Recoding of Translation in Turtle Mitochondrial Genomes: Programmed Frameshift Mutations and Evidence of a Modified Genetic Code

R. David Russell · Andrew T. Beckenbach

Abstract A +1 frameshift insertion has been documented in the mitochondrial gene *nad3* in some birds and reptiles. By sequencing polyadenylated mRNA of the chicken (*Gallus gallus*), we have shown that the extra nucleotide is transcribed and is present in mature mRNA. Evidence from other animal mitochondrial genomes has led us to hypothesize that certain mitochondrial translation systems have the ability to tolerate frameshift insertions using programmed translational frameshifting. To investigate this, we sequenced the mitochondrial genome of the red-eared slider turtle (*Trachemys scripta*), where both the widespread *nad3* frameshift insertion and a novel site in *nad4l* were found. Sequencing the region surrounding the insertion in *nad3* in a number of other turtles and tortoises reveal general mitochondrial +1 programmed frameshift site features as well as the apparent redefinition of a stop codon in Parker's snake-neck turtle (*Chelodina parkeri*), the first known example of this in vertebrate mitochondria.

Keywords Programmed translational frameshifting ·
Mitochondrial genetic code · Testudines · Recoding ·
Codon redefinition

Abbreviations

cox1, *cox3*	cytochrome oxidase subunits 1 and 3
cytb	cytochrome b
nad3, *nad4*, *nad4l*, *nad6*	NADH dehydrogenase subunits 3, 4, 4l and 6

R. D. Russell
Department of Molecular Biology and Biochemistry,
Simon Fraser University, 8888 University Drive, Burnaby,
BC, Canada V5A 1S6

A. T. Beckenbach (✉)
Department of Biological Sciences, Simon Fraser University,
8888 University Drive, Burnaby, BC, Canada V5A 1S6
e-mail: beckenba@sfu.ca

Introduction

Virtually all eukaryote cells contain mitochondria, which carry a small genome coding for some of the proteins required for electron transport. To translate these genes, mitochondrial genomes maintain a minimal translation system, separate from the eukaryote translation system used for nuclear-encoded proteins. Detailed study of mitochondrial gene function has uncovered a wide range of anomalies, including modifications of the genetic code and editing of mRNA and tRNA sequences (reviewed by Burger et al. 2003).

Recent studies have found single nucleotide insertions in certain essential mitochondrial protein coding genes, resulting in frameshifts. The first example was discovered in the NADH dehydrogenase subunit 3 gene (*nad3*) of the ostrich mitochondrial genome (Härlid et al. 1997). This insertion, at about position 174 of the 348-nucleotide (nt) gene, would appear to result in a severe truncation of the *nad3* gene product. Mindell et al. (1998) surveyed a wide range of birds, as well as a painted turtle, for *nad3* gene sequence and found that 46 of 61 bird species, and the painted turtle, carried the extra nucleotide at that site. They hypothesized that the insertion in this sequence context fortuitously programmed the ribosome to jump the extra nucleotide, allowing translation of a complete protein product. For convenience, we refer to the widespread *nad3* frameshift at position 174 in birds and turtles as N3-174.

In 1998, Zardoya and Meyer reported the complete mitochondrial genome sequence of another turtle, the

African helmeted turtle (*Pelomedusa subrufa*). This species lacks the N3-174 insertion but carries single nucleotide insertions at three other sites. One is at approximately position 135 in the *nad3* gene (N3-135); the other two are in the *nad4l* gene, at approximately positions 99 and 262 from the start of the gene.

Beckenbach et al. (2005) discovered +1 insertions at one or two sites in the cytochrome *b* (*cytb*) gene of several species of *Polyrhachis* ants. A total of four different sites were affected. They were able to identify features common to all frameshift sites that give clues to the mechanism of their translation. They presented a model for the decoding of these mitochondrial frameshift sites. Milbury and Gaffney (2005) subsequently discovered another +1 frameshift insertion at a different site in the *cytb* gene of the eastern oyster. Parham et al. (2006b), in a survey of tortoise mitochondrial sequences, found the N3-174 insertion to be present in all species examined, as well as a second site in the pancake tortoise (*Malacochersus tornieri*), in the *nad4* gene. Rosengarten et al. (2008) reported two sites, one in *cox3* and one in *nad6*, in the mitochondrial genome of a glass sponge.

These observations suggest that +1 frameshift mutations in animal mitochondrial genomes are widespread (but rare) anomalies that are evidently tolerated by the mitochondrial translation systems of some animal species. Examination of the sequences at each of these sites, across a wide range of animal groups (four phyla) and in six different genes, reveals some remarkable similarities (Fig. 1). All well-documented examples are single-nucleotide insertions, shifting the correct reading frame to the +1 frame. All require a wobble pairing in the codon at the insertion site. In most, the inserted nucleotide creates an AGN codon in the original reading frame (0-frame) at the first site downstream from the insertion. In the oyster sequence, a stop codon (TAG) is generated, with AGG in the first +1 frame (Fig. 1). These features are conserved in taxa requiring a frameshift during translation, including nucleotide sites that are variable in related taxa that lack the frameshift insertions (Mindell et al. 1998; Beckenbach et al. 2005).

A number of the requirements for efficient translational frameshifting have been experimentally determined. Well-studied examples are the release factor 2 (*prfB*) gene in *E. coli*, two different yeast transposable (Ty) elements, and the mammalian ornithine decarboxylase antizyme gene (reviewed by Farabaugh 1996; Gesteland and Atkins 1996). In these examples, as in every translational system that tolerates frameshifts, only specific sequences will frameshift. One of the most common conserved elements is a stall in translation, due to either a slowly decoding codon, a strong mRNA secondary structure (pseudoknot), or both. In *E. coli*, the *prfB* gene requires a +1 frameshift early in

Organism	Gene	0-Frame	Shift to +1 Frame
Turtles:			
Painted Turtle (*Chrysemys*)	*nad3*	CTG AGT AGC L S S	CTG A GTA GCA L V A
Pancake Tortoise (*Malacochersus*)	*nad3*	CTC AGT AGC L S S	CTC A GTA GCA L V A
	nad4	GGC AGT AGA G S *	GGC A GTA GAT G V D
Side-neck Turtle (*Pelomedusa*)	*nad3*	CTT AGA ACC L * T	CTT A GAA CCA L E P
	nad4l	CTT AGA AGG L * *	CTT A GAA GGC L E G
	nad4l	CTT AGA TAT L * Y	CTT A GAT ATA L D M
Birds:			
Ostrich (*Struthio*)	*nad3*	CTC AGT AGC L S S	CTC A GTA GCA L V A
Chicken (*Gallo*)	*nad3*	CTC AGT AGC L S S	CTC A GTA GCC L V A
Ants:			
Polyrhachis sexspinosa	*cytb*	GGT AGT AAC G S N	GGT A GTA ACC G V T
	cytb	GGG AGC AAC G S N	GGG A GCA ACC G A T
Polyrhachis phryne	*cytb*	TGG AGT ATG W S M	TGG A GTA TGG W V W
	cytb	ATA GGT AAC M G N	ATA G GTA ACC M V T
Eastern Oyster:			
Crassostrea virginica	*cytb*	TAT TAG GGG Y * G	TAT T AGG GGC G S G
Glass Sponge:			
Aphrocallistes vastus	*cox3*	TCC TGG AAT S W N	TCC T GGA AAT S G I
	nad6	GTA TGG AGC V W S	GTA T GGA GCC V G A

Fig. 1 Mitochondrially encoded frameshift mutations in animals. Sequences are shown in triplets corresponding to translation both in the 0-frame and after a shift to the +1 frame

translation to produce a complete *prfB* protein. Through amino acid sequence and mRNA comparisons, the ribosome has been shown to shift to the +1 frame over the sequence CUU UGA C (Craigen et al. 1985). Here the UGA stop codon in the 0-frame is thought to initiate a stall in translation. After successfully shifting one nucleotide, translation continues in the +1 frame, in effect reading the sequence as CUU U GAC. In yeast (*Saccharomyces cerevisiae*) the Ty1 and Ty3 elements are retrotransposons containing two genes, gag and pol. The 3′ end of the pol gene overlaps the first 38 nucleotides of gag in the +1 frame. In these elements, a gag-pol fusion polypeptide, whose production requires a frameshift, is an essential protein. In Ty1, the frameshift site is CUU AGG C, as written in codons of gag. The second codon in the frameshifting heptamer, AGG, is rarely used and is again thought to stall the ribosome (Belcourt and Farabaugh 1990; Farabaugh et al. 1993). In Ty3, the sequence surrounding

the frameshift is GCG AGU U in *gag*, again with the in-frame rarely used serine codon AGN. In the mammalian antizyme gene, a frameshift is required over the sequence UCC UGA N, where slow recognition of the UGA stop codon by the release factor produces the translational pause (Matsufuji et al. 1995).

Wobble pairing at the P-site when the stall occurs is seen in most examples as well (Baranov et al. 2002). In *prfB*, tRNA-Leu (CUN; anticodon GAG) base-pairs with CUU, a shift-prone codon (Curran 1993), in the 0-frame. Like *prfB*, the peptidyl codon in Ty1 is a leucine, decoded by the tRNA-Leu (CUN; anticodon UAG). The Ty3 element appears to be an exception: it uses a GCG as the P-site codon, which has canonical (exact Watson-Crick) pairing with its tRNA-Ala (anticodon CGC). Finally, in most cases, a rapidly decoding +1 codon that is an exact Watson-Crick match to its cognate tRNA enhances the probability of a frameshift at that site (Farabaugh 1996; Baranov et al. 2002).

These well-studied examples of programmed translational frameshifting have similarities that appear to be generic to all +1 shifts. All employ a rare or nonsense codon immediately after the last 0-frame codon. To frameshift at the required efficiency, a frameshift-capable tRNA for the P-site codon is required. In bacterial *prfB* and yeast Ty1, it is a tRNA that is able to slip easily and re-pair well in the +1 frame. In yeast Ty3, it appears either that a well-paired +1 codon:anticodon binding is not necessary or that there is some other property of the P-site tRNA that increases frameshifting levels, possibly by causing the incoming A-site tRNA to bind +1. The final element is the presence in the first +1 site of a commonly used codon that is often a Watson-Crick match to its cognate tRNA (Baranov et al. 2002). All of these requirements appear to be met in the animal mitochondria examples as well.

Gesteland et al. (1992) first suggested the term *recoding* to encompass all the events during translation and transcription that do not conform to the standard rules of decoding. These events include redefinition of codons as well as programmed frameshifts. The purpose of this paper is to examine a number of recoding events evident in vertebrate mitochondrial genomes, especially those of turtles. The first objective was to examine the mature mRNA for the *nad3* gene from a species carrying the widespread N3-174 insertion, to distinguish between translational frameshifting and RNA editing as a mechanism for decoding of this gene. We chose the domestic chicken because of easy availability of fresh tissue. This species is known to carry the N3-174 frameshift (Mindell et al. 1998). Next, we surveyed a wide range of turtle species for the presence of the N3-174 frameshift insertion, to examine its distribution in chelonians. Finally, we sequenced the entire mitochondrial genome of the

red-eared turtle (*Trachemys scripta*) to determine whether additional sites are present elsewhere in the genome of this species. This part of the study is motivated by the prediction that taxa which have a mitochondrial translation system that tolerates the N3-174 insertion may tolerate similar insertions elsewhere in the genome. The presence of multiple insertion sites has been previously reported in the African helmeted turtle (Zardoya and Meyer 1998), the pancake tortoise (Parham et al. 2006b), the *cytb* gene of two species of *Polyrhachis* ants (Beckenbach et al. 2005), and a hexactinellid sponge (Rosengarten et al. 2008).

Materials and Methods

Source of the Specimens

Blood and tissue samples of the turtles were obtained from three sources. Some specimens were collected from road-kills in the southeastern United States. Blood and tissue samples were also provided by the Reptile Refuge, in Surrey, British Columbia, and by the Empire of the Turtle, in Yalaha, Florida. Origins of these specimens and Gen-Bank accession numbers are given in Table 1. In addition, we compare these sequences to the corresponding region from published complete turtle mt genomes (Table 2). Taxon names given in these tables correspond to the recommendations of the Turtle Taxonomy Working Group (2007).

Isolation of Mitochondrial DNA and mRNA from Chicken Liver

Fresh chicken livers were obtained from a local slaughterhouse for extraction of mitochondrial DNA and mRNA. Intact mitochondria were isolated using standard methods. Briefly, liver tissue was homogenized in cold MSB buffer (210 mM mannitol, 70 mM sucrose, 50 mM Tris-HCL, pH 7.5, 10 mM EDTA), then cellular debris removed by centrifugation at 4000 rpm. The supernatant was centrifuged for 20 min at 20,000 rpm, then resuspended in MSB buffer and pelleted again. For DNA extraction, a portion of the mitochondrial pellet was resuspended in proteinase K buffer and genomic DNA extracted by standard phenol/chloroform/isoamyl alcohol procedure, followed by ethanol precipitation.

For the RNA extraction, care was taken to avoid RNase contamination. The Ambion® Inc. ToTALLY RNA RNA isolation kit was used following the manufacturer's protocol. After purification of the nucleic acids, contaminating DNA was removed by treatment with Ambion® TURBO DNase (RNase-free) using the manufacturer's protocol.

Table 1 Taxa included in this study

Turtle	Source	GenBank Accession
Trachemys scripta, red-eared slider	Reptile Refuge	FJ392294
Malaclemys terrapin, diamondback terrapin	Reptile Refuge	FJ392295
Clemmys guttata, spotted turtle	Reptile Refuge	FJ392296
Terrapene carolina, Florida box turtle	DOR, Fla.	FJ392297
Chelonoidis carbonaria, red-legged tortoise	Empire of the Turtle	FJ392298
Geochelone sulcata, African spurred tortoise	Reptile Refuge	FJ392299
Indotestudo forstenii, Travencore tortoise	Empire of the Turtle	FJ392300
Mauremys annamensis, Vietnamese pond turtle	Empire of the Turtle	FJ392301
Rhinoclemmys pulcherrima, painted wood turtle	Empire of the Turtle	FJ392302
Chelydra serpentina, common snapping turtle	DOR, Fla.	FJ392303
Macrochelys temminckii, alligator snapping turtle	Reptile Refuge	FJ392304
Staurotypus triporcatus, Mexican musk turtle	Empire of the Turtle	FJ392305
Sternotherus odoratus, common musk turtle	DOR, Fla.	FJ392306
Apalone ferox, Florida softshell turtle	DOR, Ga.	FJ392307
Batrachemys nasuta, toad-headed turtle	Empire of the Turtle	FJ392308
Chelodina parkeri, Parker's snake-necked turtle	Empire of the Turtle	FJ392309
Other reptiles		
Alligator mississipiensis, American alligator	DOR, Ga.	FJ392310
Caiman crocodilus, spectacled caiman	Reptile Refuge	FJ392311
Tantilla rubra, crowned snake	Big Bend, Tex.	FJ392312

Table 2 Published complete turtle and tortoise sequences included in this study

Species	GenBank RefSeq	Reference
Chrysemys picta, painted turtle	NC_002073	Mindell et al. (1999)
Geochelone pardalis, leopard tortoise	NC_007694	Parham et al. (2006b)
Indotestudo elongata, elongated tortoise	NC_007695	Parham et al. (2006b)
Indotestudo forstenii, Travencore tortoise	NC_007696	Parham et al. (2006b)
Malacochersus tornieri, pancake tortoise	NC_007700	Parham et al. (2006b)
Testudo graeca, Mediterranean tortoise	NC_007692	Parham et al. (2006b)
Testudo kleinmanni, Egyptian tortoise	NC_007699	Parham et al. (2006b)
Testudo horsfieldii, Horsfield's tortoise	NC_007697	Parham et al. (2006b)
Testudo marginata, marginated tortoise	NC_007698	Parham et al. (2006b)
Manouria emys, brown tortoise	NC_007693	Parham et al. (2006b)
Cuora mouhotti, keeled box turtle	NC_010973	Nie and Zhang, unpublished
Mauremys reevesi, Reeve's turtle	NC_006082	Nie, Pu, and Peng, unpublished
Mauremys mutica, Asian yellow pond turtle	NC_009330	Nie and Song, unpublished
Macrochelys temminckii, alligator snapping turtle	NC_09260	Nie and Yan, unpublished
Platysternon megacephalum, big-headed turtle	NC_007970	Parham et al. (2006a)
Chelonia mydas, green sea turtle	NC_000886	Kumazawa and Nishida (1999)
Pelomedusa subrufa, African helmeted turtle	NC_001947	Zardoya and Meyer (1998)

Reverse Transcription PCR of the Chicken *nad3* Gene

Poly(A) RNA was reverse transcribed using the Enhanced Avian HS RT-PCR Kit (Sigma), with a poly(T) primer of 24 nucleotides, paired with a primer that anneals 30–46 residues from the start of the *nad3* gene (5′-TCC TTTCTACTAAGCGC-3′). The RT-PCR reactions were prepared in RNase-free, certified 0.2-ml, thin-walled PCR tubes, with final concentrations of 200 μM of each dNTP, 3.0 mM $MgCl_2$, 0.4 μM of each primer, 0.4 unit/μl of RNase inhibitor enzyme, 0.4 μnit/μl of eAV-RT reverse transcriptase, and 0.05 unit/μl of Jumpstart AccuTaq LA DNA polymerase, in a 1 × reaction buffer with 0.4 ng of RNA extract. The RT protocol began with a 60-min incubation step at 42°C to enable reverse transcription, followed by 2 min at 94°C and 35 cycles at 94°C (15 s),

55°C (30 s), and 68°C (2 min), with a 5-min extension step at 68°C. Control reactions were run alongside the RT-PCR reactions in which each tube received 1 unit of RNase.

Analysis of the *nad3* Region in Reptiles

DNA was extracted from ground-up tissue using a standard phenol/chloroform/isoamyl alcohol procedure followed by ethanol precipitation. PCR was carried out using TaqPro and the manufacturer's recommended protocol. The N3-174 region was amplified using forward and reverse primers, 5′-CCCCATAYGAGTGYGGATTYGGATTYG ACCC and 5′-GCTCATTCTAGKCCTCCTTGRATCC. PCR cycling began with a 1.5-min denaturation at 94°C, then continued with four cycles of 20 s of denaturation at 93°C, 30 s of annealing at 45°C, and extension for 30 s at 72°C. Following the initial 4 cycles, 35 cycles were done, with the only difference being an anneal temperature of 50°C instead of 45°C. Some primer pairs produced non-specific results and required that the anneal temperature be raised. In these cases, the anneal temperature was raised to 52°C for all cycles, with 35 total cycles. All other temperatures and times were kept the same. Sequences were determined for both strands, using the amplification primers.

The *nad3* sequences were aligned by hand, using the BioEdit software package and an internally developed sequence editor. The only indels observed in this study were the single-nucleotide insertion mutations.

Sequencing of the Red-Ear Turtle Mitochondrial Genome

The mitochondrial genome was amplified in overlapping fragments using a combination of heterologous and taxon-specific primer pairs. Heterologous primers were designed using an alignment of published turtle and tortoise sequences. Once portions of the *T. scripta* genome were sequenced, sequence-specific primers were designed to amplify remaining sections. PCR cycling conditions were the same as described for amplification of the *nad3* region. All fragments were sequenced for both strands.

Analysis

The *Trachemys scripta* mitochondrial genome was aligned against other complete turtle mitochondrial genomes using BioEdit (Hall 1999). We annotated the genome based on comparisons to these and other vertebrate mtDNA sequences. All protein-coding genes were translated using the standard vertebrate mitochondrial code and examined visually for unusual features.

Results

The Chicken *nad3* Frameshift Site

To confirm the presence of the extra nucleotide reported by Mindell et al. (1998) and to determine if it is removed by an RNA editing process, a small region of the *Gallus gallus* mitochondrial genome around the *nad3* frameshift site was sequenced along with a corresponding region of the polyadenylated *nad3* mRNA transcript. Both sequences show the presence of the extra frameshift-causing nucleotide at position 174 in the *nad3* gene. The two sequences also align perfectly with the sequence reported for the chicken by Mindell et al. (1998). There was no evidence in the sequence traces of sequence lacking the frameshift mutation. This result appears to eliminate RNA editing as a possible mechanism for accurate *nad3* translation and suggests that compensation for the frameshift occurs through a translational mechanism, allowing it to be read through. To allow for the production of a functional *nad3* polypeptide, the ribosome somehow must be instructed to shift frames at this particular site and continue translation in the correct +1 frame.

The *nad3* gene in bird and turtle mitochondria is typically 348 nt or 116 codons in length. The N3-174 insertion is widespread in birds and turtles (Mindell et al. 1998). The *nad3* gene is essential, and functional translated proteins are required in all organisms. The ribosome therefore must have a relatively efficient way of translating over the frameshift disruption caused by the extra nucleotide.

The *nad3* Frameshift Region in Turtles

We wished to investigate whether there were any particular sequences or other features that are conserved in turtles having the frameshift nucleotide that may have a role in frameshift stimulation. This approach is especially powerful if we can subsequently show the absence of these elements in mitochondrial genomes without the extra nucleotide. To do this, we sequenced the region surrounding the frameshift site within the *nad3* gene in 21 different turtles, tortoises, and other reptiles (Fig. 2). The extra (frameshifting) nucleotide was present in 14 of these sequences, all chelonians. Within these taxa, all but the musk turtle (*Sternotherus odoratus*), Mexican giant musk (*Staurotypus triporcatus*), toad-headed turtle (*Batrachemys nasuta*), and the African helmeted turtle (*Pelomedusa subrufa*) showed the extra nucleotide. We also confirmed the presence of a different *nad3* frameshift site upstream from the common site in *P. subrufa*, as first reported by Zardoya and Meyer in 1998. None of the other reptiles investigated had any frameshift insertion mutations within their *nad3* genes. Close examination of the Parker's

Fig. 2 *Nad3* frameshift region in turtles and other reptiles included in this study. The sequences are shown from positions 132–180 relative to the red-eared turtle *nad3* sequence. The top sequence ("consensus") shows the nucleotide most common among the taxa, with the predicted translation below. The number scale indicates the position relative to the start of the *nad3* gene. Asterisks indicate nucleotides that are evidently skipped in taxa carrying the N3-135 and N3-174 frameshift mutations

```
                                    *     140          150          160          170        *     180
           Consensus       CTA - GAA TCT GCT CGC CTA CCA TTC TCA ATC CGA TTC TTC CTC A GTA GCA ATC TTA
                             L      E   S   A   R   L   P   F   S   I   R   F   F   L    V   A   I   L
           Chrysemys picta         TTA - GAA TCA GCT CGC TTG CCA TTC TCT ATC CGA TTC TTC CTG A GTA GCA ATC TTA
           Trachemys scripta       TTA - GAA TCA GCC CGA CTG CCA TTC TCC ATC CGG TTC TTC CTG A GTA GCA ATC TTA
           Malaclemys terrapin     TTA - GAA TCA GCT CGC CTA CCA TTC TCA ATC CGA TTC TTC CTG A GTA GCA ATC TTA
           Clemmys guttata         CTA - GAA TCA GCC CGC CTT CCA TTC TCA ATC CGA TTC TTC CTT A GTA GCA ATC TTA
           Terrapene carolina      CTA - GAA TCA GCC CGC CTT CCA TTC TCA ATC CGA TTC TTT CTC A GTA GCA ATC TTA
           Chelonoidis carbonaria  CTA - GAA TCT GCC CGT CTA CCG TTC TCA ATC CGA TTC TTT CTC A GTA GCC ATC CTA
           Geochelone sulcata      CTA - AAA TCT GCT CGT CTA CCA TTC TCA ATT CGA TTC TTC CTC A GTA GCA ATC TTA
           G. pardalis             CTA - GAA TCT GCT CGC CTA CCA TTT TCA ATC CGA TTC TTC CTC A GTA GCA ATC CTA
           Indotestudo forstenii   ATA - AAA TCT GCT CGC CTA CCA TTC TCA ATC CGA TTC TTC CTC A GTA GCA ATC TTA
           I. elongata             ATA - AAA TCT GCC CGC CTA CTA TTC TCA ATC CGA TTC TTC CTC A GTA GCA ATC TTA
           Malacochersus tornieri  CTA - GAA TCT GCT CGT CTA CCA TTC TCA ATC CAA TTT TTC CTC A GTA GCA ATC CTA
           Testudo graeca          CTA - AAA TCT GCT CGT CTA CCA TTC TCC ATC CGA TTC TTT CTT A GTA GCG ATT CTA
           T. kleimanni            CTA - AAA TCT GCC CGC CTA CCA TTC TCA ATC CGA TTC TTT CTT A GTA GCA ATT TTA
           T. horsfieldii          CTA - AAA TCT GCT CGT CTA CCA TTC TCA ATT CGA TTC TTC CTC A GTA GCA ATC TTA
           T. marginata            CTA - AAA TCT GCC CGC CTA CCA TTC TCA ATC CGA TTC TTT CTT A GTA GCA ATT TTA
           Cuora mouhotti          TTA - AAA ACC ACT CGT CCA CCA TTC TCA ATC CGA TTC ATT CTT A GTA GCA ATT TTA
           Manouria emys           CTA - GAG TCT GCT CGC CTA CCA TTC TCA ATC CGA TTC TTC CTC A GTA GCA ATT TTA
           Mauremys mutica         CTA - AAA ACC ACT CGT CCG CCG TTC TCA ATT CGA TTC TTT CTT A GTA GCA ATT CTA
           M. annamensis           CTA - AAA ACC ACC CGT CCA CCA TTC TCA ATT CGA TTC TTT CTT A GTA GCA ATT TTA
           M. reevesi              CTA - AAA ACC ACC CGT CCC CCA TTC TCA ATT CGA TTC TTC CTA A GTA GCA ATT TTA
           Rhinoclemmys pulcherrima TTA - AAA ACT ACT CGC CCA CCA TTC TCT ATC CGA TTT TTC CTC A GTA GCA ATC TTA
           Platysternon megacephalum TTA - GAA TCA GCT CGT TTA CCA TTC TCA ATC CAA TTT TTC CTA A GTA GCA ATC CTA
           Chelydra serpentina     TTA - GAA TCA GCT CAT CTA CCA TTC TCC ATC CGA TTC TTC CTC A GTA GCA ATC TTA
           Macrochelys temminckii  TTA - GAA TCA GCT CAC CTA CCA TTC TCA ATC CGA TTT TTC CTC A GTA GCA ATC TTA
           Staurotypus triporcatus TTA - GAA ACA GCT CAC CTC CCA TTT TCA GTA CGA TTT TTC TTA - ACA GTA ATC CTA
           Sternotherus odoratus   CTA - GAA CCA GCT CAC CTT CCA TTT TCA ATC CAA TTT TTC CTA - GTA GCA ATT CTA
           Chelonia mydas          TTA - GAA TCA GCT CGC CTA CCA TTC TCA ATC CGA TTC TTC CTC A GTA GCA ATT TTA
           Apolone ferox           TTA - AAA ACC ATA CGC CTA CCA TTC TCA ATC CGC TTC TTC CTC A GTA GCA ATT CTA
           Chelodina parkeri       CTT - GAA TCA ACT TGA CTG CCA TTC TCT ATT AGA TTC TTC CTT A GTA GCA ATC CTA
           Batrachemys nasuta      CTA - GAA TCA GCT TGA TTA CCT TTC TCA ATC TGA TTC TTC CTA - GTT GCT ATC CTC
           Pelomedusa subrufa      CTT A GAA CCA GCT TGA TTA CCA TTC TCA GTT CGA TTT TTC CTT - ATC GCA ATC TTA
           Alligator               CTC - GGC TCT GCC CTA CCC CTA TCA ACC CGG TTC TTT ATA - GTA GGC ATC CTA
           Caiman                  CTC - GGA TCT GCC CGC CTT CCC CTA TCA ATT CGA TTC TTC ATA - GTC GCC ATC CTA
           Tantilla coronata       ATA - GGA AAC GCC CGA ACC CCT ATC TCT ATC CAG TTT TTC CTA - GTC GCC ATT CTA
```

snake-necked turtle (*Chelodina parkeri*) reveals one additional feature. At the site of a conserved arginine codon in most other sequenced chelonians, this species has AGA (position 163-5; Fig. 2). We have confirmed this sequence from both strands and by careful examination of the sequence scans. AGA is interpreted as a termination codon in the vertebrate mitochondrial code.

We compare the 0-frame translation of these sequences with translation predicted by a model of programmed shift to the +1 frame at the N3-174 frameshift site in Fig. 3. In all taxa the 0-frame translation is altered at the frameshift site, and there is an AGA terminator in the 0-frame 11 codons downstream. The gene, which would normally produce a protein of 116 amino acids, is truncated to only 68 residues.

Fig. 3 Comparison of the 0-frame with frameshifting across the N3-174 insertion. **a** Translation in the 0-frame. **b** Translation with a +1 frameshift over position 175 in the *nad3* gene

```
                            A) Decoded in the 0-frame                          B) Translation over the extra nucleotide
           Chrysemys        ATC CGA TTC TTC CTG AGT AGC AAT CTT             ATC CGA TTC TTC CTG A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Trachemys        ATC CGG TTC TTC CTG AGT AGC AAT CTT             ATC CGG TTC TTC CTG A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Malaclemys       ATC CGA TTC TTC CTG AGT AGC AAT CTT             ATC CGA TTC TTC CTG A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Clemmys          ATC CGA TTC TTC CTT AGT AGC AAT CTT             ATC CGA TTC TTC CTT A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Terrapene        ATC CGA TTC TTT CTC AGT AGC AAT CTT             ATC CGA TTC TTT CTC A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Chelonoidis      ATC CGA TTC TTT CTC AGT AGC CAT CCT             ATC CGA TTC TTT CTC A GTA GCC ATC
                             I   R   F   F   L   S   S   H   P               I   R   F   F   L    V   A   I
           Geochelone       ATT CGA TTC TTC CTC AGT AGC AAT CTT             ATT CGA TTC TTC CTC A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Indotestudo      ATC CGA TTC TTC CTC AGT AGC AAT CTT             ATC CGA TTC TTC CTC A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Mauremys reevesi ATT CGA TTC TTT CTA AGT AGC AAT TTT             ATT CGA TTC TTT CTA A GTA GCA ATT
                             I   R   F   F   L   S   S   N   F               I   R   F   F   L    V   A   I
           M. annamensis    ATT CGA TTC TTT CTT AGT AGC AAT TTT             ATT CGA TTC TTT CTT A GTA GCA ATT
                             I   R   F   F   L   S   S   N   F               I   R   F   F   L    V   A   I
           Rhinoclemmys     ATC CGA TTT TTC CTC AGT AGC AAT CTT             ATC CGA TTT TTC CTC A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Chelydra         ATC CGA TTC TTC CTC AGT AGC AAT CTT             ATC CGA TTC TTC CTC A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Macrochelys      ATC CGA TTT TTC CTC AGT AGC AAT CTT             ATC CGA TTT TTC CTC A GTA GCA ATC
                             I   R   F   F   L   S   S   N   L               I   R   F   F   L    V   A   I
           Chelonia         ATC CGA TTC TTC CTC AGT AGC AAT TTT             ATC CGA TTC TTC CTC A GTA GCA ATT
                             I   R   F   F   L   S   S   N   F               I   R   F   F   L    V   A   I
           Apolone          ATC CGC TTC TTC CTC AGT AGC AAT TCT             ATC CGC TTC TTC CTC A GTA GCA ATT
                             I   R   F   F   L   S   S   N   S               I   R   F   F   L    V   A   I
           Chelodina        ATT AGA TTC TTC CTT AGT AGC AAT CCT             ATT AGA TTC TTC CTT A GTA GCA ATC
                             I   *   F   F   L   S   S   N   P               I   *   F   F   L    V   A   I
```

Conserved Features in *nad3* Frameshifting Genes

One of the common features of the N3-174 frameshift site in turtles appears to be the presence of two AGY serine codons immediately following the inserted nucleotide, put in-frame as a result of the insertion. These two codons, AGT followed by AGC, are conserved in all of the turtles and birds that carry the frameshift, but the corresponding nucleotides are variable in those that do not (Fig. 2). Codon usage for the six serine codons is given for three turtle species in Table 3. AGY is used for only 15–17% of serines across these species, and AGT codes only 5% of serines. The AGT appears to be the required stall-inducing, rarely used codon. The conservation of the second 0-frame codon, AGC, is particularly interesting. The 'A' corresponds to the third codon position of GTN in taxa that do not have the N3-174 frameshift mutation, and is variable in those species (Fig. 2). The conservation of the 'GC' in the second 0-frame codon either may be due to the need for an alanine codon (GCN) at the second +1 site or may indicate a role for those residues in frameshifting.

A second conserved feature found in all turtles and birds that carry the frameshift is a leucine codon as the last conserved 0-frame position, the codon that is at the P-site of the ribosome where the shift is thought to take place. The P-site codon is thought to play a critical role in stimulating a frameshift (Baranov et al. 2002). This codon is CTB in 26 of the 27 turtles known to require a frameshift,

where B is the extra nucleotide and is a T, G, or C (Fig. 2). It is the third position of this codon that disrupts the reading frame and may be the inserted nucleotide. This last position of the codon needs to induce a shift to maintain the conserved amino acid sequence of the *nad3* polypeptide. In turtles with the N3-174 insertion, this nucleotide is most often a C, occurring 15 times, but there are also eight instances of T in this position and three Gs. It does not seem to be important which nucleotide is inserted, as long as it is not an A. Analysis of the same region in all complete avian mitochondria in the database shows that the nucleotide in this position is always a C, which is consistent with the fact that birds have closer evolutionary relationships to each other than do the more divergent groups of turtles. The only example with an A in the third position of the leucine codon in species with the N3-174 frameshift is Reeve's turtle, *Mauremys reevesi* (Nie, Pu and Peng, unpublished). We were not able to obtain samples of this species to verify the sequence in this region. Aside from this species, only organisms that do not require a frameshift to translate *nad3* use an A in the third codon position. This observation is particularly interesting, as CTA is by far the most common leucine codon in turtle mitochondrial genomes (Table 3). That the CTA codon is not usually found in organisms requiring the frameshift may be due to its being a perfect match for the tRNA-Leu (anticodon TAG) that recognizes the CTN codons. This strong binding may not allow for the required level of frameshifting in most organisms.

Amino Acid Conservation

There is a high degree of conservation at the protein level in the region surrounding the frameshift site. Only one amino acid is changed between 9 positions upstream of the frameshift site and 21 positions downstream in all the turtles having the N3-174 frameshift mutation. A transversion at position 163 in *Chelodina parkeri* replaces what is normally a CGN arginine with an AGA (Fig. 2), which is defined as a stop codon in the vertebrate mitochondrial code. At this site, 28 of the 34 reptiles compared here use an arginine codon, 25 using CGA, with a single example of CGC and two of CGG. In five species, arginine is replaced by either tryptophan (TGA), as in *Batrachemys nasuta*, or glutamine (CAA), as in the common musk turtle, *Sternotherus odoratus*. It is worth noting that besides being the only two turtle species to show variation at this position, they also do not have the N3-174 frameshift insertion. The four species that were shown to lack that insertion also have five other amino acid substitutions in this area, two in each of *P. subrufa* and *S. triporcatus* and one in *B. nasuta*. This region of the *nad3* gene is quite conserved regardless of the presence of frameshift insertions, though it appears

Table 3 Codon usage in selected turtle mitochondrial genomes: total numbers and relative synonymous codon usage (RSCU) values for leucine and serine codons for the 13 protein-coding genes in three turtle species

Codon	*Trachemys*		*Chelonia*		*Pelomedusa*	
	No.	RSCU	No.	RSCU	No.	RSCU
Leucine						
TTA	167	1.62	112	1.09	175	1.69
TTG	24	0.23	17	0.17	17	0.16
CTT	69	0.67	49	0.49	96	0.92
CTC	82	0.80	82	0.80	96	0.92
CTA	256	2.49	344	3.36	220	2.12
CTG	19	0.18	10	0.10	19	0.18
Serine						
TCT	52	1.12	32	0.68	52	1.12
TCC	63	1.35	75	1.60	71	1.53
TCA	117	2.52	128	2.72	105	2.26
TCG	4	0.09	1	0.02	4	0.09
AGT	15	0.32	14	0.30	15	0.32
AGC	28	0.60	32	0.68	32	0.69

Note: Frameshift mutations in *nad3* and *nad4l* genes were removed prior to counting codons

that selection is relaxed somewhat in the absence of a need for frameshifting.

African Helmeted Turtle *nad3* Frameshift

In the African helmeted turtle (*Pelomedusa subrufa*), which lacks the common N3-174 insertion, there is a different insertion mutation farther upstream. The addition of either a C or a T between positions 133–135 in *P. subrufa*, first recorded by Zardoya and Meyer (1998), results in an AGA stop as the next downstream codon. The last in-frame codon is CTT, which is another example of a wobble-matched CTN codon decoded by the tRNA-Leu (CTN) (anticodon TAG). We were able to confirm this sequence independently from a specimen from the Empire of the Turtle in Florida. Unlike in Parker's snake-neck, where an AGA stop codon appears to be redefined as a sense codon, in the African helmeted turtle the AGA must induce a frameshift to allow for accurate decoding of *nad3*.

Complete Mitochondrial Genome Sequence of the Red-Eared Slider

To search for additional sites of potential frameshifting in a species known to tolerate the N3-174 frameshift mutation, we undertook the sequencing of the complete mitochondrial genome of a red-eared slider (*Trachemys scripta*). The red-eared slider mitochondrial genome contains the usual complement of mitochondrial genes and conforms to the typical vertebrate mitochondrial genome arrangement. It is comprised of 16,810 base pairs and contains all 13 protein-coding genes, 22 tRNA genes, and 2 ribosomal RNA genes normally found in vertebrate mitochondrial

genomes. The 13 protein-coding genes align well with previously reported turtles. Eleven of the genes translate normally, while two have frameshift insertions that disrupt the reading frame. As noted above, the *nad3* gene contains the inserted nucleotide at position 174 previously reported in other species. A second frameshift insertion in the *T. scripta* mitochondrial genome is present in the *nad4l* gene, where what is likely a C or a T is inserted somewhere between nucleotide positions 231 and 234 from the start of this gene, near the 3′ end (Fig. 4). The sequence of this novel frameshift site (CTT AGT AGC A) is virtually identical to the sequence at the N3-174 frameshift site (CTG AGT AGC A) in this species. As noted above, the last 0-frame codon is a CTB leucine, while the first two +1 frame codons (GTA GCA) are identical at both sites in *T. scripta*. These observations suggest that the entire sequence plays a role in frameshifting.

The frameshift heptamer CTB AGT A was not found anywhere else in-frame in the *T. scripta* mitochondrial genome. The AGT AGC A motif seen downstream of the frameshift insertion was also not found anywhere else in-frame—nor, for that matter, were any two consecutive AGY codons. Though there are 28 instances of consecutive serine codons, none had more than one AGY codon, and this was always in the second position. The only two places where these sequences occur in the *T. scripta* genome are the two programmed frameshift sites.

Discussion

The results of this study suggest that turtle mitochondrial translation systems tolerate a variety of recoding

Fig. 4 Frameshift region in the *nad4l* gene in turtles.
a Translation in the 0-frame results in frameshifts at about position 235 in the red-eared turtle and at about position 261 in the African helmeted turtle. Both polypeptides would be truncated to 87 residues.
b Translation after +1 frameshifts at position 235 in the red-eared turtle and position 261 in the African helmeted turtle

requirements. These requirements include correct decoding of the widespread N3-174 frameshift mutation, an identical mutation in the red-eared slider *nad4l* gene, and apparent terminators (AGA) in the African and Parker's side-necked turtles.

The Two +1 Frameshift Sites in the Red-Eared Turtle Mitochondrial Genome

Sequencing the complete mitochondrial genome of *Trachemys scripta* revealed not only the conserved programmed translational frameshift site within the *nad3* gene, but also a novel frameshift site within *nad4l*. A similar situation appears in two other turtles. The pancake tortoise (*Malacochersus tornieri*) has both the widespread N3-174 site and a novel site in the *nad4* gene (Fig. 1) (Parham et al. 2006b). In *Pelomedusa subrufa*, frameshift insertions are present at three different sites not found in other species. There is a high degree of conservation between the different frameshift sites in *T. scripta*. At the *nad3* site, the conserved reading frame shifts +1 over the sequence CTG AGT AGC A, written as codons of the original 0-frame. In *nad4l*, the change of frame occurs over CTT AGT AGC A, and it would appear likely that there are properties specific to this sequence that are essential in inducing the shift. Translation of either site gives the same result. In the 0-frame, they both translate as a leucine followed by two consecutive serines, while a leucine followed by valine and alanine is the amino acid sequence if the frameshift-causing nucleotide is skipped. The only difference between the two nucleotide sequences is the synonymous G or T in the wobble position of the leucine codon.

In the survey of the N3-174 frameshift, organisms with the insertion also showed a high degree of conservation of the frameshift sequence found in *T. scripta*. This is strong evidence that in *T. scripta*, and likely in other turtles, the sequence of CTB AGT A stimulates +1 frameshifting. It also implicates the two relevant tRNAs, both tRNA-Leu (decoding CTN) and tRNA-Ser (decoding AGY), as having roles in the frameshift mechanism of organisms where this sequence is present in-frame. Certain tRNAs have been shown elsewhere to have a major role in determining frameshift frequencies. For instance, in a study in the yeast Ty3 element, where GCG is used as the last in-frame codon, mutating it to GCA—a change that causes it to be decoded by tRNA-Ala (TGC) rather than tRNA-Ala (CGC)—completely eliminates frameshifting (Vimaladithan and Farabaugh 1994).

The codon immediately prior to the consecutive AGY codons at either frameshift site is a CTN leucine, also found at most other vertebrate mitochondrial frameshift sites, with CTG in *nad3*, or CTT in *nad4l*. These are decoded by tRNA-Leu (CTN) with an anticodon of TAG,

which wobble pairs with the CTG codon in the third position, and is a mismatch for the same position at the *nad4l* CTT. In other organisms, this last in-frame codon is rarely CTA, which would be exact Watson-Crick base-pairing to the anticodon. It is possible that cognate codons in the peptidyl site for the leucine tRNA anticodon TAG are unable to promote required levels frameshifting. This poses a dilemma with regard to any proposed mechanism. The tRNA-Leu (CTN) that recognizes the last in-frame codon does not have good pairing with TGA in the +1 frame, as a G-T wobble pairing in the first base and A-G mismatch in the middle base result, causing difficulties for any mechanism that would suggest slippage of this tRNA to the +1 frame. At the same time, codons that do not provide good binding to the leucine tRNA seem to be selected for at these frameshift sites, suggesting that incomplete recognition of the last in-frame codon is important in frameshifting.

Comparisons to Other +1 Programmed Translational Frameshift Sites

Certain features appear to be required in organisms where a frameshift is necessary for accurate translation of a gene. The +1 programmed translational frameshifts in *E. coli prfB*, yeast Ty1 and Ty3 elements, and mammalian antizyme have two such elements in common. The first is an apparent pause in translation at the shift site, caused by the slow decoding of either a rare or a nonsense codon in the next 0-frame position and possibly aided by the presence of mRNA secondary structure. In yeast Ty3 elements, the frameshift heptamer is GCG AGT T. It is the AGT serine codon that is thought to cause the required stall allowing the ribosome to shift frames (Vimaladithan and Farabaugh 1994). The AGT codon in the *nad3* and *nad4l* sites likely has a similar role. We believe, as has been theorized previously, that this stall leads to a competition among different possible outcomes, varying from termination of translation to the frameshift required to produce a functional protein (Fig. 5).

The second element is a peptidyl-site codon that has poor wobble-position pairing with the corresponding tRNA and often good pairing with the same tRNA if shifted +1. Change to this position in known *E. coli* frameshifting genes alters frameshift efficiency by up to 1000-fold (Curran 1993). Use of a common codon—or, by extension, one that is quickly decoded—in the +1 codon from the P-site codon has also been shown to aid frameshift efficiency (Hansen et al. 2003). Such is the case in the yeast Ty1 element and the majority of the frameshift sites found in the *Polyrhachis* ants. In that group of ants, however, one site, TGG AGT A, does not have good +1 pairing for the P-site tRNA. In Ty3 elements the tRNA that decodes

Fig. 5 Alternative outcomes after a ribosomal pause at the rare AGU codon in the red-eared turtle *nad3* gene. **a** Zero-frame decoding by binding of a charged tRNA^{Ser}-AGY. If this tRNA binds, translation would likely pause again at the following AGC codon. **b, c** Programmed +1 frameshift by binding of a charged tRNA^{Val} either by re-pairing of the peptidyl site tRNA-Leu (B) or by occlusion of the first position of the amino-acyl site (C). P' denotes the peptidyl, and 'A' denotes the aminoacyl, sites of the ribosome
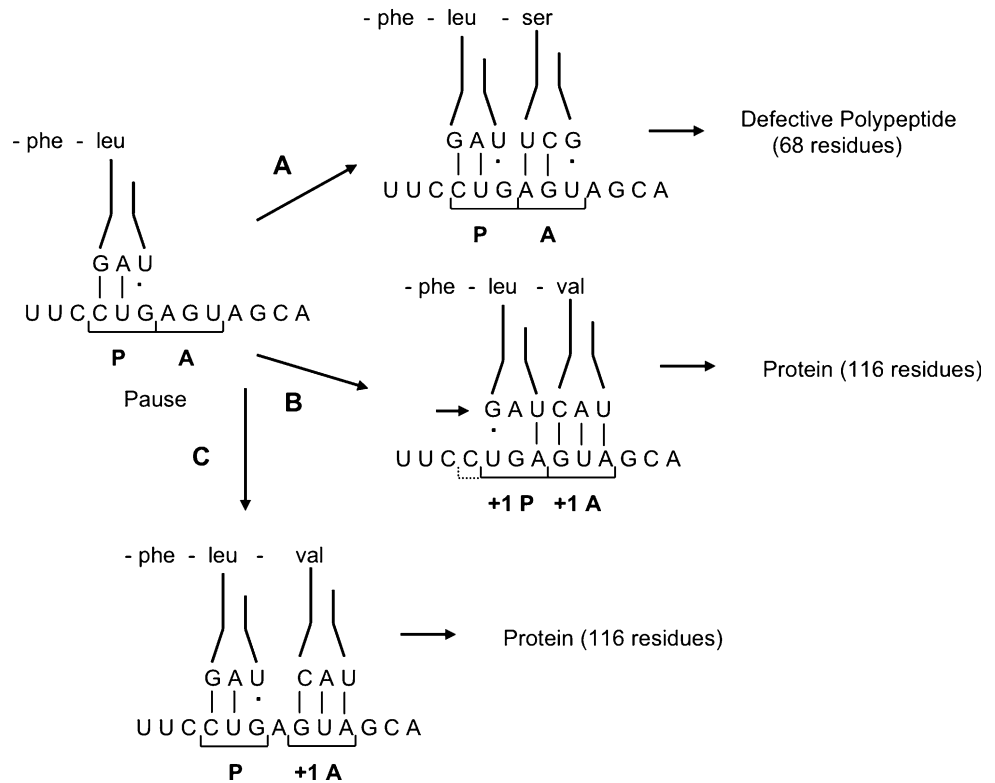
the first codon of the frameshift site GCG AGT T, the codon in the ribosomal P-site, is tRNA-Ala (GCN) (anticodon CGC), again with poor +1 binding (Vimaladithan and Farabaugh 1994). In *T. scripta*, two different leucine codons are used in the equivalent position. In *nad3* it is CTG, and in *nad4l*, CTT; both are decoded by the tRNA-Leu (CTN). In both of the *T. scripta* frameshift heptamers, the P-site +1 pairing is poor.

We present a model for decoding of the N3-174 frameshift mutation in Fig. 5. We assume that the 0-frame AGT causes a stall in processing. The nascent amino acid chain is attached to the tRNA-Leu at the ribosomal P-site. The stall produces a competition between 0-frame decoding with a charged tRNA-Ser (AGY) and +1-frame decoding with tRNA-Val, either by occlusion of the A-site or by re-pairing of the tRNA-Leu at the P-site in the +1 reading frame. If 0-frame decoding occurs, the result is a truncated polypeptide that is likely nonfunctional and will be degraded. Only translation by the second pathway, shifting to the +1 reading frame at this site can produce a full-length, functional gene product.

Nonrandom Distribution of Frameshift Sites

Sites requiring a frameshift during translation have been identified in the mitochondrial genes of only a few groups of animals: birds, turtles, a single genus of ants, the eastern

oyster, and a glass sponge. Yet within these taxa, multiple sites have been discovered in three species of turtles (*Pelomedusa* has three sites; *Trachemys* and *Malacochersus* have two each), two species of *Polyrhachis* ants (two sites in each), and the glass sponge (two sites). In the vast majority of taxa where complete mitochondrial sequences are available, there is no evidence of frameshift mutations. While it is possible that sites are occasionally missed, or dismissed as sequencing errors (and "corrected"), we believe that the overrepresentation of frameshift sites in only a few taxa is a real phenomenon. If so, it would indicate that mitochondrial translation systems vary in their susceptibility to shift reading frames during translation. We suggest that in the vast majority of animal groups, the ribosome will not frameshift over "slippery" sequences with high enough efficiency to tolerate these mutations.

Origin and Loss of the N3-174 Frameshift Mutation

The N3-174 mutation is present in a majority of bird and turtle species. The absence of this mutation in other amniotes, in particular, in other reptiles, requires either multiple origins, losses of the mutation, or both. Whether this mutation appeared in birds and turtles independently, or is derived from a single ancient mutation, is unknown. We can, however, use phylogenetic arguments at least to estimate the minimum number of independent indel

Fig. 6 Phylogenetic tree of turtles included in this study. Lineages lacking the N3-174 frameshift mutation are shown as dotted branches. The topology for this tree is taken from Krenz et al. (2005). The Cryptodira and Pleurodira are the two extant suborders of this group
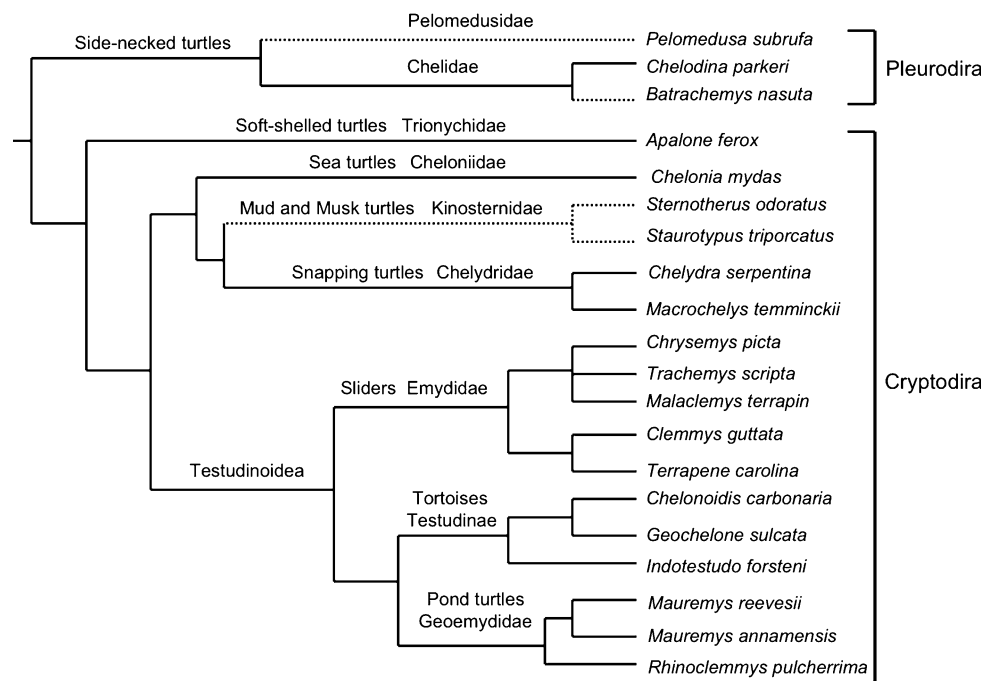
mutations that have occurred. The phylogeny of extant turtles has been extensively studied, and a robust phylogeny developed based on molecular, morphological, and paleontological evidence (Shaffer et al. 1997; Krenz et al. 2005). Mapping the mutation to this phylogeny demonstrates that at least three independent indel mutations are required to explain the current distribution of this mutation in turtles (Fig. 6). The fact that the N3-174 frameshift is absent from some turtles and birds demonstrates that it is not necessary for proper functioning of the *nad3* gene. If it has a regulatory role, that role is dispensable.

Examination of the sequence surrounding the N3-174 site allows us to make some inferences about the requirements for a single-nucleotide insertion to generate a programmed frameshift and, thus, be tolerated. First, the translation system itself must be able to slip over a nucleotide at a particular sequence with a relatively high efficiency. Second, the sequence CTN GTA appears to be required in the 0-frame at the site for a single-nucleotide insertion to generate a programmed frameshift. If the sequence at a site is CTA GTA, a single-nucleotide insertion of a C, T, or G at the third position, or a T at the second position, will produce the required CTB AGT A. If the original sequence is CTB GTA, insertion of an A at the fourth position is required. We can compare these possibilities with the requirement for loss. First, no special requirements of the translation system are necessary: translation of the mutated sequence is entirely 0-frame. The mutation CTB AGT A → CTN GTA can occur by deletion

of either the third or the fourth nucleotide. In the special case of CTT AGT A, deletion of any one of the first four nucleotides translates as Leu-Val, as required. There are more ways to lose an existing programmed frameshift site than there are to gain one de novo.

The overall picture suggests that the widespread occurrence of the N3-174 frameshift mutation places its origin prior to the separation of the Cryptodira and Pleurodira. Fossil evidence suggests that this split occurred more than 200 mya (Near et al. 2005). This age lends credence to the possibility that the mutation first occurred in a common ancestor of birds and turtles and has been retained in these species since that time. If so, its absence in some birds and turtles, and all other reptiles, represents secondary losses.

There is evidence as well for an ancient origin of programmed frameshifts in two genes in budding yeasts, ABP140 and EST3, where the mutations appear to trace back ∼ 150 myr (Farabaugh et al. 2006). The frameshift sequences are remarkably similar to the N3-174 sequence: CTT AGG C in yeast and CTB AGT A in *nad3*, implicating the same codon families.

Decoding of the AGA Codon

The AGN codon family evidently plays a central role in translational frameshifting. Most mitochondrial frameshift sites appear to require a member of this family at the 0-frame, stall-inducing A-site of the ribosome (Figs. 1, 3 and 5). More generally, this family of codons appears to be

involved in an extraordinary variety of recoding events. In the standard genetic code, AGY is decoded as serine and AGR is decoded as arginine. In the invertebrate mitochondrial code, AGN is decoded as serine. In the vertebrate mitochondrial code, AGY is decoded as serine, as in the standard genetic code, but AGR either does not appear in-frame in protein-coding genes or occurs only where we expect a terminator. Efficient termination requires recognition of the terminator by a release factor. Reassigning AGR from a sense codon to a terminator would appear to require the addition of an appropriate release factor. Where such a release factor may come from is not clear, but Ivanov et al. (2001) identified two tRNA-like structures within the large subunit of rRNA with anticodons complementary to AGA and AGG. The authors propose that these structures, which they have labeled *term-tRNAs*, are responsible for terminating translation at AGR codons in vertebrate mitochondria. If these structures are responsible for terminating translation of AGR, perturbations in their structure in African and Parker's side-necked turtles may explain the ribosome's apparently noncanonical behavior at these sites.

In the turtle mitochondrial sequences we have examined, AGA appears to be decoded in any of three ways, depending on the species and sequence context. This codon appears in-frame in only one place in the red-eared turtle mitochondrial genes, at positions 1546–1548 in the *cox1* gene, where a terminator is expected. There are no TNN triplets in this region that might be converted into a stop codon after processing and polyadenylation. In this species, the vertebrate mitochondrial code appears to provide the correct interpretation. In the African helmeted turtle, AGA appears in-frame in three places. All three are in the sequence context of CTT AGA W, where W = A or T. In this species and context, AGA is part of a 7-nt sequence that evidently stimulates +1 translational frameshifting.

Perhaps the most interesting recoding of AGA can be seen in Parker's snake-necked turtle (*Chelodina parkeri*) at positions 163–165 in the *nad3* gene (Figs. 2 and 3). It is neither a terminator nor a frameshift site. The mitochondrial translation system of this species is evidently capable of frameshifting +1 over the N3-174 site (CTT AGT AGC), only four codons downstream, but in this context (ATT AGA TTC) it must be translated as a sense codon. This site is a conserved arginine in other turtles (Fig. 3). One possible mechanism is that the first position of AGA is modified through RNA editing. An interesting alternative hypothesis is that it is decoded by tRNA-Arg with a first position wobble, using a two-of-three pairing rule, with exact Watson-Crick matches in the second and third codon positions. If so, then the tRNA-Arg must decode five codons, CGN and AGA.

These arguments suggest that the AGA at position 163 in *nad3* of *Chelodina* may be decoded by any of three different, and perhaps competing, pathways. These pathways are depicted in Fig. 7. We assume that AGA in the 0-frame induces a stall on the ribosome. If AGA is recognized by a release factor, translation is terminated. Alternatively, if tRNA-Ser (AGY; anticodon GCT) binds, translation will continue after incorporation of a serine. Finally, if tRNA-Arg (anticodon UCG) binds, an arginine residue will be incorporated. Each tRNA has one purine-purine clash, a G:A mismatch in the first position (tRNA-Arg) or third position (tRNA-Ser [AGY]). In the African helmeted turtle, a fourth pathway may be added where the AGA is part of a "slippery" sequence (Fig. 8). In this case, the pause at the AGA codon, coupled with rapid canonical matching of the GTA in the +1 position by tRNA-Val, appears to allow it to shift to the required +1 reading frame.

This observation is not the first example of an altered genetic code involving the AGR codon family. Reassignment of AGR codons has evidently occurred a number of times in the course of evolution of animal mitochondrial genomes, in particular, within arthropods (Abascal et al. 2006).

Conclusion

Turtles, for reasons that are not entirely clear, appear to exhibit a wide variety of the features requiring recoding of translation. Their mitochondrial genomes are evidently susceptible to both frameshifting and codon redefinition. Frameshift insertion mutations have now been documented at six separate sites, in three different genes. Although one site, N3-174, is widespread in turtles and birds, and may be quite ancient, five appear to be unique to particular lineages and are likely relatively recent. The fact that most cases are unique suggests that frameshifting over these sites does not have a regulatory role but is, nevertheless, tolerated under certain conditions. The conditions appear to be a specific nucleotide sequence paired with a translation system that is amenable to frameshifting. In chelonian mitochondria, and animal mitochondrial genomes in general, at minimum the sequence likely consists of an in-frame codon, which is often a CTB leucine, followed by a rare or nonsense codon overlapped in the final two nucleotides by a more frequently used sense codon. To minimize selective pressure and allow this frameshift insertion to persist, the mitochondrial translational system must allow a level of frameshifting over this context sequence to produce an adequate supply of functional protein product. Finally, in *Chelodina parkeri*, it appears that another type of recoding, codon redefinition, is required to produce a functional mitochondrial *nad3* protein from a transcript that contains what was previously thought to be a stop codon.

Fig. 7 Possible outcomes for decoding the in-frame AGA codon at position 163 in the *nad3* gene of Parker's snake-necked turtle (*Chelodina parkeri*). **a** If the AGA is recognized by a release factor (RF), translation will be terminated. **b** If tRNA^Arg binds the AGA, translation proceeds in the 0-frame after incorporation of arginine. **c** If the AGA is bound by tRNA^Ser-AGY, translation can proceed in the 0-frame after incorporation of serine
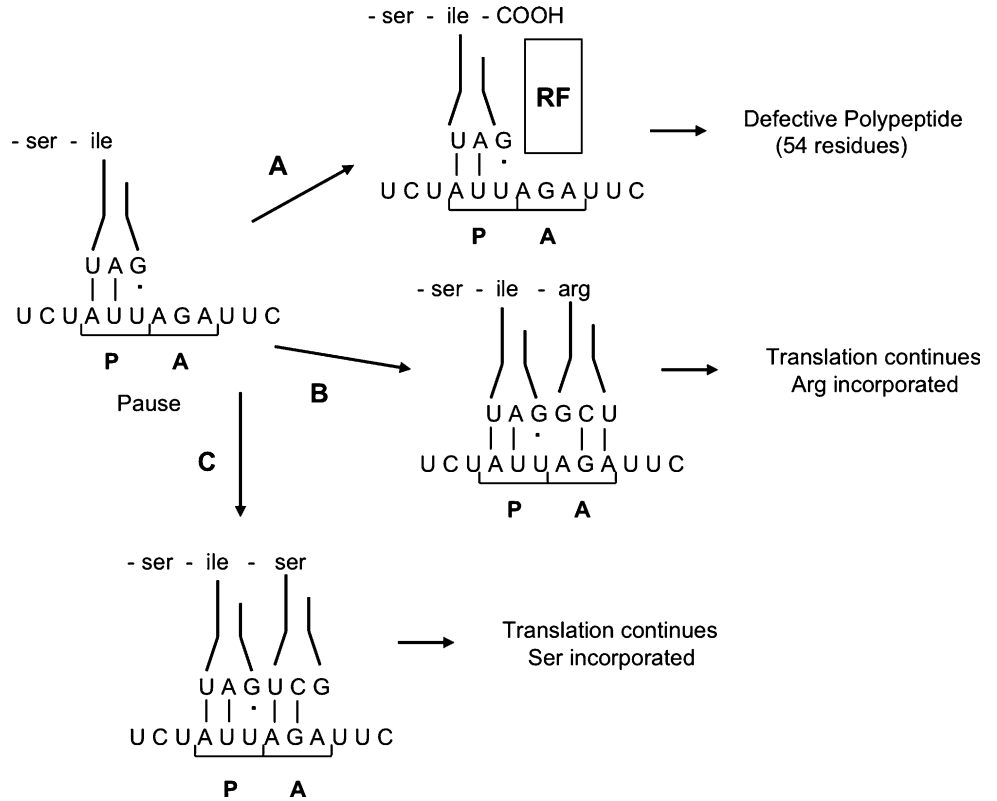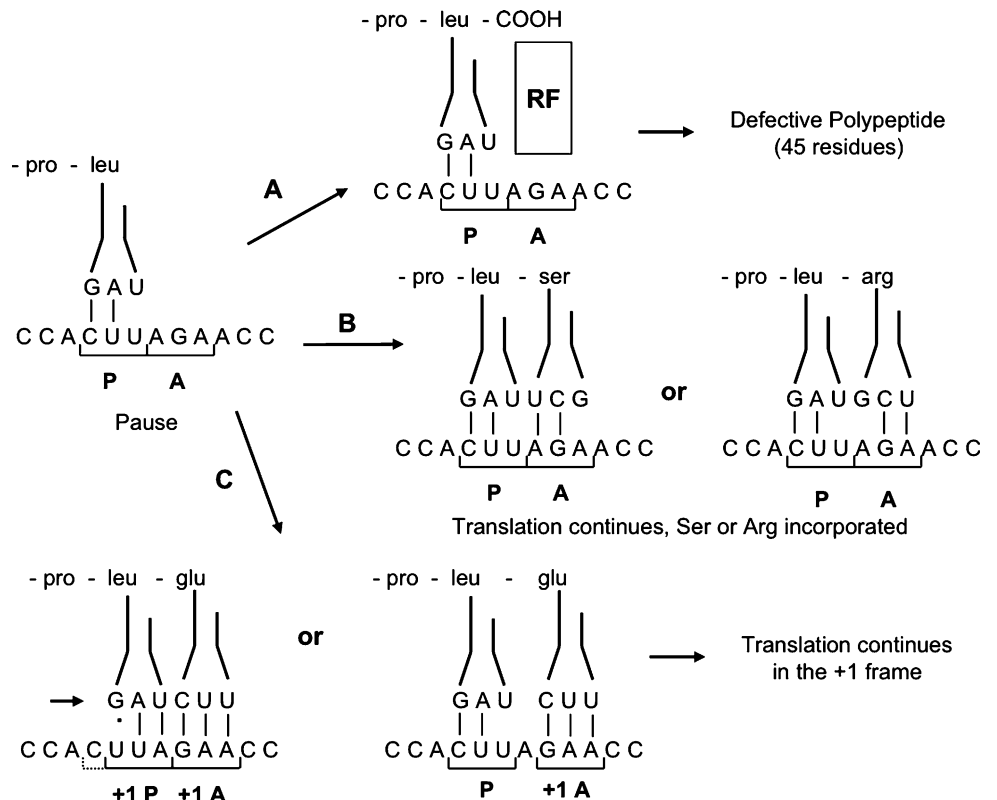

Fig. 8 Possible outcomes for decoding the in-frame AGA codon at position 135 of the African helmeted turtle (*P. subrufa*). **a** If the AGA is recognized by a release factor (RF), translation terminates. **b** If the AGA is bound in the 0-frame, either by tRNA^Ser-AGY or by tRNA^Arg, translation continues in the 0-frame, resulting in a defective polypeptide. **c** If the ribosome shifts to the +1 reading frame over the residue at position 135, either by slippage and re-pairing or by occlusion of this residue, a full-length protein of 116 residues is produced

References

Abascal F, Posada D, Knight RD, Zardoya R (2006) Parallel evolution of the genetic code in arthropod mitochondrial genomes. PLOS Biol 4:e127

Baranov PV, Gesteland RF, Atkins JF (2002) Recoding: translational bifurcations in gene expression. Gene 286:187–201

Beckenbach AT, Robson SKA, Crozier RH (2005) Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome b gene in ants of the genus *Polyrhachis*. J Mol Evol 60:141–152

Belcourt MF, Farabaugh PJ (1990) Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell 62:339–352

Burger G, Gray MW, Lang BF (2003) Mitochondrial genomes: anything goes. Trends Genet 19:709–716

Craigen WJ, Cook RG, Tate WP, Caskey CT (1985) Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc Natl Acad Sci USA 82:3616–3620

Curran JF (1993) Analysis of effects of tRNA: message stability on frameshift frequency at the *Escherichia coli* RF2 programmed frameshift site. Nucleic Acids Res 21:1837–1843

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: frameshifting without tRNA slippage. Cell 74:93–103

Farabaugh PJ, Kramer E, Vallabhaneni H, Raman A (2006) Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order Saccharomycetales. J Mol Evol 63:545–561

Gesteland RF, Atkins JF (1996) Recoding: dynamic reprogramming of translation. Annu Rev Biochem 65:741–768

Gesteland RF, Weiss RB, Atkins JF (1992) Recoding: reprogrammed genetic decoding. Science 257:1640–1641

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

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

Ivanov V, Beniaminov A, Mikheyev A, Minyat E (2001) A mechanism for stop codon recognition by the ribosome: a bioinformatic approach. RNA 7:1683–1692

Krenz JG, Naylor GJP, Shaffer HB, Janzen FJ (2005) Molecular phylogenetics and evolution of turtles. Mol Phylogenet Evol 37:178–191

Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. Mol Biol Evol 16:784–792

Matsufuji S, Matsufuji T, Miyazaki Y, Murakami Y, Atkins JF, Gesteland RF, Hayashi R (1995) Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. Cell 80:51–60

Milbury CA, Gaffney PM (2005) Complete mitochondrial DNA sequence of the eastern oyster *Crassostrea virginica*. Mar Biotechnol (NY) 7:697–712

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

Near TJ, Meylan PA, Shaffer HB (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am Nat 165:137–146

Parham JF, Feldman CR, Boore JL (2006a) The complete mitochondrial genome of the enigmatic bigheaded turtle (*Platysternon*): description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evol Biol 6:11

Parham JF, Macey JR, Papenfuss TJ, Feldman CR, Türkozan O, Polymeni R, Boore J (2006b) The phylogeny of Mediterranean tortoises and their close relatives based on complete mitochondrial genome sequences from museum specimens. Mol Phylogenet Evol 38:50–64

Rosengarten RD, Sperling EA, Moreno MA, Leys SP, Dellaporta SL (2008) The mitochondrial genome of the hexactinellid sponge *Aphrocallistes vastus*: evidence for programmed translational frameshifting. BMC Genomics 9:33

Shaffer HB, Meylan P, McKnight ML (1997) Tests of turtle phylogeny: molecular, morphological and paleontological approaches. Syst Biol 46:235–268

Turtle Taxonomy Working Group (2007) An annotated list of modern turtle terminal taxa with comments on areas of taxonomic instability and recent change. In: Shaffer HB, FitzSimmons NN, Georges A, Rhodin AGJ (eds) Defining turtle diversity: Proceedings of a workshop on genetics, ethics and taxonomy of freshwater turtles and tortoises. Chelonian Res Monogr 4:173–199

Vimaladithan A, Farabaugh PJ (1994) Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. Mol Cell Biol 14:8107–8116

Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. Proc Natl Acad Sci USA 95:14226–14231