tRNomics: A comparative analysis of *Picrophilus torridus* with other archaeal thermoacidophiles

Bibekanand Mallick*, Jayprokas Chakrabarti†, Smarajit Das and Zhumur Ghosh

Computational Biology Group, Theory Department, Indian Association for the Cultivation of Science, Kolkata 700 032, India

Received 8 December 2004; revised 4 July 2005

In the euryarchaeal thermoacidophile *Picrophilus torridus* DSM 9790, we identified a copy of rare tRNA\(^{Ile}\)(TAT) gene, along with the other 47 tDNAs with the help of our in-house program. Further, tRNAs of *P. torridus* were also compared with other archaeal thermoacidophiles *Thermoplasma acidophilum*, *T. volcanium*, *Sulfolobus solfataricus* and *S. tokodaii*.

**Keywords:** tRNomics, *Picrophilus torridus*, overlapping tRNA, base pairs, intron, Archaea, thermoacidophile

In our recent paper, we analyzed cytoplasmic tRNA genes of 22 species of 12 orders of three phyla of archaea\(^1\). This was to identify the probable identity elements for recognition of cognate amino acids. During this investigation, we found some tDNAs missing or misidentified in archaea. Archaeal tRNA genes have canonical introns (CI) (i.e., between 37 and 38) and non-canonical introns (NCI) located elsewhere. These unusually located introns in archaeal tDNAs were observed in 1987\(^2\). The introns at non-canonical sites vary in size, number and location\(^3\). These non-canonical introns make detection of tRNA genes in archaea tricky. Although several computational approaches\(^4,5\) have been developed since then to annotate tRNA genes from a given genome, but non-canonical intron strategy has been neglected.

In the present work, we revisited and extended our earlier observations, concerning the identification of the missing tDNAs in the euryarchaeal thermoacidophile *Picrophilus torridus*\(^6\). We, therefore, developed a computational approach to search tRNA genes by incorporating introns at positions, other than canonical and identified the non-annotated rare isoleucine tRNA(TAT)-gene in *P. torridus*. Interestingly, this new tRNA gene overlaps with tRNA\(^{Trp}\)(CCA) gene. Restructuring of tDNAs is based upon intron splicing at alternate position\(^7\). Previously, in mitochondrial genomes of many animal species, evidence of overlapping tRNA genes by one to six nucleotides with downstream genes on the same strand was found. tRNA\(^{Tyr}\) and tRNA\(^{Cys}\) genes in human mitochondrial genome overlaps with one another by one nucleotide which is the first base of tRNA\(^{Cys}\) and discriminator base of tRNA\(^{Tyr}\)\(^8\). But, in tDNA\(^{Ile}\)(TAT) of *P. torridus*, the domain of overlap is far wider, encompassing the entire tDNAs which is a novel finding. Further, we compared the tRNomics\(^3\) of this species with four other archaeal thermoacidophiles viz., *T. acidophilum*\(^9\), *T. volcanium*\(^10\), *S. solfataricus*\(^11\) and *S. tokodaii*\(^12\), whose complete sequences are known.

The single-stranded primary tRNA nucleotide chain folds back onto itself to form the cloverleaf secondary structure. This secondary cloverleaf structure of a tRNA has: (i) acceptor or A arm, in which the 5’ and 3’ ends of tRNA are base-paired into a stem of 7 bp; (ii) DHU or D-arm: Structurally, a stem-loop, it frequently contains the modified base dihydrouracil; (iii) anticodon or AC arm, made of a stem and a loop containing the anticodon. At 5’ end of this anticodon-loop is a pyrimidine base at 32, followed by an invariant U at 33. The anticodon triplet is located at 34, 35, and 36 in the exposed loop region; (iv) an extra arm or V arm, which is not always present. It is of variable length, and is largely responsible for the variation in length of tRNAs.
tRNA classification into types I and II depends on length of V-arm; (v) T-Ψ-C arm or T arm: It has the conserved sequence of three ribonucleotides: ribothymidine, pseudouridine and cytosine and has stem-loop structure; and (vi) tRNA terminates with CCA at 3’ end. tDNA may or may not have CCA; If absent it is added during tRNA maturation. At the corner where D- and T-loops meet (DT-region), there are several unique base pairs (bps): U54:A58; Ψ55-mediated U-turn in the T-loop; the inter-loop G18:Ψ55 and G19:C56, and the stack of four mutually intercalated purine bases A58-G18-R57-G19 (R represents a purine). These positions are largely conserved indicating their importance in tRNA functions. We have studied these positions in all tRNAs of five euryarchaeal thermoacidophiles, and looked for minor variations over the conserved bases and base pairs (bps).

**Methodology**

As of now several powerful routines\(^4,5\) accessible for identification of tDNAs or tRNA genes from genomes. Our experience\(^1,7\) has shown that it is necessary to be careful when these routines are applied to archaeal genomes, due to the presence of introns at non-canonical sites. In a recent work dealing with non-canonical introns in tRNA genes archaea\(^1\), we discussed this issue at length. We, therefore, developed a computational approach to search tRNA genes by incorporating introns at positions other than canonical and identified some non-annotated tRNA-genes in archaea\(^1,7\). About 1000 thousand tRNA-genes from archaea were studied for this purpose.

From this database of 1000 tRNA-genes, we fine-tuned the strategy to locate non-canonical introns and some of the non-annotated tRNA genes. The salient features were: (i) sequences were considered that give rise to the regular cloverleaf secondary structure; (ii) conserved elements: T8 (except Y8 in *Methanopyrus kandleri*), G18, R19, R53, T44, Y55, and A58 were considered as conserved bases for all archaeal tRNA. Further, there were tRNA-specific conserved or identity elements\(^1\) of archaea; (iii) the constraints of lengths of stems of regular tRNA A-, D-, AC- and T-arms were 7, 4, 5 and 5 bp, respectively. In few cases, the constraints on lengths of D- and AC-arms were relaxed. (iv) base positions optionally occupied in D-loop were 17, 17a, 20a and 20b; (v) the V-arm was considered for type II tRNAs. The constraint on length of V-arm: less than 21 bases (iv) Non-canonical introns were considered at any position in tRNA-genes. The introns constrained to harbour the minimal Bulge-Helix-Bulge (BHB) secondary structure for splicing-out during tRNA maturation. The minimum length of introns allowed was 6 bases; (v) archaeal box A promoter\(^13\) sequence 25 bases upstream of 5’ end tRNA genes was identified. With these features in our algorithm, we re-investigated and extracted tRNA genes in archaeal thermoacidophiles: *P. torridus* (NC_005877), *T. acidophilum* (NC_002578), *T. volcanium* (NC_002689), *S. solfataricus* (NC_002754) and *S. tokodaii* (NC_003106).

**Results and Discussion**

The general features of *P. torridus*, *T. acidophilum*, *T. volcanium*, *S. solfataricus* and *S. tokodaii* are summarized in Table 1. Words about notation: (i)

<table>
<thead>
<tr>
<th>General features</th>
<th><em>P. torridus</em>(^b)</th>
<th><em>T. acidophilum</em>(^b)</th>
<th><em>T. volcanium</em>(^10)</th>
<th><em>S. solfataricus</em>(^11)</th>
<th><em>S. tokodaii</em>(^12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habitat</td>
<td>Dry solfataric field</td>
<td>Coal refuse piles, solfataric fields</td>
<td>Coal refuse piles, solfataric fields</td>
<td>Volcanic hot springs</td>
<td>Volcanic hot springs</td>
</tr>
<tr>
<td>Morphology</td>
<td>Cell wall present</td>
<td>No rigid cell wall</td>
<td>No rigid cell wall</td>
<td>No rigid cell wall</td>
<td>No rigid cell wall</td>
</tr>
<tr>
<td>Temp. range (°C)</td>
<td>36-65</td>
<td>36-67</td>
<td>36-67</td>
<td>75-80</td>
<td>75-80</td>
</tr>
<tr>
<td>Optimum temp (°C)</td>
<td>60</td>
<td>59</td>
<td>60</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>logH (opt.)</td>
<td>0.7</td>
<td>2</td>
<td>2</td>
<td>2-4</td>
<td>2-3</td>
</tr>
<tr>
<td>O₂ requirement</td>
<td>Aerobe</td>
<td>Anaerobe</td>
<td>Anaerobe</td>
<td>Aerobe</td>
<td>Aerobe</td>
</tr>
<tr>
<td>Chromosome size (bp)</td>
<td>1,545,900</td>
<td>1,564,905</td>
<td>1,584,799</td>
<td>29922245</td>
<td>2694756</td>
</tr>
<tr>
<td>G+C Content</td>
<td>36%</td>
<td>46%</td>
<td>39.9%</td>
<td>35.8%</td>
<td>32%</td>
</tr>
<tr>
<td>No of genes</td>
<td>1535</td>
<td>1478</td>
<td>1499</td>
<td>2997</td>
<td>2826</td>
</tr>
<tr>
<td>tRNA</td>
<td>48</td>
<td>46</td>
<td>46</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>tDNAs with introns</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>18</td>
<td>22</td>
</tr>
</tbody>
</table>

\(^a\) Words about notation: (i)
tRNA^{Trp} (CCA), for example represents tryptophan tRNA with anticodon CCA; (ii) G30:C40 represents the bp between G at position 30 with C at position 40. Semi-conserved bases are ones that are almost conserved with, but a few exceptions. 3D bps represents three-dimensional base-pairings. These are essential for tertiary structure.

**Novel tRNA gene of *P. torridus***

By our computational approach we have identified 48 tRNA genes in *P. torridus*, out of which 47 was annotated by NCBI previously. Our identified tRNA set includes a newly annotated tRNA^{Ile} (TAT) gene. This tRNA gene ranges between 276477 and 276618 and has two introns, canonical of 12 bases and non-canonical between 32 and 33 of 57 bases long (secondary structure is shown in Fig. 1). The important 3D base pairs i.e. T8:A14, G18:U55, 19G:C56 and U54:A58 of archaeal tRNA^{Ile} gene are consistently found in this tRNA as well. In addition, this tRNA gene has A73, A35 and T36. These are unique to tRNA^{Ile} and key elements recognized by corresponding tRNA-synthetase for proper aminoacylation\(^1\). The introns too have proper Bulge-Helix-Bulge (BHB) motif for splicing-out during tRNA maturation process. Interestingly, this tDNA is overlapped with tRNA^{Trp}(CCA) gene. The latter actually located between bases 276477 and 276618 was mis-annotated in NCBI (lying between 276483 and 276608). Considered as tRNA^{Trp}(CCA) it has a canonical intron of length 69 bases. This tRNA also has all the key aminoacylating elements of archaeal tRNA^{Trp}, along with the correct BHB motif for intron splicing. Thus, the scenario of intron-splicing at alternative positions give rise to two tRNA genes: one choice gives tRNA^{Ile}(TAT), the other tRNA^{Trp}(CCA).

**tRNomics of thermoacidophiles**

Comparative analysis of conserved bases and bps of different thermoacidophiles is given in Table 2. We here present a few significant observations from this analysis. Despite the high G+C content\(^14\) in A-arm of archaeal tRNAs, the lack of G/C7: C/G66 at the bottom of A-arm in some of the tRNAs of thermoacidophiles, especially euryarchaeal thermoacidophiles are unusual and remarkable. Instead of either GC or CG at 7:66, A: U is present in tRNA^{Arg}(CCU), tRNA^{Ser}(GGA,GCU) and tRNA^{Trp}(CCA) of *T. acidophilum* and *T. volcanium*. tRNA^{Ala} (GGA,UGC), tRNA^{Asn} (GUU), tRNA^{Leu} (CAG,UAA), tRNA^{eMet}, tRNA^{Phe} (GAA), tRNA^{Ser} (CGA,GGA,GCU) and tRNA^{Trp} (CCA) too have A7:U66 in *P. torridus*.

Interestingly, we did not observe either A:U or U:A base pairing at 7-66 of A-arm in any tRNAs of crenarchael thermoacidophiles i.e., *S. solfataricus* and *S. tokodaii*. In tRNAs, 33\(^{rd}\) base is normally U, exceptions being C33 in tRNA^{Asn}(CGA) of *P. torridus* and tRNA^{Cy}(GCA) of *S. solfataricus*. Again, A33 is
also present in tRNA^Pro\textsuperscript{Pro}(CGG) of S. solfataricus. U33 is a characteristic signature in tRNAs located just ahead of the anticodon, which occupies 34, 35 and 36 positions. It helps to expose the anticodon to the solvent and facilitate the codon-anticodon interactions\textsuperscript{15}. These exceptions at 33\textsuperscript{rd} base in above-mentioned tRNAs may pose obstacle in codon-anticodon interactions. The bp 54:58, a 3D bp helps to keep the tertiary L-shape of tRNAs, along with other conserved 3D bps, namely 8-14, 15-48, 18-55, 19-56. The 54-58 is universally conserved with U:A in most of the archaeal tRNAs, but tRNA\textsuperscript{Thr} (CGU,GGU,UGU) of P. torridus and tRNA\textsuperscript{Thr} (CGU,UGU) of T. volcanium have deviations at this position. These tRNAs have A:A at 54:58 instead of conserved reverse hoogsten U:A bp. This peculiarity at 54:58 might have some impact on its tertiary structure formation and needs to be investigated further.

### Conclusion

Non-canonical introns in archaeal tRNAs at various locations make the detection of tRNA genes difficult. By our NCI-containing tRNA gene identification program, we were able to identify the missed-out tRNA genes in some of the archaeal species\textsuperscript{1,7}. While

<table>
<thead>
<tr>
<th>Base pairs</th>
<th>S. solfataricus</th>
<th>S. tokodaii</th>
<th>P. torridus</th>
<th>T. acidophilum</th>
<th>T. volcanium</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-72</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
</tr>
<tr>
<td>2-71</td>
<td>GC/Cg</td>
<td>GC/Cg</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>3-70</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>4-69</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>5-68</td>
<td>--</td>
<td>CG/Cg</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>6-67</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>7-66</td>
<td>GC</td>
<td>GC</td>
<td>--</td>
<td>--</td>
<td>GC</td>
</tr>
<tr>
<td>8-14</td>
<td>UA</td>
<td>UA</td>
<td>UA</td>
<td>UA</td>
<td>UA</td>
</tr>
<tr>
<td>10-25</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>11-24</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>12-23</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>13-22</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>27-43</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>28-42</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>29-41</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>30-40</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
</tr>
<tr>
<td>31-39</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>49-65</td>
<td>--</td>
<td>CG</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>50-64</td>
<td>CG</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>51-63</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
</tr>
<tr>
<td>52-62</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
</tr>
<tr>
<td>53-61</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
<td>GC</td>
</tr>
<tr>
<td>33</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
</tbody>
</table>

The bases/bp labelled in normal font denote conserved; fonts labelled in bold denote semi-conserved bases/bp. The bps in italic denote that these are equally present at the corresponding base positions.
doing further analysis, we identified in *P. torridus*, a copy of tRNA\textsuperscript{Ile}(TAT) gene harbouring one canonical and one non-canonical intron. This tRNA\textsuperscript{Ile}(TAT) gene is overlapped with tRNA\textsuperscript{Trp}(CCA) gene. Furthermore, the comparative analyses of the four archaeal thermoacidophiles with *P. torridus* revealed the conserved, semi-conserved bases and base pairs, and of course, some peculiarity in them which may have important implications in codon-anticodon interaction or maintenance of tertiary structure of tRNAs. A further investigation on the effects of these bp deviations, especially at 33\textsuperscript{rd} and 54:58 on the corresponding tRNA’s tertiary structure is in progress in our laboratory.

References