The origin of the tRNA molecule

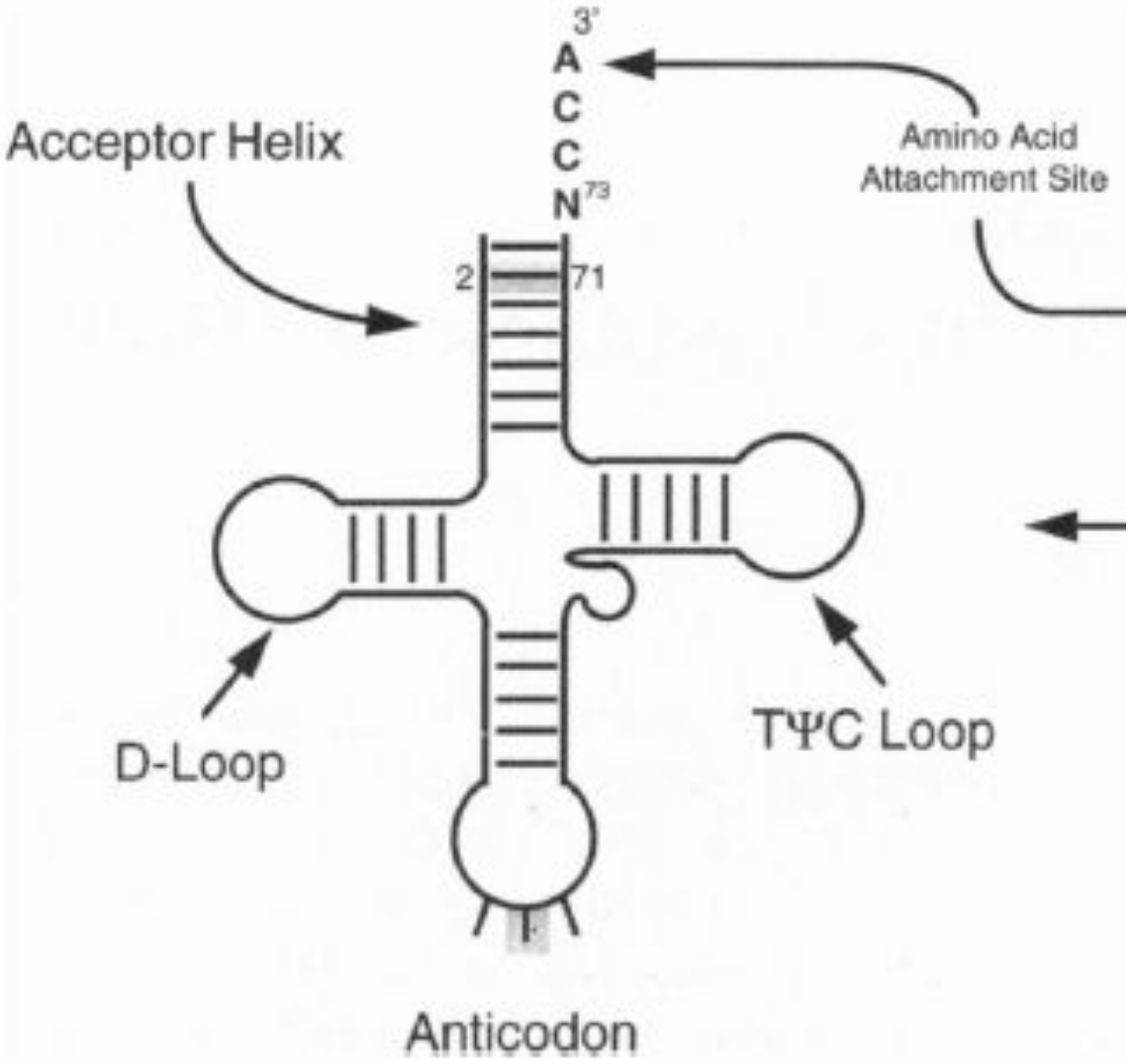
Massimo Di Giulio

Institute of Genetics and Biophysics ‘Adriano Buzzati-Traverso’, Naples, Italy

What is Evolution? Bicentennial of Charles Darwin's BirthDate

October 15-18, 2009

Kyoto, Japan


3'
A
C
C
N73
Acceptor Helix
Amino Acid
Attachment Site
2
71
D-Loop
TΨC Loop
Anticodon

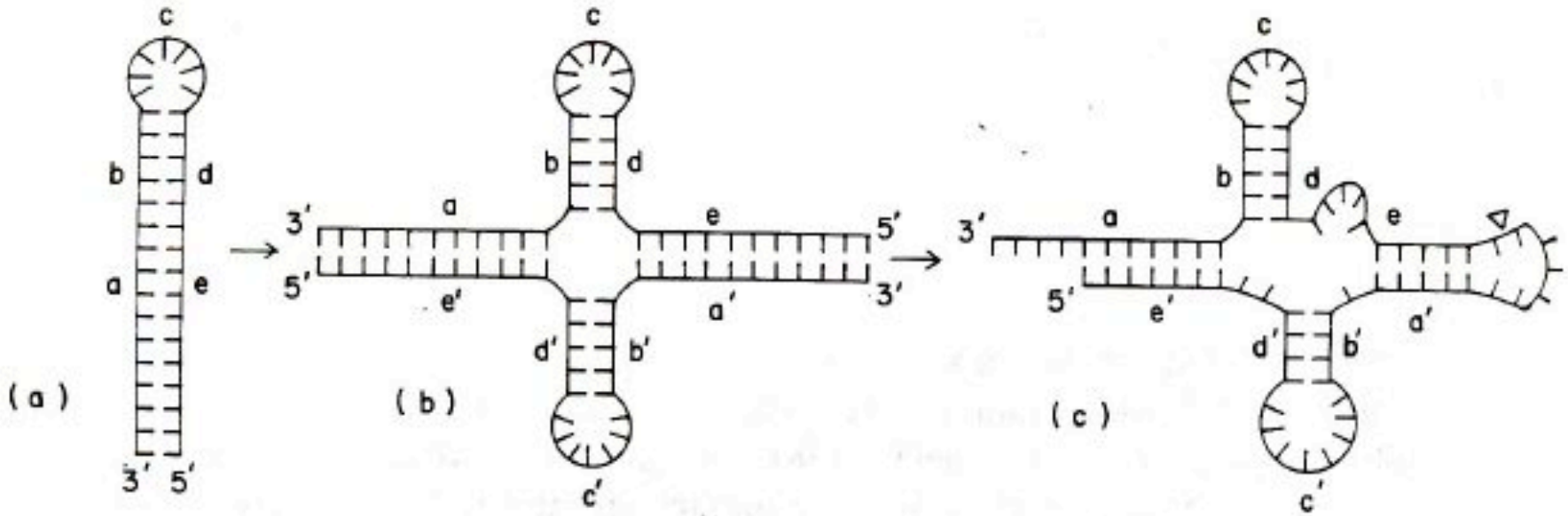


FIG. 1. Model for the origin of the tRNA molecule. In (a) a hairpin structure is represented. The dashes indicate the bases characterizing both the loop (c) and the stem defined by the complementary regions a–e and b–d. (b) Shows a cloverleaf structure as a dimer of the hairpin molecule and as a precursor of the molecule of the present-day tRNAs. This cloverleaf structure must also be considered as covalently closed at one of the free ends. The formation of the anticodon loop and, therefore, the closure of the molecule and its transition towards the final structure (c) should probably be placed in relation with the presence of introns (triangles) in this position of the tRNA genes (see Discussion). The figure seems to suggest that in the transition towards the final molecule, the 5′-half of the tRNA underwent a shift towards the anticodon region, thus leading to the formation of the variable loop. In (c) the secondary structure of the present-day tRNAs is shown. The various regions are only approximately given in scale. The letters indicate the following regions moving from 5′ to 3′ of the molecule: e′–a = aminoacyl-stem, d′–b′ = D-stem, c′ = D-loop, a′–e = anticodon-stem, d–b = Tψ-stem, c = Tψ-loop. For discussion and further information see text.

J. theor. Biol. (1992) **159**, 199–214

On the Origin of the Transfer RNA Molecule

MASSIMO DI GIULIO

†*International Institute of Genetics and Biophysics, CNR, Via G. Marconi* 10, 80125 *Naples, Italy*

(*Received on* 10 *October* 1991, *Accepted in revised form on* 23 *March* 1992)

Data and arguments are given in favour of the hypothesis that the primitive tRNA molecule may have originated from a direct duplication event involving one of the two halves of the tRNA molecule. It seems that a molecule capable of assuming a hairpin structure was involved as a precursor in this duplication. The two halves of the present tRNAs could, therefore, be considered as paralogous.

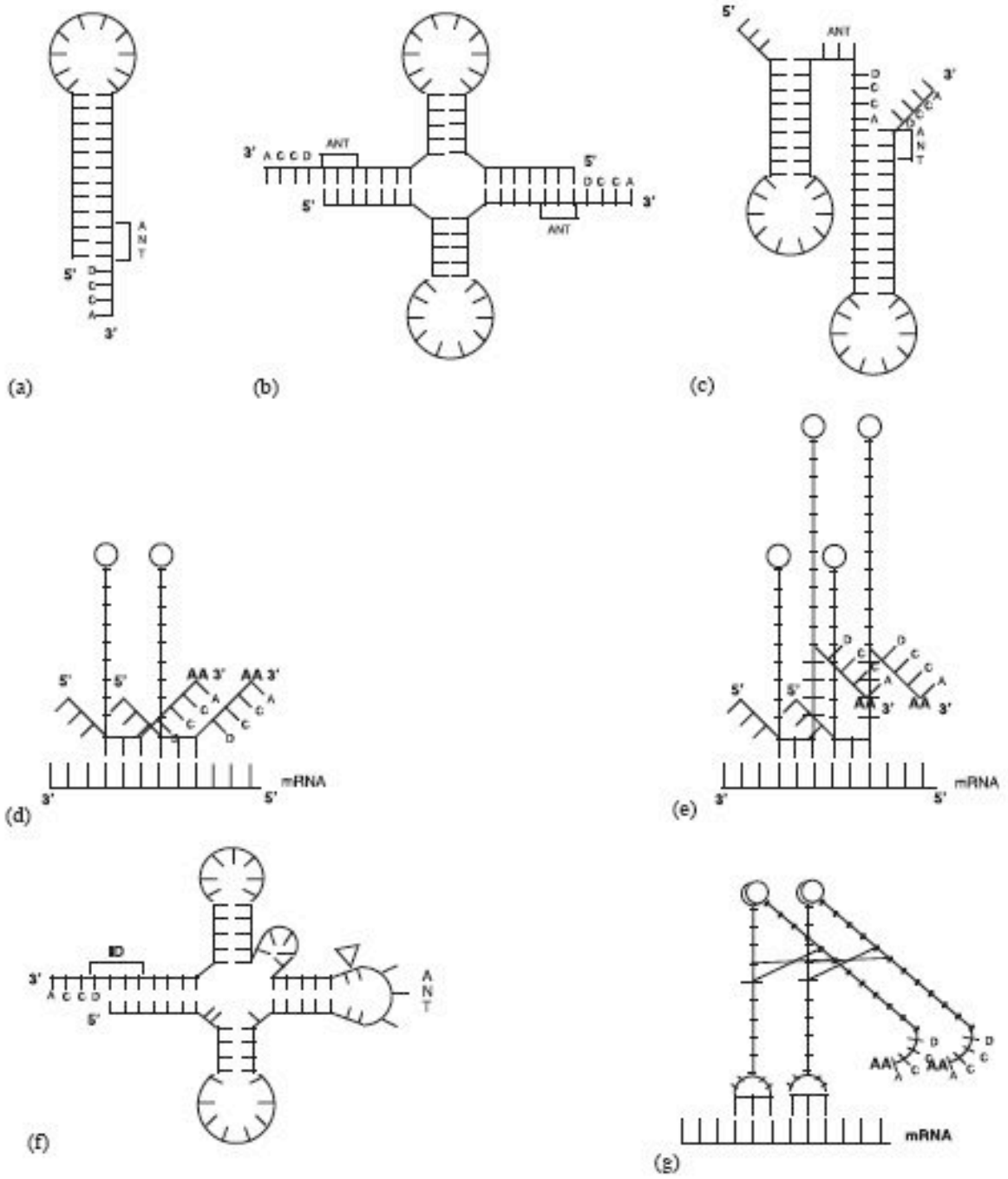
1. Introduction and Hypothesis

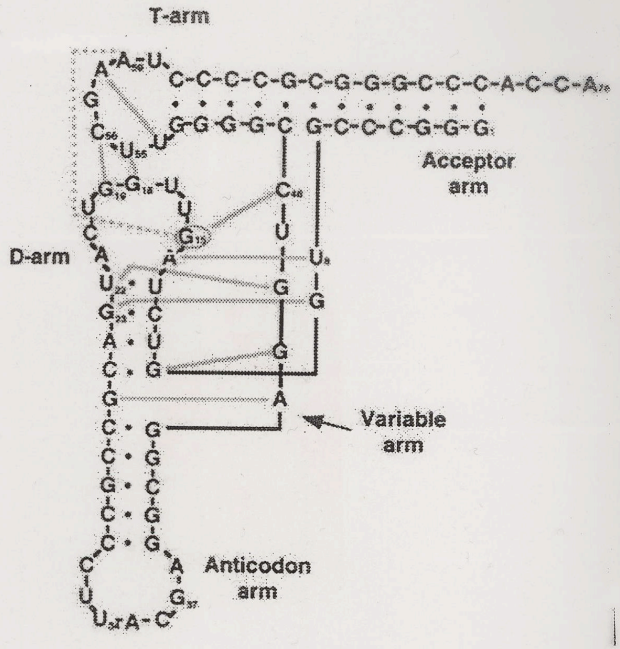
The transfer RNA (tRNA) molecule is one of the most important molecules ever to have appeared on this planet because it has the extraordinary ability of relating two languages: the language of nucleic acids, with their great versatility to duplicate themselves and hence act as the depositary of genetic information, and the stereochemical language of proteins, with as great an ability to be highly efficient catalysts. An understanding of how this molecule originated may, therefore, help clarify how life developed on our planet and, in particular, provide information on how the genetic code evolved.

There are various forms of ancestral tRNA that have been proposed as precursors of the actual molecule. Some hypotheses consider the primitive tRNA as being formed of a mononucleotide (Lacey *et al.*, 1975). Dillon (1978) suggests that the primordial tRNA was formed only of the CCA terminus. Crick *et al.* (1976) propose a primitive tRNA form of an anticodon loop comprising seven bases, five of which must have interacted with mRNA, even if the reading module was always the triplet.

structures (Figs. 1b and c). In particular, direct duplication of the gene coding the hairpin structure

(Fig. 1c) (Tanaka and Kikuchi, 2001). Simple evolutionary redistribution of the pairings of the four regions




T-arm
Acceptor arm
D-arm
Variable arm
Anticodon arm

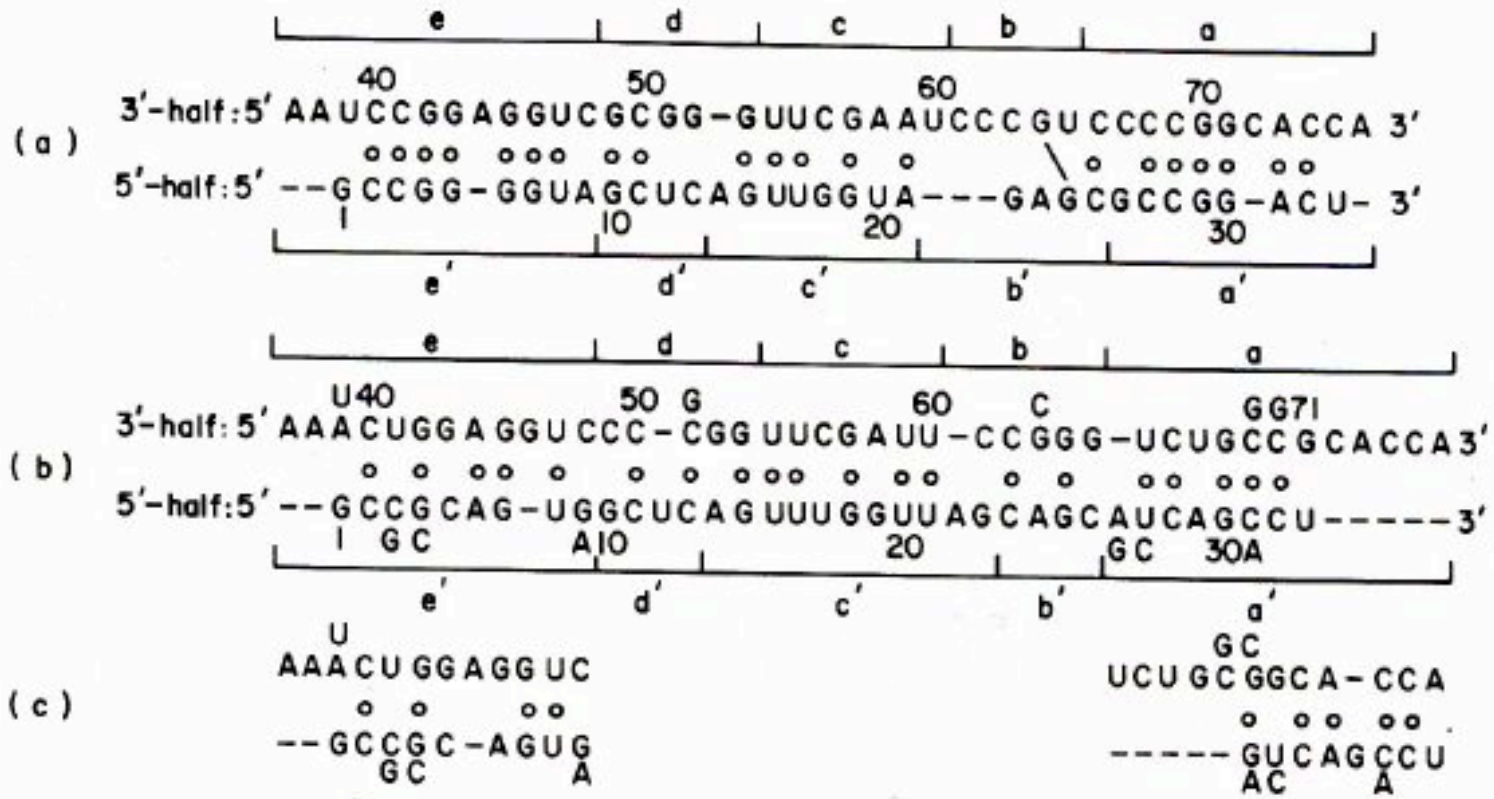


FIG. 2. The self-alignments of the ancestral tRNA sequences are shown considering the anticodon as a region dividing the molecule in two halves: the 3'-half and the 5'-half. (a) Shows the self-alignment of the sequence of Eigen & Winkler-Oswatitsch (1981*a*: 228, fig. 12). The oblique line identifies the nucleotides 64 and 24 which might be homologous (see text). (b) Shows the self-alignment of the sequence of Fitch & Upper (1987: 765, fig. 6). (c) Gives other self-alignment patterns of the sequence shown in (b) under the corresponding regions. The circles represent the matches identified between the two halves of the ancestral tRNA molecule. The gaps are indicated by a dash. The letters identify the regions visualized in Fig. 1(c) and the boundary between these regions is defined according to the secondary tRNA structure (Sprinzl *et al.*, 1989: r^2, fig. 1). The numbers identify the nucleotide positions in the secondary tRNA structure according to the official nomenclature [see e.g. Sprinzl *et al.* (1989)]

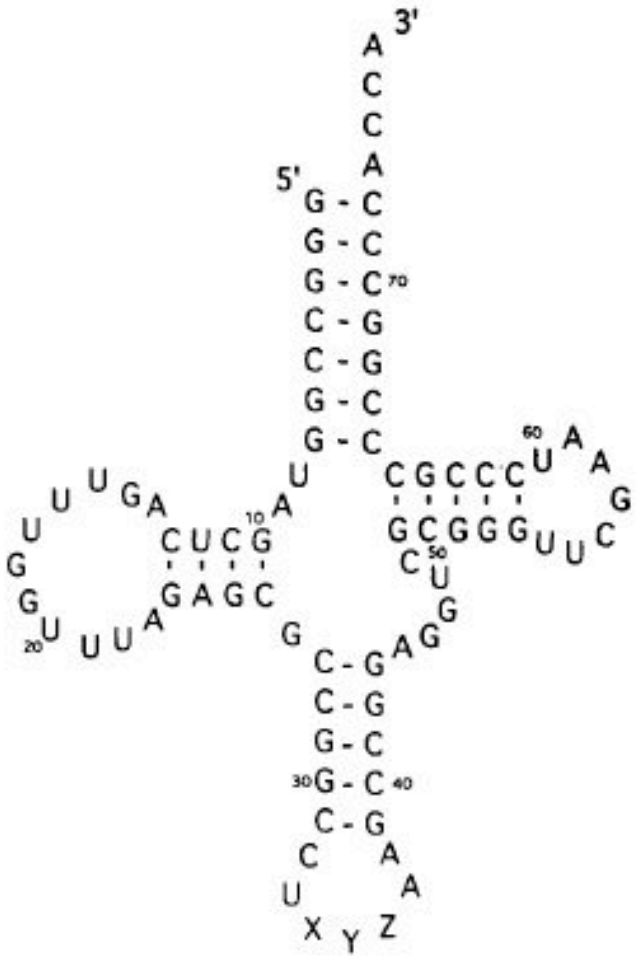


FIG. 2. Secondary structure of the ancestral sequence of tRNA obtained as a simple consensus from the sequences in Fig. 1. The numbers refer to nucleotide positions in agreement with the official nomenclature (see for example Sprinzl *et al.*, 1989); the anticodon is indicated with XYZ. See text for further information.

synthesis (see Introduction) and, in particular, this structure has also been suggested as the precursor of the tRNA molecule (Hopfield, 1978; Eigen & Winkler-Oswatitsch, 1981*a*; Bloch *et al.*, 1985; Di Giulio, 1992; Maizels & Weiner, 1993, 1994; Schimmel *et al.*, 1993). It can thus be expected that at least one of the two halves of the ancestral sequence (Fig. 2) is capable of reconstructing a hairpin structure. Figure 4 shows the hairpin structure reconstructed using the 5′-half of the ancestral sequence (Fig. 2). While the hydrogen bonds between the bases indicated by a dash (Fig. 4) are the ones imposed by the secondary tRNA structure (Fig. 2), the hydrogen bonds indicated by a dot are the ones actually reconstructed (Fig. 4). Of these, six out of a total of eight possible pairings are observed (Fig. 4). The probability of observing at least six pairings out of eight possible ones, calculated using binomial distribution is 0.0042 $[=\Sigma_k^n \binom{n}{k} \pi^k (1-\pi)^{n-k},\ \pi = 0.25,\ n = 8,\ k = 6,\ 7,\ 8]$. [This probability does not vary if the expected value of a random pairing (0.25) is substituted by the value calculated using the base composition of the whole hairpin structure (Fig. 4). If we consider G-U pairing to be possible, which can be observed in several actual tRNAs (Sprinzl *et al.*, 1989), then the hairpin structure pairings (Fig. 4) are in practice 100% reconstructed. These observations (Fig. 4) are highly suggestive as they seem to imply that a hairpin structure was effectively the precursor of the tRNA molecule (Hopfield, 1978; Di Giulio, 1992).

The alignment between the two halves of tRNA

a

```
          anticodon variab. TψC  TψC   TψC aminoacyl
               stem  loop stem  loop  stem  stem
              aaaaaaaeeeeessssslllllllssssmmmmmmmmmmmm
                40         50         60         70
3'-half: 5' AAGCCGGAGGUCGCGGGUUCGAAUCCCGCCCCGGCCCACCA 3'
              :::::  : : : : :: :  :   ::   :::  ::
5'-half: 5' GGGCCGGUAG-CUCAGUUUGGUUUAGAGC---GCCGGCCU 3'
                   10           20            30
              mmmmmmmmms sssllllllllllssss   aaaaaaaa
              aminoacyl  D       D       D    anticodon
               stem    stem    loop    stem    stem
```

b

```
                B
                K M  W                          SGSS
3'-half: 5' AAGCCGGUGGUCGCGGGUUCGAAUCCCGCCCCGGCCACCA 3'
              : :::: : : : : :: :  :   :: ::::::
5'-half: 5' GGGGCGGUAG-CUCAGUUUGGUUUAGAGCGCCGGCCU--- 3'
                S S                            K M
                  B
```

FIG. 3. This shows two self-alignments between the 3'- and 5'-halves of different ancestral sequences of tRNA. The various regions of the tRNA molecule are shown in a self-explanatory way. The numbers refer to nucleotide positions in agreement with the official nomenclature (see for example Sprinzl *et al.*, 1989). The colon indicates a match between the bases of the two halves of the molecule. (b) Also shows, above the corresponding base, the positions presenting more than one base, specifying these using the standard abbreviations. See text for further information.

(iii) it would be more difficult to evaluate the statistical significance because of the non-independence of some of the matches observed, and (iv) because Fig. 3b shows a very similar alignment of the ends in question, albeit shifted by one base.

Discussion

statistical significance at least as far as the particular problem tackled here is concerned. Therefore, the empirical data are, on the whole, in favour of the previously presented model (Di Giulio, 1992) as they cannot prove it to be false.

The model of the origin of the tRNA molecule (Di Giulio, 1992) also seems to have a good explanatory power as it can rationalize a great deal of

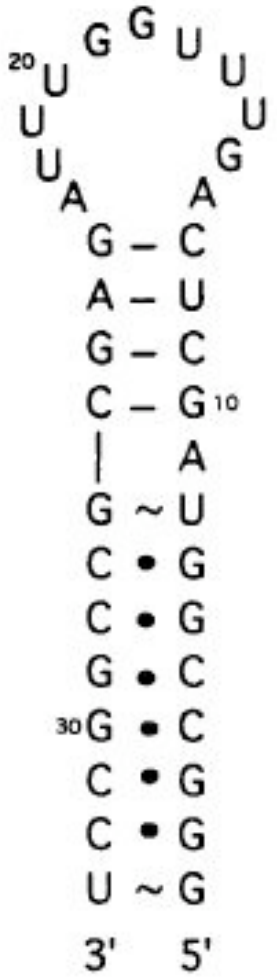


FIG. 4. This shows the hairpin structure reconstructed using the 5'-half of the ancestral tRNA molecule (Fig. 2). The standard hydrogen bonds between the bases are symbolized in different ways in order to highlight their different meanings in terms of the reconstruction of this structure (see text). The possible G-U pairings are also symbolized. The numbers refer to nucleotide positions of the tRNA molecule in agreement with the official nomenclature (see for example Sprinzl *et al.*, 1989).

J Mol Evol (2005) 61:524–530
DOI: 10.1007/s00239-004-0315-1

JOURNAL OF MOLECULAR EVOLUTION

tRNA Creation by Hairpin Duplication

Jeremy Widmann,[1] Massimo Di Giulio,[2] Michael Yarus,[3] Rob Knight[1]

[1] Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[2] International Institute of Genetics and Biophysics, CNR, Via G. Marconi 10, Naples 80125, Italy
[3] Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

Received: 4 November 2004 / Accepted: 3 May 2005 [*Reviewing Editor:* Dr. Niles Lehman]

Abstract. Many studies have suggested that the modern cloverleaf structure of tRNA may have arisen through duplication of a primordial hairpin, but the timing of this duplication event has been unclear. Here we measure the level of sequence identity between the two halves of each of a large sample of tRNAs and compare this level to that of chimeric tRNAs constructed either within or between groups defined by phylogeny and/or specificity. We find that actual tRNAs have significantly more matches between the two halves than do random sequences that can form the tRNA structure, but there is no difference in the average level of matching between the two halves of an individual tRNA and the average level of matching between the two halves of the chimeric tRNAs in any of the sets we constructed.

tRNAs depends only on the acceptor stem for certain amino acids (Schimmel and Henderson 1994), and aminoacyl tRNA synthetases can even charge minihelices that resemble only one half of the tRNA molecule (Tamura and Schimmel 2001). These charged minihelix structures have been shown to function in peptide synthesis and may have been part of the primordial protein synthesis machinery (Dick and Shamel 1995). It has also been suggested that the top half of modern tRNAs has an ancient origin in replication and is recognized separately by RNaseP, the CCA-adding enzyme, telomerase, and aminoacyl-tRNA synthetases (Weiner and Maizels 1987; Maizels and Weiner 1994). The 3′ half of modern tRNAs has been proposed to be older than the 5′ half due to its base composition and repetitive sequence

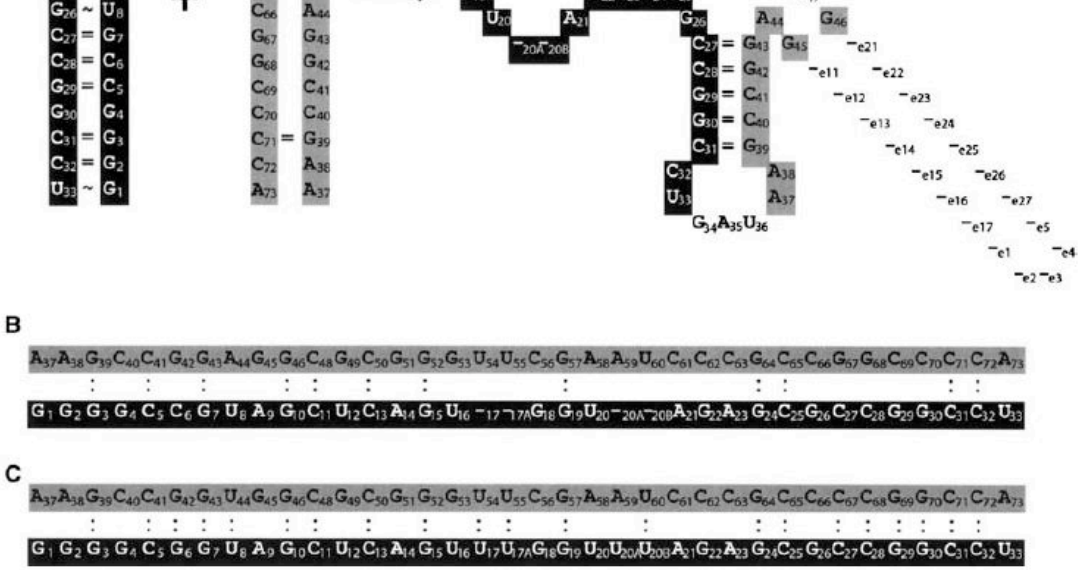


Fig. 2. Matches between the two halves of the modern cloverleaf structure, possibly produced by hairpin duplication. a Fusion of two hairpins to form the modern cloverleaf. Bases are numbered as in the Sprinzl database. The most frequent base is shown at each position. b Matches between the two halves of the consensus sequence from the Sprinzl database. c Matches between the two halves of the reconstructed ancestral tRNA sequence (Di Giulio 1995).

database. For each tRNA sequence in the database, we made one list containing each unpaired base in that tRNA and a second list containing each base pair. Each list was shuffled to randomize the order. This shuffling used the Yates–Fisher algorithm and the Mersenne Twister random number generator as implemented in the Python 2.3 package. We reconstructed the sequence from these two lists so that the structure and base composition were the same as the original tRNA, although the sequence was randomized.

Comparing Matches for Real and Random tRNAs

For each tRNA sequence in the Sprinzl database, we counted the number of times that the corresponding positions in the two halves of the tRNA (as defined in Fig. 2) matched. We repeated this procedure for the set of randomized tRNAs.

Comparing Matches for Real and Chimeric tRNAs

We organized the tRNA sequences in the Sprinzl database into (overlapping) groups as follows: all tRNAs regardless of domain and specificity, all tRNAs in the same domain, all tRNAs with same amino acid specificity, and all tRNAs with the same domain and same specificity. Within each group, we joined the first half of each tRNA to the second half of another, randomly chosen, tRNA. We then counted the number of matches between the first and the second halves of the new, chimeric tRNAs. We compared the distribution of matches from each group of these chimeric tRNAs to that of the actual tRNAs, as identified above.

Results

Real tRNAs have significantly more matches between their two halves than do random tRNAs (Fig. 3) ($t = 15.8$, df $= 11898$, $p = 4.67 \times 10^{-55}$, paired two-sample t-test). The distributions of matches between the two halves of chimeric tRNAs from any combination of domain and/or specificity are essentially identical to the distributions of matches between the halves of real tRNA sequences (Fig. 3).

We also tested whether the specific positions within the tRNA that contributed most to matches

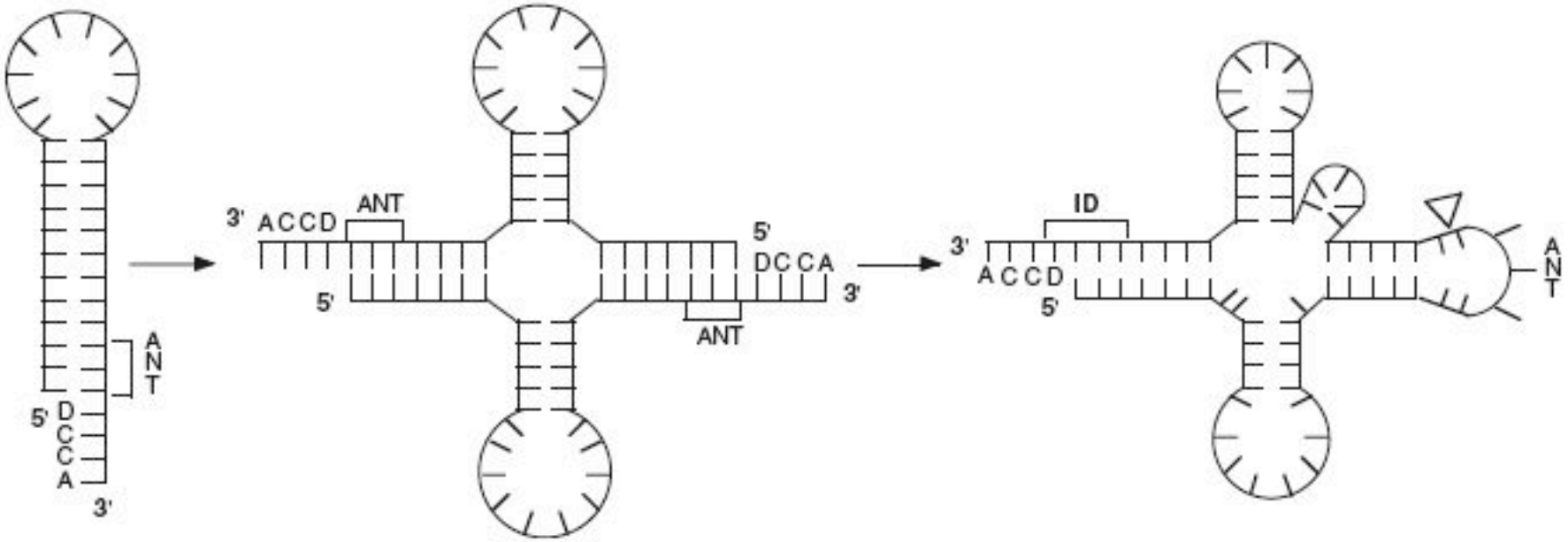
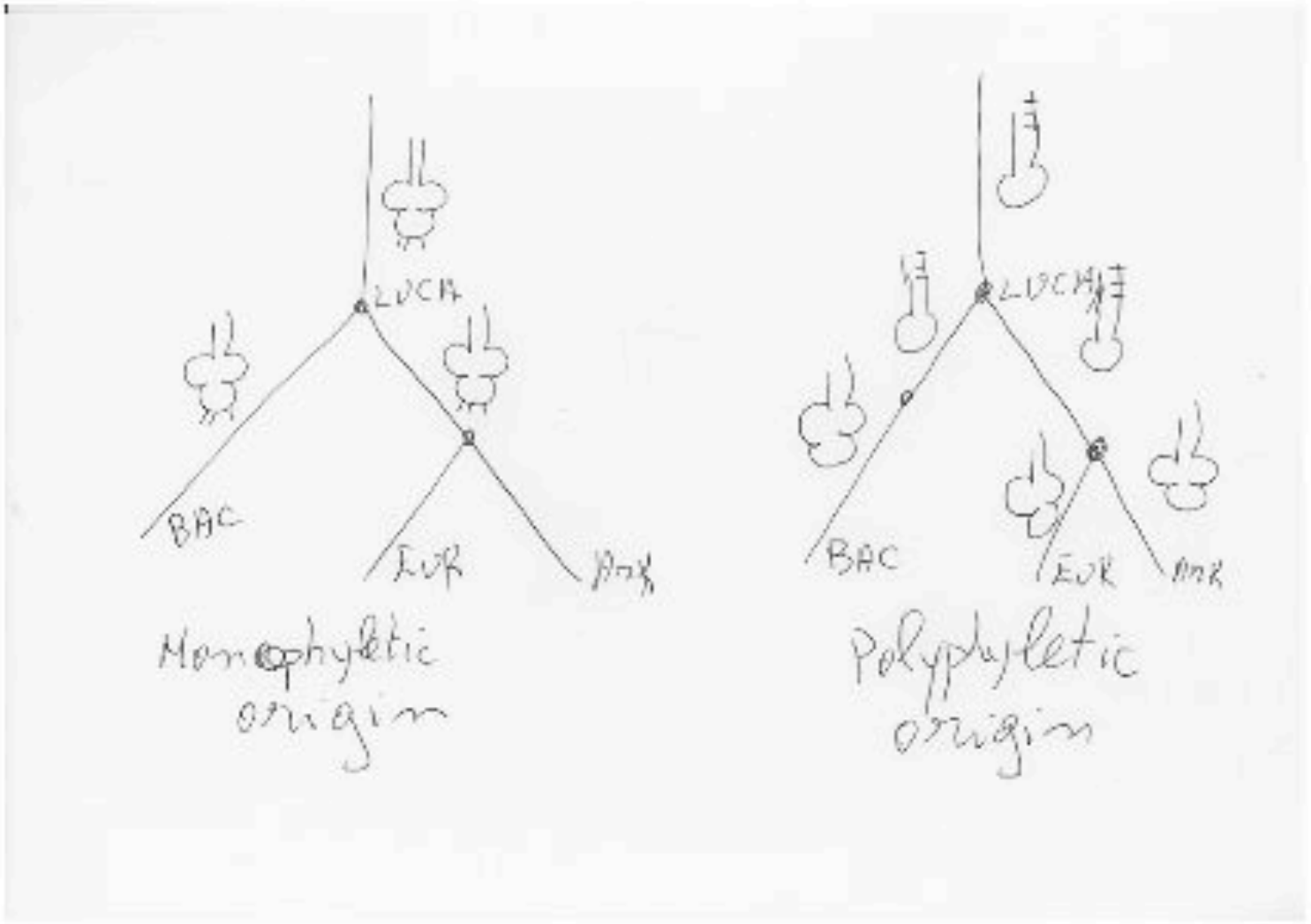


Fig. 1. Model for the origin of the tRNA molecule. For all details see Di Giulio (1992, 2004). The dimer of the two hairpin structures should be regarded as being covalently closed between one 3′ end and the other nearby 5′ end (Di Giulio, 1992). ANT means anticodon, ID are the nucleotides determining the identity of the tRNAs, whereas the triangle represents the position where the intron is normally found in tRNA genes (Di Giulio, 2004).

the introns were present in the tRNA genes, they would have to be located in the anticodon loop of the tRNA molecule because in this position they would cut this molecule into two halves corresponding to the original two single hairpin structures (Di Giulio, 1992, 1995, 1999, 2004). In other words, by means of the exon theory of genes, the tRNA model would envisage that minigenes codifying for the hairpin structures must have existed and their successive evolution must have led to the formation of modern tRNA genes (Di Giulio, 1992, 1995, 1999, 2004). Therefore, the introns in these genes would remember the in this section. Consequently, in the following sections, I will attempt to clarify why the observations of Randau et al. (2005a) are so important for what is here introduced regarding the origin and evolution of the tRNA molecule.

2. An important finding: in *Nanoarchaeum equitans* six tRNA genes have the 5′ and 3′ halves of this molecule codified in two completely separate and non-contiguous genes in the genome


LUCA
BAC
EUK
ARK
Monophyletic origin
LUCA
BAC
EUK
ARK
Polyphyletic origin

J. theor. Biol. (1999) **197,** 403–414
Article No. jtbi.1998.0882, available online at http://www.idealibrary.com on IDEAL

The Non-monophyletic Origin of the tRNA Molecule

MASSIMO DI GIULIO*

International Institute of Genetics and Biophysics, CNR, Via G. Marconi 10, 80125 *Naples, Italy*

(*Received on* 1 *October* 1998, *Accepted on* 26 *November* 1998)

The hypothesis that the tRNA molecule may have originated from the assembly of two similar RNA hairpin structures is utilised to understand the evolutionary period in which this molecule originated. Consistent with the exon theory of genes is the observation that the introns in tRNA genes are found almost exclusively in the anticodon loop and "stitched together" the two halves of the molecule, which originally may have been simply two hairpin structures and which can still be observed in the three-dimensional structure of tRNAs. This theory therefore considers these hairpin structures as minigenes on which complex protein synthesis may have been achieved. This in turn leads to the belief that the organisation of

xtensive experimental analysis aiming to see whether there are still remains of this atypical origin.

Materials and Methods

The sequences of tRNAs or tRNA genes and heir relative alignment were taken from an existing data base (Steinberg *et al*., 1993). The sequences of tRNAs containing numerous nodified nucleotides were transformed into DNA sequences using the table of modified nucleotides in this data base (Steinberg *et al*., 993).

Phylogenetic analysis was carried out using programs from the PAUP package (Swofford, 993). In particular, bootstrap analysis (Felsenstein, 1985) was conducted using all the default parameters with the *Steepest Descent* option Swofford, 1993). The bootstrap percentages efer to at least 1000 replications. In some cases he bootstrap analysis conducted using parsinony rules was confirmed using distances-based nethods, i.e. using the programs from the PHYLIP package (Felsenstein, 1993): SEQBOOT, DNADIST, NEIGHBOR and CONSENSE. Here, too, bootstrap analysis was carried out on 1000 replications.

The presence of a significant phylogenetic signal in a set of tRNA sequences was ascertained using the g_1 statistic (Hillis & Huelsenbeck, 1992). The significance of g_1 was established using the table of the relative critical values (Hillis & Huelsenbeck, 1992), whereas, the values of g_1 were estimated using the *Random Trees* option in PAUP (Swofford, 1993). Five

EVIDENCE FROM LITERATURE

As the tRNA molecule is one of the molecules that have been most extensively studied over the last 30 years and if, as is suggested here, this molecule was assembled late on in evolution, then the literature regarding tRNAs should refer to data substantiating a non-monophyletic origin of this molecule. The following data is in agreement with such an origin: (i) a comparison of the nucleotide sequences of a tRNA of Gly and Val belonging to *E. coli* gave an identity percentage of 78% (Squires & Carbon, 1971). This similarity was referred to as the highest between two tRNAs with different amino acid specificities (Squires & Carbon, 1971). The sequence of a tRNA of Val was shown to be more correlated to that of a tRNA of Gly than to another tRNA of Val (Squires & Carbon, 1971); (ii) analysis showed that some tRNAs specific for different amino acids present high and unexpected similarities (Holmquist *et al*., 1973); (iii) the convergence of ancestral sequences independently built from different branches of a phylogentic tree was used as a homology test for several tRNA sequences (Cedergren *et al*., 1980). This criterion led to the conclusion that the tRNAs specific for Tyr and belonging to eukaryotes and prokaryotes might have independent origins (Cedergren *et al*., 1980) and thus they might not be strictly homologous. Moreover, the close relationship between tRNAs specific for Gly and Val in some prokaryotes was confirmed (Cedergren *et al*., 1980); (iv) a high similarity (about 70%) in the tRNA sequences of Tyr and Phe was repeatedly observed (Muller & Clarkson, 1980; Green & Jones, 1986; Wong,

Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5′- and 3′-halves

Lennart Randau[1,2], Richard Münch[2], Michael J. Hohn[1,3], Dieter Jahn[2] & Dieter Söll[1,4]

[1]*Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA*
[2]*Institut für Mikrobiologie, Technical University Braunschweig, P.O. Box 3329, D-38023 Braunschweig, Germany*
[3]*Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany*
[4]*Department of Chemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA*

Analysis of the genome sequence of the small hyperthermophilic archaeal parasite *Nanoarchaeum equitans*[1,2] has not revealed genes encoding the glutamate, histidine, tryptophan and initiator methionine transfer RNA species. Here we develop a computational approach to genome analysis that searches for widely separated genes encoding tRNA halves that, on the basis of structural prediction, could form intact tRNA molecules. A search of the *N. equitans* genome reveals nine genes that encode tRNA halves; together they account for the missing tRNA genes. The tRNA sequences are split after the anticodon-adjacent position 37, the normal location of tRNA introns. The terminal sequences can be accommodated in an intervening sequence that includes a 12–14-nucleotide GC-rich RNA duplex between the end of the 5′ tRNA half and the beginning of the 3′ tRNA half. Reverse transcriptase polymerase chain reaction and aminoacylation experiments of *N. equitans* tRNA demonstrated maturation to full-size tRNA and acceptor activity of the $tRNA^{His}$ and $tRNA^{Glu}$ species predicted *in silico*. As the joining mechanism possibly involves tRNA *trans*-splicing, the presence of an intron might have been required for early tRNA synthesis.

The origin of the tRNA molecule is the subject of continuing

$tRNA^{His}$, $tRNA_i^{Met}$, $tRNA^{Trp}$ and two $tRNA^{Glu}$ species (Fig. 1). Further analysis of the tRNA half genes revealed several striking features. First, the location of the sequence separation that generated all nine tRNA half genes is after position 37, one nucleotide downstream of the anticodon and the common location of tRNA introns[9]. Second, a consensus sequence matching the highly conserved archaeal Box A promoter element[10] was found upstream of all 5′ tRNA halves. Third, this same consensus sequence (5′-TTTT/ATAAA-3′) was located 17–25 base pairs (bp) further upstream of the 3′ tRNA halves, resulting in a transcript with a 12–14-bp-long GC-rich leader sequence. Last, it is remarkable that this leading sequence is in all cases the exact reverse complement to a sequence following the corresponding 5′ tRNA half.

The existence of three $tRNA^{Glu}$ half genes was most exciting. Two 5′ tRNA halves were identified that differed solely by one anticodon base (isoacceptors with UUC and CUC anticodon), whereas only one 3′ $tRNA^{Glu}$ half gene was found. Both 5′ $tRNA^{Glu}$ half genes were followed by the identical 14-bp sequence that was the exact reverse complement of the single 3′ $tRNA^{Glu}$ half upstream sequence. All identified split tRNA genes contained the consensus bases of all archaeal elongator tRNAs[6], namely U8, A14, G15, G18, G19, C32, U33 and the T-loop GTTCA/GAATC (53–61), with the exception of the putative $tRNA^{Trp}$ harbouring an unusual GG sequence preceding the anticodon. The identified $tRNA_i^{Met}$ displays the consensus sequences of archaeal initiator tRNAs such as the anticodon stem/loop nucleotides (nt) 29–41 (GGGCU-CAUAACCC) and the R11:Y24 base pair (G11:C24), which is the reverse of the Y11:R24 base pair found in elongator tRNAs including the annotated *N. equitans* $tRNA^{Met}$. Therefore we define the split $tRNA_i^{Met}$ as the missing initiator tRNA. The sequences also reveal characteristic nucleotides in the respective tRNA species needed for recognition by the cognate aminoacyl-tRNA synthetase. For example, the $tRNA^{His}$ half genes encode the unique G-1:C73 base pair required for aminoacylation of $tRNA^{His}$ by histidyl-tRNA synthetase[11], and the $tRNA^{Glu}$ isoacceptors contain the characteristic D-loop nucleotides 20a and 20b and the deletion of base 47 essential for making the 'augmented D-helix'[12].

We performed reverse transcriptase polymerase chain reaction (RT–PCR) analysis of *N. equitans* total tRNA to verify the computationally predicted sequence of the newly discovered joined tRNAs. Our sequencing results confirmed the sequences for $tRNA^{Glu}$

(UUC), $tRNA^{Glu}$ (CUC) and $tRNA_i^{Met}$ (Fig. 2a). Despite extensive efforts we could not amplify the full-length $tRNA^{Trp}$ and $tRNA^{His}$ (even though its existence was shown by aminoacylation; see below); this might have been due to the extreme thermostability of the GC-rich *N. equitans* tRNAs containing modified nucleosides. Nevertheless, we confirmed the presence of six tRNA half transcripts by RT–PCR and sequence analysis (Fig. 2d). The primary transcripts of these tRNA half genes include the intervening complementary sequences at the position of separation. In addition, RT–PCR of anchor-ligated tRNA (Fig. 2c) revealed that the primary transcript of the 5′ $tRNA^{His}$ half terminates at the AT-rich region following the complementary downstream sequence found in all tRNA half genes.

For a tRNA to participate in protein biosynthesis it must carry a 3′-terminal CCA sequence to which the amino acid will be esterified. In *N. equitans* and most Archaea, this CCA sequence is not encoded in the tRNA genes (including the split tRNA genes) but is added post-transcriptionally by the ATP(CTP):tRNA nucleotidyltransferase[13,14], an enzyme probably encoded by the still uncharacterized NEQ152 gene. By using a RT–PCR approach involving circularization of the tRNA[15] we were able to identify the 5′ and 3′ ends of the mature tRNA. Our sequencing results show size-maturation of the joined $tRNA^{Glu}$, as a CCA sequence is indeed added to the 3′ end of both $tRNA^{Glu}$ isoacceptors after transcription (Fig. 2b). A final requirement for tRNA functionality *in vivo* is the ability to serve as a substrate for amino acid attachment by

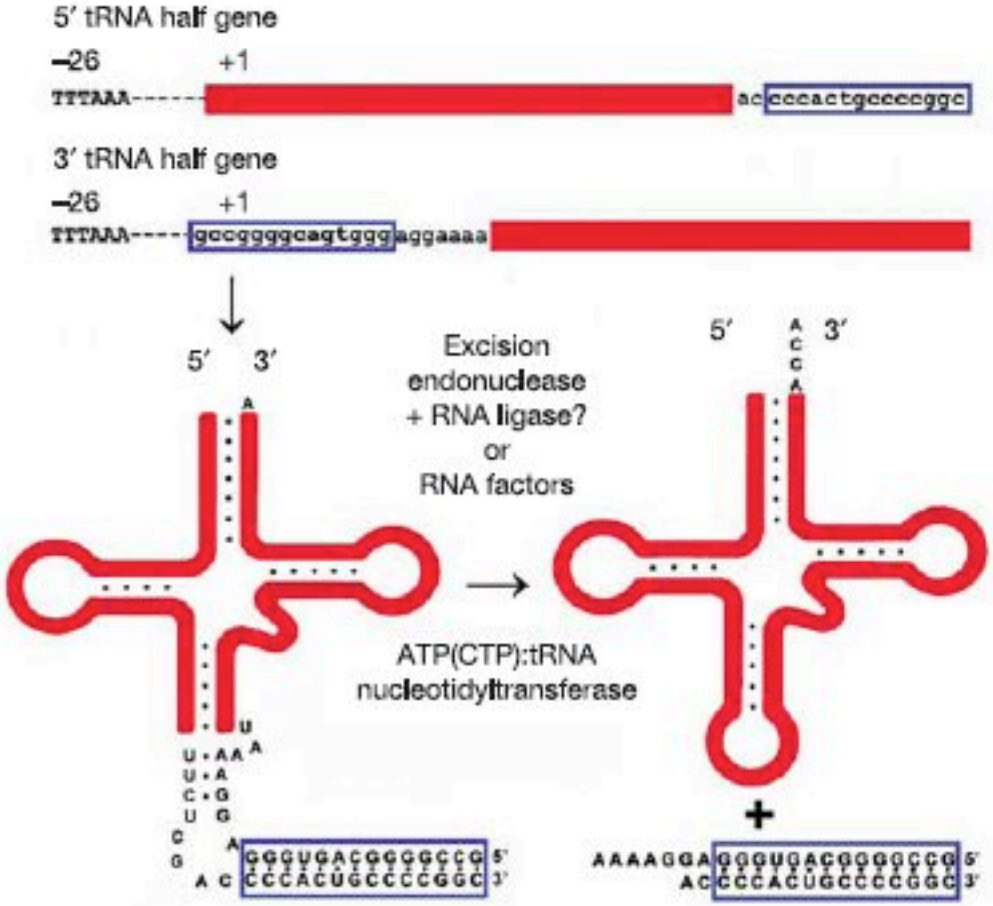


Figure 4 Schematic representation of a 5′ tRNA half gene ($tRNA^{Glu}$) and the corresponding 3′ tRNA half gene found in *N. equitans*. The archaeal RNA polymerase III promoter consensus sequence (TTTAAA), the tRNA half genes (red) and the intervening reverse complementary sequences that are supposed to facilitate joining of the halves (blue) are indicated.

contain leader sequences upstream of the 5′ end of tRNA. In absence, there would not be a need for the presence of RNase P organism.

An extensive search of the available bacterial and arc genome sequences did not reveal split tRNA genes in other o isms. Future sequences of other very small or compact gen should reveal whether split tRNA genes are signs of a very genome[2] or whether they are created in a later process of ge size reduction. The sequencing of other Nanoarchaeota genor therefore eagerly awaited.

Methods

Computational method for tRNA identification

tRNA genes were predicted by use of a new bioinformatics approach and the pro Virtual Footprint (http://www.prodoric.de/sts/). Position weight matrices were ge from both a conserved, continuous 3′ region of tRNA genes (nt 54–76) and a 5′ re tRNA genes (nt 1–16) in an alignment of more than 4,000 tRNA genes (taken from For this purpose the information content was used as scoring function[28] with slig modifications. tRNA gene searches were performed with these position weight mat a genome scale with the highest sensitivity (the threshold score was taken from the scoring sequence of the training set). The information that the 3′ region contains a stretch of 7 nt to a reverse complementary part in the 5′ region (the tRNA's accepto was used to identify matching pairs of tRNA gene halves. Using this approach, al previously annotated tRNAs were identified, including nine additional tRNA hal fell into the threshold range of the annotated tRNAs.

Cell culture and tRNA isolation

N. equitans cells were grown in a 300-l fermenter in a simultaneous culture with *Ig* sp. and purified by gradient centrifugation as described[1]. The cell pellet was lysed chemical digestion with 2% SDS, followed by the isolation and purification of total described[29]. The tRNA was further purified by MonoQ HR 5/5 anion-exchange chromatography to eliminate residual genomic DNA contamination. The tRNA v eluted with a linear 60-ml gradient of 0–1 M NaCl in 20 mM MOPS pH 6.2.

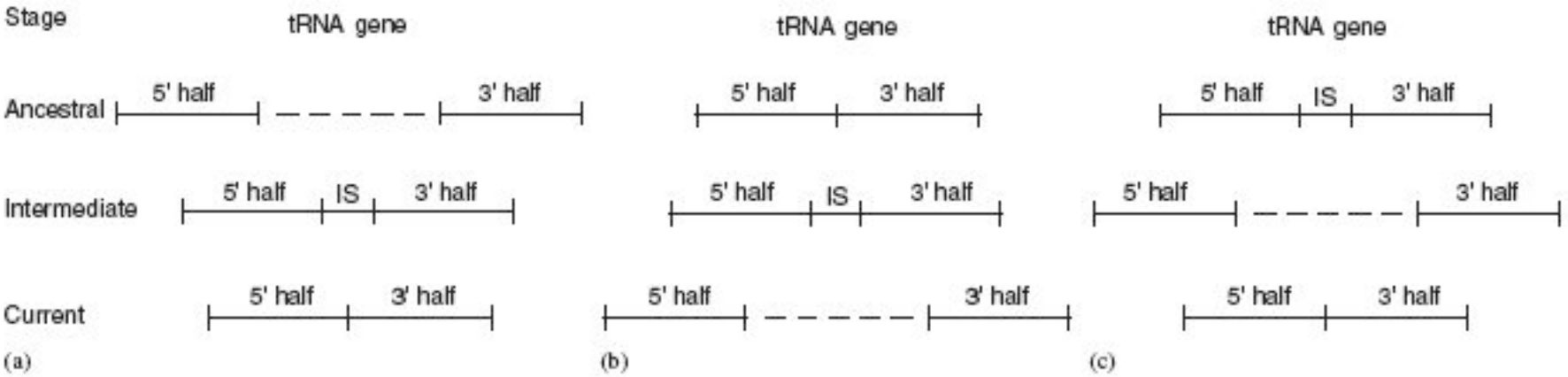


Fig. 2. This represents the three evolutionary stages through which the evolution of tRNA genes is thought to have passed. (a) From top to bottom, schematically and respectively shows: (i) the two completely separate genes codifying the 5′ and 3′ halves of tRNAs (ancestral stage); (ii) the gene for the tRNA in which the 5′ and 3′ halves are interrupted by an intron (IS) (intermediate stage); (iii) the gene for tRNA continuously codifying for the 5′ and 3′ halves (current stage). (b) and (c) show the same evolutionary stages but in different successions.

Table 1 Effect of the logical implications of the reasoning on the number of all the possible successions of the three evolutionary stages

Number of the possible or remaining permutations of the three evolutionary stages:	6
A1. The genes of tRNAs with introns have an evolutionary link with the split genes of tRNAs:	4
B1. The split genes of tRNAs did not confer a selective advantage;	
B2. Genetic drift did not fix the split genes of tRNAs;	
B3. Some successions of evolutionary stages lead to an increase in complexity:	3
C1. Four successions of evolutionary stages are not natural successions;	
C2. The observed absolute frequencies of the three different evolutionary stages categorically exclude that these four successions might be part of natural successions of evolutionary stages;	
C3. Two of these four successions are also falsified by the fact that the split genes of tRNAs have an evolutionary link with the genes of tRNAs with introns (see A1):	1

It can be seen how the implications progressively reduce the number of successions of evolutionary stages compatible with them until it is possible to identify the only one compatible with the entire reasoning (Fig. 1a). See text for a complete comment

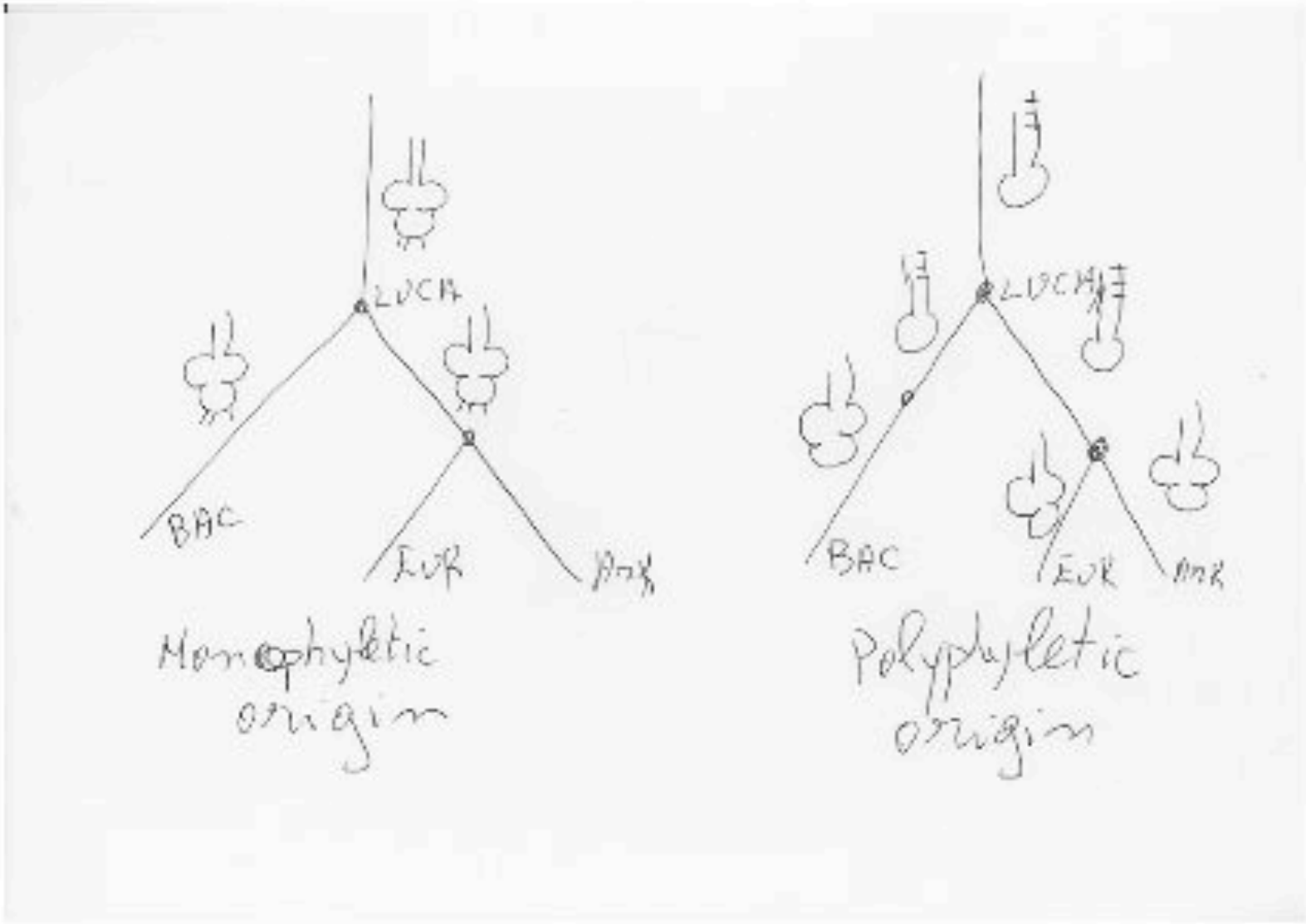
There is only one succession of evolutionary stages, the one which identifies the split genes of tRNAs as a plesiomorphic character (Fig. 1a)

4. Implications for the origin of the tRNA molecule and for the origin of protein synthesis

Having shown that the completely separate tRNA genes codifying for the 5′ and 3′ halves of this molecule (Randau et al., 2005a) might have been the ancestral form from which modern tRNA genes continuously codifying for this molecule evolved (Fig. 2a), we can ask what the implications of this are for the model of tRNA origin (Fig. 1). There are some arguments that lead us to believe that this model (Fig. 1) is strongly corroborated by the assertions made in the present paper. Indeed, the tRNA genes codifying for the 5′ and 3′ halves of this molecule (Randau et al., 2005a) by definition still codify for structures which are almost complete hairpins, and hence in agreement with the tRNA model (Fig. 1) because this would seem to imply that the ancestral forms of these structures were actually complete hairpins proper. Moreover, as predicted by the model, the 5′ halves of these molecules have anticodons positioned at their very 3′ ends (Randau et al., 2005a). This provides strong evidence in favour of the model because, as already discussed (Di Giulio, 1995, 1999, 2004), the anticodons positioned in the stem of the hairpin structure might have been transferred by means of duplication into the anticodon loop during the origin of the tRNA molecule. Therefore, these two observations identify the hairpin structure reported in Fig. 1 as the probable precursor of the tRNA molecule, thus making the model itself very likely. Moreover, this hairpin structure would have also had to be an intermediary in the origin of protein synthesis because any intermediary of the origin of the tRNA molecule must be so (Di Giulio, 1999, 2004). In other words, a complex protein synthesis must have taken place on the hairpin structures (Di Giulio, 1999, 2004). Indeed, the mere existence of genes separately codifying for the 5′ and 3′ halves of tRNAs together with their ancestrality seems to imply that these genes originally

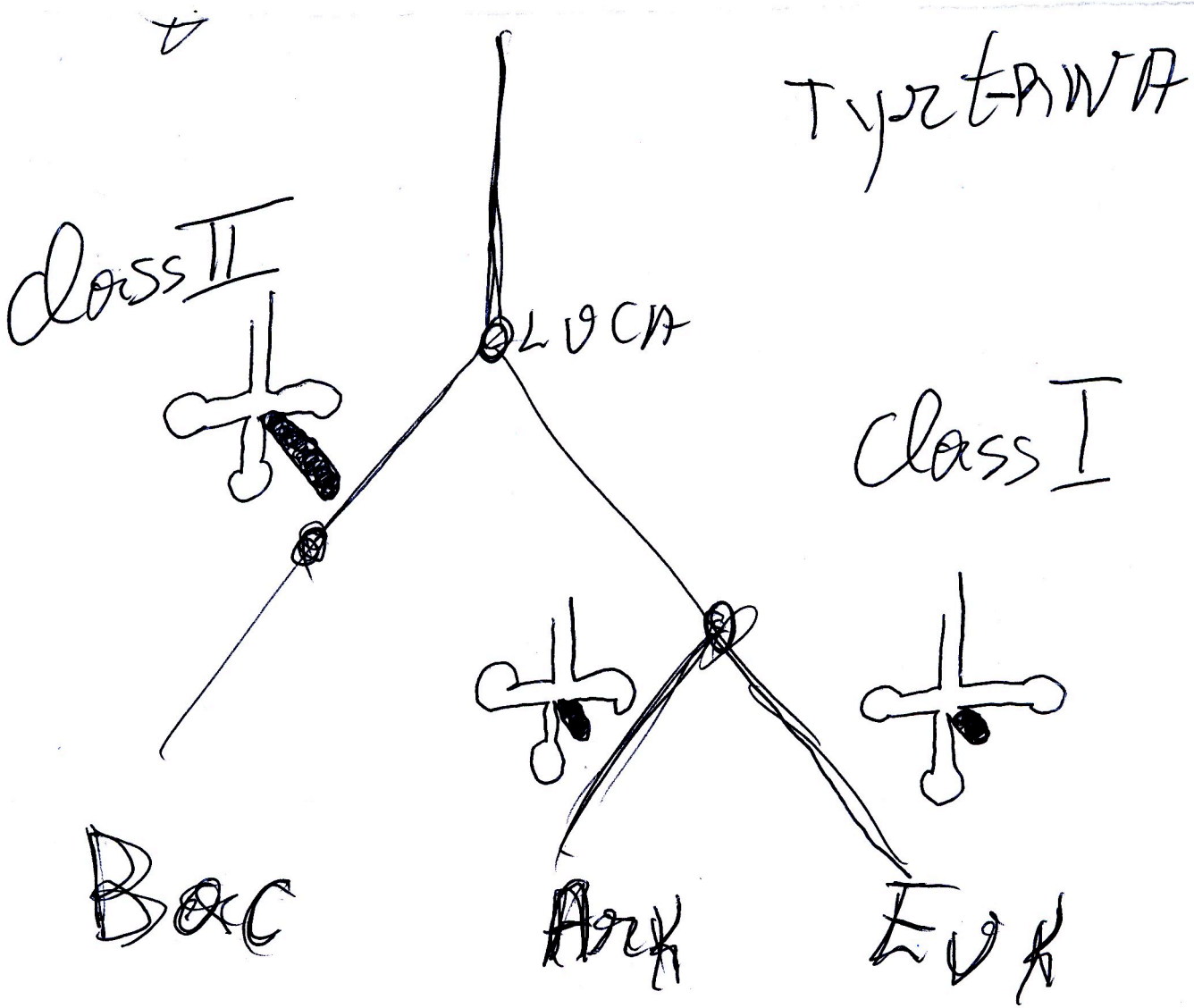
The theorem of the polyphyletic origin of tRNA genes

"*If the half genes of tRNAs of N. equitans were the plesiomorphic form of tRNA genes, then the mere observation of the existence of these genes in an organism implies that the hypothesis of the monophyletic origin of tRNA is false, and thus the polyphyletic origin of tRNA genes is true*".


LUCA
BAC
EUK
ARK
Monophyletic origin
LUCA
BAC
EUK
ARK
Polyphyletic origin

In a previous section, I have shown that the completely separate genes codifying for the 5′ and 3′ halves of tRNAs (Randau et al., 2005a) represent the ancestral stage through which the modern genes continuously codifying for tRNAs might have passed and from which they might have evolved (Fig. 2a). This, in addition to the very existence of these genes, implies that the origin of the tRNA molecule was non-monophyletic; in other words, this molecule was definitively assembled only after the establishment of the main lines of divergence leading to the domains of life. Indeed, if this origin had been monophyletic, i.e. the complete tRNA molecule had appeared prior to or during the stage of the LUCA, then all the genes for tRNA molecules would have had to be, by definition, complete genes for tRNAs, that is to say they could not have been codified in two separate halves because we are reasoning under the hypothesis of a monophyletic origin

> which implies that all the genes for tRNA molecules were already completely developed and evolved at this stage. Given that we have seen that (i) some tRNA genes are completely separate codifying for the 5′ and 3′ halves of this molecule and that (ii) this represents the ancestral stage of this molecule, it follows that the origin of the tRNA molecule must therefore have been non-monophyletic. In other words, if this origin had been monophyletic we should not have observed in a real organism the existence of completely separate genes codifying for the 5′ and 3′ halves of tRNAs as the monophyleticity hypothesis does not admit the existence of these genes. This is because, under the hypothesis of monophyletic origin, at the LUCA stage all the tRNA molecules must have already been entirely developed and complete and, therefore, they could not in a successive evolutionary stage have still been codified in completely fragmented genes separately codifying for the 5′ and 3′ halves of this molecule. Clearly, this points to a non-monophyletic origin for this molecule.


Type-A
class II
class I
B&C
A
E

Conclusions:

1.The tRNA molecule originated by means of the assembly of two hairpin-like structures

2.The origin of tRNA genes occurred very late in evolution i. e. this origin was polyphyletic.

These conclusions are strong because are based on molecular fossils of ancient tRNA genes