraints have been accepted as the major cause of such evolutionary processes. Two kind of benefits have been suggested which influence the overall performance of biosynthetic machinery (immediate benefits) and influence the phenotypic variation in population and modulate the strength of selection on a lineage (variational benefits). Later are indicative of phylogenomic selection of such processes (Rogozin et al., 2002; Fukuda et al., 2003; Johnson and Chisholm, 2004; Makalowska et al., 2005; Belshaw et al., 2007; Cock and Whitworth, 2007; Lillo and Krakauer, 2007; Sabath et al., 2008).

An interesting outcome of the hidden stops usage analyses is the occurrence of taxonomic biases with reference to significant positive correlation. Unicellular, invertebrates and plants genomes manifest such strong, statistically significant positive correlation, whereas, none of the vertebrates, mammals, primates and rodents genomes. Similarly for mt genomes significant correlation exists only in invertebrates, yeasts, and plant chloroplasts.

In order to statistically evaluate this taxonomic bias and to estimate a plausible phylogenomic trend we categorize our data into two classes, with and without significant positive correlation between hidden stop codons and codon usage frequencies. We

Table 4
Test for independence of variables from different levels of taxonomic categories data sets. Three taxonomic levels have been checked for significant correlation of codon usage frequency and contribution of codons to hidden stops.

Taxonomic category	Taxonomic level	χ^2	Degrees of freedom	χ^2 distribution (P)
Bacteria	Phylum	42.52649	3	3.10195E–09
	Class	91.79372	6	1.2844E–17
	Family	70.24941	9	1.3605E–11
Viruses	Phylum	–	–	–
	Class	25.69775	5	0.000102122
	Family	67.71966	25	8.34509E–06
Bacteriophages	Phylum	–	–	–
	Class/Category	10.15644	2	0.00623099
	Family	19.84658	4	0.000535455
Invertebrates	Phylum	16.37914	8	0.037263981
	Class	57.31249	10	1.16438E–08
	Family	15.66719	9	0.074165287
Plants	Phylum	311.5995	4	3.40701E–66
	Class	223.5239	3	3.47483E–48
	Family	48.80149	6	8.17009E–09

employed chi-square (χ^2) test for independence of variables on two categories of our data set, based on significance level (Table 2). The null hypothesis is that the two variables are independent, i.e. the number of organisms with and without significant correlation.

We analyzed data from several taxonomic groups such as bacteria, viruses, bacteriophages, invertebrates and plants for different taxonomic levels such as phylum, class and families where significant positive correlation exists. The *P*-value is less than the significance level (0.05) in almost all cases (Table 4) therefore; we rejected the null hypothesis and conclude that the two categories of positively correlated organisms are interrelated.

Outcome of chi-square (χ^2) test indicates biased distribution of hidden stops selection in diverse taxonomic groups and levels. Also results are statistically significant ($P < 0.05$; Table 4). The distribution of hidden stops usage has been limited to specific classes, families, phyla, in analyzed taxonomic groups, where correlations were positive. Based on this analysis we propose the phylogenomic selection of ambush mechanism.

Recently it has been proposed that mutation pattern is a major player in determining frame-shift stop codon frequencies rather than selection. Such conclusion has been proposed in an extensive study with reference to overlapping genes (Rogozin et al., 2002; Fukuda et al., 2003; Johnson and Chisholm, 2004; Makalowska et al., 2005; Belshaw et al., 2007; Cock and Whitworth, 2007; Lillo and Krakauer, 2007; Sabath et al., 2008). Similarly, previous studies on microsatellite (bi-, tri-, tetra-, penta-, and hexa-nucleotide) repeats proposed that these repeats arise from specific selection against frame-shift mutations. Additionally, the excess of tri- and hexa-nucleotide coding repeats have been suggested to be controlled primarily by mutation pressure (Cox and Mirkin, 1997; Metzgar et al., 2000).

These data collectively illustrate how selection and mutation pressures could be examined and employed to identify several evolutionary aspects of genomic data. Phylogenomic selection of hidden stops is just one such mechanism. Such controversial hypotheses on adaptive-selective and neutral aspects indicates that there is a need to carry out extensive analyses on compositional factors such as length distribution of overlaps, proper use of start and stop codons in all phases, frequency of regulatory elements, microsatellite repeats, and other related evolutionary aspects, with the availability of additional genomic data.

4. Final Remarks

We propose the phylogenomic selection of hidden stops based on our statistical analyses on nuclear and mitochondrial genomic data. It has been found that hidden stops have been under positive selection for minimization of frame-shifted translational errors. Strongest impact of this event was found in viruses and bacteria. Invertebrates and plants also indicated a phylogenetic trend up to some extent. Least impact was observed in several categories of vertebrates, indicates, that this ancient putative event has been transpired and been utilized in very early stages of the evolution. However, some recent controversial hypotheses put an effort to perform an all-embracing neutralist/selectionist debate on hidden stops, and their evolutionary implications.

Acknowledgements

We thank Dr. Tal Pupko for suggesting this problem. We acknowledge Nimrod Rubinstein for help during statistical analyses. We thank Dr. A.K. Konopka and the three reviewers for their insightful comments.

References

- Archetti, M., 2004. Codon usage bias and mutation constraints reduce the level of error minimization of the genetic code. *J. Mol. Evol.* 59 (2), 258–266.
- Baranov, P.V., Gurvich, O.L., Fayet, O., Prere, M.F., Miller, W.A., Gestel, R.F., Atkins, J.F., Giddings, M.C., 2001. RECODE: a database of frame-shifting, bypassing and codon redefinition utilized for gene expression. *Nucleic Acids Res.* 29, 264–267.
- Belshaw, R., Pybus, O.G., Rambaut, A., 2007. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* 17, 1496–1504.
- Bernardi, G., 2007. The neoselectionist theory of genome evolution. *Proc. Natl. Acad. Sci.* 104 (20), 8385–8390.
- Bernardi, G., Olofsson, B., et al., 1985. The mosaic genome of warm-blooded vertebrates. *Science* 228, 953–957.
- Bidou, L., Hatin, I., Perez, N., Allamand, V., Panthier, J.J., Rousset, J.P., 2004. Premature stop codons involved in muscular dystrophies show a broad spectrum of readthrough efficiencies in response to Gentamicin treatment. *Gene Ther.* 11, 619–627.
- Brooks, D.A., Muller, Viv, J., Hopwood, J.J., 2006. Stop-codon read-through for patients affected by a lysosomal storage disorder. *Trends Mol. Med.* 12/8, 367–373.
- Castresana, J., Feldmaier-Fuchs, G., Paabo, S., 1998. Codon reassignment and amino acid composition in hemichordate mitochondria. *Proc. Natl. Acad. Sci.* 95, 3703–3707.
- Chavancy, G., Garel, J.P., 1981. Does quantitative tRNA adaptation to codon content in mRNA optimize the ribosomal translation efficiency? Proposal for a translation system model. *Biochimie* 63/3, 187–195.
- Cock, P.J.A., Whitworth, D.E., 2007. Evolution of gene overlaps: relative reading frame bias in prokaryotic two-component system genes. *J. Mol. Evol.* 64, 457–462.
- Cox, R., Mirkin, S.M., 1997. Characteristic enrichment of DNA repeats in different genomes. *Proc. Natl. Acad. Sci.* 94, 5237–5242.
- Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436, 588–592.
- Elzanowski, A., Ostell, J., Leipe, D., Sousoff, V., 2000. The genetic codes. NCBI. <http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c>.
- Farabaugh, P.J., 1996. Programmed translational frame-shifting. *Annu. Rev. Genet.* 30, 507–528.
- Farabaugh, P.J., Björk, G.R., 1999. How translational accuracy influences reading frame maintenance. *EMBO J.* 18 (6), 1427–1434.
- Fernández, V., Zavala, A., Musto, H., 2001. Evidence for translational selection in codon usage in *Echinococcus* spp. *Parasitology* 123, 203–209.
- Frank, A.C., Lobry, J.R., 1999. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* 238, 65–77.
- Freeland, S.J., Hurst, L.D., 1998. The genetic code is one in a million. *J. Mol. Evol.* 47, 238–248.
- Freeland, S., Knight, R.D., Landweber, L.F., Hurst, L.D., 2000. Early fixation of an optimal genetic code. *Mol. Biol. Evol.* 17 (4), 511–518.
- Fukuda, Y., Nakayama, Y., Tomita, M., 2003. On dynamics of overlapping genes in bacterial genomes. *Gene* 323, 181–187.
- Gesteland, R.F., Atkins, J.F., 1996. Recoding, dynamic reprogramming of translation. *Annu. Rev. Biochem.* 65, 741–768.
- Gillis, D., Massar, S., Cerf, N.J., Rooman, M., 2001. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* 2 (11), 0049.1–0049.12.
- Gouy, M., Gautier, C., 1982. Codon usage in bacteria, correlation with gene expressivity. *Nucleic Acids Res.* 10/22, 7055–7074.
- Giulio, M.D., 2005. The origin of the genetic code: theories and their relationships, a review. *Bio Syst.* 80, 175–184.
- Henry, I., Sharp, P.M., 2007. Predicting gene expression level from codon usage bias. *Mol. Biol. Evol.* 24/1, 10–12.
- Itzkovitz, S., Alon, U., 2007. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. *Genome Res.* 17, 405–412.
- Johnson, Z.I., Chisholm, S.W., 2004. Properties of overlapping genes are conserved across microbial genomes. *Genome Res.* 14 (11), 2268–2272.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–625.
- Kimura, M., 1983. *The Neutral Theory of Evolution*. Cambridge University Press, Cambridge, UK.
- Knight, R.D., Freeland, S.J., Landweber, L.F., 2001. Rewriting the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* 2, 49–58.
- Kurland, C.G., 1991. Codon bias and gene expression. *FEBS* 285/2, 165–169.
- Lillo, F., Krakauer, D.C., 2007. A statistical analysis of the three-fold evolution of genomic compression through frame overlaps in prokaryotes. *Biol. Direct* 2, 22.
- Lovett, P.S., Ambulos Jr, N.P., Mulbry, W., Noguchi, A., Rogers, E.J., 1991. UGA can be decoded as Tryptophan at low efficiency in *Bacillus subtilis*. *J. Bacteriol.* 173/5, 1810–1812.
- Makalowska, I., Lin, C.F., Makalowski, W., 2005. Overlapping genes in vertebrate genomes. *Comp. Biol. Chem.* 29, 1–12.
- Metzgar, D., Bytof, J., Wills, C., 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.* 10, 72–80.
- Miller, W.A., Brown, C.M., Wang, S., 1997. New punctuation for the genetic code, *Luteovirus* gene expression. *Semin. Virol.* 8, 3–13.
- Morton, B.R., Wright, S.I., 2007. Selective constraint in codon usage of nuclear genes from *Arabidopsis thaliana*. *Mol. Biol. Evol.* 24/1, 122–129.
- Muto, A., Osawa, S., 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci.* 84, 166–169.
- Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon usage tabulated from the international DNA sequence databases, status for the year 2000. *Nucleic Acids Res.* 28/1, 292.

- Osawa, S., Jukes, T.H., 1989. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* 28, 271–278.
- Parker, J., 1989. Errors and alternatives in reading the universal genetic code. *Microbiol. Rev.* 53 (3), 273–298.
- Randall, L.L., Josefsson, L.G., Hardy, S.J.S., 1980. Novel intermediates in the synthesis of maltose-binding protein in *Escherichia coli*. *Eur. J. Biochem.* 107, 375–379.
- Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., Humphreys, G., 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res.* 12/17, 6663–6671.
- Rocha, E.P.C., Danchin, A., Viari, A., 1998. Universal replication biases in bacteria. *Mol. Microbiol.* 32, 11–16.
- Rogozin, I.B., Spiridonov, A.N., Sorokin, A.V., Wolf, Y.I., Jordan, I.K., Tatusov, R.L., Koonin, E., 2002. Purifying and directional selection in overlapping prokaryotic genes. *Trends Genet.* 18 (5), 228–232.
- Sabath, N., Graur, D., Landan, G., 2008. Same-strand overlapping genes in bacteria: compositional determinants of phase bias. *Biol. Direct* 3, 36.
- Santos, M.A.S., Cheesman, C., Costa, V., Morads-Ferreira, P., Tuite, M.F., 1999. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in *Candida* spp. *Mol. Microbiol.* 31/3, 937–947.
- Santos, M.A.S., Moura, G., Massey, S.E., Tuite, M.F., 2004. Driving change: the evolution of alternative genetic codes. *Trends Genet.* 20/2, 95–102.
- Schultz, D.W., Yarus, M., 1994a. Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.* 235/5, 1377–1380.
- Schultz, D.W., Yarus, M., 1994b. tRNA structure and ribosomal function. I. tRNA nucleotide 27–43 mutations enhance first position wobble. *J. Mol. Biol.* 235/5, 1381–1394.
- Seligmann, H., Pollock, D.D., 2004. The ambush hypothesis: hidden stop codons prevent off-frame gene reading. *DNA Cell Biol.* 23, 701–705.
- Sharp, P.M., Matassi, G., 1994. Codon usage and genome evolution. *Curr. Opin. Genet. Dev.* 6, 851–860.
- Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci.* 48, 2653–2657.
- Sueoka, N., 1995. Intrastrand parity rules of DNA base composition and usage biases synonymous codons. *J. Mol. Evol.* 40, 318–325 (erratum: 42, 323).
- Suzuki, T., Ueda, T., Watanabe, K., 1997. The 'polysemous' codon, a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J.* 16, 1122–1134.
- Tuite, M.F., McLaughlin, C.S., 1982. Endogenous read through of a UDA termination codon in a *Saccharomyces cerevisiae* cell-free system, evidence for involvement of both a mitochondrial and a nuclear tRNA. *Mol. Cell Biol.* 2/5, 490–497.
- Varenne, S., Lazdunski, C., 1986. Effect of distribution of unfavorable codons on the maximum rate of gene expression by an heterologous organism. *J. Theor. Biol.* 120/1, 99–110.
- Wagner, A., 2005. Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* 22 (6), 1365–1374.
- Wan, H., Wootton, J.C., 2000. A global compositional complexity measure for biological sequences: AT-rich and GC-rich genomes encode less complex proteins. *Comp. Chem.* 24, 71–94.
- Woese, C., 1998. The universal ancestor. *Proc. Natl. Acad. Sci.* 95, 6854–6859.