

Single Nucleotide +1 Frameshifts in an Apparently Functional Mitochondrial Cytochrome b Gene in Ants of the Genus *Polyrhachis*

Andrew T. Beckenbach,¹ Simon K.A. Robson,² Ross H. Crozier²

¹ Department of Biological Sciences, Simon Fraser University, Burnaby B.C., Canada V5A 1S6

² School of Tropical Biology, James Cook University, Townsville, QLD, Australia

Received: 25 November 2003 / Accepted: 1 October 2004 [Reviewing Editor: Dr. W. Ford Doolittle]

Abstract. Twelve of 30 species examined in the ant genus *Polyrhachis* carry single nucleotide insertions at one or two positions within the mitochondrial cytochrome b (*cytb*) gene. Two of the sites are present in more than one species. Nucleotide substitutions in taxa carrying insertions show the strong codon position bias expected of functional protein coding genes, with substitutions concentrated in the third positions of the original reading frame. This pattern of evolution of the sequences strongly suggests that they are functional *cytb* sequences. This result is not the first report of +1 frameshift insertions in animal mitochondrial genes. A similar site was discovered in vertebrates, where single nucleotide frameshift insertions in many birds and a turtle were reported by Mindell et al. (Mol Biol Evol 15:1568, 1998). They hypothesized that the genes are correctly decoded by a programmed frameshift during translation. The discovery of four additional sites gives us the opportunity to look for common features that may explain how programmed frameshifts can arise. The common feature appears to be the presence of two consecutive rare codons at the insertion site. We hypothesize that the second of these codons is not efficiently translated, causing a pause in the translation process. During the stall the weak wobble pairing of the tRNA bound in the peptidyl site of the ribosome, together with an exact Watson–Crick codon–anticodon pairing in the +1 position, allows translation to continue in the +1 reading frame. The

result of these events is an adequate level of translation of a full-length and fully functional protein. A model is presented for decoding of these mitochondrial genes, consistent with known features of programmed translational frameshifting in the yeast TY1 and TY3 retrotransposons.

Key words: Frameshifts — Programmed translational frameshifting — animal mitochondrial genomes — *Polyrhachis* ants

Introduction

In their analysis of a complete mitochondrial genome of an ostrich, Härlid et al. (1997) reported an apparent truncation of the *nad3* (NADH dehydrogenase subunit 3) gene, due to the insertion of a single nucleotide at position 174 in the gene, altering the reading frame. Only the first 58 of 116 residues of the *nad3* protein would be translated correctly. The rest of the gene is present, but not accessible in the starting reading frame. Mindell et al. (1998) examined that region in the mitochondrial genomes of a large number of bird species and in the painted turtle. They found the extra base in 46 of 61 bird species and in the turtle. More surprising, the entire reading frame could be restored in all taxa by deleting the extra nucleotide from the data file. There was nothing else wrong with these sequences. They examined the pattern of evolution among the taxa and concluded

Table 1. List of primers used for PCR amplification and sequencing of *Polyrhachis* mitochondrial genes

Primer name ^a	Gene	Sequence (5'–3')
TL-J3033	<i>tRNA^{Leu-UUR}</i>	TAATATGGCAGATTAGTGCA
C2-J3396	<i>cox2</i>	ACAATTGGTCATCAATGATA
C2-J3624	<i>cox2</i>	ACTCCTGGACGATTAAYCA
C2-N3665	<i>cox2</i>	CCACAAATTTCTGAACATTG
TD-N3873	<i>tRNA^{ASP}</i>	AGAATATTGTTAAATTGACA
C3-J4792	<i>cox3</i>	GTTGATTATAGACCWTGRCC
N3-N5731	<i>nad3</i>	TTAGGGTCAAATCCRCAYTC
N3-J5741	<i>nad3</i>	TCACTTTTGAATGYGGNTT
TF-N6384	<i>tRNA^{Phe}</i>	TATATTTAGAGYATRAYAYTGAAG
N4-J8944	<i>nad4</i>	GGAGCTTCAACATGAGCTTT
CB-J10601	<i>cytb</i>	CTCCCAACTCCTATTAATATTTCTT
CB-J10621	<i>cytb</i>	TATTGATGAAAYTTTGGNTC
CB-J10864	<i>cytb</i>	CATAATACTTGAATAATTGG
CB-N10755	<i>cytb</i>	CCATTGATATGAATATTATG
CB-J10933	<i>cytb</i>	GTCTACCTTGAGGNCAAARTC
CB-N11010	<i>cytb</i>	TTAATAGARAATCCYCCYCA
CB-N11322	<i>cytb</i>	AGAAAGTATCATTGAGGTTG
CB-N11370	<i>cytb</i>	AACTCCTCTAGTTTATTNGG
TS-N11683	<i>tRNA^{Ser-UCN}</i>	TATTTCTTTATTATGTTTTCAAAC
LR-J13901	<i>lsr</i>	TGATAAACCCCTGATACAMAAGG
SR-N14588	<i>ssr</i>	CAAACCAGGATTAGATACCCTGTTAT

^aPrimers are numbered according to the position of the 3' nucleotide in the *Drosophila* genome (Clary and Wolstenholme 1985). J refers to the majority coding strand; N to the minority strand.

that the sequences are functional. Evidently the extra nucleotide is skipped during translation. They suggested that the site was subject to a +1 programmed frameshift during translation, allowing correct translation of a full-length, functional protein. Their report was the first evidence of programmed translational frameshifting in mitochondrial genomes.

During a phylogenetic study of 30 species of the ant genus *Polyrhachis*, using the mitochondrial *cytb* (cytochrome b) gene, we discovered a single nucleotide insertion of a G in *P. dives* at approximate position 333 within the gene. As the study proceeded, a total of 12 species were found to have single nucleotide insertions at one or two sites in the *cytb* gene, with four different sites affected. Two of the sites are present in more than one species, allowing examination of the pattern of substitution in the molecule subsequent to the insertion.

In this paper, we report the finding of four new sites of +1 frameshift insertions in animal mitochondrial genes. The finding of additional sites gives us the opportunity to examine the sequences in the region of the insertions for common features that may give some insight into the mechanism. The results give additional evidence that the genes are translated in the original reading frame and point to +1 translational frameshifting as the mechanism that restores the correct reading frame at each insertion site. We present evidence that these sequences are functional and provide a model for their translation, consistent with the mechanisms hypothesized for translation of programmed frameshift sites in the TY1 and TY3 retrotransposons in yeast.

Materials and Methods

DNA Extraction, Amplification, and Sequencing

Total genomic DNA was extracted from a single leg of a representative of each species using organic solvents and ethanol precipitation, as described by Beckenbach et al. (1993). Selected regions of the mitochondrial genome were amplified using primers listed in Table 1. PCR products were purified either by gel isolation, followed by a freeze/squeeze extraction (Tautz and Renz 1983), or by using the QIAquick PCR Purification Kit (QIAGEN). The purified products were sequenced directly using the amplification primers. The PCR products were sequenced manually using ³³P-labeled dideoxy cycle sequencing with Thermal Sequenase (Amersham) or using the automated sequencing services of the DNA Sequencing Lab, University of Calgary Core DNA & Protein Services.

All sequences were determined from both strands. Those showing insertions were sequenced from at least two different overlapping PCR products, using different PCR primer pairs. The gels (for manual sequencing) or sequencing scans were carefully examined for accuracy in the region of the insertions. In no case was a conflicting result observed, nor was there evidence of double sequence, as might be observed if multiple copies of the gene were present.

Analysis

Sequences were aligned manually. Aside from the single nucleotide insertions at four sites in the *cytb* gene, there were no indels in either of the genes. Protein coding sequences were translated using the *Drosophila* mitochondrial code (Clary and Wolstenholme 1985). Regions surrounding the frameshift insertions, including 4 codons upstream, the insert region, and 12 codons downstream (17 codons; 51 nucleotides [nt]), were scanned for potential secondary structure at 25°C in the mRNA using the mfold 2.3 web server (<http://www.bioinfo.rpi.edu/applications/mfold> [Zuker 2003]). A sliding window of fragments of the same length (51 nt), at 20-nt

Table 2. Occurrence of frameshift insertions in the cytochrome b gene of 12 species of *Polyrhachis*

Subgenus	Species	FS position ^a	1st stop ^b	Product length ^c
<i>Campomyrma</i>	<i>P. phryne</i>	474, 773	499	166
<i>Hagiomyrma</i>	<i>P. trapezoidea</i>	474	499	166
<i>Hemioptica</i>	<i>P. boltoni</i>	333	427	142
	<i>P. scissa</i>	333	460	153
<i>Myrmatopa</i>	<i>P. nr. floricornis</i>	333	427	142
	<i>P. lombokensis</i>	333	427	142
	<i>P. ulysses</i>	333	457	152
<i>Myrmhopla</i>	<i>P. dives</i>	333	427	142
	<i>P. sexspinosa</i>	333, 411	433	144
<i>Myrmotherinax</i>	<i>P. nr. abnormis</i>	333	496	165
<i>Polyrhachis</i>	<i>P. bihamata</i>	333	427	142
	<i>P. ypsilon</i>	333	427	142

Note. The full length of the gene *P. turneri* is 1116 nt, encoding a predicted product of 372 residues.

^aApproximate nucleotide position of the frameshift insertion(s).

^bPosition of first inframe stop codon.

^cLength of predicted truncated product.

intervals, was also scanned for potential secondary structure at 25°C for the entire *cytb* gene of *Polyrhachis turneri*, to provide a basis for interpreting those results.

Three of the tRNA genes associated with the frameshift sites, tRNA^{Gly}, tRNA^{Val}, and tRNA^{Ala}, were identified and folded using tRNAscan-SE (Lowe and Eddy 1997). We also sequenced the tRNA^{Ser-AGN} genes from *Polyrhachis phryne* and from an outgroup ant species, *Lasius alienus*. The outgroup species was included to confirm identification of the gene. The genes were identified based on position in the genome, potential for forming a secondary structure consistent with those observed for each gene in other insects, and examination of the substitution pattern between the species. The sequences for tRNA^{Ser-AGN} were folded using mfold 2.3 at 25°C with default settings for other parameters (Zuker 2003).

Results

We determined the complete sequences of the *cytb* and *cox2* genes for *Polyrhachis turneri*. The *cytb* gene consists of 1116 nt, coding for a predicted product of 372 amino acids, while *cox2* is 678 nt coding for 226 amino acids. Both genes are terminated by an inframe TAA stop encoded in the DNA sequence. We examined a 750-bp fragment of *cytb* and the first 579 bp of *cox2* in 29 other species of the genus *Polyrhachis*. Each species gave a single, unique sequence for both genes. In all cases the *cox2* sequences translated as expected. In 12 species, however, the *cytb* gene is not an open reading frame. In those species, a single nucleotide insertion at one or two sites disrupted the reading frame. These insertions were evident on both strands.

One insertion, at approximate position 333 in the *cytb* gene, was first discovered in *Polyrhachis dives*. After confirming it by sequencing both strands, we extracted DNA from three additional specimens and repeated the analysis. We designed additional PCR primers to examine the region, based on conserved regions in the *Polyrhachis cytb* sequences. The extra nucleotide was always present. As the survey of *cytb*

sequences in *Polyrhachis* progressed, nine other species showed a single nucleotide insertion at the same site, eight of them an extra G. In all cases the extra nucleotide was present in sequences from both strands, and confirmed from at least two PCR products using different primer combinations.

In the course of the survey of *Polyrhachis* species, a total of 12 species showed single nucleotide insertions at one or two sites in the *cytb* gene (Table 2). A total of four sites were involved. If translated in the expected reading frame (0-frame), they would produce products of between 142 and 166 amino acid residues, less than half the expected length of the wild-type protein. The four sites are numbered according to the approximate position of the inserted nucleotide.

Description of the Sites

The sequences at the four sites discovered in *Polyrhachis*, along with the site observed by Mindell et al. (1998) for several representative vertebrates, are given in Fig. 1. Included is a hypothetical “premutation” sequence obtained by removing the extra nucleotide.

One site (FS333) at approximate position 333 within the gene was encountered in 10 species (Table 2). The sequence of the affected codon in 17 of the 20 other species is GGA, with third position G, T, and C appearing once each in the other three species. In nine of the FS333 species there is a third position G, so that the inferred mutation was GGA → GGG A (Fig. 1). The other FS333 species, *P. sexspinosa*, has an inserted T: GGA → GGT A. The insertion in this species could represent a recent G → T transversion, following an original insertion of a G or an independent insertion of a T at the same site. Species having an inserted nucleotide at this site include representatives of 5 of the 12 subgenera of the genus

Species	Site*	"Premutation"	Actual Sequence
<i>Polyrhachis</i> (5 species)	FS333	GGA GTA TTA G V L	GGG AGT ATT G S I
<i>Polyrhachis</i> (2 species)	FS333	GGA GTA TCA G V S	GGG AGT ATC G S I
<i>P. dives</i>	FS333	GGA GTA ATA G V M	GGG AGT AAT G S N
<i>P. ypsilon</i>	FS333	GGA GTA ACA G V T	GGG AGT AAC G S N
<i>P. sexspinoso</i>	FS333	GGA GTA ACC G V T	GGT AGT AAC G S N
	FS411	GGA GCA ACC G A T	GGG AGC AAC G S N
<i>P. trapezoidea</i>	FS474	TGA GTA TGA W V W	TGG AGT ATG W S M
<i>P. phryne</i>	FS474	TGA GTA TGG W V W	TGG AGT ATG W S M
	FS773	ATA GTA ACC M V T	ATA GGT AAC M G N
Turtle (<i>Chrysemys</i>)	FS174	CTA GTA GCA L V A	CTG AGT AGC L S S
Ostrich (<i>Struthio</i>)	FS174	CTA GTA GCT L V A	CTC AGT AGC L S S
Falcon (<i>Falco</i>)	FS174	CTA GTA GCA L V A	CTC AGT AGC L S S

*Approximate position of insertion, from start of gene.

Fig. 1. Sequence context of +1 frameshift sites in *Polyrhachis* species (this study) and selected vertebrates (Mindell et al. 1998). The "premutation" is the sequence that results after removal of the inserted nucleotide, based on sequence comparisons with species not carrying the frameshift.

Polyrhachis. We analyzed two representatives of subgenera *Hemioptica* and *Polyrhachis* and three species of subgenus *Myrmatopa*. All seven members of these subgenera carried the insertion. Only one representative of subgenus *Myrmotherinax* showed the insertion; the other, *P. queenslandica*, did not. This subgenus is believed to be monophyletic (Rudy Kohout, personal communication). If so, the absence of FS333 in *P. queenslandica* must represent a secondary loss of the insertion. In addition, two of four representatives of subgenus *Myrmhopla* carried the FS333 insertion. This subgenus is divided into several species groups and is likely polyphyletic. A molecular phylogenetic study of these species will be published elsewhere.

In addition to the frameshift insertion at position 333, *P. sexspinoso* showed a second frameshift insertion at approximate position 411 (FS411). The consensus codon of the other 29 *Polyrhachis* species is GGA, although the third position is variable. The inferred insertion was GGN → GGG A (Fig. 1). The insertion could have been either a G prior to a third position A or an A following a GGG codon.

Two species showed a single nucleotide insertion at approximate position 474 (FS474; Fig. 1). These mutations appear to be an insertion of a G. The codon at that position is TGA in the other 28 species, so the inferred mutation was TGA → TGG A. One of these two species, *P. phryne*, also carries a second insertion at approximate position 773 (FS773). The consensus at that site is ATT but the site is variable. The mutation appears to be GTA → GGT A.

Potential for Secondary Structure Near the Frameshift Sites

We examined the regions surrounding the insertion sites for the potential to form stable secondary structures. Optimal folding energies at 25°C for 51-nt fragments, including 12 nt (4 codons) prior to the insertion site, the codon sustaining the insertion, and 36 nt (12 codons) downstream from each insertion, ranged from -8.2 to -12.4 kcal/mol for the 10 FS333 sites. The optimal folding energies for the other sites were -8.2 for FS411 of *P. sexspinoso*, -9.5 and -10.3 for FS473, and -7.9 for FS773 of *P. phryne*. For comparison, we examined the folding energies at 25°C for fragments of the same length in a 20-nt sliding window along the length of the entire *cytb* gene of *P. turneri*. Optimal folding energies across the gene ranged from 0.9 to -17.8, with a mean of -7.0 (more negative values indicating increased stability). Although the stabilities of the regions surrounding the frameshift insertions were above average, the optimal folding structures and energies varied among species. We conclude that there is no evidence for unusual secondary structure in the regions immediately downstream from the insertion sites that might account for translational stalls.

Codons Involved in the +1 Insertion Sites

Examination of the sequences in the regions of the four frameshift sites in ants, and that in vertebrates, reveals several similarities (Fig. 1). The insertion at all five sites occurred just prior to a GUA or GCA codon (or both). At four of the five sites, the insertion results in an AGY as the first downstream 0-frame codon. The frameshift insertion observed in vertebrates at approximate position 174 in the *nad3* gene (FS174; Fig. 1) has AGY as the first two downstream 0-frame codons.

The likely site of the insertion in two of the ant examples (FS333 and FS411) is a glycine (GGA) codon. The site of the FS474 insertion appears to be a second or third position of tryptophan (UGA). We therefore focus on these codons and the tRNA genes responsible for their translation.

tRNA Genes

We sequenced the tRNA gene involved in translation of GGN (glycine) codons from two *Polyrhachis* species. One species, *P.nr. abnormis*, carries a frameshift at position 333, in a glycine codon. The other species, *P. thusnelda*, is apparently free of frameshift mutations. We were particularly interested in asking whether there were any differences between the glycine tRNAs that might account for alternative

tRNA ^{Gly}		***
<i>P. thusnelda</i>	ATTTATATAGTATAAAATAAATCTATTACATTTAATTTCCAATTAATAAAATTTAAATATCTTAATATAAATA	
<i>P. nr. abnormis</i>	ATTTATATAGTATAAAT-----TATTACACTTAATTTCCAATTAATAAAATTTAAACAT-TTAATATAAATA	
tRNA ^{Val}		***
<i>P. thusnelda</i>	AAAAATTTTAGTTTAAATAAAA-TAATTCATTTACATAGAAGAGATTAATTAATTCATTAAATTTTA	
<i>P. nr. abnormis</i>	AAGA-TTTTAGTTTAAATAAAAATAATTCATTTACATAGAATAGATTAAT--ATTTTATTAATTTT	
tRNA ^{Ala}		***
<i>P. phryne</i>	AGGATTTTAGTTAATTAATAACATGATAGTCATTTAGAAATAATTTAATATCTACATATAAATCTTT	
<i>Lasius alienus</i>	AGGATTTTAGTTAAATTATAACATTTAAATTCATTTAAATTTATATTTTTTTT---ATATAAATCTTA	

Fig. 2. Sequences of mitochondrial tRNA^{Gly}, tRNA^{Val}, and tRNA^{Ala} genes in ants.

translation of these codons. Suppressor mutations of +1 frameshifts have been studied in *Salmonella* (Riddle and Roth 1970). In their study, at least four of six genes identified involve altered tRNA structures, and two of those are alterations in tRNA^{Gly-GGG} genes (Riddle and Carbon 1973). The tRNA^{Gly} sequences of the *Polyrhachis* species are compared in Fig. 2. In both species, the anticodon gives an exact Watson-Crick match to GGA codons. No differences in the anticodon loops were found. The few differences in other parts of the molecules give no clues concerning a possible mechanism for +1 frameshift read-through. This result is not surprising, of course, since this tRNA must correctly translate glycine codons throughout the mitochondrial protein coding genes.

We were interested in the valine and alanine tRNAs since GUA and GCA may be involved in the +1 translation at these sites. In the *Polyrhachis* mitochondrial genome, the tRNA^{Val} anticodon in both of the species we sequenced is UAC (5' → 3'), giving an exact codon-anticodon match to GUA valine codons (Fig. 2). The anticodon of tRNA^{Ala} in the two ant species we examined is UGC (Fig. 2), giving an exact match to GCA codons.

In most of the mitochondrial +1 frameshift examples in ants and vertebrates, codons AGY appear in the immediate downstream position(s). The tRNA responsible for decoding these codons is somewhat challenging to locate, since it lacks the DHU stem and does not form a standard cloverleaf structure (Clary and Wolstenholme 1985; Crozier and Crozier 1993; Stewart and Beckenbach 2003). Structures representing the minimum energy folds for this gene from *P. phryne* and *Lasius alienus* are shown in Fig. 3. These species differ by over 23% of nucleotide sites in the *cox2* and *cytb* genes. Comparison of these sequences shows only a single substitution (G ↔ A transition) in the paired stem regions, yielding a G-U versus an A-U pair at the base of the anticodon stem. There are no changes in the anticodon loop, but the other unpaired regions are not alignable between these species. The apparent structural conservation between these fairly distant species and the appro-

priate UCU anticodon predicted from these structures indicate that these sequences are the genes responsible for decoding AGN codons in ants.

Discussion

We have presented evidence for single nucleotide frameshift insertions at four sites within the mitochondrial *cytb* gene of 12 species of the ant genus, *Polyrhachis*. The principal questions are whether these sequences are functional mitochondrial *cytb* genes, rather than pseudogenes, and if they are functional, what the likely mechanism for their translation is.

Evidence That the Frameshifted Sequences Are Functional

Two of the insertion sites, FS333 and FS474, were observed in 10 and 2 species, respectively. Nine of the ten FS333 insertions are G. The possibility that these insertions are independent insertions of G at this site during the recent creation of a pseudogene in each species is remote. Thus we assume the FS333 sequences are descendants of a single sequence in a common ancestor of these 10 species. The evolution observable among these sequences must have occurred subsequent to the insertion event. Evolution of functional protein coding genes is easily distinguishable from evolution of pseudogenes. Thus the pattern of substitutions subsequent to the frameshift mutation provides clear evidence of their function. We use a similar logic for the FS474 sequences.

We compare the sequence evolution in the two genes, *cox2*, where no indels were observed, and *cytb*, in species with and without frameshift insertions. Table 3 gives the pairwise distances for these two genes, corrected for multiple substitutions using Jukes-Cantor. Across the genus, *cox2* shows a mean of about 18% sequence divergence between species, while *cytb* shows about 21%. The variation observed in *cytb* is higher than that observed in *cox2* in 393 of 435 comparisons. The pattern of substitution in both

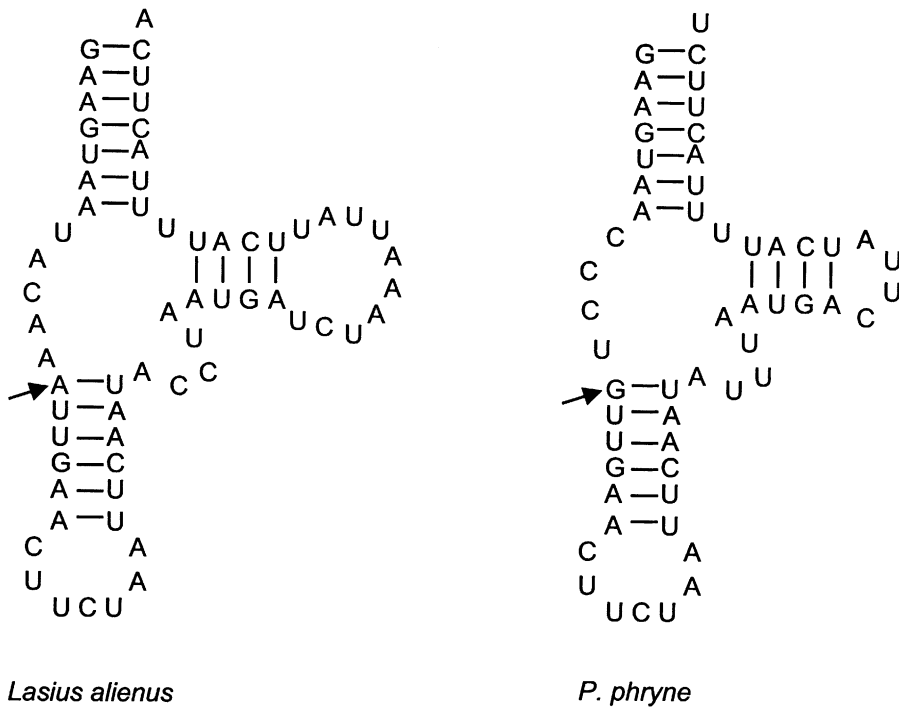


Fig. 3. Inferred secondary structures of tRNA^{Ser-AGN} of two species of ants. Optimal folds for *Lasius alienus* ($\delta G = -13.4$ kcal/mol) and *P. phryne* ($\delta G = -13.2$ kcal/mol) predicted by mfold (Zuker 2003).

Table 3. Mean pairwise Jukes–Cantor distances between *Polyrhachis* species

Gene	Comparison	<i>N</i>	Overall	1st position	2nd position	3rd position
<i>cox2</i>	FS474	2	13.39	6.49	1.57	36.69
	FS333	10	18.36	10.61	4.78	46.12
	All FS species	12	17.83	9.72	4.51	45.72
	Non-FS species	18	16.28	9.58	2.87	42.33
<i>Cytb</i>	FS474					
	Start ^a	2	19.52	9.13	5.68	51.99
	End ^a	2	16.92	10.64	1.67	45.08
	Entire region ^c	2	18.17	9.82	3.69	48.28
	FS333					
	Start ^b	10	22.19	13.74	5.87	57.20
	End ^b	10	21.83	12.05	3.06	63.98
	Entire region ^c	10	21.83	12.52	3.94	60.93
	All FS species ^c	12	21.50	12.19	4.02	59.67
	Non-FS species	18	20.13	11.70	3.36	55.34

Note. *cytb* sequences of species carrying frameshift insertions are separated into the region prior to the frameshift codon (start), and the region beyond the main frameshifts, in the putative original reading frame.

^aBased on segments of 384 sites prior to position 472 and 363 sites after position 475.

^bBased on segments of 243 sites prior to position 331 and 504 sites after position 334.

^cEntire region of 750 nucleotides, after “correction” to restore the original reading frame.

genes is similar, with a ratio of about 2:1:4 by codon position in the normal reading frame for both sequences. This pattern is typical of insect mitochondrial protein coding genes evolving under constraints required to maintain a full-length, functional protein product (Beckenbach and Borkent 2003).

The observed genetic distance between the two species carrying FS474, *P. trapezoidea* and *P. phryne*, is 13.4% for *cox2* and 18.2% *cytb* (Table 3). Evidently a substantial amount of sequence evolution has occurred in both genes since these two species separated. If we “repair” the *cytb* gene by deleting the

extra nucleotide from the data file to restore the original reading frame, and examine the pattern of substitution by codon position, it is evident that the two genes have a similar pattern of substitution. In particular, the substitution pattern in *cytb* is approximately 2:1:4. Thus the sequence evolution in *cytb* has continued in these sequences under constraints expected in the original reading frame.

A similar argument can be made for the ten species in the FS333 group (Table 3). Comparable bias of substitutions at the three codon positions is evident in both genes, suggesting that the sequences are the

functional mitochondrial genes for *cytb*. The original reading frame appears to be maintained despite the frameshift insertion(s).

To rule out the possibility that a truncated portion of the gene, coding for either the amino-terminal or the carboxy-terminal regions, is evolving under constraints of a functional reading frame we also separate out the regions prior to the insertions and those beyond the insertions, in the original reading frame (Table 3). Similar substitution bias is evident both upstream and downstream from the insertion sites, for both FS333 and FS474. Therefore the entire gene is evolving as a functional protein coding gene requiring a shift to the +1 reading frame at the site of the insertion(s).

Mutational Effects of Frameshifts

There are several consequences of frameshift mutations that are distinct from those of replacement substitutions. Frameshift mutations affect not only the site of the mutation, but the translation of all residues downstream. A frameshift insertion results in a defective polypeptide product unless it is corrected either by processing of the mRNA to remove the frameshift or through some mechanism to read through the frameshift during translation, shifting to the correct reading frame. In the frameshifted sequences observed in this study, the predicted products would be truncated to less than half the length of the normal *cytb* protein if translated normally (Table 1). For the FS333 species, only the first 111 codons of the 372 required for a wild-type product would be translated correctly. In contrast, if a stall during translation causes a +1 translational frameshift at the site of the insertion, the entire 372-residue protein may be translated correctly, yielding a *wild-type* product. This result is quite different from the effect of a nonsynonymous nucleotide substitution, where translation of the mutated message produces no wild-type product.

In this context, it is important to note that under normal conditions, translation is error prone. A processivity error rate of 2.5×10^{-4} per codon has been estimated for wild-type *E. coli* under ideal laboratory growth conditions (Kurland 1992). For a protein of 372 residues, about 9% of products would fail to terminate correctly, through premature termination or read-through of the terminator. These errors include spontaneous translational frameshifts, which have been estimated to occur at about 5×10^{-5} per codon (Kurland 1992). Improperly terminated products are subject to proteolytic degradation and are part of normal gene function. The frequency of frameshift errors during translation is strongly sequence dependent. Specific sequences ("programmed

frameshift sites") may undergo translational frameshifting at rates three or four orders of magnitude higher than sequences at nonprogrammed sites and may exceed 0-frame translation (Farabaugh 1996; Gesteland and Atkins 1996). A frameshift mutation in a protein coding sequence that generates a compensating (programmed) translational frameshift will produce at least some wild-type product. The efficiency of the frameshift translation depends on both the sequence and the physiological state of the cell. Experimental manipulation of the abundance of the cognate tRNA for 0-frame decoding of the first downstream codon can alter the efficiency of translational frameshifting for certain sequences in yeast, from a few percent to almost 70% (Sundararajan et al. 1999).

As noted above, a unique feature of frameshift mutations, compared to missense mutations, is that any full-length products generated from a compensating translational frameshift are wild-type. That is, all full-length, properly terminated products should be fully functional, not mutant. Products produced when the translation fails to shift into the +1 reading frame will truncate prematurely and be subject to degradation. Thus the fitness effect depends on the efficiency of the programmed frameshift in taxa carrying a frameshift, relative to the accuracy of translation of the wild-type message in taxa with a full-length open reading frame.

A Model For Decoding of the FS333 Sequences

In all of the well-studied systems, including the TY1 and TY3 frameshifts in yeast, there are two requirements for relatively efficient translational frameshifting (Sundararajan et al. 1999). First, there must be a pause at the frameshift site during translation. Second, the tRNA bound at the peptidyl site (P-site), carrying the nascent polypeptide, must have a poor wobble pairing in the third codon position. The translational pause may be caused by slow decoding of the codon in the amino acyl site (A-site) or by the presence of stable secondary structure just downstream from the frameshift site (Ivanov et al. 1998). We found little evidence, however, for unusually stable secondary structure in the mRNA sequences downstream from the *Polyrhachis* frameshift sites. Thus we focus on the codon-anticodon interactions.

While the conditions noted above are required, the efficiency of a frameshift during translation may be enhanced if the tRNA bound at the P-site is able to slip to a +1 position having exact cognate recognition, followed by rapid decoding of the codon in the +1 position at the A-site (Hansen et al. 2003). Most of the *Polyrhachis* and vertebrate mtDNA frameshift sites meet these requirements. A model for decoding

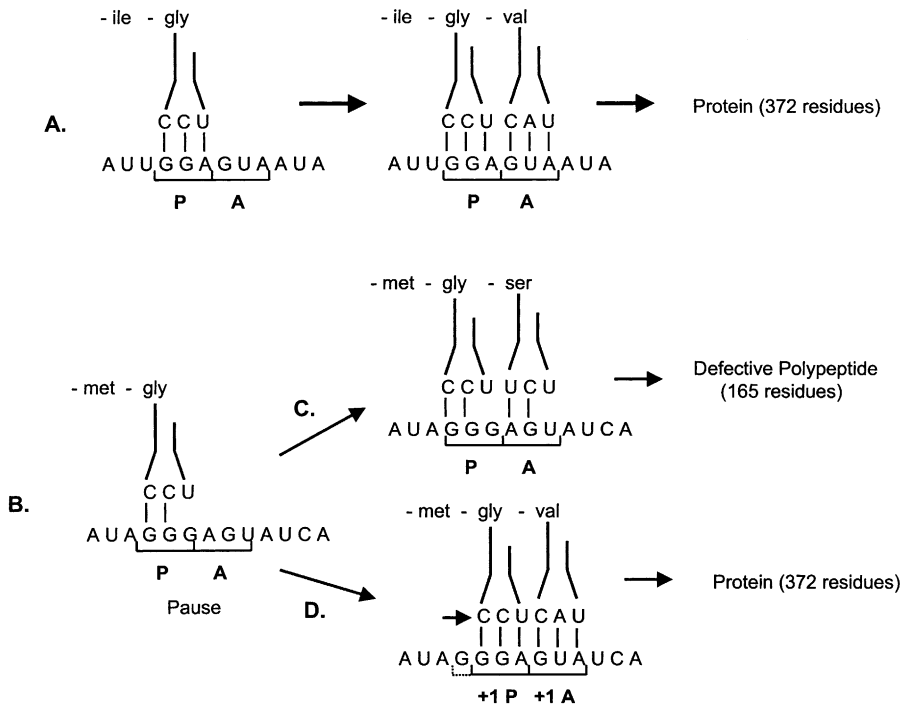


Fig. 4. Model for translational frameshifting: alternative outcomes of translation. **A** In species without the frameshift, cognate recognition at both the peptidyl (P) and the amino acyl (A) sites allows rapid continuation of translation in the current reading frame. **B** Species carrying the FS333 insertion require a near-cognate decoding of the GGG codon at the P-site, followed by near-cognate decoding of the poorly recognized AGU at the A-site. The resulting pause allows two competing paths. **C** Decoding of AGU at the A-site leads to a continuation in the 0 reading frame, resulting in a truncated product of only 165 residues. **D** Slippage of the ribosome to the +1 frame allows cognate decoding of the GGA at the P-site, and rapid cognate decoding of the GUA (valine), yielding a wild-type protein of 372 residues.

of the FS333 sites is depicted in Fig. 4. This model follows those presented in Sundararajan et al. (1999) and Hansen et al. (2003). In species without the frameshift, exact cognate recognition of the GGA (glycine) codon by tRNA^{Gly} (anticodon, UCC), followed by rapid decoding of the next codon (GUA) in the normal reading frame (0-frame), allows translation to proceed through the site without pause (Fig. 4A).

In nine of the FS333 species, the frameshift site is GGG AGU A (Fig. 1). The GGG codon must be decoded by tRNA^{Gly} (anticodon, UCC). We hypothesize that slow decoding of the AGU codon in the A-site generates a pause (Fig. 4B) and sets up two competing pathways (Figs. 4C and D). If a charged tRNA^{Ser-AGN} binds at the A-site, translation continues in the 0 reading frame, resulting in a defective polypeptide product (Fig. 4C). That product would be degraded. Alternatively, if the tRNA^{Gly} bound in the P-site slips to a +1 position prior to binding of a tRNA in the A-site, the next (+1) codon will be rapidly decoded by exact cognate recognition of tRNA^{Val} (Fig. 4D). Translation can then continue in the +1 reading frame to produce a fully wild-type product. The fitness effect of the mutation must depend on the efficiency of the programmed frameshift—the proportion of translations that follow path 4D.

An almost-identical model can account for +1 decoding of the FS411 site (sequence GGG AGC A; Fig. 1), except that rapid cognate decoding of the +1 A-site requires tRNA^{Ala} instead of tRNA^{Val}. Decoding of FS474 (sequence, UGG AGU A, in both species; Fig. 1) requires some modification of the

model, since the +1 P-site codon (GGA) is not an exact match for tRNA^{Trp}. This site may be comparable to the frameshift site in the yeast TY3 retrotransposon. Alternative models have been proposed for decoding of the frameshift in TY3, where the tRNA bound at the P-site cannot slip to an exact cognate binding in the +1 position. Farabaugh et al. (1993) suggested that out-of-frame (+1) binding of an abundant tRNA at the A-site would allow frameshifted read-through of this sequence without P-site slippage, while Hansen et al. (2003) argued that slippage was the likely mechanism even when the tRNA bound at the P-site can form only two Watson-Crick base pairs at the +1 codon. Either of these models can explain translational frameshifting at the FS474 site.

A critical part of the model is the role of the tRNAs. For efficient +1 translational frameshifting, the tRNA that has already bound and shifted to the P-site (which therefore cannot be exchanged), must have a suboptimal codon-anticodon pairing. The codon in the A-site at that stage in translation must be slowly recognized to generate a stall. This second requirement has been experimentally studied by Vimaldithan and Farabaugh (1994) in yeast, where elimination of the tRNA^{Arg-AGG} gene required for cognate recognition of the AGG codon in the A-site increased the efficiency of programmed frameshifting by more than threefold. Thus translation pauses until a “near-cognate” tRNA binds or a +1 frameshift allows translation to continue in the new reading frame. In insect mitochondria there are only 22 tRNAs to decode 62 sense codons. Since most (or all)

Table 4. Codon usage in *Polyrhachis* and representative vertebrates

Amino acid	Codon	<i>cox2</i>	<i>cytb</i> ^a	Ostrich ^b	Turtle ^b
Valine	GUU	34	75	48	27
	GUC	17	26	39	24
	GUA	42	76	72	100
	GUG	3	5	22	16
Serine—UCN	UCU	220	220	41	44
	UCC	89	69	69	62
	UCA	160	163	106	116
	UCG	6	4	13	7
Serine—AGN	AGU	18	12	16	14
	AGC	8	3	41	29
	AGA	65	84	—	—
	AGG	0	3	—	—
Tryptophan	UGA	143	203	82	101
	UGG	7	9	27	12
Glycine	GGU	51	82	40	47
	GGC	27	67	58	55
	GGA	64	255	83	97
	GGG	11	54	38	28

Note. Frequencies of selected codons from a 579-nt region of cytochrome oxidase subunit 2 (*cox2*) and a 750-nt region of cytochrome b (*cytb*) are given for 30 species of *Polyrhachis*. Corresponding codon usage frequencies for a bird and turtle are given for all protein coding genes. Codons decoded by cognate recognition of the tRNA are in bold face.

^aThe 12 species showing frameshift insertions were “corrected” prior to translation.

^bTotals for 13 protein coding genes.

of these sense codons appear in mitochondrial genes, suboptimal (near-cognate) recognition is actually more common than exact cognate pairing (see Table 4 for examples). For the translation of nuclear genes there is a correlation between the availability of charged tRNAs and the number of copies of each tRNA gene in the genome (Percudani et al. 1997). In animal mitochondrial genomes there is exactly one copy of each tRNA gene per genome. Thus we cannot fall back on stoichiometry at the genic level to define “hungry” (slowly decoding) codons. Instead we have to look at relative codon usage under the assumption that codon usage reflects translational efficiency. Table 4 gives codon usage for codon families relevant to these sites for the *cox2* and *cytb* genes of *Polyrhachis* and for all genes in two vertebrates. The codons that are decoded by exact cognate binding (where known) are in boldface.

The first point to note is that serine is more commonly coded by UCN rather than AGN (about a five fold difference in *Polyrhachis cox2* and *cytb*). We also note that GGG (glycine—FS333 and FS411) and UGG (tryptophan—FS474) are rarely used. The frameshift sites themselves present an interesting contrast to these general observations. Among FS333 species there are 15 glycine (GGN) codons within the region of *cytb* studied here (Table 5). Across these species, the third position is most often A (53%), giving exact cognate recognition by the *Polyrhachis* glycine tRNA. There is, however, considerable variation at 14 of these sites among the FS333 species. The exception is the frameshift site itself (FS333). At

this site (and at glycine FS411) the cognate codon (GGA) is not present in any of the sequences. Nine of the ten FS333 have GGG and the other, *P. sexspinosa*, carries GGU at the FS333 site and GGG at the FS411 site. Under the model for efficient programmed translational frameshifting (Fig. 4) a third position mismatch is *required* at these sites. A simple G → A transition would eliminate that mismatch and strongly favor 0-frame translation (Fig. 4C). This mutation would likely not be viable and is not observed at any of the sites. Evidently the conservation of a mismatch at the third position is required for efficient decoding of these *cytb* genes.

The same argument can be made for the FS474 site (Table 6). A third position G is generally avoided in the UGR codons at the 12 sites in the two genes examined here. Both FS474 species have UGG at the frameshift site. Although we did not sequence the tRNA responsible for decoding tryptophan in ants, in most insects and vertebrates the anticodon is UCA, with cognate recognition of UGA codons. The conservation of UGG where frameshifting is required for translation of these genes is consistent with the requirement for “near-cognate” codon–anticodon interactions at the P-site at this position.

Fitness Considerations

In all of the well-studied cases of translational frameshifting, the requirement for a frameshift during translation appears to serve a functional role: to

Table 5. Third position variation at GGN codons in 30 species of *Polyrhachis*

Gene	Position	FS333 species				Nonframeshift species			
		U	C	A	G	U	C	A	G
<i>Cytb</i>	99	0	1	9	0	6	0	13	1
	129	0	1	9	0	1	1	17	1
	213	2	2	6	0	6	9	5	0
	243	4	1	2	3	6	0	7	7
	288	1	2	6	1	5	2	7	6
	333 ^a	1	0	0	9	1	1	17	1
	375	2	0	8	0	1	2	12	5
	393	2	2	6	0	4	2	9	5
	411 ^b	1	0	5	4	4	2	9	5
	456	3	5	1	1	11	6	3	0
	483	2	0	5	3	1	1	12	6
	486	1	2	5	2	0	1	19	0
	597	0	1	7	2	0	5	15	0
	615	2	3	5	0	6	2	12	0
	678	1	2	6	1	7	7	5	1
Total		22	22	80	26	59	41	162	38

Note. Ten of the species carry +1 frameshift insertion(s) in the GGN codon(s) of *cytb* (FS333); the other 20 are designated nonframeshift species.

^aFrameshifts in all 10 taxa of the FS333 group.

^bIncludes a frameshift in *P. sexspinosa* which has GGG at position 411.

Table 6. Third position variation at UGR codons in 30 species of *Polyrhachis*

Gene	Position	FS474 species		Nonframeshift species		
		A	G	A	G	
<i>Cytb</i>	214	2	0	28	0	
	322	2	0	28	0	
	388	2	0	25	3	
	406	2	0	28	0	
	472 ^a	0	2	28	0	
	478	1	1	26	2	
	799	2	0	26	2	
	<i>cox2</i>	12	2	0	27	1
		192	2	0	26	2
309		2	0	28	0	
315		2	0	28	0	
486	1	1	25	3		
Total		20	4 ^b	323	13	

Note. Two of the species carry +1 frameshift insertion(s) in the UGR codon(s) of *cytb* (FS474); the other 28 are designated nonframeshift species.

^aFrameshifts in both FS474 taxa.

^bIncludes two frameshifts in the FS474 taxa.

down-regulate the gene except under certain physiological conditions. These examples include the vertebrate Ornithine Decarboxylase Antizyme gene, the *E. coli* Release Factor 2 gene, and the TY1 and TY3 retrotransposons in yeast (reviewed by Farabaugh 1996; Gesteland and Atkins 1996). Gurvich et al. (2003) recognized that the sequences that promote elevated levels of translational frameshifting may occur by chance in genes that do not require a

frameshift. In these cases, any frameshift products would be nonfunctional, and the presence of such sequences may actually reduce fitness. They showed that such sequences undergo significant levels of translational frameshifting and are underrepresented *E. coli* genes. An alternative coding for the same amino acid sequence is overrepresented. The implication is that these sequences generate high levels of frameshifted product, whether desirable or not.

It is difficult to see a functional role for programmed frameshifts in animal mitochondrial genes. The same set of genes is present in nearly all animal mitochondrial genomes. The only examples that have been identified, where a translational frameshift is required to translate these genes, are in the *nad3* gene of birds and a turtle and the *cytb* gene of ants. These cases appear to be exceptional. As well, there is phylogenetic evidence that the frameshift insertion has been lost multiple times during evolution. These observations argue against a functional role for the frameshifts we have observed. We suggest, instead, that certain sequence contexts can tolerate a single nucleotide frameshift insertion, without excessive harm. Since the mitochondrial genome is haploid, the organism and her descendants must live with it. For eusocial insects, such as ants, that includes the queen, her colony, and any descendant colonies.

Conclusions

The discovery of a single nucleotide +1 frameshift insertion at a single site in many birds and a turtle

(Mindell et al. 1998) confronts us with two apparently unlikely alternatives: either (1) the mutation arose as a unique event in a common ancestor of birds and turtles or (2) identical deleterious mutations have been fixed two or more times independently during evolution. In the first case, we must infer that the mutation has been retained in many birds and a turtle for at least 250 million years, despite apparent harmful effects. The alternative appears even less likely: the fixation of a deleterious insertion at precisely the same site in a molecule of more than 16 kb, in two or more lineages independently.

Since the vast majority of frameshift mutations in essential genes are rejected by selection, the sites where a +1 insertion produces a viable result must be limited to sequences where the gene can still be correctly translated. We scanned the protein coding sequences of the ostrich (Haddrath and Baker 2001; GenBank Accession NC 002785) and painted turtle (Mindell et al. 1999; NC 002073) for occurrences of consecutive codons of NNA GTA GYN, where a single nucleotide insertion prior to the first A will generate consecutive AGU AGY codons immediately downstream in the transcribed mRNA. Aside from the *nad3* site present in the putative premutation sequence, there are seven occurrences in the ostrich and eight in the turtle. If these are the only sites that can generate a *viable* frameshift mutation, then frameshifts occurring at all other sites will be filtered out by evolution.

The evidence presented here shows that four different sites in *cytb* of ants have independently sustained +1 insertions in the relatively brief history of the genus *Polyrhachis*. Evidently +1 mutations do arise from time to time, but the only ones we can observe will be those that can be correctly translated to produce an adequate level of full length, functional product.

A corollary to our hypothesis that consecutive AGU AGY codons can generate a high incidence of +1 frameshift products during translation in birds and turtles is that this sequence should be avoided in all mitochondrial genes. We examined all of the protein coding genes of both ostrich and turtle sequences. This sequence occurs in frame in only one place, in the ostrich *nad4l* gene. A +1 frameshift product of this gene would be severely truncated and probably not functional.

An unexpected result of our phylogenetic study of *Polyrhachis* was the finding of not just one, but multiple sites carrying a +1 frameshift insertion in different species of a single genus. We suggest that frameshift mutations continuously arise, but are rarely fixed. Ohta (1977) has shown that the likelihood of fixation of mildly deleterious mutations is inversely proportional to the effective population size. Eusocial insect groups, such as ants of the genus *Polyrhachis*, with an inherently small effective population size (Pamilo and Crozier 1997; Wilson 1963),

may be much more susceptible to the fixation of deleterious, but survivable, mutations. Thus close study of these species may provide us with a wealth of examples that demonstrate the limits and constraints of evolution at the molecular level.

Supplementary Material

The sequences for *P. turneri* have been deposited in GenBank under accession numbers AY437886 (*cytb*) and AY437887 (*cox2*); partial sequences of the two genes for the other 29 *Polyrhachis* species are available from GenBank under accession numbers AY442226–AY442254 (*cytb*) and AY443390–AY443418 (*cox2*). Voucher specimens are retained in the collections of the Queensland Museum and S.K. Robson.

Acknowledgments. This work was funded by a grant from the NSERC of Canada (to A.T.B.) and grants from the Australian Research Council and James Cook University (to S.K.A.R. and R.H.C.). We wish to thank Rudy Kohout, Queensland Museum, for his assistance in providing samples and identifying material and M.J. Smith for comments on the manuscript. We thank two anonymous reviewers for providing valuable insight and considerable clarification of underlying mechanisms of programmed translational frameshifting.

References

- Beckenbach AT, Borkent A (2003) Molecular analysis of the biting midges (Diptera: Ceratopogonidae), based on mitochondrial cytochrome oxidase subunit 2. *Mol Phylogenet Evol* 27:21–35
- Beckenbach AT, Wei YW, Liu H (1993) Relationships in the *Drosophila obscura* species group, inferred from mitochondrial cytochrome oxidase II sequences. *Mol Biol Evol* 10:619–634
- Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code. *J Mol Evol* 22:252–271
- Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: Complete sequence and genome organization. *Genetics* 133:97–117
- Farabaugh PJ (1996) Programmed translational frameshifting. *Microbiol Rev* 60:103–134
- Farabaugh PJ, Zhao H, Vimaladithan A (1993) A novel programmed frameshift expresses the *POL3* gene of retrotransposon Ty3 of yeast. *Cell* 74:93–103
- Gesteland RF, Atkins JF (1996) Recoding: Dynamic reprogramming of translation. *Annu Rev Biochem* 65:741–768
- Gurvich OL, Baranov PV, Zhou J, Hammer AW, Gesteland RF, Atkins JF (2003) Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli*. *EMBO J* 22:5941–5950
- Haddrath O, Baker AJ (2001) Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proc R Soc Lond B Biol Sci* 268:939–945
- Hansen TM, Baranov PV, Ivanov IP, Gesteland RF, Atkins JF (2003) Maintenance of the correct open reading frame by the ribosome. *EMBO Rep* 4:499–504

- Härlid A, Janke A, Arnason U (1997) The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. *Mol Biol Evol* 14:754–761
- Kurland CG (1992) Translational accuracy and the fitness of bacteria. *Annu Rev Genet* 26:29–50
- Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964
- Mindell DP, Sorenson MD, Dimcheff DE (1998) An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol Biol Evol* 15:1568–1571
- Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Syst. Biol* 48:138–152
- Ohta T (1977) Extension to the neutral mutation random drift hypothesis. In: Kimura M (ed) *Molecular evolution and Polymorphism*. National Institute of Genetics, Mishima, pp 148–167
- Pamilo P, Crozier RH (1997) Population biology of social insect conservation. *Mem Mus Victoria*. 56:411–419
- Percudani R, Pavesi A, Ottonello S (1997) Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J Mol Biol* 268:322–330
- Riddle DL, Carbon J (1973) Frameshift suppression: A nucleotide addition in the anticodon of a glycine transfer RNA. *Nature New Biol* 242:230–234
- Riddle DL, Roth JR (1970) Suppressors of frameshift mutations *Salmonella typhimurium*. *J Mol Biol* 54:131–144
- Stewart JB, Beckenbach AT (2003) Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle, *Crioceris duodecimpunctata*. *Mol Phylogenet Evol* 26:513–526
- Sundararajan A, Michaud WA, Qian Q, Stahl G, Farabaugh PJ (1999) Near-cognate peptidyl-tRNAs promote +1 translational frameshifting in yeast. *Mol Cell* 4:1005–1015
- Tautz D, Renz M (1983) An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. *Anal Biochem* 132:14–19
- Vimaladithan A, Farabaugh PJ (1994) Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. *Mol Cell Biol* 14:8107–8116
- Wilson EO (1963) Social modifications related to rareness in ant species. *Evolution* 17:249–253
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406–3415