Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 345 (2005) 13–20

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes

Xuhua Xia*

Department of Biology, University of Ottawa, 150 Louis, P.O. Box 450, Station A, Ottawa, Ontario, Canada, K1N 6N5

Received 14 July 2004; received in revised form 14 October 2004; accepted 7 November 2004
Available online 19 December 2004
Received by U. Bastolla

Abstract

The H-strand of vertebrate mitochondrial DNA is left single-stranded for hours during the slow DNA replication. This facilitates C→U mutations on the H-strand (and consequently G→A mutations on the L-strand) via spontaneous deamination which occurs much more frequently on single-stranded than on double-stranded DNA. For the 12 coding sequences (CDS) collinear with the L-strand, NNY synonymous codon families (where N stands for any of the four nucleotides and Y stands for either C or U) end mostly with C, and NNR and NNN codon families (where R stands for either A or G) end mostly with A. For the lone ND6 gene on the other strand, the codon bias is the opposite, with NNY codon families ending mostly with U and NNR and NNN codon families ending mostly with G. These patterns are consistent with the strand-specific mutation bias. The codon usage biased towards C-ending and A-ending in the 12 CDS sequences affects the codon–anticodon adaptation. The wobble site of the anticodon is always G for NNY codon families dominated by C-ending codons and U for NNR and NNN codon families dominated by A-ending codons. The only, but consistent, exception is the anticodon of tRNA-Met which consistently has a 5′ -CAU-3′ anticodon base-pairing with the AUG codon (the translation initiation codon) instead of the more frequent AUA. The observed CAU anticodon (matching AUG) would increase the rate of translation initiation but would reduce the rate of peptide elongation because most methionine codons are AUA, whereas the unobserved UAU anticodon (matching AUA) would increase the elongation rate at the cost of translation initiation rate. The consistent CAU anticodon in tRNA-Met suggests the importance of maximizing the rate of translation initiation.

Keywords: Codon–anticodon adaptation; Anticodon loop; tRNA; Deamination; Strand asymmetry

1. Introduction

Since the discovery of the correlation between codon usage and tRNA abundance in *Escherichia coli* (Ikemura, 1981; Gouy and Gautier, 1982) and *Saccharomyces cerevisiae* (Bennetzen and Hall, 1982), much progress has been made in understanding codon usage and codon–anticodon adaptation (Bulmer, 1987, 1991) in the context of maximizing transcription and translation rates (Eyre-Walker, 1996; Xia, 1996, 1998; Akashi, 2003). However, the role played by mutation has been neglected to such an extent that almost all publications consider mutation as disruptive to the evolution and maintenance of codon usage bias and the associated codon–anticodon adaptation (Akashi, 1995; Berg and Martelius, 1995; Berg, 1996; Xia, 1996; Akashi, 1997). In other words, while selection is supposed to be the main force driving and maintaining the evolution of synonymous codons towards maximizing the codon that matches the anticodon of the most abundant tRNA, mutation is thought to reduce codon usage bias and disrupt codon–anticodon adaptation and is invoked whenever one fails to see strong codon usage bias or codon–anticodon adaptation. The relative abundance of different tRNA species is often, albeit implicitly, taken as prefixed, and this tRNA bias then drives codon usage bias. To my knowledge, in spite of studies on the effect of mutation spectrum on GC content and amino acid usage (Sueoka,

Abbreviations: mtDNA, mitochondrial DNA; CDS, coding sequence; AA, amino acid; rNTP, ribonucleotides; AC, anticodon.

* Tel.: +1 613 562 5800x6886; fax: +1 613 562 5486.
E-mail address: xxia@uottawa.ca.

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.11.019

1961; Lobry, 2004), there has been no empirical documentation of mutation pressure that maintains codon usage bias, neither is there any report demonstrating that codon usage bias drives tRNA bias.

In this paper, I will show that, in vertebrate mitochondria, the codon usage bias is largely maintained by the strand-specific mutation pressure and the codon usage bias drives the evolution of tRNA anticodon. I will first review the conceptual framework underlying these two findings and then present empirical evidence substantiating these two findings.

Mammalian mitochondrial genome has two strands of different buoyant densities and consequently named the H-strand and the L-strand. The H-strand is the sense strand for 1 protein-coding gene (ND6) and 8 tRNA genes and the L-strand is the sense strand for 12 protein-coding genes, 2 rRNA genes, and 14 tRNA genes. The two strands have different nucleotide frequencies, with the H-strand rich in G and T and the L-strand rich in A and C (Jermiin et al., 1995; Perna and Kocher, 1995). This asymmetrical distribution of nucleotides has been explained as follows (Tanaka and Ozawa, 1994; Reyes et al., 1998), based on the strand-displacement model of mitochondrial DNA (mtDNA) replication (Clayton, 1982, 2000; Shadel and Clayton, 1997; Bogenhagen and Clayton, 2003).

During mtDNA replication, the L-strand is first used as a template to replicate the daughter H-strand, while the parental H-strand was left single-stranded for an extended period because the complete replication of mtDNA takes nearly 2 h (Clayton, 1982, 2000; Shadel and Clayton, 1997). Spontaneous deamination of both A and C (Sancar and Sancar, 1988; Lindahl, 1993) occurs frequently in human mitochondrial DNA (Tanaka and Ozawa, 1994). Deamination of A leads to hypoxanthine that forms stronger base pair with C than with T, generating an A.T→G.C mutation. Deamination of C leads to U, generating C.G→U.A mutations. Among these two types of spontaneous deamination, the C→U mutation occurs more frequently than the A→G mutation (Lindahl, 1993). In particular, the C→U mutation mediated by the spontaneous deamination occurs in single-stranded DNA more than 100 times as frequent as double-stranded DNA (Frederico et al., 1990). Note that these C→U mutants will immediately be used as a template to replicate the daughter L-strand, leading to a G→A mutation in the L-strand after one round of DNA duplication. Therefore, the H-strand, left single-stranded for an extended period during DNA replication, tends to accumulate A→G and C→U mutations and become rich in G and T while the L-strand will become rich in A and C.

When C→U, and the associated G→A, mutations happen at the first or the second codon positions, they are mostly nonsynonymous. Such nonsynonymous mutations are typically purged off by purifying selection. However, if such mutations happen at the third codon position, then they are synonymous and tend to accumulate under the biased mutation pressure. This would also lead to an increased frequency of A-ending codons. Similarly, if an N→C mutation happens at the third codon position in the L-strand, it tends to stay because the L-strand is not left single-stranded during DNA replication. So the net effect of this strand-specific mutation should lead to (1) most codons in the 12 CDS sequences (that are collinear with the L-strand) ending with A or C, (2) the codon bias in the ND6 gene on the opposite strand should be the opposite, and (3) the 8 tRNA sequences collinear with the H-strand should be richer in G and T than the 14 tRNA sequences collinear with the L-strand. While the first two predictions are expected from research on eubacterial genomes (Lobry, 1996; McInerney, 1998; Lobry and Sueoka, 2002), the last has not been tested. Our first objective is to test these three predictions of the vertebrate mitochondrial codon usage.

The main objective of this paper is to study the evolution of tRNA anticodon in response to (1) the strand-specific mutation bias itself and (2) the overall codon usage bias maintained by the strand-specific mutation bias. If the strand-specific mutation bias dominates the evolution of tRNA anticodons, then we should expect the wobble site of the anticodon, as well as other nonessential sites, in the 14 tRNA sequences collinear with the L-strand to evolve towards A or C but the anticodon wobble site in the eight tRNA sequences on the other strand should evolve towards U or G. On the other hand, given that 12 out of 13 CDSs are expected to have codons mostly ending with A or C due to the strand-specific mutation bias, we expect selection to favor anticodons to evolve its wobble site towards U or G to match the A-ending and C-ending codons. Such anticodon evolution would increase the translation efficiency for 12 protein-coding genes collinear with the L-strand but reduce the translation efficiency for the lone ND6 gene collinear with the H-strand. If the selection for translation efficiency is strong, then the 12 protein-coding genes would "outvote" the lone ND6 gene and the wobble site of the anticodon should evolve towards U and G. Our main objective is to evaluate the relative importance of the envisioned mutation and selection pressure on the evolution of tRNA anticodons.

2. Materials and methods

I retrieved all 382 vertebrate genomes by using NCBI Entrez. The codon usage pattern and tRNA anticodons are very similar among all 382 vertebrate species and I will only present data from two teleost fish, *Erpetoichthys calabaricus* (GenBank accession: NC_005251) and *Masturus lanceolatus* (NC_005837) and two mammalian species, *Mus musculus* (NC_005089) and *Bos taurus* (NC_001567). There is no particular reason for choosing these species except for an effort to capture the rather limited diversity of vertebrate mitochondrial genomes.

The tRNA and CDS sequences were extracted and analyzed by using DAMBE (Xia, 2001; Xia and Xie,

2001). The anticodon in almost all tRNA sequences from all species share the regular feature of being flanked by two nucleotides on either side to form a loop that is held together by a stem. For example, the anticodon loop (AC loop) of tRNA-Ala in *M. musculus* is 24AUUGA**UU**UUGC**AU**UCAAU40 where the starting and ending numbers indicate the position of the AC loop in the tRNA sequence (numbered from 1), with the anticodon (5′-UGC-3′) flanked by two nucleotides on either side (in bold) to form a loop that is held together by a stem made of the first and the last four nucleotides. Such a regular AC loop and its anticodon can be easily identified by dynamic programming. A few tRNA sequences have an anticodon flanked by three nucleotides, e.g., tRNA-Val in *E. calabaricus* and tRNA-Ser1 in the blue whale, *Balaenoptera musculus*. Some tRNA sequences have a suspicious AC loop. For example, the AC loop of tRNA-Trp is 26GAGC**CU**UCA**AA**GCCC42 with a stem that has a mismatch. For such tRNA sequences with an irregular AC loop, DAMBE will flag them out and the AC loop is identified by aligning the tRNA sequences against other isoaccepting tRNA sequences with a regular AC loop.

In all analyses, the two groups of Leu codons (CUN and UUR) were treated as two separate synonymous codon families, so are the two groups of Ser codons (AGY and UCN). In a separate paper, I will demonstrate differential selection on CUN and UUR Leu codons, on AGY and UCN Ser codons, and on their corresponding tRNA anticodons.

3. Results and discussion

3.1. Codon usage bias maintained by strand-asymmetrical mutation bias

I have previously mentioned that the strand-specific mutation bias favors A and C on the L-strand. The codon usage of the 12 CDS sequences collinear with the L-strand (Table 1) is consistent from this mutation bias, with the third codon position of the most frequent codon in each synonymous codon family (refereed to simply as codon family hereafter) being either A or C. In particular, NNY codon families are dominated by the C-ending codons, and NNR and NNN codon families are dominated by the A-ending codons. There is little variation in codon usage among the four species, or between the two fish species and the two mammalian species (N in Table 1). The remarkable consistency in this pattern from teleost fish to mammals demonstrates the power of the AC-biased mutation on the L-strand.

The observation that NNN codon families are dominated by NNA codons, instead by both NNA and NNC codons, might have adaptive significance (Xia, 1996), based on the observation that cellular concentration of ATP is much higher than that of the other three rNTPs (Colby and Edlin, 1970). For example, in the exponentially proliferating chick embryo fibroblasts in culture, the concentration of ATP, CTP, GTP, and UTP, in the unit of (mol$\times 10^{-12}$ per 10^6 cells), is 1890, 53, 190, and 130, respectively, in 2-h culture, and 2390, 73, 220, and 180, respectively, in 12-h culture. The transcription hypothesis of codon usage (Xia, 1996) states that, with the high availability of A and relatively low availability of the other three rNTPs, the transcription efficiency can be increased by maximizing the use of A in the third codon position of protein-coding genes.

In contrast to the L-strand with mutations favoring A and C, the H-strand is expected to accumulate mutations in the opposite direction, i.e., favoring G and T. Consequently, we should predict that the third codon position of the ND6 gene, which is the only 1 of the 13 protein-coding sequences collinear with the H-strand, should be either G or U. This prediction is also borne out by the empirical evidence (Table 2). In particular, NNY codon families are dominated by the U-ending codons, and NNR and NNN codon families are dominated by the G-ending codons.

The strand-specific mutation bias is also visible in tRNA sequences (Table 3), with the eight tRNA sequences collinear with the H-strand being rich in (G+T) and the 14 tRNA sequences collinear with the L-strand being rich in (A+C). This pattern is consistent from the teleost fish to mammalian species (Table 3).

3.2. Anticodon evolves to adapt to codon usage bias

Given the strand-specific mutation bias in vertebrate mitochondrial genomes, what can we predict about the anticodon evolution of the tRNA sequences? There are only 22 tRNA genes in vertebrate mitochondrial genome and each tRNA anticodon essentially has to wobble to recognize two or four synonymous codons. This suggests that the wobble position may not be strongly constrained and may be shaped by the strand-specific mutation bias. We can make a specific prediction if the strand-specific mutation pressure is the dominant force in shaping anticodon evolution. For the 14 tRNA sequences collinear with the L-strand, the wobble position of the anticodon should be either C or A. Similarly, for the 8 tRNA sequences collinear with the H-strand, the wobble position of the anticodon should be either G or U. This will be referred to hereafter as the mutation hypothesis of anticodon evolution.

In contrast to the prediction above assuming a dominant role for mutation, we can also make inferences of anticodons if selection plays a significant role in shaping codon–anticodon adaptation. Given the codon usage bias in 12 of the 13 CDS sequences maintained by the strand-specific mutation pressure, it is easy to see from Tables 1 and 2 that the overall codon usage bias at the genomic level is (1) C-ending codons most frequent in NNY codon families and (2) A-ending codons most frequent in NNR and NNN codon families. Such a codon usage may drive the wobble sites of the anticodon towards either G or U, regardless of which strand the tRNA gene is on. This will be referred to hereafter as the selection hypothesis of anticodon adaptation.

Table 1
Observed codon frequencies (N) and relative synonymous codon usage (RSCU; which is scaled to have an expectation of 1 when a codon is not over-used or under-used, see Sharp et al., 1986) for the 12 protein-coding genes collinear with the L-strand of the mitochondrial genomes from two teleost species, *Erpetoichthys calabaricus* (E. c.) and *Masturus lanceolatus* (M. l.) and two mammalian species, *Bos taurus* (B. t.) *Mus musculus* (M. m.), sorted by the one-letter representation of amino acids (AA)

Codon	AA	E. c.		M. l.		B. t.		M. m.	
		N	RSCU	N	RSCU	N	RSCU	N	RSCU
AGA	*	1	0.40	0	0.00	1	0.44	0	0.00
AGG	*	1	0.40	0	0.00	0	0.00	0	0.00
UAA	*	6	2.40	5	3.33	7	3.11	7	3.11
UAG	*	2	0.80	1	0.67	1	0.44	2	0.89
GCA	A	104	1.48	86	1.05	102	1.69	94	1.67
GCC	A	92	1.31	160	1.95	90	1.49	83	1.48
GCG	A	4	0.06	14	0.17	1	0.02	5	0.09
GCU	A	82	1.16	68	0.83	48	0.80	43	0.76
UGC	C	16	1.33	21	1.91	16	1.52	18	1.50
UGU	C	8	0.67	1	0.09	5	0.48	6	0.50
GAC	D	31	0.89	50	1.39	46	1.44	43	1.27
GAU	D	39	1.11	22	0.61	18	0.56	25	0.74
GAA	E	76	1.75	71	1.50	73	1.70	77	1.86
GAG	E	11	0.25	24	0.51	13	0.30	6	0.15
UUC	F	88	0.82	139	1.22	130	1.16	133	1.15
UUU	F	126	1.18	89	0.78	95	0.84	98	0.85
GGA	G	85	1.68	72	1.32	93	1.93	101	2.17
GGC	G	57	1.13	90	1.65	60	1.24	39	0.84
GGG	G	24	0.48	31	0.57	19	0.39	20	0.43
GGU	G	36	0.71	25	0.46	21	0.44	26	0.56
CAC	H	49	1.00	74	1.40	63	1.36	63	1.30
CAU	H	49	1.00	32	0.60	30	0.65	34	0.70
AUC	I	89	0.52	117	0.91	160	1.03	138	0.77
AUU	I	255	1.48	139	1.09	151	0.97	221	1.23
AAA	K	72	1.76	53	1.45	88	1.81	101	1.98
AAG	K	10	0.24	20	0.55	9	0.19	1	0.02
CUA	L	152	1.61	153	1.43	283	2.94	266	2.84
CUC	L	44	0.47	189	1.77	95	0.99	64	0.68
CUG	L	28	0.30	49	0.46	29	0.30	27	0.29
CUU	L	130	1.37	138	1.29	61	0.63	84	0.90
UUA	L	191	2.02	95	0.89	100	1.04	114	1.22
UUG	L	23	0.24	18	0.17	10	0.10	7	0.08
AUA	M	175	1.58	77	1.09	214	1.71	213	1.81
AUG	M	46	0.42	64	0.91	37	0.30	23	0.20
AAC	N	55	0.76	88	1.44	102	1.29	107	1.31
AAU	N	89	1.24	34	0.56	56	0.71	56	0.69
CCA	P	112	2.19	54	1.00	85	1.79	129	2.67
CCC	P	20	0.39	91	1.69	63	1.33	34	0.71
CCG	P	19	0.37	16	0.30	3	0.06	2	0.04
CCU	P	54	1.05	55	1.02	39	0.82	28	0.58
CAA	Q	81	1.78	76	1.58	79	1.84	78	1.93
CAG	Q	10	0.22	20	0.42	7	0.16	3	0.07
CGA	R	41	2.34	33	1.81	42	2.71	34	2.16
CGC	R	22	1.26	20	1.10	11	0.71	18	1.14
CGG	R	0	0.00	11	0.60	3	0.19	3	0.19
CGU	R	7	0.40	9	0.49	6	0.39	8	0.51
AGC	S	25	0.61	40	1.08	42	0.96	35	0.75
AGU	S	13	0.32	5	0.14	9	0.21	11	0.24
UCA	S	89	2.17	60	1.61	98	2.23	145	3.10
UCC	S	53	1.29	77	2.07	64	1.46	47	1.00
UCG	S	6	0.15	8	0.22	4	0.09	3	0.06
UCU	S	60	1.46	33	0.89	47	1.07	40	0.85
ACA	T	123	1.78	87	1.15	150	2.00	157	2.08
ACC	T	66	0.95	129	1.70	95	1.27	86	1.14
ACG	T	5	0.07	25	0.33	14	0.19	5	0.07
ACU	T	83	1.20	63	0.83	41	0.55	54	0.72
GUA	V	83	2.08	52	1.06	82	1.96	69	1.87
GUC	V	18	0.45	77	1.57	46	1.10	33	0.89
GUG	V	10	0.25	22	0.45	9	0.22	8	0.22
GUU	V	49	1.23	45	0.92	30	0.72	38	1.03
UGA	W	108	1.93	86	1.50	91	1.82	93	1.90
UGG	W	4	0.07	29	0.50	9	0.18	5	0.10
UAC	Y	44	0.77	69	1.35	72	1.13	60	1.03
UAU	Y	70	1.23	33	0.65	56	0.88	56	0.97

Our empirical data (Table 4) strongly support the selection hypothesis of anticodon adaptation. There are two points worth highlighting in Table 4. First, for each tRNA, the anticodon is the same in all vertebrates from teleost fish to mammals. This implies that the selection at the wobble site must be very strong. Second, for tRNAs recognizing NNY codon families, the wobble site is always G to match the most frequently used C-ending codon. For tRNAs recognizing NNR and NNN codon families, the wobble site is always U to match the most frequently used

Table 2
Codon frequencies for the two teleost species combined (Teleost) and the two mammalian species combined (Mammal) for the ND6 gene collinear with the H-strand in the mitochondrial genome

Codon	AA	Teleost	Mammal	Codon	AA	Teleost	Mammal
AGA	*	1	0	AUA	M	12	8
AGG	*	0	0	AUG	M	10	14
UAA	*	0	2	AAC	N	1	1
UAG	*	1	0	**AAU**	**N**	**3**	**8**
GCA	A	9	3	CCA	P	2	1
GCC	A	2	1	CCC	P	2	0
GCG	A	5	3	CCG	P	2	0
GCU	**A**	**16**	**8**	**CCU**	**P**	**3**	**5**
UGC	C	0	0	CAA	Q	0	0
UGU	**C**	**4**	**5**	CAG	Q	1	1
GAC	D	1	0	CGA	R	1	1
GAU	**D**	**3**	**10**	CGC	R	0	0
GAA	E	2	8	CGG	R	6	0
GAG	**E**	**10**	**9**	CGU	R	3	1
UUC	F	4	3	AGC	S	1	1
UUU	**F**	**19**	**23**	**AGU**	**S**	**4**	**6**
GGA	G	9	12	UCA	S	2	3
GGC	G	3	2	UCC	S	0	1
GGG	**G**	**18**	**21**	UCG	S	2	2
GGU	G	14	17	**UCU**	**S**	**16**	**6**
CAC	H	1	0	ACA	T	0	3
CAU	H	0	0	ACC	T	0	1
AUC	I	0	1	ACG	T	3	2
AUU	**I**	**12**	**27**	**ACU**	**T**	**4**	**8**
AAA	K	0	2	GUA	V	6	10
AAG	K	0	3	GUC	V	2	2
CUA	L	4	2	GUG	V	17	9
CUC	L	0	0	**GUU**	**V**	**24**	**22**
CUG	L	4	0	UGA	W	4	4
CUU	**L**	**8**	**4**	UGG	W	4	5
UUA	**L**	**26**	**25**	UAC	Y	4	2
UUG	L	14	14	**UAU**	**Y**	**13**	**17**

The most frequently used codons are in bold for large N.

Table 3
Frequency of (A+C) for the four vertebrate species

tRNA	*Bos*	Erpetoichthys	Masturus	Mus
Pro	0.379	0.429	0.414	0.373
Gln	0.347	0.380	0.437	0.380
Glu	0.406	0.420	0.377	0.377
Ala	0.377	0.391	0.420	0.449
Ser1	0.394	0.394	0.437	0.435
Asn	0.397	0.466	0.397	0.437
Cys	0.463	0.470	0.448	0.455
Tyr	0.500	0.451	0.423	0.522
Ile	0.522	0.493	0.514	0.493
Arg	0.507	0.486	0.551	0.485
Asp	0.507	0.536	0.507	0.557
Thr	0.565	0.521	0.514	0.507
Leu1	0.547	0.533	0.541	0.533
Met	0.522	0.529	0.536	0.551
Gly	0.507	0.557	0.521	0.574
Leu2	0.521	0.548	0.548	0.563
His	0.571	0.493	0.536	0.612
Ser2	0.500	0.591	0.529	0.525
Lys	0.522	0.562	0.573	0.538
Val	0.582	0.549	0.556	0.580
Trp	0.582	0.571	0.583	0.582
Phe	0.612	0.620	0.603	0.574

The first eight tRNA sequences are collinear with H-strand and tend to be AC-poor.

A-ending codons (with only one exception). This is consistent regardless of which strand the tRNA sequence is on. Thus, the selection is sufficiently strong to eliminate the effect of strand-specific mutation bias.

The only, but stubbornly consistent, exception is the anticodon of tRNA-Met which has a 5′ -CAU-3′ anticodon base-pairing with the AUG codon (the translation initiation codon) instead of the more frequent AUA codon. The rate of protein synthesis depends on both the rate of translation initiation (Liljenstrom and von Heijne, 1987; Bulmer, 1991) and the rate of peptide elongation (Varenne et al., 1984; Bulmer, 1987; Xia, 1998). While the observed CAU anticodon in tRNA-Met would increase the initiation rate, it is expected to reduce the elongation rate because most methionine codons are AUA. In contrast, the unobserved UAU anticodon for tRNA-Met would have increased the elongation rate but would decrease the initiation rate. The observation that the anticodon of the tRNA-Met gene is universally CAU suggests that increasing initiation rate is far more important than increasing the elongation rate.

The interpretation above assumes no nucleotide modification at the anticodon site. However, it is known that the C in anticodon CAU may be modified to 5-formylcytidine which may allow it to pair with both A and G (Moriya et al., 1994; Matsuyama et al., 1998). Whether such a modification would increase both the initiation rate and elongation rate (by efficiently pairing with both AUG and AUA codons) has not been elucidated by experimental studies.

Given the observed pattern of tRNA anticodons (Table 4), it is quite clear that the ND6 gene (the only one collinear with the H-strand) exhibits little codon–anticodon adaptation. The tRNA bias in vertebrate mitochondria is far stronger than any other translation system, either in prokaryotes or in the nucleus of eukaryotes. For example, in *E. coli*, there are six tRNA-Gly genes, with four (glyW, glyV, glyX, and glyY) having the GCC anticodon to form Watson–Crick pairing with the GGC codon, one (glyU) with the CCC anticodon to pair with GGG and one (glyT) with the TCC anticodon to pair with GGA. So the selection against the use of GGG and GGA are not strong because there are perfect tRNA-Gly adaptors, albeit a bit rarer, for them. In vertebrate mitochondrial genome, there is only one tRNA-Gly gene with the TCC anticodon to pair with GGA and there is no perfect tRNA-Gly adaptor for any other Gly codons. Thus, the selection against the other three Gly codons should be very strong. Yet the supposedly strong selection is entirely powerless in shaping the codon usage of the ND6 gene collinear with the H-strand. It is reasonable to suggest that, at least in vertebrate mitochondria, codon usage does not evolve as a response to tRNA bias. Instead, it is the tRNA anticodon that evolves as a response to mutation-maintained codon usage bias.

3.3. An alternative hypothesis of selection on anticodon versatility

P. Higgs and R.W. DeBry (personal communication) independently suggested an alternative selection hypothesis on the evolution of the anticodon wobble site which I will

Table 4
Anticodon (AC) of the 22 tRNA genes from the four species (Table 1) and their associated synonymous codon families (SCF)

tRNA	Strand	SCF	AC
Ala	C	GCN	UGC
Arg		CGN	UCG
Gly		GGN	UCC
Leu		CUN	UAG
Pro	C	CCN	UGG
Ser	C	UCN	UGA
Thr		ACN	UGU[a]
Val		GUN	UAC
Ser		AGY	GCU
His		CAY	GUG
Ile		AUY	GAU
Asn	C	AAY	GUU
Asp		GAY	GUC
Cys	C	UGY	GCA
Phe		UUY	GAA
Tyr	C	UAY	GUA
Gln	C	CAR	UUG
Glu	C	GAR	UUC
Leu		UUR	UAA
Lys		AAR	UUU
Met		AUR	CAU
Trp		UGR	UCA

"C" stands for "complementary strand", i.e., not on the same strand as the 12 protein-coding genes. Note that the first nucleotide of the anticodon (AC) is the wobble site.

[a] GGU in *Mus musculus*, which might be due to sequencing error because the anticodon loop is irregular.

term the selection hypothesis of anticodon versatility. The hypothesis has been implicitly presented before (e.g., Agris, 2004). Given that each synonymous codon family is translated by a single tRNA species in vertebrate mitochondria, the versatility of this single tRNA in translating two or four synonymous codons are important for the translation machinery. For two-fold degenerate codons ending with C or U, then the obvious anticodon wobble site should be G because G can pair with both C and U. This is consistent with our findings. For two-fold degenerate codons ending with A or G, then the obvious anticodon wobble site should be U. For four-fold degenerate codon families, a wobble U might confer the anticodon greater versatility than other nucleotides [Yokoyama and Nishimura, 1995 #12903; Sibler et al., 1986 #12902; Inagaki et al., 1995 #12864; Yokobori et al., 2001 #13596; Andachi et al., 1989 #13419; Barrell et al., 1980 #13384]. According to this hypothesis, the use of U and G at the wobble site of the tRNA anticodon can be predicted with no reference to codon usage, although codon–anticodon adaptation can then evolve as secondary adaptation given that wobble U and G are strictly maintained by selection. This selection hypothesis of anticodon versatility, which was also implicitly mentioned in (Tong and Wong, 2004), is perfectly consistent with our data of codon usage and tRNA anticodons in vertebrate mitochondrial genomes.

The vertebrate mitochondrial data are unable to distinguish between the selection hypothesis of anticodon versatility and the selection hypothesis of anticodon adaptation that I have mentioned before because both hypotheses have the same predictions for the anticodon wobble site. However, the codon usage of the four-fold degenerate arginine codons (CGN) in the mitochondrial genome of four species: *Caenorhabditis elegans* (nematode), *Marchantia polymorpha* (plant), *Pichia canadensis* (fungi), and *S. cerevisiae* (fungi) sheds light on resolving these two hypotheses (Table 5). In these four mitochondrial genomes, the four synonymous CGN codons, with CGU being the most dominant, are translated by a single tRNA just as in vertebrate mitochondria. The selection hypothesis for anticodon versatility would have predicted a "versatile" U in the tRNA anticodon wobble site for these four-fold degenerate codons. However, this is not true because the wobble anticodon site is A instead of U in all four mitochondrial genomes. This means that (1) a wobble U in the tRNA anticodon may not necessarily confer greater versatility than a wobble A or other nucleotides or that (2) the selection for anticodon versatility is weak. On the other hand, given that the CGU codon is the most dominant of the four synonymous arginine codons in all four mitochondrial genomes (Table 5), the hypothesis of anticodon adaptation would predict an A at the anticodon wobble site, which is true for all four species. More detailed study may be needed to further evaluate the relative significance of these selection hypotheses on shaping anticodon evolution.

Table 5
Codon usage of the four-fold degenerate arginine codons in four species

Species	Accession	CGA	CGC	CGG	CGU
C. elegans	NC_001328	1	0	1	29
M. polymorpha	NC_001660	260	165	118	286
P. Canadensis	NC_001762	0	1	0	19
S. cerevisiae	NC_001224	0	2	1	18

It is important to highlight the fact that none of the species in Table 5 is a vertebrate. Even if my interpretation of the result in Table 5 is correct, it is not necessarily generalizable to vertebrates. Furthermore, I should also emphasize here that my result does not reject the hypothesis of anticodon versatility, although it does cast some doubts on its necessity, especially with reference to mitochondrial codon–anticodon adaptation.

I should mention here that, although the conceptual framework that leads to the prediction of strand-asymmetry in mutation spectrum is based on the classical strand-displacement model of mtDNA replication (Clayton, 1982; Shadel and Clayton, 1997; Bogenhagen and Clayton, 2003), the prediction can also be derived from the strand-coupled model of bidirectional mtDNA replication (Holt et al., 2000; Yang et al., 2002; Holt and Jacobs, 2003). Many studies have documented an excess of (G+T) in the leading strand and an excess of (A+C) in the lagging strand in most prokaryotic genomes examined (Perriere et al., 1996; Francino and Ochman, 1997; Freeman et al., 1998; Grigoriev, 1998; McLean et al., 1998), and spontaneous deamination has also been invoked as the main factor contributing to the strand asymmetry (Lobry and Sueoka, 2002). Thus, if the H-strand is the leading strand, and the L-strand the lagging one, then we would also predict an excess of (G+T) in the H-strand and of (A+C) in the L-strand, just as we would expect from the strand-displacement model of mtDNA replication.

One limitation of this study is that it cannot be generalized to invertebrate mitochondrial genomes although they also have about 13 protein-coding genes and 22 tRNA genes. There are several major differences between vertebrate and invertebrate mtDNA. First, invertebrate mitochondrial genomes are generally extremely AT-rich and the distribution of the protein-coding genes is less asymmetrical between the two strands than in vertebrate mitochondrial genomes. Take the common honey bee mtDNA for example. 9 of the 13 CDSs are collinear with the L-strand and 4 are collinear with the H-strand, in contrast to 12 CDSs collinear with the L-strand and only 1 collinear with the H-strand in vertebrate mitochondrial genomes.

In summary, I have presented the first case in which (1) codon usage bias is maintained by strand-specific mutation bias and (2) the biased codon usage drives the evolution of tRNA anticodons. At least in vertebrate mitochondria, it is the codon usage that drives the evolution of tRNA anticodons. In contrast, almost all current literature on codon–anticodon adaptation assumes that it is tRNA bias that drives the evolution of codon usage. This assumption is probably questionable.

Acknowledgment

I thank Hiroshi Akashi, David H. Ardell, Stephane Aris-Brosou, Ron W. Debry, Donal Hickey, Paul Higgs, Lars S. Jermiin, Jean R. Lobry, and Jeff Thorne for their comments that clarified several points and, in particular, brought my attention to a few highly relevant publications by European scientists. This study is supported by a grant from University of Ottawa and from the discovery and strategic grants from NSERC-Canada.

References

Agris, P.F., 2004. Decoding the genome: a modified view. Nucleic Acids Res. 32, 223–238.

Akashi, H., 1995. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA [see comments]. Genetics 139, 1067–1076.

Akashi, H., 1997. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. Gene 205, 269–278.

Akashi, H., 2003. Translational selection and yeast proteome evolution. Genetics 164, 1291–1303.

Andachi, Y., Yamao, F., Muto, A., Osawa, S., 1989. Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*. Resemblance to mitochondria. J. Mol. Biol. 209, 37–54.

Barrell, B.G., et al., 1980. Different pattern of codon recognition by mammalian mitochondrial tRNAs. Proc. Natl. Acad. Sci. U. S. A. 77, 3164–3166.

Bennetzen, J.L., Hall, B.D., 1982. Codon selection in yeast. J. Biol. Chem. 257, 3026–3031.

Berg, O.G., 1996. Selection intensity for codon bias and the effective population size of *Escherichia coli*. Genetics 142, 1379–1382.

Berg, O.G., Martelius, M., 1995. Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella typhimurium* and their relationship to gene expression and selection pressure. J. Mol. Evol. 41, 449–456.

Bogenhagen, D.F., Clayton, D.A., 2003. The mitochondrial DNA replication bubble has not burst. Trends Biochem. Sci. 28, 357–360.

Bulmer, M., 1987. Coevolution of codon usage and transfer RNA abundance. Nature 325, 728–730.

Bulmer, M., 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129, 897–907.

Clayton, D.A., 1982. Replication of animal mitochondrial DNA. Cell 28, 693–705.

Clayton, D.A., 2000. Transcription and replication of mitochondrial DNA. Hum. Reprod. 15, 11–17.

Colby, C., Edlin, G., 1970. Nucleotide pool levels in growing, inhibited, and transformed chick fibroblast cells. Biochemistry 9, 917.

Eyre-Walker, A., 1996. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy? Mol. Biol. Evol. 13, 864–872.

Francino, M.P., Ochman, H., 1997. Strand asymmetries in DNA evolution. Trends Genet. 13, 240–245.

Frederico, L.A., Kunkel, T.A., Shaw, B.R., 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 29, 2532–2537.

Freeman, J.M., Plasterer, T.N., Smith, T.F., Mohr, S.C., 1998. Patterns of genome organization in bacteria. Science 279, 1827a.

Gouy, M., Gautier, C., 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10, 7055–7064.

Grigoriev, A., 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. 26, 2286–2290.

Holt, I.J., Jacobs, H.T., 2003. Response: the mitochondrial DNA replication bubble has not burst. Trends Biochem. Sci. 28, 355–356.

Holt, I.J., Lorimer, H.E., Jacobs, H.T., 2000. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. Cell 100, 515–524.

Ikemura, T., 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. 151, 389–409.

Inagaki, Y., Kojima, A., Bessho, Y., Hori, H., Ohama, T., Osawa, S., 1995. Translation of synonymous codons in family boxes by *Mycoplasma capricolum* tRNAs with unmodified uridine or adenosine at the first anticodon position. J. Mol. Biol. 251, 486–492.

Jermiin, L., Graur, D., Crozier, R., 1995. Evidence from analyses of intergenic regions for strand-specific directional mutation pressure in metazoan mitochondrial DNA. Mol. Biol. Evol. 12, 558–563.

Liljenstrom, H., von Heijne, G., 1987. Translation rate modification by preferential codon usage: intragenic position effects. J. Theor. Biol. 124, 43–55.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. Nature 362, 709–715.

Lobry, J.R., 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. Mol. Biol. Evol. 13, 660–665.

Lobry, J.R., 2004. Life history traits and genome structure: aerobiosis and G+C content in bacteria. Lect. Notes Comput. Sci. 3039, 679–686.

Lobry, J.R., Sueoka, N., 2002. Asymmetric directional mutation pressures in bacteria. Genome Biol. 3, 1–14 (research58).

Matsuyama, S., Ueda, T., Crain, P.F., McCloskey, J.A., Watanabe, K., 1998. A novel wobble rule found in starfish mitochondria. Presence of 7-methylguanosine at the anticodon wobble position expands decoding capability of tRNA. J. Biol. Chem. 273, 3363–3368.

McInerney, J.O., 1998. Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. Proc. Natl. Acad. Sci. U. S. A. 95, 10698–10703.

McLean, M.J., Wolfe, K.H., Devine, K.M., 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. 47, 691–696.

Moriya, J., et al., 1994. A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. Biochemistry 33, 2234–2239.

Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358.

Perriere, G., Lobry, J.R., Thioulouse, J., 1996. Correspondence discriminant analysis: a multivariate method for comparing classes of protein and nucleic acid sequences. Comput. Appl. Biosci. 12, 519–524.

Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957–966.

Sancar, A., Sancar, G.B., 1988. DNA repair enzymes. Annu. Rev. Biochem. 57, 29–67.

Shadel, G.S., Clayton, D.A., 1997. Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. 66, 409–435.

Sharp, P.M., Tuohy, M.F., Mosurski, K.R., 1986. Codon usage in yeast cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14, 5125–5143.

Sibler, A.P., Dirheimer, G., Martin, R.P., 1986. Codon reading patterns in *Saccharomyces cerevisiae* mitochondria based on sequences of mitochondrial tRNAs. FEBS Lett. 194, 131–138.

Sueoka, N., 1961. Correlation between base composition of deoxyribonucleic acid and amino acid composition of proteins. Proc. Natl. Acad. Sci. U. S. A. 47, 1141–1149.

Tanaka, M., Ozawa, T., 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics 22, 327–335.

Tong, K.L., Wong, J.T., 2004. Anticodon and wobble evolution. Gene 333, 169–177.

Varenne, S., Bug, J., Lloubes, R., Lazdunski, C., 1984. Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Biol. Chem. 180, 549–576.

Xia, X., 1996. Maximizing transcription efficiency causes codon usage bias. Genetics 144, 1309–1320.

Xia, X., 1998. How optimized is the translational machinery in *E. coli*, *S. typhimurium*, and *S. cerevisiae*? Genetics 149, 37–44.

Xia, X., 2001. Data analysis in Molecular Biology and Evolution. Kluwer Academic Publishers, Boston.

Xia, X., Xie, Z., 2001. DAMBE: software package for data analysis in molecular biology evolution. J. Heredity 92, 371–373.

Yang, M.Y., et al., 2002. Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. Cell 111, 495–505.

Yokobori, S., Suzuki, T., Watanabe, K., 2001. Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. J. Mol. Evol. 53, 314–326.

Yokoyama, S., Nishimura, S., 1995. Modified nucleotides and codon recognition. In: Soll, D., RajBhandary, U. (Eds.), tRNA: Structure, Biosynthesis and Function. ASM Press, Washington, pp. 207–223.