Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the ‘Q’ link to go to the location in the proof.

<table>
<thead>
<tr>
<th>Location in article</th>
<th>Query / Remark: click on the Q link to go</th>
<th>Please insert your reply or correction at the corresponding line in the proof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Reference(s) given here were noted in the reference list but are missing from the text – please position each reference in the text or delete it from the list.</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>The reference given here is cited in the text but is missing from the reference list – please make the list complete or remove the reference from the text: Freeland et al. (1998).</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>Please check the address for the corresponding author that has been added here, and correct if necessary. Ref. Freeland et al. (1998) is cited in the text but not provided in the reference list. Please provide it in the reference list or delete this citation from the text.</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>Please check the hierarchy of the section headings.</td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>Table 4 has been cited in the text, but the corresponding table and caption have not been provided. Please check.</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td>Uncited references: This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section. Please check the caption of Table 1.</td>
<td></td>
</tr>
<tr>
<td>Q7</td>
<td>Please provide the significance of bold values in Table 1.</td>
<td></td>
</tr>
<tr>
<td>Q8</td>
<td>Please provide the significance of italic values in Table 2.</td>
<td></td>
</tr>
<tr>
<td>Q9</td>
<td>Please provide the significance of footnote ‘h’ in Table 2.</td>
<td></td>
</tr>
<tr>
<td>Q10</td>
<td>The footnote indicator ‘d’ has been given for two footnotes in Table 2. Please check.</td>
<td></td>
</tr>
</tbody>
</table>

Thank you for your assistance.
Error compensation of tRNA misacylation by codon–anticodon mismatch prevents translational amino acid misinsertion

Hervé Seligmann∗
Error compensation of tRNA misacylation by codon–anticodon mismatch prevents translational amino acid misinsertion

Hervé Seligmann

► For a majority of tRNAs, non-cognate amino acids with the greatest probability of being misloaded, match the codons with the greatest probability of being mismatched by the tRNA’s anticodon. ► This property of error compensation is more stronger for regular polymorphisms than pathogenic ones. ► It is stronger in lizards with low developmental instability. ► It is stronger for rarely used codons and anticodons.
Error compensation of tRNA misacylation by codon--anticodon mismatch prevents translational amino acid misinsertion

Hervé Seligmann\(^{a,b,c,}\)∗

\(^{a}\) Department of Evolution, Systematics & Ecology, Hebrew University of Jerusalem, 91904 Jerusalem, Israel
\(^{b}\) Center for Ecological and Evolutionary Synthesis, Department of Biological Sciences, University of Oslo, Blindern, N-0316 Oslo, Norway
\(^{c}\) Department of Life Sciences, Ben Gurion University, 84105 Beer Sheva, Israel

Abstract

Codon--anticodon mismatches and tRNA misloadings cause translational amino acid misinsertions, producing dysfunctional proteins. Here I explore the original hypothesis whether mismatches tend to compensate misacylation, so as to insert the amino acid coded by the codon. This error compensation is promoted by the fact that codon--anticodon mismatch stabilities increase with tRNA misacylation potentials (predicted by ‘tfam’) by non-cognate amino acids coded by the mismatched codons for most tRNAs examined. Error compensation is independent of preferential misacylation by non-cognate amino acids physico-chemically similar to cognate amino acids, a phenomenon that decreases misinsertion impacts. Error compensation correlates negatively with (a) codon/anticodon abundance (in human mitochondria and Escherichia coli); (b) developmental instability (estimated by fluctuating asymmetry in bilateral counts of subdigital lamellae, in each of two lizard genera, Anolis and Sceloporus); and (c) pathogenicity of human mitochondrial tRNA polymorphisms. Patterns described here suggest that tRNA misacylation is sometimes compensated by codon--anticodon mismatches. Hence translation inserts the amino acid coded by the mismatched codon, despite mismatch and misloading. Results suggest that this phenomenon is sufficiently important to affect whole organism phenotypes, as shown by correlations with pathologies and morphological estimates of developmental stability.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The genetic code is optimal in relation to several properties important for coding and translation. This suggests that the genetic code is not frozen (Sella and Ardell, 2006), but evolves towards a multi-functional optimum (Bollenbach et al., 2007). For example, the genetic code might have minimized codon length (Baranov et al., 2009), and seems to minimise mutation impacts (Freeland et al., 2008; Freeland et al., 2000; Gilis et al., 2001; Sella and Ardell, 2002) as well as costs of accidental ribosomal framshifts during protein synthesis (Seligmann and Pollock, 2004; Seligmann, 2007), while maximizing the potential for secondary structure formation (Itzkovitz and Alon, 2007).

Presumably, the evolution of the genetic code involved codon reassignments (Osawa and Jukes, 1989; Knight et al., 2001), and created alternative genetic codes (Santos et al., 2004), perhaps because needs for optimization of different properties differ among organisms (as for optimizing numbers of off frame stops (Singh and Pardasani, 2009)). Early observations that physico-chemical properties of amino acids correlate with properties of codons, and especially anticodons (Jungck, 1978), suggest that the structure of the genetic code coevolved with properties of tRNAs (Chechethkin, 2006) and of the tRNA synthetases that aminoacylate the tRNAs (Jestin and Soulé, 2007). Indeed, the two major groups of tRNA synthetases, class I and II, seem to minimize impacts of misinserted amino acids in protein sequences by tRNAs that were misloaded (predicted by ‘tfam’) by non-cognate amino acids coded by the mismatched codons for most tRNAs examined. Error compensation is independent of preferential misacylation by non-cognate amino acids physico-chemically similar to cognate amino acids, a phenomenon that decreases misinsertion impacts. Error compensation correlates negatively with (a) codon/anticodon abundance (in human mitochondria and Escherichia coli); (b) developmental instability (estimated by fluctuating asymmetry in bilateral counts of subdigital lamellae, in each of two lizard genera, Anolis and Sceloporus); and (c) pathogenicity of human mitochondrial tRNA polymorphisms. Patterns described here suggest that tRNA misacylation is sometimes compensated by codon--anticodon mismatches. Hence translation inserts the amino acid coded by the mismatched codon, despite mismatch and misloading. Results suggest that this phenomenon is sufficiently important to affect whole organism phenotypes, as shown by correlations with pathologies and morphological estimates of developmental stability.

© 2011 Elsevier Ltd. All rights reserved.


Q1 Correspondence address: Department of Evolution, Systematics & Ecology, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel.
E-mail addresses: herves@bio.uio.no, podarcissicula@gmail.com

1476-9271/‐ see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2011.03.001
Despite tRNA editing, tRNA misloading occurs, and typically results in amino acid misinsertion. This means that the amino acid that is added to the elongating peptide is not the one that is coded by the mRNA's codon. Misinsertion does not only result from misloading. Correct insertion depends on that the tRNA is loaded by the amino acid that matches the tRNA's anticodon, and that the anticodon complements a codon that codes for the tRNA's cognate. Therefore, each tRNA misloading and codon–anticodon mismatch cause misinsertions.

Even misloaded tRNAs can still, occasionally, transfer the amino acid to the ‘right’ position, and that is when the anticodon of that misloaded tRNA is mismatched with a codon that codes for the misloaded, non-cognate amino acid. The original working hypothesis of this study is that tRNA potentials for amino acid misloading correspond to potentials for mismatching codons coding for the misloaded amino acid. If this is the case, part of the translational activity by misloaded tRNAs does not result in misinsertions. The process is termed here error compensation. Error compensation occurs by increasing the frequency of adequate combinations of mismatches and misloadings. This optimization of the genetic code and of the translational machinery would minimize the frequency of translational errors. This mechanism has to be contrasted with existing evidence for the genetic code's optimization to minimize the effects of replicational and translational errors (termed here misinsertion impact), as suggested earlier (Sonneborn, 1965; Woese, 1965a,b; Massey, 2008).

In this context, it is important to note that two different processes result in misloading: either the tRNA synthetase that is adequate for the tRNA loads an amino acid that is not its cognate, hence errors result from similarities between cognate and non-cognate amino acids; or a tRNA synthetase that is not the adequate one for the tRNA loads its cognate amino acid to the tRNA’s acceptor stem, and errors result from similarities between tRNAs. Previous analyses about optimization of the genetic code and the structure of the translational machinery (Torabi et al., 2007) address only the mechanism by which amino acid similarities cause misacylations. This mechanism is a major component of the phenomenon that minimizes misacylation impacts and hence yields identical predictions with the hypothesis of error compensation. Part of the analyses presented here account for amino acid similarities, and hence specifically test the error compensation hypothesis in the context of the mechanism where tRNA synthetases confuse tRNAs, not amino acids. These analyses also make sure that the phenomenon described is due to error compensation, and is not an indirect result of the known phenomenon that minimizes misacylation impacts.

The working hypothesis produces several testable predictions, some tested below in 4 independent, different datasets. (1) Are mismatches and misacylation correlated (tested for the most frequent (modal) tRNA sequences from human mitochondrial genomes and Escherichia coli tRNAs)? (2) Is error compensation weaker in pathogenic human polymorphisms of these mitochondrial tRNAs than in unpathogenic tRNAs? (3) Does error compensation of mitochondrial tRNAs increase developmental stability in lizards (two independent tests, for iguanid genera Anolis and Scelopus, expecting more developmental stability in species with high error compensation)? Mitochondrial genomes were chosen because ample comparative data is available (within a single species), for human tRNA mutation data, making comparisons between pathogenic and unpathogenic tRNA mutations possible, and because of availability of sequence data corresponding to lizard species for which data on developmental stability is also available (Seligmann, 1998, 2000, 2006; Seligmann et al., 2003a,b, 2008). Analyses confirm that error compensation occurs in a wide majority of tRNAs and codons, but mainly in rare ones, and that this property affects whole organism properties: error compensation is weaker in tRNA mutations that cause pathologies, and in species with high developmental instability.

2. Materials and methods

I explored the working hypothesis for the 22 mitochondrial human tRNAs and their polymorphisms, using tRNA sequences from the appendix in Seligmann (2008), which was updated using Mitomap (as accessed in early 2009, for pathogenic polymorphisms) (Ruiz-Pesini et al., 2007), and mtDB (http://www.genpat.uu.se/mitDB/ for unpathogenic polymorphisms (Ingman and Gyllensten, 2006). The stability (ΔG) of RNA duplexes formed by each of the 22 mitochondrial anticodons and all 64 codons was predicted by the online available DinaMelt server (http://www.genpat.uu.se/, Taquist et al., 2007). This software estimates the quality of alignments of the focal, input tRNA sequence with sequences of tRNAs whose cognate has been experimentally determined. The output of tafm yields alignment quality scores for each tRNA functional group, hence for the tRNA with the same cognate as the focal tRNA and all 19 tRNA species loaded by non-cognates. Note that analyses presented here arguably assume that these alignment scores estimate the aminoacylation potential of the focal tRNA with the amino acid that is the cognate of the tRNA groups used as references. This issue is in part addressed and justified by previous analyses and discussion (Seligmann, 2010a).

In order to avoid inconsistencies between annotations of sequences in Genbank, tRNA sequences were extracted using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/, Lowe and Eddy, 1997), using its organellar tRNA model option, and the vertebrate genetic code. The vertebrate mitochondrial tRNAs for the two-fold codon family of serine frequently lack a D-arm (Shimada et al., 2001), preventing detection by tRNAscan SE, which is based on an algorithm searching for patterns of covariation between regions of the sequence that match the regular cloverleaf secondary structure, which includes a D-arm. I used Arwen (http://130.235.46.10/ARWEN/, Laslett and Canback, 2008) to detect mitochondrial tRNA Ser GCU.

Statistical tests follow standard procedures. I used weighted linear regressions between misacylation and mismatch potentials, weighing by the frequencies of the corresponding codons in the human mitochondrial protein coding genes. Standard t-tests between independent samples were used when adequate. In specific cases, partial correlation analyses were done, using standard procedures for that calculation. Each test was done on each tRNA. Therefore, statistical trends over the complete set of tRNAs (meta-analyses) were evaluated by combining P-values of the independent tests from the 22 mitochondrial tRNAs, using Fisher’s method to combine P. This method sums over all k tests – 2 × log P, where i ranges from 1 to k (for example, in human mitochondria, there are 22 tRNAs, hence k = 22). This sum is a chi-square statistic with 2 × k degrees of freedom. In order to detect whether associations are significant for specific tRNAs, while considering multiple testing, I used the Benjamini–Hochberg readjustment of Bonferroni’s adjustment for multiple testing (Benjamini and Hochberg, 1995), as Bonferroni’s method is over-conservative (Perneger, 1998).

3. Results

3.1. Misacylation potentials of mitochondrial tRNAs

The code for aminoaclayion specificity is not yet well understood, especially for mitochondria (Taquist et al., 2007). This means that for mitochondrial tRNAs, tfam does not necessarily predict the cognate amino acid as having the greatest potential for loading the tRNA. Indeed, for the 22 mitochondrial tRNAs, the aminoaclayion potential (Table 2 in Seligmann (2010a)) predicted for the 19 non-cognate amino acids is greater than that of the cognate in 144 among 418 (34%) combinations of tRNAs and non-cognate amino acids. The scores in that Table 2 (Seligmann, 2010a) are log-odds of similarity measures between the input tRNA sequence and each tRNA functional group used by tfam. Positive values indicate greater than random similarity, negative values indicate similarity lower than for random input sequences. The z-transformed alignment score for the cognate amino acid is calculated by subtracting the mean score across all columns from the score for the column with the cognate, and dividing this difference by the standard deviation of these scores. This score is positive for 15 among 22 tRNAs (68%, mean z = 0.37 ± 1.17). This means that the cognate’s aminoaclayion potential is greater than average in the majority of tRNA species (P < 0.03345 according to a one sided sign test). Nevertheless, if tfam’s output estimates aminoaclayion potentials, an apparently unreasonable amount of tRNA misloading would occur. Hence tfam’s alignment scores might be inadequate to predict aminoaclayion potentials, especially for mitochondrial tRNAs (tfam predicts much better tRNA function for non-organellar tRNAs). These results are produced by an analysis of tfam’s output that assumes that correct aminoaclayion of a specific tRNA results from competition between tRNA synthetases, however, it is also possible that different tRNAs compete for a given tRNA synthetase. Indeed, analysing tfam’s output along the latter principle yields aminoaclayion specificity that is greater than random, and tends to compensate for low aminoaclayion specificity due to competition among tRNA synthetases (Seligmann, 2010a). Hence tfam’s output apparently yields relatively valid estimates of aminoaclayion potentials, even when classical straightforward analyses of these estimates do not detect the cognate amino acid that should match the tRNA’s anticodon according to the genetic code. It is less the estimates of tfam’s output than their interpretation that have to be reconsidered with caution, as the conundrum of aminoaclayion specificity by competition between tRNA synthetases versus competition between tRNAs reveals.

3.2. Posttranscriptional mitochondrial tRNA modifications

Another important point is to note that the various alignment analyses (tfam as well as the one presented below) do not take into account posttranscriptional tRNA modifications. However, for human mitochondrial tRNAs, such modifications have been reported for (only) 6 among the 22 tRNAs (tRNA followed by the number of modifications): tRNA Ile 5; tRNA Leu UUR 9; tRNA Lys 6; tRNA Pro 8; tRNA Ser AGY 2; and tRNA Ser UCN 5 (Florentz et al., 2003). Estimates of tendencies for cognate acylation (the score estimating the tendency for aminoaclayion of a tRNA by a given amino acid according to the software tfam) (http://tfam.lcb.uu.se; Taquist et al., 2007) tend to decrease with the number of posttranscriptional modifications on that tRNA (P < 0.05). Hence these modifications probably decrease our ability to detect and estimate the tRNA’s amino-acylation tendency for its cognate. However, the effect seems weak, and general principles deduced from the results are probably little affected by these posttranscriptional modifications. Results in various sections will be analysed and discussed according to this information on presence or absence of posttranscriptional modifications.

3.3. Mismatches between codons and mitochondrial anticodons

The 22 anticodons found in mitochondrial tRNAs interact with different numbers of codons, and these interactions yield different stabilities. For example, the Dinamelt server predicts thermodynamically viable interactions between tRNA Gly’s anticodon UCC and 12 among 64 potential codons. Four codon for Gly, and the mean ΔG of their interaction with UCC is −0.025 kcal per mole (ΔG = −1.1 kcal/mol for the GCC-GGA anticodon–anticodon pair). Anticodon UCC has the potential to mismatch only 8 codons (mean ΔG = 0.6125 kcal/mol). The positive ΔG indicates that a small amount of energy has to be invested for the codon–anticodon interaction to occur. Univable interactions require unreasonable energy investment. Dinamelt displays for these 99 kcal/mol. For the anticodon of tRNA His, GUG, there are 46 codons that yield thermodynamically viable interactions according to Dinamelt, among which 44 do not code for His (mean ΔG = 1.35 kcal/mol) and 2 for His (mean ΔG = 0.75 kcal/mol). ΔG = 0.2 kcal/mol for the GUG-CAC anticodon–codon RNA duplex. For Gln, no codon formed a more stable duplex with the anticodon than those coding for Gln, but for His, 7 codons that code for other amino acids form more stable codon–anticodon RNA duplexes. Table 1 describes such data for all 22 anticodons found in vertebrate mitochondrial tRNAs. Note that various wobble position modifications have been detected for 6 vertebrate mitochondrial tRNAs (tRNA Glu, tRNA Gln, tRNA Leu UUR, tRNA Lys, tRNA Met, and tRNA Trp, from Watanabe, 2007). These might affect the stabilities of the codon–anticodon mismatches used here. Results are discussed and evaluated considering this issue.

3.4. Codon–anticodon mismatches and potentials for mitochondrial tRNA misacylations

The aminoaclayion potentials as determined by tfam (i.e. in Table 2 from Seligmann (2010a), can be paired with codon–anticodon duplex stabilities according to the identities of the codons and the amino acids. Hence, for each tRNA, one can test whether stabilities of codon–anticodon interactions correlate with potentials for aminoaclayion of the tRNA with that anticodon by the amino acid coded by the codon mismatched by that anticodon. Because occurrences of codon–anticodon interactions depend on codon frequencies in mRNAs, correlation analyses weight data according to frequencies of corresponding codons in human mitochondrial protein coding genes (counted using the software at http://www.kazusa.or.jp/codon/countcodon.html). Table 1 presents rw, the weighted linear regression coefficient of determination between misacylation potentials and codon–anticodon ΔGs for each mitochondrial tRNA, according to row- and column analyses of tfam’s output (see previous section, row analysis assumes that aminoaclayion specificity results from competition among tRNA synthetases for tRNA aminoacylation; column analysis assumes that that aminoacylation specificity results from competition among tRNAs for tRNA synthetases). In both cases, directions of most rw’s are as expected by the hypothesis of avoidance of misinsertions by error compensation (matching misacylations with mismatches): more stable (more negative) ΔGs associate with relatively high aminoaclayion potentials (from Table 2 in Seligmann (2010a)). Hence 15 and 16 rw’s are negative for row- and column-based analyses, respectively (68% and 73%, which is significant according

Table 1

<table>
<thead>
<tr>
<th>tRNA</th>
<th>Anti</th>
<th>N</th>
<th>ΔC</th>
<th>Cogn</th>
<th>Compl</th>
<th>&lt;s</th>
<th>Row rw</th>
<th>P</th>
<th>Part</th>
<th>Col rw</th>
<th>P</th>
<th>Part</th>
<th>Row ptfam</th>
<th>Col ptfam</th>
<th>E coli ptfam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>TCC</td>
<td>38</td>
<td>1.512</td>
<td>0.675</td>
<td>0.11</td>
<td>0</td>
<td>0.23</td>
<td>0.16</td>
<td>-0.31</td>
<td>-0.43</td>
<td>0.007</td>
<td>-0.43</td>
<td>-0.04</td>
<td>0.15</td>
<td>-0.10</td>
</tr>
<tr>
<td>Arg</td>
<td>TCT</td>
<td>24</td>
<td>1.005</td>
<td>0.075</td>
<td>-0.2</td>
<td>4</td>
<td>0.52</td>
<td>0.01</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.769</td>
<td>-0.52</td>
<td>-0.03</td>
<td>0.29</td>
<td>0.37</td>
</tr>
<tr>
<td>Asn</td>
<td>GCT</td>
<td>44</td>
<td>1.791</td>
<td>1.650</td>
<td>-0.1</td>
<td>8</td>
<td>0.18</td>
<td>0.25</td>
<td>-0.22</td>
<td>0.27</td>
<td>0.079</td>
<td>0.29</td>
<td>0.18</td>
<td>0.26</td>
<td>0.13</td>
</tr>
<tr>
<td>Asp</td>
<td>GTC</td>
<td>40</td>
<td>1.605</td>
<td>0.250</td>
<td>0.3</td>
<td>0</td>
<td>-0.01</td>
<td>0.942</td>
<td>-0.05</td>
<td>-0.11</td>
<td>0.497</td>
<td>-0.08</td>
<td>0.22</td>
<td>0.04</td>
<td>0.39</td>
</tr>
<tr>
<td>Cys</td>
<td>GCA</td>
<td>21</td>
<td>-0.368</td>
<td>-0.350</td>
<td>1</td>
<td>7</td>
<td>0.37</td>
<td>0.101</td>
<td>0.34</td>
<td>0.28</td>
<td>0.217</td>
<td>0.41</td>
<td>-0.13</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>Glu</td>
<td>TCT</td>
<td>44</td>
<td>1.957</td>
<td>1.300</td>
<td>1.3</td>
<td>3</td>
<td>-0.59</td>
<td>0.0001</td>
<td>-0.05</td>
<td>-0.03</td>
<td>0.840</td>
<td>-0.01</td>
<td>0.20</td>
<td>0.18</td>
<td>0.34</td>
</tr>
<tr>
<td>Glu</td>
<td>TCC</td>
<td>20</td>
<td>2.067</td>
<td>0.800</td>
<td>0.8</td>
<td>0</td>
<td>0.19</td>
<td>0.43</td>
<td>-0.20</td>
<td>-0.48</td>
<td>0.030</td>
<td>-0.57</td>
<td>0.01</td>
<td>-0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>Gly</td>
<td>CTC</td>
<td>12</td>
<td>0.613</td>
<td>0.025</td>
<td>-1.1</td>
<td>0</td>
<td>0.02</td>
<td>0.942</td>
<td>-0.15</td>
<td>0.35</td>
<td>0.263</td>
<td>0.20</td>
<td>0.01</td>
<td>0.14</td>
<td>0.39</td>
</tr>
<tr>
<td>His</td>
<td>GTG</td>
<td>46</td>
<td>1.348</td>
<td>0.750</td>
<td>0.2</td>
<td>7</td>
<td>0.48</td>
<td>0.0001</td>
<td>-0.44</td>
<td>-0.11</td>
<td>0.474</td>
<td>-0.02</td>
<td>0.40</td>
<td>0.24</td>
<td>0.33</td>
</tr>
<tr>
<td>Ile</td>
<td>CAT</td>
<td>26</td>
<td>1.608</td>
<td>1.500</td>
<td>0.9</td>
<td>5</td>
<td>0.60</td>
<td>0.009</td>
<td>0.52</td>
<td>0.37</td>
<td>0.064</td>
<td>0.41</td>
<td>-0.18</td>
<td>0.22</td>
<td>-0.30</td>
</tr>
<tr>
<td>Leu</td>
<td>TAG</td>
<td>24</td>
<td>2.106</td>
<td>1.867</td>
<td>1.3</td>
<td>0</td>
<td>0.26</td>
<td>0.217</td>
<td>0.25</td>
<td>-0.38</td>
<td>0.070</td>
<td>-0.38</td>
<td>0.14</td>
<td>0.14</td>
<td>0.52</td>
</tr>
<tr>
<td>Leu</td>
<td>TAA</td>
<td>19</td>
<td>2.134</td>
<td>2.300</td>
<td>2</td>
<td>6</td>
<td>0.20</td>
<td>0.418</td>
<td>0.38</td>
<td>&gt;0.17</td>
<td>0.476</td>
<td>0.09</td>
<td>-0.14</td>
<td>0.14</td>
<td>0.60</td>
</tr>
<tr>
<td>Lyt</td>
<td>TCT</td>
<td>20</td>
<td>2.189</td>
<td>2.200</td>
<td>2.2</td>
<td>7</td>
<td>-0.16</td>
<td>0.492</td>
<td>-0.18</td>
<td>-0.07</td>
<td>0.762</td>
<td>-0.07</td>
<td>0.21</td>
<td>0.23</td>
<td>0.12</td>
</tr>
<tr>
<td>Met</td>
<td>CAT</td>
<td>20</td>
<td>2.050</td>
<td>1.800</td>
<td>1.4</td>
<td>5</td>
<td>0.41</td>
<td>0.075</td>
<td>-0.41</td>
<td>-0.09</td>
<td>0.712</td>
<td>-0.16</td>
<td>-0.16</td>
<td>0.18</td>
<td>0.07</td>
</tr>
<tr>
<td>Phe</td>
<td>GAA</td>
<td>1.4</td>
<td>1.477</td>
<td>1.150</td>
<td>0.8</td>
<td>5</td>
<td>0.44</td>
<td>0.119</td>
<td>0.16</td>
<td>0.48</td>
<td>0.086</td>
<td>0.10</td>
<td>-0.10</td>
<td>0.40</td>
<td>-0.03</td>
</tr>
<tr>
<td>Pro</td>
<td>TGG</td>
<td>46</td>
<td>1.407</td>
<td>2.200</td>
<td>-0.6</td>
<td>0</td>
<td>-0.26</td>
<td>0.081</td>
<td>-0.02</td>
<td>-0.18</td>
<td>0.229</td>
<td>0.39</td>
<td>0.19</td>
<td>0.38</td>
<td>0.35</td>
</tr>
<tr>
<td>Ser</td>
<td>GCT</td>
<td>23</td>
<td>0.110</td>
<td>0.350</td>
<td>0.5</td>
<td>5</td>
<td>0.50</td>
<td>0.014</td>
<td>0.29</td>
<td>0.07</td>
<td>0.739</td>
<td>0.31</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.29</td>
</tr>
<tr>
<td>Ser</td>
<td>TCA</td>
<td>40</td>
<td>1.489</td>
<td>0.425</td>
<td>0</td>
<td>0</td>
<td>0.20</td>
<td>0.207</td>
<td>-0.15</td>
<td>-0.01</td>
<td>0.936</td>
<td>0.17</td>
<td>-0.12</td>
<td>-0.02</td>
<td>0.10</td>
</tr>
<tr>
<td>Thr</td>
<td>TCT</td>
<td>46</td>
<td>1.760</td>
<td>0.625</td>
<td>0.2</td>
<td>0</td>
<td>0.28</td>
<td>0.064</td>
<td>-0.23</td>
<td>-0.20</td>
<td>0.180</td>
<td>-0.05</td>
<td>-0.09</td>
<td>0.19</td>
<td>0.29</td>
</tr>
<tr>
<td>Trp</td>
<td>TCA</td>
<td>18</td>
<td>0.675</td>
<td>0.450</td>
<td>0</td>
<td>5</td>
<td>0.51</td>
<td>0.032</td>
<td>-0.52</td>
<td>-0.16</td>
<td>0.525</td>
<td>-0.18</td>
<td>0.08</td>
<td>0.22</td>
<td>0.25</td>
</tr>
<tr>
<td>Tyr</td>
<td>GTA</td>
<td>38</td>
<td>1.256</td>
<td>1.450</td>
<td>0.98</td>
<td>12</td>
<td>0.46</td>
<td>0.004</td>
<td>-0.43</td>
<td>-0.24</td>
<td>0.149</td>
<td>-0.29</td>
<td>-0.03</td>
<td>0.21</td>
<td>0.31</td>
</tr>
<tr>
<td>Val</td>
<td>TAC</td>
<td>19</td>
<td>2.093</td>
<td>1.350</td>
<td>0.9</td>
<td>0</td>
<td>0.58</td>
<td>0.009</td>
<td>-0.59</td>
<td>-0.45</td>
<td>0.056</td>
<td>-0.49</td>
<td>-0.12</td>
<td>-0.02</td>
<td>-0.40</td>
</tr>
<tr>
<td>Chi²</td>
<td></td>
<td>114</td>
<td>1 × 10⁸</td>
<td>74</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a Cognate and b anticodon identities, c numbers of codons forming duplexes with that anticodon, d mean stability of mismatched codon–anticodon duplexes, e mean ΔG for duplexes with codons coding for the cognate, f ΔG of the canonical complementary codon–anticodon duplex, g number of mismatches that are more stable than the canonical complementary codon–anticodon duplex, h weighted correlation coefficients between codon–anticodon duplex ΔG and aminoacylation potentials by amino acids coded by the codon, according to row, and i column analysis of tfam’s output (Table 2 in Seligmann (2010a)), followed by respective one tailed statistical significances; j, k partial correlation coefficients after controlling for evolutionary exchangeabilities between tRNA cognate and the amino acid coded by the codon; correlations coefficients between evolutionary exchangeability of cognate and misacylated amino acids according to the PAM 250 amino acid substitution matrix and tfam’s aminocaylation potentials, for row- and column-based analyses of mitochondrial tRNAs, and m row analyses of E. coli tRNAs with the same anticodon as the corresponding mitochondrial tRNA.
3.6. Mitochondrial wobble modifications

Posttranscriptional modifications are a potential confounding variable not only for estimates of misacylation, but also in relation to estimates of codon–anticodon interactions, when they involve wobble position modifications (the anticodon’s first position, which corresponds to the codon’s third position). Different tRNAs undergo different modifications, according to which organism (Knight et al., 2001), but in general, these modifications alter the thermodynamic properties of the codon–anticodon interactions (Vendeix et al., 2008). These effects are not yet included in predictions of standard algorithms for predicting hybridization, as the one used here (Mfold). Hence results are discussed according to potential effects of wobble modifications, but these are not yet included in predictions. Various wobble position modifications have been detected for 6 vertebrate mitochondrial tRNAs (Watanabe, 2007). These modifications are typically designed to expand the recognition of anticodons to the whole (or most) codons within the same codon family, hence for codons assigned to the same amino acid as the codon complementing the unmodified anticodon. On the other hand, some modifications seem to restrict codon recognition, hence increase anticodon specificities to the ‘correct’ codons. Both effects could influence codon–anticodon mismatches, and alter results of analyses for error compensation. In principle, and when taking into account all tRNAs, not accounting for these effects most probably adds noise to the data. Hence patterns detected despite not taking them into account would most probably be strengthened if these were taken into account, but they are unlikely to be the cause of...
the general patterns observed (that error compensation occurs in most tRNAs). Indeed, error compensation as estimated by rws from Table 1 does not differ significantly between tRNAs where wobble modifications occur, versus those for which no such modifications have been reported.

3.7. Amino acid mutability as a confounding factor

The structure of the genetic code could be at the origin of the positive correlation between mismatch stability and misacylation potential. Indeed, physico-chemical properties of amino acids tend to correlate with properties of their codons and anticodons, and as a result, similar amino acids have similar codons and anticodons. In addition, the nucleotide content of tRNAs tends to resemble that of the anticodon (Seligmann and Amzallag, 2002), which could affect alignments in a way that similarities between complete tRNA sequences reflect similarities between anticodons. These phenomena result in optimizations of the genetic code that have been already described: replication and translation are robust to errors, because codon-anticodon mismatches, respectively misacylated tRNAs, preferentially misinsert amino acids that have physico-chemical properties that are similar to those of the cognate amino acid. This would ensure that errors on average do not change much the protein’s properties. This adaptation of the genetic code promotes conservative misinsertions, it minimizes misinsertion impacts despite mismatches and misacylations. Note again that in contrast, the hypothesis tested here is about minimizing the frequency of misinsertions by adequately combining misacylation with codon-anticodon mismatch.

In the context of this study, it is important to test whether the negative correlations described in the previous section for rw are due to the latter, known phenomena that minimize the effects of misinsertions, or whether they also reflect a previously unknown adaptation, which avoids misinsertions by error compensation.

Amino acids that have similar physico-chemical properties are exchangeable. This is reflected at evolutionary level by the structure of amino acid replacement matrices, such as the commonly used PAM matrix. The frequency of such exchanges is proportional to physico-chemical similarities between amino acids, such as Grantham’s distance (Grantham, 1974). I used the PAM matrix as calculated for 250 generations (http://www.bioinformatics.nl/tools/pam.html). In order to control for similarity/evolutionary exchangeability between amino acid pairs, I calculated partial correlation coefficients between codon-anticodon stabilities and t-fm’s scores for aminoacylation potentials. Unlike in the analyses described in the previous sections that yield rw, partial correlation controls for effects of other variables. Here I controlled for PAM’s scores for evolutionary exchangeability between the cognate amino acid and the presumed misinserted amino acid. Note that also here data points were weighted proportionally to codon frequencies in human mitochondrial mRNAs. These weighted partial correlation coefficients are shown in Table 1 for each row- and column analyses, next to the respective regular weighted correlation coefficients rw described in a previous section. Their columns are headed ‘Part’. For row analyses, results are qualitatively similar for regular and partial correlation analyses: 15 among 22 weighted partial correlation coefficients are negative, as for rw. This means that the positive association between mismatch and misacylation exists independently of similarities between the cognate and potentially misinserted non-cognate amino acids, and hence independently of the known optimization to minimize the impact of misinsertions.

For analyses of t-fm’s output by columns, only 13 among 22 partial correlation coefficients are negative, which is less than the 16 found by regular correlation analyses. This difference between row and column analyses means that similarities between amino acids affect misacylation of the pool of tRNAs by specific tRNA synthetases, but that it does not affect much misacylation of specific tRNAs by the pool of tRNA synthetases. In other words, competition between tRNAs is affected by the similarity between their cognates, but competition between tRNAs synthetases is not. This result was predictable: it is reasonable to consider that tRNA synthetases have higher specificities for their cognate amino acid than tRNAs. Hence error rates by tRNA synthetases are less affected by physico-chemical similarity between amino acids than those of tRNAs. The results confirm, at least for row-analyses (which reflect competition between tRNA synthetases for a specific tRNA (Seligmann, 2010a)), that mismatches and misacylations are matched to avoid misinsertions. For column analyses (which reflect competition between tRNAs for a specific tRNA synthetase (Seligmann, 2010a)), results suggest that much of the phenomenon of misacylation relates to minimization of misacylation impacts. Hence between the two mechanisms that promote aminoacylation specificity, competition between tRNA synthetases minimizes the frequency of misinsertions, while competition between tRNAs minimizes more the disruptive impact of misinsertions.

3.8. Misacylation in pathogenic versus unpathogenic tRNA polymorphisms

The previous sections indicate that potentials of tRNAs for misacylation and the stability of interactions between tRNA anticodons and codons coding for the misacylated amino acids are proportional. This means that misacylated tRNAs will tend to mismatch with codons that code for the misloaded amino acid. This error compensation presumably decreases amino acid misinsertions. Misinsertions probably disrupt protein function, and cumulation of dysfunctional proteins causes ageing-related degenerative mitochondrial diseases (Selkoe, 2004). About 30% of all pathogenic mitochondrial mutations are from tRNA sequences (from Florentz et al. (2003), but note that the collection of polymorphisms used here is updated using mitomap, http://www.mitomap.org/bin/view/MITOMAP/MutationsRNA (Ruiz-Pesini et al., 2007)) for pathogenic polymorphisms and mtDB (http://www.genpat.uu.se/mtDB/ (Ingman and Gyllensten, 2006) for other polymorphisms. Hence it is plausible that tRNA misacylation and the resulting amino acid misinsertions are involved in diseases associated with tRNA mutations.

Table 2 shows the t statistics of comparisons between mean t-fm cognate amino acid potentials of pathogenic and unpathogenic tRNAs, for each of the 22 tRNAs (column marked t-tfam in Table 2). Meta-analysis of Table 2 using Fisher’s method for combining Ps of t-tfam does not detect that t-fm’s aminoacylation potentials for cognates are lower for tRNAs bearing pathogenic polymorphisms, as compared to homologous tRNAs with unpathogenic mutations: for some tRNAs, notably for tRNA Asn and tRNA Lys, pathogenic mutations on average decrease aminoacylation potentials for cognates, but the opposite is observed for other tRNAs (tRNA Arg and tRNA Trp). Overall, no clear trend is detected, and statistical significances for single tRNAs are not strong enough to detect with certainty such an effect at the level of any single tRNA when considering the number of tests performed: none of the Ps for t-tfam in Table 2 approaches P = 0.05 after adjustments for multiple tests.

The variation between tRNA species in mean effects of pathogenic mutations as compared to unpathogenic ones on t-fm’s misacylation scores correlates with the number of posttranscriptional modifications (from Fig. 3 in Florentz et al., 2003) (patterns are unchanged if information on post-transcriptional modifications from more recent literature (Moellers et al., 2005; Hao and Wang, 2006; Voigt-Hoffmann et al., 2007) is included): on average, misacylation scores of pathogenic mutants are greater than those of
unpathogenic polymorphisms for mitochondrial tRNA species with post-transcriptional modifications (mean t-tfam = −0.85 ± 0.88, t = −2.37, one tailed P = 0.032), while the opposite is the case for those lacking modifications (mean t-tfam = 0.58 ± 1.25, t = 1.86, one tailed P = 0.042). These calculations tested whether t-tfam is on average below and above zero, respectively, using one sample t-tests, for tRNA species where modifications are absent, and, respectively, present. Comparing these two groups, t-tfam is less negative (pathogenic mutations with less misacylations than unpathogenic ones) in the group of tRNA species lacking modifications (t-test of difference between two independent samples, t = 2.56, two tailed P = 0.019). Hence, while no clear overall effect of pathogenic mutations on t-fam misacylation scores is detectable per se, associations between misacylation and pathogenicity interact with the presence/absence of post-transcriptional modifications of that specific human tRNA species. Results do not indicate any particular effect of modifications at the wobble position, and hence effects of modifications on the presented results are more likely causing noise in the estimates of misacylation than of mismatches due to wobble position modifications.

3.9. Less error coordination in pathogenic tRNAs

Further analyses suggest that the ‘absolute’ potential for cognate acylation (previous section) is less relevant to pathogenicity than error compensation, as estimated by rw (calculated as presented in Fig. 1 and Table 1), the association between codon–anticodon mismatch stability and misacylation potentials. Table 1 presents rw for the most frequent human mitochondrial tRNA polymorphisms. Data in Table 2 show results of tests of differences between mean rw of homologous pathogenic tRNA polymorphisms and mean rw of homologous unpathogenic tRNA polymorphisms. Note that codon–anticodon mismatch stabilities are identical for all polymorphisms besides the very few ones where the polymorphism affects the anticodon, which were excluded from these analyses. The estimated rw is calculated only for row analyses of t-fam’s output. The column headed t-rw in Table 2 shows that the average rw of pathogenic tRNAs is less negative than the average rw for homologous unpathogenic tRNAs in 14 among 22 tRNAs. Hence error compensation is weaker in pathogenic tRNAs than unpathogenic ones. The difference, according to t-tests, was significant at P < 0.05 for 4 tRNAs (tRNA Gln, tRNA Ile, tRNA Lys, and tRNA Pro). This difference remains significant for tRNA Lys after Bonferroni’s adjustment for multiple tests. Hence for this tRNA known to have a particularly high proportion of pathogenic polymorphisms, it seems that error compensation is particularly important to avoid disruption of mitochondrial function. The meta-analysis of the results in Table 2 confirms that this is the case overall, for the pool of all 22 mitochondrial tRNAs: Fisher’s multiple test statistic of the results in Table 2 is P < 0.00136. Hence the disparity in rw between pathogenic and unpathogenic tRNAs is significant at P < 0.05 after Bonferroni correction for multiple tests.

Table 2

Comparisons between misacylation of pathogenic and unpathogenic polymorphisms of 22 human mitochondrial tRNAs.

<table>
<thead>
<tr>
<th>tRNA&lt;sup&gt;y&lt;/sup&gt;</th>
<th>N&lt;sup&gt;b&lt;/sup&gt;</th>
<th>t-tfam&lt;sup&gt;f&lt;/sup&gt;</th>
<th>2 tail&lt;sup&gt;i&lt;/sup&gt;</th>
<th>t-rw&lt;sup&gt;e&lt;/sup&gt;</th>
<th>2 tail&lt;sup&gt;j&lt;/sup&gt;</th>
<th>t-pfam&lt;sup&gt;g&lt;/sup&gt;</th>
<th>2 tail&lt;sup&gt;h&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>10-4</td>
<td>−0.321</td>
<td>0.753</td>
<td>−0.759</td>
<td>0.462</td>
<td>−0.024</td>
<td>0.981</td>
</tr>
<tr>
<td>Arg</td>
<td>11-2</td>
<td>0.674</td>
<td>0.022</td>
<td>−1.19</td>
<td>0.259</td>
<td>1.578</td>
<td>0.143</td>
</tr>
<tr>
<td>Asn</td>
<td>7-4</td>
<td>2.303</td>
<td>0.047</td>
<td>1.097</td>
<td>0.301</td>
<td>0.441</td>
<td>0.67</td>
</tr>
<tr>
<td>Asp</td>
<td>12-1</td>
<td>0.286</td>
<td>0.78</td>
<td>0.561</td>
<td>0.586</td>
<td>0.347</td>
<td>0.735</td>
</tr>
<tr>
<td>Cys</td>
<td>18-4</td>
<td>0.587</td>
<td>0.207</td>
<td>0.57</td>
<td>0.605</td>
<td>1.558</td>
<td>0.21</td>
</tr>
<tr>
<td>Gin</td>
<td>16-4</td>
<td>0.43</td>
<td>0.672</td>
<td>−0.885</td>
<td>0.383</td>
<td>−1.775</td>
<td>0.093</td>
</tr>
<tr>
<td>Gly</td>
<td>9-7</td>
<td>0.214</td>
<td>0.834</td>
<td>0.402</td>
<td>0.688</td>
<td>2.322</td>
<td>0.039</td>
</tr>
<tr>
<td>Gly</td>
<td>13-5</td>
<td>0.904</td>
<td>0.379</td>
<td>−2.273</td>
<td>0.037</td>
<td>2.208</td>
<td>0.042</td>
</tr>
<tr>
<td>His</td>
<td>11-3</td>
<td>0.912</td>
<td>0.38</td>
<td>0.145</td>
<td>0.887</td>
<td>0.159</td>
<td>0.876</td>
</tr>
<tr>
<td>Ile</td>
<td>10-13</td>
<td>0.798</td>
<td>0.434</td>
<td>2.154</td>
<td>0.043</td>
<td>0.253</td>
<td>0.803</td>
</tr>
<tr>
<td>Leu CUN</td>
<td>11-10</td>
<td>0.231</td>
<td>0.819</td>
<td>0.440</td>
<td>0.658</td>
<td>−0.197</td>
<td>0.846</td>
</tr>
<tr>
<td>Leu UIR</td>
<td>7-23</td>
<td>0.553</td>
<td>0.584</td>
<td>0.78</td>
<td>0.442</td>
<td>1.056</td>
<td>0.217</td>
</tr>
<tr>
<td>Lys</td>
<td>9-14</td>
<td>−1.731</td>
<td>0.098</td>
<td>−3.494</td>
<td>0.041</td>
<td>1.272</td>
<td>0.217</td>
</tr>
<tr>
<td>Met</td>
<td>6-3</td>
<td>1.064</td>
<td>0.123</td>
<td>0.294</td>
<td>0.777</td>
<td>0.139</td>
<td>0.893</td>
</tr>
<tr>
<td>Phe</td>
<td>12-7</td>
<td>1.192</td>
<td>0.25</td>
<td>0.13</td>
<td>0.896</td>
<td>−0.06</td>
<td>0.96</td>
</tr>
<tr>
<td>Pro</td>
<td>12-5</td>
<td>−1.165</td>
<td>0.262</td>
<td>−2.007</td>
<td>0.063</td>
<td>3.293</td>
<td>0.005</td>
</tr>
<tr>
<td>Ser AGY</td>
<td>19-3</td>
<td>0.278</td>
<td>0.784</td>
<td>0.04</td>
<td>0.968</td>
<td>0.355</td>
<td>0.726</td>
</tr>
<tr>
<td>Ser UCN</td>
<td>13-8</td>
<td>1.671</td>
<td>0.111</td>
<td>−0.792</td>
<td>0.438</td>
<td>1.471</td>
<td>0.158</td>
</tr>
<tr>
<td>Thr</td>
<td>7-23</td>
<td>1.171</td>
<td>0.249</td>
<td>0.382</td>
<td>0.705</td>
<td>−0.651</td>
<td>0.519</td>
</tr>
<tr>
<td>Trp</td>
<td>15-6</td>
<td>2.023</td>
<td>0.057</td>
<td>1.417</td>
<td>0.173</td>
<td>2.217</td>
<td>0.039</td>
</tr>
<tr>
<td>Tyr</td>
<td>7-4</td>
<td>1.744</td>
<td>0.115</td>
<td>0.925</td>
<td>0.379</td>
<td>1.833</td>
<td>0.1</td>
</tr>
<tr>
<td>Val</td>
<td>12-5</td>
<td>0.681</td>
<td>0.53</td>
<td>1.106</td>
<td>0.330</td>
<td>−2.098</td>
<td>0.053</td>
</tr>
<tr>
<td>Chi&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>52.63</td>
<td>0.18</td>
<td>1.49</td>
<td>0.00136</td>
<td>68.47</td>
<td>0.011</td>
</tr>
</tbody>
</table>

analyses show that, when comparing different, non-homologous tRNA species, the more modifications a tRNA usually undergoes, the less error compensation is detected. But the analyses in this section compare polymorphisms of homologous tRNAs, and effects of modifications are weaker, because homologous tRNAs are likely to have less variation in that respect. Some of these polymorphisms probably affect modifications, but comparisons of average effects are probably little affected by this, as most polymorphisms probably do not affect modifications. Therefore, effects detected for modifications are weaker than those indicated in other sections, because comparisons between homologous tRNAs account to a larger extent for modifications. This point is important, because it suggests that analyses that yield the positive result that there is less error compensation in pathogenic polymorphisms than in unpathogenic ones has a greater level of certainty considering the caveats associated with not accounting for posttranscriptional tRNA modifications than for previous analyses.

3.10. Pathogenic mutations decrease minimization of misinsertion impact

It is interesting to test whether pathogenic mutations differ from regular polymorphisms also in respect to minimization of misinsertion impacts on proteins. For that purpose, I calculated correlation coefficients between PAM’s exchangeability scores and tfams’ aminoacylation potentials. This correlation is positive for most tRNAs (Table 1, p-tfam), especially for column-based analyses of tfam’s output. This indicates that the translational machinery minimizes the impact of misinsertions, by preferably misinserting amino acids with physico-chemical properties that resemble those of the amino acid originally coded by the gene. Hence the effects of misinsertions on protein function are minimized. This is especially the case for misacylations due to competition among tRNAs, as already suggested by analyses in a previous section. Table 2 shows that this correlation (which presumably estimates minimization of misinsertion impact) is overall weaker for pathogenic tRNA mutations than regular non-pathogenic mutations, but the differences between pathogenic and unpathogenic tRNAs are never strong enough to be significant at the level of specific tRNAs. Hence while minimization of misinsertion impact seems involved in the ontogeny of mitochondriopathies, because effects seem ubiquitous, these are overall weak. Most notably, these pathogenic effects are generally weaker than the pathogenic effects associated with error compensation (t-rw in Table 2), hence minimizing the number of misinsertions seems more important for proper protein function than minimizing misinsertion impacts, at least in relation to slow cumulative ageing-associated diseases. This also means that competition between tRNAs is less important as a pathogenic mechanism than competition between tRNA synthetases (results above show that the former mechanism primarily minimizes misinsertion impact, the latter promotes primarily error compensation).

3.11. Developmental stability in lizards and error compensation

Previous studies showed that pathologies associate with effects of mutations on properties of tRNA genes that are relevant to the secondary function of tRNA genes as alternative light strand replication origins (Seligmann, 2008, 2010b). This secondary function seems also to affect developmental stability, not only as expressed by the ontogeny of human mitochondrial diseases, but also by small perturbations of late embryonic development of lizards, estimated by the random component (fluctuating asymmetry) of differences between left and right counts of subdigital lamellae (these counts
Developmental stability in bilateral counts of subdigital lamellae under the 4th toe in *Sceloporus* as a function of mean error compensation for 8 tRNAs in these species. Correlations of developmental stability with error compensation in each of the 8 tRNAs are also indicated. Bold values are for correlations that are significant at \( P < 0.05 \) after Bonferroni correction for multiple tests.

![Developmental stability in bilateral counts of subdigital lamellae under the 4th toe in *Sceloporus*](image)

**Fig. 3.** Developmental stability in bilateral counts of subdigital lamellae under the 4th toe in *Sceloporus* as a function of mean error compensation for 8 tRNAs in these species. Correlations of developmental stability with error compensation in each of the 8 tRNAs are also indicated. Bold values are for correlations that are significant at \( P < 0.05 \) after Bonferroni correction for multiple tests.

**3.1. tRNAs in *Escherichia coli***

As noted above, analyses presented here rely on alignment scores of mitochondrial tRNAs as produced by tfam. The software is least reliable for predicting tRNA cognates for mitochondrial tRNAs, and is better adapted for bacterial tRNAs. In order to show the biological generality of the results, as well as their robustness despite the relative inadequacy of tfam for mitochondrial tRNAs, analyses similar to those done for mitochondrial tRNAs in *Table 1* are done for tRNAs from the genome of *E. coli* O157H7, chosen because it is the strain with the largest tRNA number detected by tRNAscan-SE (Lowery and Eddy, 1997, [http://lowelab.ucsc.edu/GtRnAdb/#bacteria](http://lowelab.ucsc.edu/GtRnAdb/#bacteria)).

For each codon, one (randomly chosen) tRNA sequence per synonymous anticodon was analysed. Here too, I used weighted correlation analyses, weighting datapoints by codon usage frequencies in genome-wide *E. coli* protein-coding genes. Information on mRNA frequencies was not used to improve the realism for the weights used in the analyses for *E. coli*, in order for analyses to be as similar as possible to those done for mitochondria, for which such information on mRNA abundances is not readily available, keeping in mind that the major point of the *E. coli* analyses is to confirm results from mitochondria. *Table 3* presents results of these analyses for tRNAs with the same anticodons as those in vertebrate mitochondria, then for anticodons differing from the mitochondrial anticodon at the wobble position (corresponding to the third codon position) by a transition \( (U \leftrightarrow C \text{ or } A \leftrightarrow G) \). The only other anticodons found in *E. coli* had \( G \) at the wobble position, besides an exception with \( A \) indicated in *Table 3* for tRNA Arg UGC. Results are similar to those found for mitochondrial tRNAs: a majority of correlations between stabilities of codon–anticodon interactions and aminoacylation potentials are negative, meaning that relatively stable codon–anticodon mismatches have a relatively high probability to occur when the tRNA is misloaded with a non-cognate amino acid that matches the mismatched codon (only row-based analyses were done for *E. coli* tRNAs).

*Table 3* shows that there are 29 such negative rw’s, among 39 tRNAs (74%, \( P = 0.0008 \) according to a one sided sign test). *E. coli* tRNAs with the same anticodon as those in mitochondria have also similar rw’s: the correlation between the *E. coli* rw’s and the corresponding rw’s of the 22 mitochondrial tRNAs is positive (\( r = 0.56 \)).

In these analyses, only tfam’s output differs between analyses of *E. coli* and mitochondrial tRNAs (codon–anticodon interactions are considered as following canonical Crick–Watson interactions, and hence are identical [problems raised by this simplification, which does not account for wobble position modifications, some differing between mitochondria and *E. coli*, are discussed below]). This means that results are not due to tfam’s artefacts specific to mitochondrial tRNAs. Partial correlation analyses, controlling for evolutionary exchangeability between cognate and non-cognate amino acids, decreases the number of negative rw’s: 25 among 39 *E. coli* tRNAs (64%) yield negative rw’s, but the general tendency remains statistically significant (\( P = 0.027 \) according to a one sided sign test). Hence error compensation exists in *E. coli* tRNAs, also independently of minimization of misinsertion impact.

Though tfam’s output may be more trustable for bacterial tRNAs, estimates of stabilities of codon–anticodon mismatches remain problematic, as they do not account for effects of modifications at the wobble position, which is quite common in *E. coli*, but typically of a different nature than in mitochondria (Knight et al., 2001). A
3.13. Error compensation for codons

Analyses of error compensation focused until now on error compensation specific to anticodons and their tRNAs. But similarly to the analysis of tfam outputs that compare tRNA synthetases for the same tRNA (row analysis; see previous sections and Seligmann (2010a)), or compare tRNAs for the same tRNA synthetase (column analysis; see analyses and Fig. 6 in Seligmann (2010a)), error compensation can also be studied for specific codons, rather than anticodons. This means that rather than comparing matches and mismatches of a given anticodon with all possible codons, analyses now compare mismatching of a given codon with all possible (22) mitochondrial tRNA anticodons. This is then repeated for each of the 60 codons that code for amino acids in vertebrate mitochondria, and error compensation for each is calculated using the same tfam output from Table 2 in Seligmann (2010a). There is evidence for error compensation for a majority of codons, considering human mitochondrial tRNAs: the correlation between mismatch stability and misacylation was negative in 80% of the codons. Fig. 5 plots these codon-based levels of error compensation as a function of stable synonymous anticodon proportion.

Fig. 4. Error compensation, estimated by rw from Table 3, of E. coli tRNAs as a function of the relative proportion of tRNA genes with that anticodon for a given codon family. rw is the weighted correlation coefficient between codon–anticodon duplex stability and the aminoacylation potential of the tRNA with the amino acid matching the codon according to tfam. Negative rw’s promote error compensation, which avoids amino acid misinsertions by matching codon–anticodon mismatches and tRNA misacylations. rw is more negative in tRNAs with anticodon GGC exists, data indicated are for a tRNA with anticodon ACG.

predict the outcome of experiments designed to quantify misacylations.

4.2. Posttranscriptional modifications of anticodons

An important caveat of the analyses, beyond the difficulty presented by estimating aminoacylation tendencies based on sequence similarities, as done by tfam, is that the wobble position of anticodons sometimes undergoes modifications to 'non-canonical' taurine-derived nucleotides (i.e., mitochondrial tRNAs Lys and Leu UUR (Suzuki et al., 2002)). Additional wobble modifications apparently exist for 4 other mitochondrial tRNAs (Watanabe, 2007). This caveat is also problematic for the rest of the sequence, and would probably affect tfam's analyses if these could take such nucleotides into account, but this is not yet the case.

Some analyses in the sections on human pathogenic mutations consider the presence/absence of post-transcriptional modifications (presence in 6 among 22 mitochondrial tRNAs), and show that these interact with aminoacylation potentials detected by tfam, so that analyses, though not including information on such modifications, can be considered as qualitatively valid despite that they do not account for these modifications.

The problem remains for Mfold, the algorithms used to calculate duplex stabilities of RNA triplets, though only a subset of human mitochondrial tRNA wobble positions undergo such posttranscriptional modifications. In addition, even for the four regular nucleotides, it is not sure that the free energy values of formation of duplexes by RNA triplets are precise, especially that the stabilities are very low. For this and the former (modifications) reason, the free energy values used here can only be considered as rough estimates and probably include a large random component due to inaccuracy. However, these factors cannot account for general tendencies observed in the analyses. In addition, wobble position modifications are different for tRNAs from E. coli (the modifications usually use inosine- or uridine-based derivatives) (Knight et al., 2001; Björk et al., 1987), however, again, results are overall congruent for mitochondrial and E. coli tRNAs.

The reason for this congruence, despite that analyses of codon/anticodon mismatches do not account for wobble modifications, and that the presence and nature of these modifications even vary among anticodons and organisms, might be related to the following observation. Wobble modifications presumably increase the capacity of specific anticodons to read all or most synonymous codons of a codon family (Agris, 2008). However, even if one does not account for wobble modifications, the pattern of codon/anticodon mismatches is usually quite similar for pairs of anticodons differing only by a transition (U→C or G→A) at their wobble position. This is the case for the mitochondrial anticodons: the correlation between stabilities of codon/anticodon mismatches of the regular anticodon and the stabilities of mismatches for the anticodon after a transition at its wobble position is, on average across all 22 tRNA, high: mean \( r = 0.57 \). Most different (the correlation between mismatch stabilities of the anticodon pairs is least positive) are the two anticodons of Met, and those of Ile, third in the list is Lys (results not shown). Results of this analysis for Leu UUA are at the other extreme, mismatch patterns for its two anticodons are most similar. These observations are in line with the special status of the codon/anticodon family of Met as initiation codons and its versatility among alternative genetic codes with coding for Ile.

This discussion suggests that various analyses of results, while taking into account presence or absence of wobble modifications, do not show any clear pattern. Hence, together with the congruence of estimates of error compensation for mitochondrial and E. coli tRNAs, and between anticodons from the same codon family, it is probable that wobble modifications have little qualitative impact.
on the general conclusions presented here. This is also valid for inaccuracies in predictions of duplex formation stabilities by RNA triplets, independently of wobble modifications: errors are more likely to prevent detection of patterns than to create them.

This is not only true in relation to predictions of codon–anticodon interactions, but also for predictions of mismatches. Indeed, analyses taking to some extent into account modifications in the tRNA beyond the wobble position show that detection of error compensation decreases with numbers of posttranscriptional modifications in the tRNA. As error compensation is generally detected without accounting for modifications, conclusions are also qualitatively unaffected by modifications beyond the wobble position, but suggest that these modifications have a greater impact on (mis)acylation than on codon–anticodon mismatches, at least in relation to error compensation. Wobble position modifications reflect very ancient evolutionary events, at the root of the tree of life (Tong and Wong, 2004), but they may have relatively little impact for more recent events: variation analysed here to detect factors associated with error compensation is between closely related species (Figs. 2 and 3), or even variation between individuals of the same species (Table 2).

4.3. Genetic code optimization: minimizing misacylation impacts versus frequencies

The working hypothesis supported by the analyses presented here is that the structure of the translational machinery maximizes the probability of assignments of misacylations and codon–anticodon mismatches that result into transferring the non-cognate amino acid that is coded by the mismatched codon. This contrasts with previously described optimizations of the same machinery for minimizing impacts of mismatches and/or misacylations, which suggest that errors tend to be conservative (misinsertions usually involve amino acids with relatively similar physico-similar properties). The partial correlation analyses presented here suggest that minimization of misinsertion frequency by error compensation exists independently of minimization of impacts of misinsertions, at the level of row analyses of tfam’s output, but not so much at the level of column analyses of tfam’s output. This suggests that competition between tRNA synthetases (row analyses) minimizes misinsertion frequencies by assorting mismatches and misacylations, and that competition between tRNAs (column analyses) minimizes misinsertion impacts, but there is little evidence for minimizing misinsertion frequency independently of optimizing misinsertion impact at this level.

The above situation fits with the known properties of these molecules: tRNA synthetases are complex proteins which probably evolved high capacities to distinguish between different amino acids, hence similarity between amino acids is not a major accuracy-determining factor, and their evolution can optimize other factors, such as misinsertion frequency. The opposite is true for the relatively simple tRNAs, for which only general physico-chemical properties of anticodons and acceptor stems are matched with amino acid properties. Hence, at the level of comparisons between different tRNAs, similarities between amino acids are a major factor, and as a result, it is difficult to detect other, finer levels of regulation of aminoacylation accuracy.

It is also probable that different organisms use more or less error compensation versus minimization of misinsertion impact. I also calculated the correlation between PAM’s evolutionary exchange-compensation versus minimization of misinsertion impact. I also calculated the correlation between PAM’s evolutionary exchange-compensation versus minimization of misinsertion impact (misinsertion implies preferably similar amino acids). Comparisons between mitochondria from closely related species but located differently along the r–K gradient could test this. This would yield interesting results regarding overall ecological strategies of molecular processes, and would be a further test for the validity of the working hypothesis. This principle was also suggested for another translational property, that of greater protection by secondary structure formation of parts of mRNAs that code for the more essential parts of proteins (Seligmann and Pollock, 2003; Krishnan et al., 2004), and this more particularly for K- than r-selected organisms (Krishnan et al., 2008).

4.4. tRNA versatility at the origins of the genetic code

It is possible that some of the results presented here suggest that organisms with small genomes and hence a limited pool of genes templating for tRNAs preferentially use anticodons that are versatile, meaning transfer several amino acids. This is surely the case for mitochondria which possess only one tRNA per codon family: a single anticodon has to match all synonymous codons in that codon family. This anticodon versatility probably also increases the number of non-synonymous codons with which a specific anticodon can form (imperfect) codon–anticodon RNA duplexes, because anticodons used in mitochondria form more such interactions than synonymous anticodons altered only by a transition at the wobble position (see corresponding columns with N in Tables 2 and Table 3). This means that thanks to both minimization of misinsertion impacts, and to error compensation, a given tRNA can be used to transfer routinely more than one amino acid. Such versatility could have been particularly useful at the origin(s) of the translational machinery, where relatively inaccurate translation could have relied on very few tRNA genes to build relatively complex proteins. The results presented here show that this might still be relevant in modern organisms (to much more limited extents), and that selection is still affecting this property. Note that non-canonical wobble modifications are very ancient, but were supposedly absent in the earliest molecular systems (Tong and Wong, 2004). This strengthens the hypothesis that error compensation was probably most important in early cellular or subcellular systems, because the patterns described here suggest reasonably small effects of wobble modifications on error compensation.

4.5. Pathogenic versus unpunitive tRNA polymorphisms

It is interesting to note that no clear patterns arise from comparing cognate aminoacylation potentials of homologous pathogenic and unpunitive tRNAs. One would have expected at least a general trend towards lower correct amino acylation potential for pathogenic mutants, as mitochondrialopathies associated with mutant tRNAs seem frequently to affect aminoacylation (Levinger et al., 2004). However, clear patterns emerge when analysing rw, the estimate used to measure error compensation. This suggests that the factor relevant to whether a mutation is pathogenic is not so much the absolute potential for cognate aminoacylation, but more its relative value as compared to other amino acids. This might be the case, independently of matching codon–anticodon interactions, but analyses below, which do not integrate codon–anticodon mismatches, do not confirm this.

The z scores for cognate aminoacylation, quantify the potential for cognate loading in relation to the mean and the variation in loading potentials of the tRNA’s non-cognate amino acids. Hence they do not relate to absolute potentials for correct amino acylation, but
relative to non-cognates. The t-tests for differences between mean z scores of pathogenic and unpathogenic tRNAs yield patterns that are very similar to those observed with absolute aminoacylation potentials, but t statistics are systematically weaker than for ‘absolute’ cognate aminoacylation potentials (results not shown). Hence the stronger patterns observed for error compensation (t-rw) than misacylation (t-tfam) suggest that the aminoacylation property that is most relevant to whether tRNA mutations are pathogenic or not relates to the analysis of the distribution of aminoacylation potentials, specifically in relation to potentials for codon–anticodon mismatch. It is also worth noting that when partial correlation analysis controls for evolutionary exchangeability between the amino acid loaded to the tRNA and the amino acid matched according to the genetic code with the mismatched codon, partial rw’s of pathogenic tRNAs resemble regular rw’s, but again, statistical significances drop systematically. This suggests that pathogenic effects are due to disruptions of both minimization of misinsertion impact, and of its frequency by error compensation.

The systematic differences between mean rw’s of pathogenic and unpathogenic tRNAs are an important result also in relation to showing that results are not artefacts, and are not only a result of ancient constraints inherent to the genetic code and the translational machinery. This is because these analyses compare groups of homologous tRNA polymorphisms, where each polymorphism differs by a single point mutation. This shows that error compensation, and to a lesser extent, also minimization of misinsertion impacts, are still under natural selection. These properties are therefore worth considering when searching for pathogenic mechanisms associated with tRNAs. These principles also apply to analyses of associations between error compensation and developmental stabilities of phenotypes of morphological bilateral characters in lizards (Figs. 2 and 3).

4.6. Anticodon usage and avoidance of misinsertions: speed-accuracy trade-off?

As discussed in a previous section, analyses of rw in E. coli tRNAs confirm the general patterns observed for mitochondrial tRNAs, suggesting that patterns in mitochondrial tRNAs are not due to tfam inaccuracies specific to mitochondrial tRNAs. Figs. 4 and 5 suggest that, when considering specific codon family anticodons that are found on large proportions of the tRNA genes compensate less errors than rarer anticodons/codons. This might suggest that the balance between needs for aminoacylation specificities, minimization of misinsertion impacts and those of misinsertion frequencies differ between rare and common anticodons. This rationale also parallels the one used for r- versus K-selected organisms, but at the molecular level, rare anticodons matching K, and common on r-selected organisms. As with comparisons between tRNA synthetases and tRNAs, where the former seem to minimize more misinsertion frequency, it is possible that tRNAs bearing rare anticodons evolved more elaborate and accurate aminoacylation signals, while those bearing more common anticodons might maximize translational rate at the expense of accuracy.

4.7. Error compensation versus coordination: misacylation might change mismatch stability?

The analyses done do not assume that misacylation alters codon–anticodon interactions, but that these are the same for any tRNA molecule with the same anticodon. This may not be true: codon–anticodon interactions depend on electro-magnetic properties of nucleotides. These are mainly related to the specific nucleotides at the relevant positions, but may be affected by electro-magnetic properties of the rest of the tRNA molecule, including the loaded amino acid. Hence charge properties of amino acids might slightly alter the tRNA’s electro-magnetic properties, including the anticodon, and hence alter stabilities of canonical codon–anticodon interactions as well as non-canonical codon–anticodon mismatches. Misloaded amino acids might slightly increase stabilities of mismatches with codons that code for the mismatched amino acid. This venue has been termed error coordination (Seligmann, 2010c). In this respect, tRNA dipole moments might be a relevant property, as dipole moments affect evolutionary processes (Seligmann, 2006) and can be predicted in silico for complex molecules (Felder et al., 2007; Kantardjiev and Atanasov, 2009).

Hence it is possible that prevention of amino acid misinsertion occurs beyond the levels suggested here, because codon–anticodon mismatches may be conditioned by misacylation. Error compensation does not account for effects of misacylation on codon–anticodon interactions.

5. General conclusions

The relatively simple analyses described here answer in the affirmative the question whether codon–anticodon mismatches and misacylation potentials tend to favour mismatch and misacylation assortments that do not result in amino acid misinsertions in proteins, termed here as error compensation. Analyses suggest that such patterns exist more frequently than expected, but more quantitative estimations are missing. This requires converting stabilities of codon–anticodon duplexes into probabilities, but also to convert tfam’s aminoacylation scores into probabilities, which might be at least difficult. Such an endeavour would also answer the question of how many misinsertions are avoided, whether this is in the order of magnitude of one per life cycle of a mRNA, of a tRNA, or of a cellular replication cycle.

If some adequate data on misacylation probabilities from direct in vivo or in vitro experiments were available, this could help in this respect. Other limitations are inherent to the methods used and our current knowledge: analyses could not take into account effects of posttranscriptional modifications on tRNA misacylation tendencies, nor on codon–anticodon mismatch tendencies (and inaccuracies in predictions of mismatches), but, at least qualitatively, patterns seem robust to these factors, so that the conclusion that error compensation occurs is probably trustworthy. Despite various limitations discussed above and due to posttranscriptional modifications of tRNAs, results yield relatively clear conclusions in relation to several points:

1. the genetic code and the translational machinery minimize misinsertion frequencies by assorting codon–anticodon mismatches with misacylation by non-cognates matching mismatched codons according to the genetic code, a phenomenon termed error compensation;
2. this property exists independently of the minimization of misinsertion impacts by the same systems, which favours conservative errors over those misinserting very different amino acids;
3. error compensation is most apparent for competition among mitochondrial tRNAs, results yield relatively clear conclusions in relation to several points:
4. pathogenic mitochondrial tRNA mutations compensate less misacylation (t-tfam) suggest that the aminoacylation property potentials, but
References


