Tetracoding increases with body temperature in Lepidosauria

Hervé Seligmann a,b,c,*, Antonieta Labra b,d

a National Natural History Museum Collections, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
b Center for Ecological and Evolutionary Synthesis, Department of Bioscience, University of Oslo, Blindern, N-0316 Oslo, Norway
c Department of Life Sciences, Ben Gurion University, 84105 Beer Sheva, Israel
d Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Chile

A R T I C L E   I N F O
Article history:
Received 11 July 2013
Received in revised form 4 September 2013
Accepted 5 September 2013

Keywords:
Suppressor tRNAs
Genetic code expansion
Synonymous codon usage
tRNA cloverleaf structure
Developmental stability

A B S T R A C T
Codons expanded by a silent position (quadruplet or tetracodons) may solve the conundrum that at life’s origins, the weak tricodon–anticodon interactions could not promote translation in the absence of complex ribosomes. Modern genomes have isolated tetracodons resulting from insertion mutations. Some bioinformatic analyses suggest that tetracoding stretches overlap with regular mitochondrial protein coding genes. These tetragenes are probably decoded by (antisense) tRNAs with expanded anticoncodons. They are GC-rich, which produce stronger basepairs than A:T interactions, suggesting expression at high temperatures. The hypothesis that tetracoding is an adaptation to high temperatures is tested here by comparing predicted mitochondrial tetracoding in Lepidosauria (lizards, amphibians, and Sphenodon), in relation to body temperature, expecting more tetracoding in species with high body temperature. The association between tRNAs with expanded anticoncodons and tetracoding previously described for mammals and Drosophila is confirmed for Lepidosauria. Independent evidence indicates that tetracoding increases with body temperature, supporting the hypothesis that tetracoding is an adaptation for efficient translation when conditions (temperature) make triplet codon-anticodon interactions too unstable to allow efficient protein elongation.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction
The origin of life is inherently connected with the early formation of the complex molecular machinery that codes and produces proteins. Ribosome-free translation of mRNAs seems impossible because interactions between the triplet codon and the tRNA’s matching anticodon are too weak for efficient peptide elongation (Baranov et al., 2009). This point is even more extreme when a thermostable origin of the genetic code is assumed (Di Giulio, 2000, 2003). However, it seems unlikely that at the origins of life, structures as complex as ribosomes were available. This conundrum seems circumvented if codons were longer, by at least one nucleotide, and if the ancestral genetic code was based on quadruplet codons or tetracodons (Baranov et al., 2009). Hypothetically, this ancestral genetic code used a subset of 64 tetracodons, called the tesserae, among the 256 potential tetracodons, and chosen based on principles of codon symmetry and error prevention (Gonzalez et al., 2012). This seems justified because error prevention is more important than usually assumed (Warneke and Hurst, 2011). In fact, error prevention explains the modern genetic code’s structure at the level of impacts of substitution mutations (e.g. Di Giulio, 1989; Haig and Hurst, 1991; Szathmary and Zintzaras, 1992; Freeland and Hurst, 1998; Freeland et al., 2000; Woese et al., 2000; Sella and Ardell, 2006; Novozhilov et al., 2007), deletion mutations (Jestin and Kempf, 1997), amino acid misinsertions due to tRNA misloading and codon-anticodon mismatch (Seligmann, 2010a, 2011a, 2012a), protein folding (Guilloux and Jestin, 2012) and mis- sense translation (Seligmann and Pollock, 2003, 2004; Itzkovitz and Alon, 2007; Seligmann, 2007, 2010b, 2012a; Pienaar and Viljoen, 2008).

Some tRNAs have the ability to decode tetracodons, and were first discovered as frameshift mutation suppressors (Riddle and Carbon, 1973; Sroga et al., 1992; Tuohy et al., 1992). It has been suggested that these have the ability to read occasional ‘extra’ nucleotides in mitochondrial genomes of birds and turtles (Mindell et al., 1998), and other Metazoans (review in Seligmann, 2012b). In addition, in ciliate genomes, tetracodons are very frequent (euplodonts, Klobutcher and Farabaugh, 2002; Klobutcher, 2005). The mechanisms for tetracodon decoding seem multiple, and are not straightforward (Atkins and Bjork, 2009), as tRNAs with expanded anticoncodons are in some cases involved (Walker and
Frederick, 2006), but also some tRNAs with regular anticodons have that ability of decoding tetracodons (O’Connor et al., 1989; Dunham et al., 2007). Some comparative analyses of mitochondrial genomes suggest that large loops of tRNA sidearms are involved in tetradecoding, putatively through crossovers between the anticodon and a sidearm, or because tRNA halves also function in translation (Seligmann, 2013a). This hypothesis of anticodons in tRNA sidearms is in line with comparative analyses of split tRNA genes (Di Giulio, 1992, 1995, 1999, 2004, 2006, 2008a,b, 2009, 2012, 2013), the existence of armless tRNAs such as in roundworm Enoplea mitochondria (Jühling et al., 2012), and the fact RNA transcribed from the mitochondrial light strand replication origin forms a stem-loop hairpin that is aminoacylated and whose 3’ extremity is extended by the standard CCA extension of mitochondrial tRNAs (Yu et al., 2008). Expanded loops in some tRNA sidearms might be the answer why some regular tRNAs have the ability to read tetracodons.

The possibility of excision of the extra nucleotide, in some cases, cannot be excluded. This would not imply tRNA decoding, but some unknown mechanism of RNA editing. This mechanism is irrelevant to the working hypothesis developed here, as it would not involve interactions between tetracodons and tetra-anticodons.

The natural history of tetracoding has not yet been studied, and the occurrence of tetratyses is not yet confirmed by direct experiments. However, the computational evidence is coherent and in artificial setups, tetracoding actually occurs: tRNAs with expanded anticodons have been used for biotechnological applications to insert unnatural amino acids in proteins (Moore et al., 2000a,b; Maglery et al., 2001; Wang et al., 2001, 2010; Anderson et al., 2002; Rodriguez et al., 2007; Chen and Schindlinger, 2010; Neumann et al., 2010). Note that ribosomes were also artificially selected to improve tetradecoding (Wang et al., 2007, 2010; Neumann et al., 2010).

The hypothesis that natural encoding of polypeptides by presumably continuous stretches of quadruplet codons was only recently tested by comparative analyses of mitochondrial genomes of mammals and Drosophila (Seligmann, 2012b). Tessera tetracodons are overrepresented in the predicted mammal mitochondrial tetratyses, a verification of the tesserae hypothesis (Gonzalez et al., 2012). The tesserae hypothesis is tailored to the vertebrate genetic code due to its inherent symmetry properties (Gonzalez et al., 2012). It is unclear to what extents it is adequate for other genetic codes. These predicted tetratyses are relatively GC-rich as compared to the rest of the genome (Seligmann, 2012b). This suggests that tetratyses are adapted for efficient translation at high temperatures, and is in line with the rationale that four pairs of interacting nucleosides yield more stable codon–anticodon interactions than three pairs of interacting nucleosides.

The number of predicted tetratyses per mitochondrial genome coevolves with numbers of predicted antisense tRNAs with expanded anticodons in the same genome, for each a sample of mammal and of Drosophila mitochondria (Seligmann, 2012b). Similarly, numbers of predicted antisense tRNAs with expanded anticodons predict the frequency of isolated quadruplet codons within regular mitochondrial protein coding genes of turtles (Seligmann, 2012c), and in some other groups, notably birds (Seligmann, 2012b).

Computational and comparative analyses suggest translational activity by antisense tRNAs (Seligmann, 2010c). These include the observation that antisense tRNAs with anticodons matching stop codons are avoided (Seligmann, 2010d), that predicted antisense tRNA anticodon numbers increase with genomic codon usages (Seligmann, 2010c, 2013a, 2011), that predicted numbers of antisense tRNA properties matching translational activity increase with observed antisense tRNA abundances (in Drosophila mitochondria, Seligmann, 2012d), and that associations exist between mutation pathogenicity and their effects on the antisense, rather than the sense, tRNA’s cloverleaf folding capacity (Seligmann, 2011c). However, antisense tRNA translational activity has not yet been demonstrated by direct observations (Brzezniak et al., 2011). Nevertheless, besides the coevolution between tetratyses and antisense tRNAs with expanded anticodons (Seligmann, 2012b), antisense tRNAs with anticodons matching stops coevolve with overlapping genes that include stop codons and could not be expressed otherwise (Faure et al., 2011; Seligmann, 2011d, 2012c,d,e, 2013b). This suggests that several genetic coding systems coexist in parallel to the regular, recognized coding system, putatively adapted for specific conditions that remain unknown, but would associate with the expression of antisense tRNAs. Further results also indicate that systematic nucleotide permutations, by exchanging during polymerization (or RNA editing) between nucleosides, reveals unsuspected coding potential in mitochondrial genomes. Empirical results indicate that symmetric nucleotide exchanges (i.e. A → C, Seligmann, 2013c) are more frequent than those involving the potentially more complex asymmetric exchanges (i.e. A → C → G → A, Seligmann, 2013d).

Tetracodons would be an additional coding system increasing the coding density of polynucleotide sequences. The fact that tetracodon hybridization with tetra-matching tetra-anticodons, yields more stable duplexes, suggests that this system is adapted for expression at high temperatures, where tricodons are less efficient due to their less stable interactions with anticodons. The association between tetracoding and increased GC content (Seligmann, 2012b) is a further indication that tetracoding is an adaptation of translation to high temperatures.

The working hypothesis examined here is that tetracoding increases with body temperatures. The adequate combination of complete mitochondrial genome and body temperature data exists for Lepidosaurians (lizards, amphisbaenians and Sphenodon), an ectothermic group where body temperatures vary among taxa. The simple prediction that tetracoding increases with body temperature is tested using species comparisons within this group, and the coevolution of tetracoding with antisense tRNAs possessing expanded anticodons is independently tested for Lepidosaurians, considering the previous information for mammals and Drosophila (Seligmann, 2012b).

2. Materials and methods

The complete Lepidosaurian mitochondrial genomes (lizards, amphisbaenians, and Sphenodon) available in Genbank in November 2012 were listed. The literature was searched for mean field body temperatures for each of these species. Table 1 presents the species for which both complete genomes and mean field body temperature data are available.

2.1. Expanded anticodons in antisense tRNAs

Each mitochondrial genome listed in Table 1 was searched for antisense tRNAs with expanded anticodon loops, as was done previously (Seligmann, 2012b). Each genome was analyzed by the online available software tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/, Lowe and Eddy, 1997; Schattner et al., 2005), setting the cut off score for COVE at −20. COVE scores are log ratios of capacities to form the classical cloverleaf secondary structure in the folic sequence, versus random sequences, hence negative values for the structural component of COVE indicate lower capacities to form cloverleaf structures than random sequences. All tRNAs matching the 22 regular mitochondrial tRNA sequences were extracted, and each of these sequences was again analyzed by tRNAscan-SE, setting the COVE cut off
score at –200 for this second search. The cloverleaf secondary structure of each antisense tRNA, as predicted by tRNAscan-SE, was examined recording visually from the secondary structure predicted by tRNAscan-SE anticodon loop sizes.

2.2. Tetratogene detection

The next step was to estimate extents of predicted tetracoding within regular mitochondrial protein coding genes in the genomes listed in Table 1, following previous analyses (Seligmann, 2012b). For each genome, the 13 protein coding genes were extracted and translated according to the regular vertebrate genetic code, assuming that each fourth nucleotide is silent. Accordingly, each sequence has four alternative frames, and four polypeptide sequences were translated. Each of these hypothetical polypeptide sequences (four per gene, hence 52 for the 13 mitochondrial genes) was then analyzed by Genbank's BLAST (Altschul et al., 1997, 2005) using standard default search parameters, searching with proteins existing already in GenBank. Results were recorded and considered as putative polypeptides coded by candidate tetracoding genes.

2.3. Phylogenetic contrasts and statistical analyses

In analyses based on comparisons between species, different species cannot be considered as independent observations because the phylogenetic relationships among them constraint the data. Hence, statistical patterns may result from random phylogenetic constraints, rather than the studied phenomenon. Therefore, species were paired according to the principle of pairwise phylogenetic contrasts that control for common ancestry. The phylogenetic contrasts consist of pairs of species chosen to be as closely related as possible. The pairs are chosen so that the phylogenetic pathway between a pair does not cross the pathway between any other pair (Felsenstein, 1985). That way, phylogenetic independence of analyses is ascertained. Numbers of predicted tetracodons in the
species with the lower average field body temperature are subtracted from numbers in the other species of the pair, for each gene and each of the four tetracoding frames of each gene. Pairs of species were formed using phylogenetic information from Macey et al. (2008) and Vidal and Hedges (2009), and they have the same superscript number in Table 1. For example, numbers of tetracods in genes and frames of Anolis were subtracted, expecting more tetracods in A. cybotes than A. carolinensis because the latter has a lower mean field body temperature.

To determine relation between variables, we performed parametric Pearson correlation analyses and non-parametric Spearman rank correlation analyses. We also use sign tests based on the binomial distribution to test whether frequencies of qualitative results differ from the expected \( P = 0.5 \) if there is no phenomenon biasing the qualitative result in one direction.

3. Results and discussion

3.1. Expanded anticodons in antisense tRNAs

Table 1 shows the number of antisense tRNAs (tR) for which the anticodon loop is expanded (>7, see for example the antisense tRNA Val from the amphibiaenian Diplometopon zarudnyi in Fig. 1). In few cases (five among 81 antisense tRNAs with expanded anticodons), unusual secondary structures make it difficult to determine on which arm the anticodon is. In these cases, when comparing two or three arms, if only a single loop was >7, this tRNA was scored as half or one third, the reason why some numbers in Table 1 are integers. Numbers of antisense tRNAs range from zero to six (in the cordyloid Smaug warreni). As few species have more than two antisense tRNAs with predicted expanded anticodons, all complete mitochondrial genomes from genera for which no temperature data are available were analyzed to find genomes with more than two predicted antisense tRNAs with expanded anticodon. This adds five species to the list with more than two antisense tRNAs with predicted expanded anticodons, one with four antisense tRNAs with expanded anticodons (the anguid Abronia graminea) and four species with each three such tRNAs (Table 1).

3.2. Candidate tetragenes in Lepidosauria

The putative tetragenes coding for polypeptides aligning with existing proteins for Anolis carolinensis and Anolis cybotes are described in Table 2. Despite the phylogenetic proximity between these two anoles, only two predicted tetracoding regions coincide between these species, in terms of gene and frame. This is one quarter of the predicted tetragenes in A. carolinensis, and one sixth of those in A. cybotes. None of these two 'homologous' tetragenes is predicted to code for the same protein, they align with different GenBank proteins. However, two non-homologous predicted tetragenes align with a histidine kinase, one in the second frame of CO3 (A. carolinensis), and one in the third frame of ND3 (A. cybotes). A situation where overlapping genes are not evolutionarily conserved between closely related species is already known from viruses (Rancurel et al., 2009) and makes it difficult to detect false positives based on evolutionary conservation. On the other hand, it is unlikely that the presence of a histidine kinase among the predicted tetragenes in both species is coincidental, for two different, yet adjacent, genes. Indeed, considering that histidine kinases are less than 1.2% of all proteins in Genbank, this observation has a probability of \( P = 0.012 \times 0.012 = 0.000144 \). Taking into account the fact that the two predicted histidine kinase tetragenes are on adjacent genes (two neighbours per gene, hence in total three possibilities among 13 genes), this \( P \) value decreases even more: \( P = 3/13 \times 0.000136 = 0.000033 \).

Overall, the distribution of functions of putative homologues of tetracoded polypeptides seems non-random. For example, there are several alignments with membrane transporter proteins. This fits with the nature of 'regular' mitochondrion-encoded proteins, which in vertebrates are uniquely membrane-embedded proteins.
Table 2
Predicted overlapping tetracoding genes in Anolis mitochondrial protein coding genes. Columns indicate: regular gene and frame, (Loc) location of first and last tetracodon coding for polypeptide aligning with GenBank protein, alignment length, (S) similarity, (Ni) number of stop codons in aligning sequence, (Te) entry number of protein aligning with tetracoded polypeptide and (Origin) GenBank protein origin and function.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Loc</th>
<th>S</th>
<th>Si</th>
<th>Te</th>
<th>Id</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anolis carolinensis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI 2</td>
<td>59–108</td>
<td>50</td>
<td>50</td>
<td>2</td>
<td>EDS13320</td>
<td>Hypoth. prot. BACTAE 03499, Bacteroides stercoris</td>
</tr>
<tr>
<td>CO2 2</td>
<td>59–160</td>
<td>112</td>
<td>46</td>
<td>1</td>
<td>UA34313</td>
<td>Integral membrane sensor signal transduction histidine kinase, Variovorax paradoxus</td>
</tr>
<tr>
<td>CO3 4</td>
<td>92–143</td>
<td>54</td>
<td>56</td>
<td>3</td>
<td>EGGO8593</td>
<td>Hypoth. prot. Meliandra 105018, Melaspora larici-populina</td>
</tr>
<tr>
<td>ND1 1</td>
<td>36–85</td>
<td>53</td>
<td>51</td>
<td>1</td>
<td>EDS38765</td>
<td>Sorbitol dehydrogenase, Calix quinquefaciatus</td>
</tr>
<tr>
<td>ND3 1</td>
<td>19–56</td>
<td>38</td>
<td>63</td>
<td>1</td>
<td>EJ15009</td>
<td>Hypoth. prot. B224 0264, Aeromonas medi was</td>
</tr>
<tr>
<td>ND3 3</td>
<td>29–64</td>
<td>36</td>
<td>50</td>
<td>2</td>
<td>EJ12184</td>
<td>Hypoth. prot. Selimodra 446761, Selaginella moellendorfii</td>
</tr>
<tr>
<td>ND14 4</td>
<td>21–64</td>
<td>52</td>
<td>48</td>
<td>1</td>
<td>ADJ27122</td>
<td>H10933 family prot., Nitrosococcus watsoni C-113</td>
</tr>
<tr>
<td>NDE 1</td>
<td>4–57</td>
<td>60</td>
<td>50</td>
<td>0</td>
<td>XP003701682</td>
<td>Translocation prot. SECG-2-like, Megachile rotondata</td>
</tr>
<tr>
<td>NDE 2</td>
<td>3–103</td>
<td>101</td>
<td>43</td>
<td>1</td>
<td>AA244659</td>
<td>Putative PTS system, N-acetylglucosamine-specific II ABC component, Mycoplasma hypomunonea</td>
</tr>
<tr>
<td>Anolis cybotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDI 1</td>
<td>28–164</td>
<td>137</td>
<td>42</td>
<td>4</td>
<td>CCB85707</td>
<td>ABC transporter ATP-binding prot. HI,1051, Parachlamydia acanthoamae</td>
</tr>
<tr>
<td>CO1 1</td>
<td>326–380</td>
<td>61</td>
<td>56</td>
<td>1</td>
<td>CAG07775</td>
<td>Unnamed prot., Tetradon nigroviridis</td>
</tr>
<tr>
<td>CO2 1</td>
<td>78–157</td>
<td>96</td>
<td>46</td>
<td>1</td>
<td>ACS35443</td>
<td>Myosin G, Phaeacanthum tricornutum</td>
</tr>
<tr>
<td>AT2 2</td>
<td>105–167</td>
<td>64</td>
<td>61</td>
<td>1</td>
<td>EEB6922</td>
<td>Hypoth. Prot. ANHYDRO 00159, Anaerococcus hydrogenalis</td>
</tr>
<tr>
<td>AT3 3</td>
<td>66–127</td>
<td>62</td>
<td>48</td>
<td>1</td>
<td>EAA00787</td>
<td>AGAP012311-PA, Anopheles gambiae str. PEST</td>
</tr>
<tr>
<td>CO3 1</td>
<td>101–162</td>
<td>62</td>
<td>50</td>
<td>1</td>
<td>EIM67035</td>
<td>SAM-dependent methytransferase, Spingobium indicum</td>
</tr>
<tr>
<td>CO2 2</td>
<td>3–114</td>
<td>131</td>
<td>37</td>
<td>1</td>
<td>EDW69948</td>
<td>G11841, Drosophilia virilis</td>
</tr>
<tr>
<td>ND3 3</td>
<td>47–72</td>
<td>27</td>
<td>74</td>
<td>0</td>
<td>CRB27127</td>
<td>Two-component sensor histidine kinase, Oscillatoria sp.</td>
</tr>
<tr>
<td>ND4 3</td>
<td>5–53</td>
<td>49</td>
<td>55</td>
<td>3</td>
<td>EGU76346</td>
<td>Hypoth. prot. FOXB,13147, Fusarium oxysporum Fo5176</td>
</tr>
<tr>
<td>ND4 4</td>
<td>2–38</td>
<td>37</td>
<td>57</td>
<td>1</td>
<td>EIM65622</td>
<td>Hypoth. prot. PSE3,13384, Mortella sp. PSE3</td>
</tr>
<tr>
<td>NDE 4</td>
<td>4–85</td>
<td>82</td>
<td>50</td>
<td>2</td>
<td>EDO49531</td>
<td>Predicted prot. Nematostella vectensis</td>
</tr>
<tr>
<td>CytB</td>
<td>227–277</td>
<td>51</td>
<td>45</td>
<td>0</td>
<td>ACJ32539</td>
<td>Transposase, Anoxybacillus flavithermus WX1</td>
</tr>
</tbody>
</table>

(Anderson et al., 1981). Despite all this, the list of predicted tetragenes most probably includes some false positives. However, it is also a priori as likely that some tetragenes are not detected by BLAST. Overall, results such as those in Table 2 yield an approximate estimate of tetracoding in a mitochondrial genome.

Numbers of predicted tetracoded overlapping genes vary from three (Chamaeleo africanus) to 19 (Hemidactylus) (mean across species = SD = 10.36 ± 3.10), so that the percentage of the 52 hypothetical polypeptides that yield a candidate tetrage gene ranges from 5.7% to 36.5%. The total number of predicted tetracodons ranges from 19175 (Chamaeleo africanus) to 1110 (Chalarodon) (6765 ± 206.7). The shortest average length of candidate tetracoded genes is 38.5 (Hemidactylus), while the longest average size is 85.4 (Chalarodon) (mean across all species 65.9 ± 9.5). Total numbers of tetracodons increase with the total number of tetragenes (r = 0.816), but the mean tetrage length does not increase with numbers of tetragenes (r = −0.179).

3.3. Coevolution between tetracoding and expanded anticodons in Lepidosauria

It is understandable that the reality of natural tetragenes may be doubted, as at this point, only computational evidence indicates it (Seligmang, 2012b, 2013a). However, among the numerous independent and converging computational evidences that justify tetracoding as a working hypothesis and presented in Seligmang (2012b), the strongest evidence for the translation of these tetragenes is that numbers of predicted tetracodons coevolve with numbers of predicted antisense tRNAs with expanded anticodons, in mammals and Drosophila (Seligmang, 2012b). A statistically non significant positive association between the total number of tetracodons and antisense tRNAs with expanded anticodons is also observed for the Lepidosaurian genomes listed in Table 1 (r = 0.207, P = 0.076; rs = 0.196, P = 0.084, one tailed tests, see Fig. 2).

The highest number of predicted antisense tRNAs with expanded anticodons is for Smaug warreni, for which no specific body temperature data is available, and therefore is excluded from this analysis. If, as a proxy, one uses the average body temperature of other cordylids (32.71 °C, from Truter, 2011), the strength of the association between expanded anticodons and the total number of tetracodons slightly increases: r = 0.23, P = 0.06. This is speculative, but suggests that including actual temperature data from Smaug might strengthen the results.

This result, though not statistically significant, tentatively confirms the existence of tetracoding in general. This is because the trend, though not significant, is qualitatively in the expected positive direction. This enables to calculate the combined statistical significance for coevolution between numbers of expanded anticodons and tetracodons over all three taxonomic groups (mammals, Drosophila and Lepidosauria), using Fisher’s method to combine P values that sums −2 × P i where i ranges from 1 to k (k = 3 in this case), a statistic that has a χ² distribution with 2 × k degrees of freedom (Fisher, 1950). This yields, when combining only P values from mammals and Drosophila P = 0.081 and P = 0.11 for parametric Pearson and Spearman correlation, respectively. Including Lepidosauria (one tailed parametric P = 0.076 and nonparametric P = 0.084) as a third test, combined statistical significances are P = 0.0365 and P = 0.052 for parametric and non-parametric tests, respectively.

3.4. Coevolution between tetracoding and field temperatures in Lepidosauria

The principle that four nucleotides interacting during codon-anticodon hybridization yield more stable duplexes than when codons and anticodons are each three nucleotides long leads to the prediction that tetracoding is more frequent in species with high field body temperatures. This principle is tentatively confirmed by the comparative analysis shown in Fig. 3, where predicted overlap coding tetracodons in regular mitochondrial protein coding genes increase with body temperatures (r = 0.383, P = 0.005; rs = 0.39, P = 0.0048, one tailed tests). Hence in principle, numbers of predicted tetracoding regions increase with body temperatures in Lepidosauria. However, this result could be the random result
of confounding effects, notably phylogeny. Therefore, we applied pairwise phylogenetic contrasts to control for common ancestry.

We proceed in this respect to a number of statistical tests, using the species pairs indicated by superscript numbers in Table 1. In the example of A. cybotes and A. carolinensis, when considering the sum of predicted tetracodons for all four frames of a gene, there were more tetracodons in A. cybotes than A. carolinensis in seven among nine genes, 78% (four genes lacked predicted tetracodones in both species). This is statistically more than the 50% that would be expected under the null hypothesis that tetracoding is not more developed in A. cybotes than A. carolinensis according to a one tailed sign test ($P = 0.045$).

Other comparisons among congenic species are particularly useful in this respect. For example, the middle European wall lizard P. muralis has a slightly lower body temperature than its close relative, the Italian ruin lizard P. siculus. Correspondingly, there are more tetracodons in the northern species in only two among ten genes (80%) for which tetracoding could be compared between these species, which is statistically significant according to a one tailed sign test ($P = 0.027$). The species list enables five more congreneric comparisons, two among Chamaeleo, and one each among Eremias, Phrynocephalus and Varanus, where percentages of genes with more predicted tetracodons in the species with higher mean body temperature are 67%, 60%, 75%, 75% and 30%, respectively. These seven contrasts are independent and hence their combined $P$ value can be calculated using Fisher's method to combine $P$ values, which yields $\chi^2 = 32.66$ and $P = 0.00323$.

The other comparisons involve non-congeneric species of the same family. The following compared pairs are followed by percentages of genes with more predicted tetracodons in the species with higher mean body temperature and the total number of comparable genes with tetracodones: Agamidae, Calotes-Pogona 67, 9 and Chlamydosaurus-Pseudotrapelus 80, 10; Amphibiania: Bipes-Blanus 50, 10; Diplometoptera-Rhineura 38, 8; Anguidae: Anguis-Ophiurus 42, 12; Gekkonidae: Hemidactylus-Tarentola 44, 9; Gekko-Heteronotia 70, 10; Lacertidae: Lacerta-Takydromus 60, 10; Iguanidae: Gambelia-Sceloporus 50, 12; Chalarodon-Polyurus 67, 9 and Iguana-Plica 75, 8. Four pairs are formed by species from different families (Gekkonidae-Scincidae 56, 9; Helodermatidae-Iguanidae 56, 9; Lepidophytha-Shinisaurus 57, 7; Varranidae-Sphenodontidae, 44, 9).

Note that the percentages are lower when pairs combine distant species (between families) than closely related species (within genus), indicating that biologically more realistic comparisons yield clearer patterns of association between tetracoding and body temperatures.

In total, there were 15 phylogenetic pairs where more than half of the genes had more predicted tetracodons in the species with higher mean temperature than in the other species. There are 22 pairs, two of them have exactly 50% of the genes with more tetracodons in the high than the low temperature species and have to be excluded from the test. Hence qualitatively, the results fit predictions for 15 among 20 (75%) of the pairs, which is statistically significantly more than 50% expected under the null hypothesis using a one tailed sign test ($P = 0.01$). In addition, the tendency was significant for three specific pairs, the above noted Anolis and Podarcis, and the comparison between the agamids Pseudotrapelus and Chlamydosaurus ($P = 0.027$). None of the five cases where the
Fig. 3. Number of predicted tetracodons in Lepidosaurian mitochondrial genomes as a function of the mean field body temperature. Codon-anticodon interactions based on four nucleotides are more stable than those formed by codons and anticodons of three nucleotides. The increase in tetracodon number with temperature is predicted by the necessity for more stable codon-anticodon interactions at high temperature. The Pearson correlation coefficient is \( r = 0.322 \), one tailed \( P = 0.022 \). The Spearman rank correlation coefficient is \( r_s = 0.343 \), one tailed \( P = 0.016 \).

A third possibility exists: rather than summing across all four frames of a gene tetracodons and comparing each gene in a species pair, one can compare each frame of each gene for each species pair. This yields a majority of genes with more tetracodons per frame in the species with the higher mean body temperature in 14 among 20 pairs (70%), which is a significant majority of pairs according to a one tailed sign test \( (P = 0.029) \). At levels of single species pairs, the comparisons between Podarcis species, and Chamaeleo afric anus and C. calyptratus, between Pseudotrapelus and Chlamydosaurus are significant in this case (one tailed sign tests, \( P = 0.0296, P = 0.016 \) and \( P = 0.024 \), respectively). These remain significant at \( P < 0.05 \) after applying the Benjamini–Hochberg adjustment. Hence, the result that tetracoding increases with body temperature holds independently of the method used to estimate tetracoding.

4. Conclusions

The patterns presented confirm the prediction that tetracoding increases with body temperature in Lepidosauria (Fig. 3). Phylogenetically controlled analyses confirm this observation, which independently confirms the evidence presented earlier that suggests that tetracoding occurs (Seligmann, 2012b). Tetracoding is also independently strengthened by an additional test of one of the predictions of the tetracoding hypothesis, that numbers of tetra-anticodons coevolve with numbers of tetracodons. This test does not yield statistically significant results \( (P = 0.07 \) or 0.06 if Smaug is included in analyses) but strengthens previous results in others, so that the combined significance of this test for all groups is statistically significant. Hence, extents of tetracoding also increase with predicted numbers of antisense tRNAs with predicted expanded


Di Giulio, M. 2008b. Transfer RNA genes in pieces are an ancestral character. EMBO Rep. 9, 820.

Di Giulio, M. 2009. Formal proof that the split genes of tRNAs of nanoarchaeum equitans are an ancestral character. J. Mol. Evol. 69, 505–511.

Di Giulio, M. 2012. The ‘recently’ split transfer RNA genes may be close to merging the two halves of the tRNA rather than having just separated them. J. Theor. Biol. 310, 1–2.


