Journal of Theoretical Biology 480 (2019) 99–103

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtb

A comparison between two models for understanding the origin of the tRNA molecule

Massimo Di Giulio[a,b]

[a] The Ionian School, Genetic Code and tRNA Origin Laboratory, Via Roma 19, 67030 Alfedena (L'Aquila), Italy
[b] Institute of Biosciences and Bioresources, National Research Council, Via P. Castellino, 111, 80131 Naples, Italy

ARTICLE INFO

Article history:
Received 31 December 2018
Revised 6 April 2019
Accepted 30 July 2019
Available online 31 July 2019

Keywords:
Molecular fossils
Ancient tRNA genes
Split tRNA genes
tRNA models

ABSTRACT

I respond to criticisms raised by Kim et al. (2018) to my model concerning the origin of the tRNA molecule. In particular, their model would hypothesize the tRNA originated due to the ligation of three hairpin structures followed by two deletions, while my model predicts that this molecule derived from the assembly of only two hairpin-like structures. Thus, using the Ockham razor, the latter model would be chosen because it required fewer hypotheses. Furthermore, the predictions on homology between the different regions of the tRNA molecule as predicted by my model would be statistically more significant than those predicted by their model. Moreover, it would be above all the existence of molecular fossils - i.e. the split tRNA genes - to corroborate the model of the assembly of only two hairpins. These fossils would be completely absent from the Kim et al. (2018) model.

I believe that there is a highly corroborated model concerning the origin of the tRNA molecule. This model postulates that the tRNA molecule originated after the assembly of two hairpin-like molecules (Di Giulio, 1992, 1995, 1999, 2004, 2006a). In fact, there is a very simple way to construct the tRNA molecule (Fig. 1). Starting from a hairpin precursor of RNA (Hopfield, 1978; Di Giulio, 1992) if two of these hairpins are assembled as shown in Fig. 1, then the covalent union of these two hairpins would give rise - passing through the intermediate stage of the double hairpin (not shown in Fig. 1) (Tanaka and Kikuchi, 2001; Di Giulio, 2004; Branciamore and Di Giulio, 2011) - to the complete tRNA molecule (Di Giulio, 1992, 1995, 2004, 2006a). The model has several properties and there are strong arguments in its favor (Di Giulio, 1992, 1995, 1999, 2004, 2006a, 2009a, 2012), which are summarized below:

(i) If an RNA hairpin was indeed the precursor of the tRNA molecule, then this necessarily had to participate in the ancestral protein synthesis (Di Giulio, 1994, 1999, 2004, 2006a). If you place the anticodon near the 3′ end of the hairpin (Fig. 1) then the assembly of two hairpins is such that the anticodon is brought from the stem of the hairpin structure to the anticodon loop region and the covalent closure of the molecule would therefore lead to the formation of the anticodon loop (Moller and Janssen, 1992; Di Giulio, 1994, 1995, 2004, 2006a). That is to say, the model is able to explain the transfer of the anticodon from the 3′ end of the hairpin to the anticodon loop with its relative formation (Di Giulio, 1995, 2004, 2006a). Moreover, the model is able to explain why many determinants of the identity of tRNAs are located in the acceptor stem of these. Indeed, the model is able to make evolutionarily equivalent the anticodon loop region to that of the acceptor stem where many of identity determinant nucleotides of tRNAs are located (Giege et al., 1998; Branciamore et al., 2018). Namely, an anticodon of the hairpin would give rise to the real anticodon while the other would be the evolutionary precursor of determinants of the identity of tRNAs (Di Giulio, 1995, 2004, 2006a), i.e., the second genetic code (de Duve, 1988; Schimmel et al., 1993). That is to say, the model is able to relate the two functional sites of the tRNA molecule in an evolutionary relation: the anticodon in the anticodon loop and the determinants of the identity of tRNAs in the region of the acceptor stem (Di Giulio, 1995, 2004, 2006a). If all this were only the result of chance, then it would be truly unique. To reinforce that this is not due to chance, but it is the result of how the tRNA molecule originated, it is the following really amazing argument. According to the exon theory of the origin of genes, the genes originated by assembling pieces of smaller genes – i.e. the exons - and this assembly would have been mediated by introns (Gilbert, 1978; Doolittle, 1978; Gilbert et al., 1997). Therefore, a prediction of the exon gene theory when applied to the tRNA model would

E-mail address: massimo.digiulio@ibbr.cnr.it

https://doi.org/10.1016/j.jtbi.2019.07.020
0022-5193/

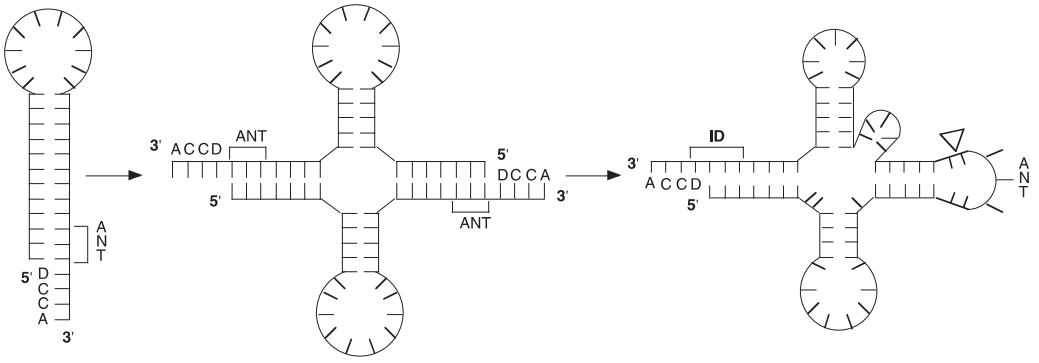

Fig. 1. The origin of the tRNA based on two hairpin-like structures. The image was taken by Di Giulio (2006a). The cruciform structure in the center must also be understood as covalently closed between only one free 3′ end and a free 5′ end (Di Giulio, 1992). In this sense the cruciform structure should also be interpreted as that of the double hairpin (Tanaka and Kikuchi, 2001; Di Giulio, 2004; Branciamore and Di Giulio, 2011), which would represent the thermodynamically more stable structure and most likely the evolutionary stage biologically active intermediate through which it passed to arrive at the final tRNA molecule (Tanaka and Kikuchi, 2001; Di Giulio, 2004; Branciamore and Di Giulio, 2011).

be that the introns in tRNA genes should be located in the anticodon loop because in this position they would remember the way with which the tRNA molecule was assembled for the first time (Di Giulio, 1992, 1995, 2004, 2006a,b). In fact, in this way the intron would cut the tRNA molecule into two almost equal parts (Di Giulio, 1992, 1995, 2004, 2006a). The introns of tRNAs genes are localized in the anticodon loop between the 37th and the 38th nucleotides (Sprinzl et al., 1998). The position of these introns in tRNA genes is conserved in tRNA genes of organisms from all three domains of life (Sprinzl et al., 1998). Thus, it would appear that the position of these introns in genes of tRNAs is the most conserved and therefore the oldest known case (Di Giulio, 1992, 1995, 2004, 2006a,b). Therefore, the intron in this position would have mediated the union between the two hairpin structures called to originate the tRNA molecule (Di Giulio, 1992, 1995, 2004, 2006a). However, the amazing results would not stop here. One of the main predictions of this model is that given that the tRNA molecule would be assembled by two RNA hairpin structures, then there should have been an evolutionary stage in which tRNA genes existed encoding only one hairpin (Di Giulio, 1995, 1999, 2004, 2006a). This prediction is incredibly confirmed. Indeed, the split genes of tRNAs were identified in which the 5′ and 3′ halves are coded on completely separate genes and occupying regions that are also very distant on the chromosome (Randau et al., 2005). These genes have been shown to be the plesiomorphic forms, that is, ancestral forms of genes for tRNAs (Di Giulio, 2006a,b, 2008a,b,c, 2009a,b). Moreover, these broken genes are interrupted at the same point between the 37th and 38th nucleotides, i.e. at the same point where the introns of tRNA genes are located (Randau et al., 2005). This established a clear evolutionary relationship between the genes of tRNAs with introns and those completely broken up and, in the latter, it would identify the plesiomorphic forms of these genes (Di Giulio, 2006b, 2009b). It would then appear that these split genes of tRNAs are the oldest known genes (Di Giulio, 2006a,b, 2009b). The existence of these split genes of tRNAs would strongly support the model presented in Fig. 1 because their ancestrality (Di Giulio, 2006a, 2006b, 2009b) would be a direct proof that this model might have been really operational since even today the trans-splicing reaction takes place between two hairpin-like RNA structures encoded by these very ancient broken genes (Di Giulio, 2004, 2006a,b, 2009b). Of course this would also imply that the hairpins were evolutionary precursors of the tRNA molecule (Di Giulio, 1995, 1999, 2004, 2006a).

(ii) The three-dimensional structure of tRNAs, the famous L-shaped structure, is made up of two domains that are basically two hairpin structures (Steinberg and Cedergren, 1994). Given that the secondary and tertiary structures of molecules of structured RNAs are considered to be better conserved than their primary structures, and given that in agreement with the model the tRNA molecule was assembled using two hairpin structures, then the prediction would be that in the tertiary structure of tRNAs the number of RNA domains should be equal to two as those actually observed. This would represent a further corroboration in favor of the model (Maizels and Weiner, 1994; Di Giulio, 1995, 2004), given that the number of RNA domains envisaged by the model would coincide with those actually observed.

(iii) Another prediction of the model is that by deriving the tRNA molecule from the assembly of two hairpin structures taken from a hairpin population of homologous structures, then the two halves of tRNA molecules should be similar in the sequence (Di Giulio, 1992, 1995, 1999). This forecast is fully confirmed. Indeed, there are several works that indicate that the two halves of tRNAs are in their statistically similar sequences (Di Giulio, 1992, 1995, 1999; Widmann et al., 2005; Branciamore and Di Giulio, 2011). Furthermore, this prediction was also confirmed by using the reconstructed ancestral sequences of tRNA genes (Di Giulio, 1992, 1995).

On the contrary, the model of Kim et al. (2018) (Fig. 2) is based on the ligation of three 31-nt mini-helices followed by two symmetrical 9-nt deletions within ligated 3′- and 5′-acceptor stems (see also: Root-Bernstein et al., 2016; Pak et al., 2017). This model posits both that the anticodon loop is homologous to the T-loop and class I and class II V loops are homologous to acceptor stems (Kim et al., 2018).

From a general point of view, the model based on the direct duplication of an RNA hairpin structure (Fig. 1) (Di Giulio, 1992, 1995, 2004, 2006a) - for the birth of the tRNA molecule - would certainly be more parsimonious than that by Kim et al. (2018) based, instead, formally on ligations of three hairpin followed at least by a deletion (Fig. 2). That is to say, according to Ockham's razor (Panaccio, 2004), the first model would be chosen because it would require fewer hypotheses to explain the origin of the tRNA molecule than the second which would require at least two more.

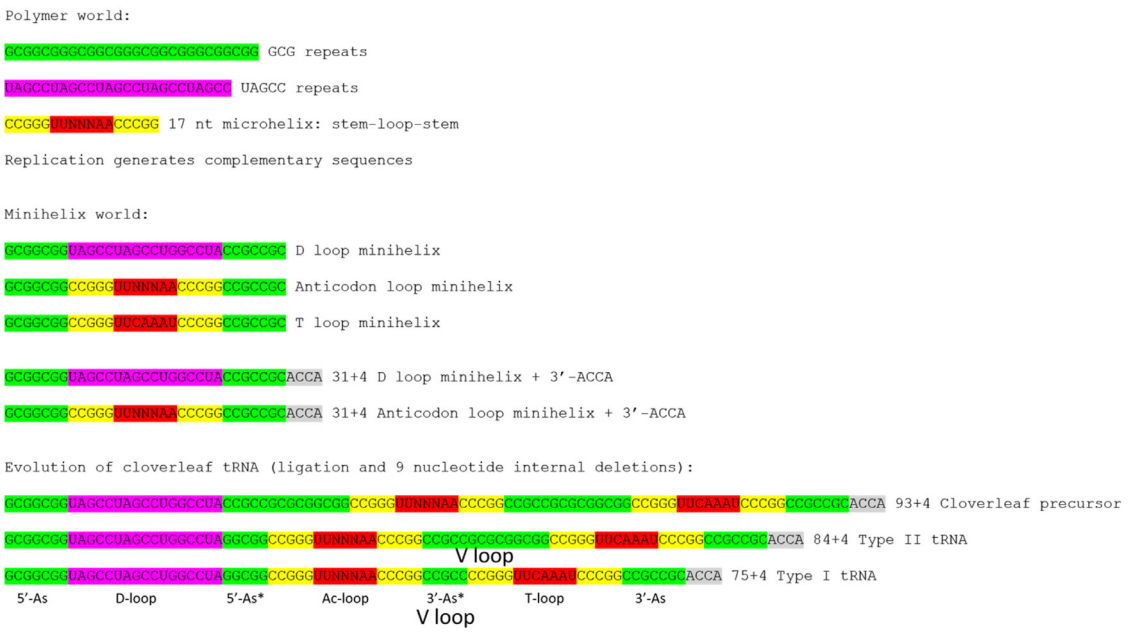

Fig. 2. Models for the evolution of class-I and class-II tRNAs. The image was taken by Kim et al. (2018). For comments see text; while for more information on the model see Kim et al. (2018), Root-Bernstein et al. (2016), and Pak et al. (2017).

A critique that can be moved to this model - which predicts that the tRNA molecule of class I and that of class II derive from a longer cloverleaf precursor (Fig. 2) - is that it fails to explain the observed distribution of amino acids to class I and class II. For the Ser, Leu and Sec amino acids only class II tRNAs were observed, while for Tyr, both class I tRNAs in archaea and eukaryotes and class II were found in bacteria (Sprinzl et al., 1998). If the hypothesis that the class I and II tRNAs derived from a longer cloverleaf precursor (Fig. 2) were true, then we would have to observe that the class I tRNAs should have been observed with a higher frequency - in tRNAs of these four amino acids - precisely because, among other things, the class I tRNAs are the most frequent ones (Sprinzl et al., 1998). What has not been observed. In other words, it would appear that the number of tRNAs of class I for tRNAs of these four amino acids is less than that expected on the basis of their absolute frequency in the population of all tRNAs. Indeed, the probability of the distribution of the class I tRNAs in the class II tRNAs for these four amino acids is equal to $P = 0.015$.[1] While a more numerous presence of class I tRNAs would have resulted in not statistically significant distributions. This probability defining an unlikely event is in favor of the hypothesis that the model of Kim et al. (2018) is not correct because the distribution of the class I tRNAs reflected in this probability does not adequately represent this model. Indeed, we would have had to observe more class I tRNAs on tRNAs of these four amino acids in order to corroborate the hypothesis that the class I and II tRNAs would derive from a longer cloverleaf precursor. This would be expected based on the model of Kim et al. (2018) since this would predict an evolution from the long cloverleaf precursor towards the class II tRNA and then to the class I tRNA (Fig. 2), with an unavoidable accumulation - on the basis of their model, which is not observed - of tRNAs of class I in organisms that would also have class II tRNAs. In other words, this relative distribution of the class I tRNAs and class II tRNAs of these four amino acids would seem better to reflect the hypothesis that these class II tRNAs are not evolving towards those of class I because otherwise we would have observed a higher number of tRNAs of class I among tRNAs of these four amino acids. This supports a model of non-evolutionary transition between the class II tRNAs and those of class I (Di Giulio, 2013), such as that of Fig. 1.

Kim et al. (2018) claimed that: "One proposed two minihelix model is based too heavily on analysis of tRNA introns in the Ac loop of one archaeal species. Introns are found in many sites of archaeal tRNAs, not just in the Ac loop". This is absolutely incomprehensible. The model (Di Giulio, 1992, 1995, 2004, 2006a) is not based on the analysis of introns of tRNAs of the Ac (anticodon) loop of one or few archaeal species but on the fact that the position of the intron between the 37th and 38th nucleotides of tRNAs genes is unique in that it is conserved in organisms from all three domains of life: indeed that in one or some archaeal species! Therefore, the position of this intron could have been a witness of the origin of the tRNA molecule because the intron of this position might have mediated the union between the two hairpin structures called to form the tRNA molecule, given its secure antiquity (see above; Di Giulio, 1992, 1995, 2004, 2006a, 2009b). It is true that there are other introns in genes of tRNAs in other positions but these are not absolutely relevant to the model. Their presence in no way alters the conclusions concerning the intron of the loop of the Ac because these introns are most likely of the acquired traits but not very ancient as that of the loop of the Ac (Di Giulio, 1992, 1995, 1999, 2004, 2006a, 2009b).

Kim et al. (2018) claimed that: "Our model makes strong sequence predictions, which are all justified by statistical tests. So far as we can judge, two minihelix models do not make strong sequence predictions that can be justified by any analysis we can apply". The observations of Kim et al. (2018) are statistically less significant (see also: Root-Bernstein et al., 2016; Pak et al., 2017) than those of our analyses that have comparably used thousands of tRNA sequences (Di Giulio, 1995; Widmann et al., 2005; Branciamore and Di Giulio, 2011). Thus, it is to be seen whether their statistical tests corroborate how much their model assumes. In

[1] The probability was calculated with the Barnard test with parameters: $a = 0$, $b = 9$, $c = 2$, $d = 1$; while Fisher's exact test provides a $P = 0.045$. The 2×2 contingency table has the following four categories: 1. tRNAs of class I ($a = 0$), 2. tRNAs of class II ($b = 9$), 3. amino acids with only one class of tRNAs ($c = 2$), 4. amino acids with the two classes of tRNAs ($d = 1$).

contrast, my model not only makes precise predictions regarding the homology between different regions of tRNA sequences, showing for example, which is the two halves of tRNAs being homologous (Di Giulio, 1992, 1995); but also that in an analysis used 6810 tRNAs homology was corroborated both on the 5′ and 3′ halves of tRNAs as they are integer in the statistical analysis, and on individual regions that the model could be homologous such as, for example, the homology between the D loop and Tψ loop (Branciamore and Di Giulio, 2011). In this regard, Kim et al. (2018) claimed: "We show clearly that the Ac loop and T loop are homologs. In a two minihelix model, however, the Ac and the T stem-loop-stem cannot be homologous, because the Ac loop must be bisected to make the comparison, spoiling the alignment. Rather, a two minihelix model predicts that the D loop and T loop should be similar in sequence, which they clearly are not". In Di Giulio (1992) we can read: "In conclusion, even if there seems to be some doubt as to which pairs of nucleotides are really homologous, what we can certainly say is that the D-loop and Tψ-loop of the present tRNAs have a very high degree of similarity, even in different base homology patterns, and are the most convincing evidence that the tRNA molecule was able to originate following a duplication event (Di Giulio, 1992)". This conclusion was subsequently confirmed (Di Giulio, 1995; Widmann et al., 2005; Branciamore and Di Giulio, 2011). Indeed, the invariance of some nucleotides in the loops and neighboring regions (Eigen et al. 1989) is such as to make the similarity of these regions unquestionable (Di Giulio, 1992) even if it was not easy to identify the truly homologous nucleotides in the two loops (Di Giulio, 1992) and this might, at least partly, justify the contrasting behaviour of these regions (Branciamore and Di Giulio, 2011). Therefore, we verified the robustness of the complementarity in the loop regions and we find that it transforms into homology when the compared regions of the Tψ-loop are less rigid and allow two or three nucleotides from the nearby stem to be included in the analysis (Branciamore and Di Giulio, 2011). In conclusion, it is extremely probable that the D loop is homologous to the Tψ-loop, while it is really unlikely that the Ac loop can be homologous to the T stem-loop-stem as predicted by the Kim et al. (2018). Furthermore, Kim et al. (2018) claimed that: "As we show here, and as we have shown previously, the D-loop microhelix is based on a UAGCC repeat, which cannot be similar in sequence to a CCGGGUUCAAAUCCCGG T stem-loop-stem. In Figure 4B, we show two perfect UAGCC repeats in the D loop, indicating the UAGCC repeat". As just recalled it is really unlikely that the Ac loop can be homologous to the T stem-loop-stem as predicted by Kim et al. (2018). Furthermore, the UAGCC repeat is not properly part of the D loop or at least these two repeats cover only less than half of the D stem-loop-stem (see Fig. 4B of Kim et al. (2018)) while our analysis compared more precisely the regions of D stem-loop-stem and that of the T stem-loop stem and clearly showed that these are similar, i.e. homologous regions (Di Giulio, 1992, 1995; Widmann et al., 2005; Branciamore and Di Giulio, 2011).

Kim et al. (2018) claimed that: "another criticism of the two minihelix models is that they appear to require unlikely sequence and structural convergence of the 7-nt U-turn Ac and T loops. If the homology of the Ac and T stem-loop-stems is accepted (Fig. 2), only the three minihelix model makes sense". My model of the two minihelices (Fig. 1) does not require any structural convergence in the 7-nt sequences U-turn Ac and T loops because it does not establish any evolutionary relationship between these two regions of the tRNA molecule. Indeed, this model rejects the homology between the region of the Ac and that of the T stem-loop-stem, suggesting instead that the T-stem-loop-stem is homologous to the D stem-loop-stem on the basis that this last homology is corroborated by the strong sequence similarity between these two regions with probability values of several orders of magnitude more significant (Branciamore and Di Giulio, 2011) than those found in the works supporting instead the homology between the region of the Ac and that of the T stem-loop-stem (Kim et al., 2018; Root-Bernstein et al., 2016; Pak et al., 2017).

In conclusion, I am not only convinced that the model by Kim et al. (2018) fails to explain the origin of the tRNA molecule appropriately, but that the model of Bloch et al. (1985) - very similar to the model by Kim et al. (2018) - based on three replication cycles of a small hairpin, is able to give a most parsimonious description of the tRNA origin and therefore better than that provided by the Kim et al. (2018), but however - it seems to me - not superior to that of the model favored here (Di Giulio, 2012, 2013).

References

Bloch, D.P., McArthur, B., Mirrop, S., 1985. tRNA–rRNA sequence omologies: evidence from an ancient modular format by tRNAs and rRNAs. BioSystems 17, 209–225.

Branciamore, S., Di Giulio, M., 2011. The presence in tRNA molecule sequences of the double hairpin, an evolutionary stage through which the origin of this molecule is thought to have passed. J. Mol. Evol. 72, 352–363.

Branciamore, S., Gogoshin, G., Di Giulio, M., Rodin, A.S., 2018. Intrinsic properties of tRNA molecules as deciphered via Bayesian network and distribution divergence analysis. Life 8 (1), 5.

de Duve, C., 1988. The second genetic code. Nature 33, 117–118.

Di Giulio, M., 1992. On the origin of the transfer RNA molecule. J. Theor. Biol. 159, 199–214.

Di Giulio, M., 1994. On the origin of protein synthesis: a speculative model based on hairpin RNA structures. J. Theor. Biol. 171, 303–308.

Di Giulio, M., 1995. Was it an ancient gene codifyng for a hairpin RNA that, by means of direct duplication, gave rise to the primitive tRNA molecule? J. Theor. Biol 177, 95–101.

Di Giulio, M., 1999. The non-monophyletic origin of tRNA molecule. J. Theor. Biol. 197, 403–414.

Di Giulio, M., 2004. The origin of the tRNA molecule: implications for the origin of protein synthesis. J. Theor. Biol. 226, 89–93.

Di Giulio, M., 2006a. The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). J. Theor. Biol. 240, 343–352.

Di Giulio, M., 2006b. Nanoarchaeum equitans is a living fossil. J. Theor. Biol. 242, 257–260.

Di Giulio, M., 2008a. Permuted tRNA genes of Cyanidioschyzon merolae, the origin of the tRNA molecule and the root of the Eukarya domain. J. Theor. Biol. 253, 587–592.

Di Giulio, M., 2008b. Split genes, ancestral genes. In: Tze-Fei Wong, J, Lazcano, Antonio (Eds.), Prebiotic Evolution and Astrobiology, Chap 13. Landes Bioscience Publisher, Texas, USA.

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

Di Giulio, M., 2009a. A comparison among the models proposed to explain the origin of the tRNA molecule: a synthesis. J. Mol. Evol. 69, 1–9.

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

Di Giulio, M., 2012. The origin of the tRNA molecule: independent data favor a specific model of its evolution. Biochimie 94, 1464–1466.

Di Giulio, M., 2013. A polyphyletic model for the origin of tRNAs has more support than a monophyletic model. J. Theor. Biol. 318, 124–128.

Doolittle, W.F., 1978. Genes in pieces: were they ever together? Nature 272, 581–582.

Eigen, M., Lindemann, B.F., Tietze, M., Winkler-Oswatitsch, R., Dress, A., Von Haeseler, A., 1989. How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 244, 673–679.

Giege, R., Sissler, M., Florentz, C., 1998. Universal rules and idiosyncratic features in tRNA identity. Nucleic Acid Res. 26, 5017–5035.

Gilbert, W., 1978. Why genes in pieces? Nature 271, 501.

Gilbert, W., de Souza, S.J., Long, M., 1997. Origin of genes. Proc. Natl. Acad. Sci. USA 94, 7698–7703.

Hopfield, J.J., 1978. Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence and kinetic proofreading. Proc. Natl. Acad. Sci. USA 75, 4334–4338.

Kim, Y., Kowiatek, B., Opron, K., Burton, Z., 2018. Type-II tRNAs and evolution of translation systems and the genetic code. Int. J. Mol. Sci. 19 (10), 3275.

Maizels, N., Weiner, A.M., 1994. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc. Natl. Acad. Sci. USA 91, 6729–6734.

Moller, W., Janssen, G.M.C., 1992. Statistical evidence for remnants of primordial code in the acceptor stem of prokaryotic transfer RNA. J. Mol. Evol. 34, 471–477.

Pak, D., Root-Bernstein, R., Burton, Z.F., 2017. tRNA structure and evolution and standardization to the three nucleotide genetic code. Transcription 8, 205–219.

Panaccio, C., 2004. Ockham on Concepts, (Ashgate Studies in Medieval Philosophy). Ashgate, Aldershot ISBN.

Randau, L., Munch, R., Hohn, M., Jahn, D., Soll, D., 2005. Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'- and 3'-halves. Nature 433, 537–541.

Root-Bernstein, R., Kim, Y., Sanjay, A., Burton, Z.F., 2016. tRNA evolution from the proto-tRNA minihelix world. Transcription 7, 153–163.

Schimmel, P., Giege, R., Morras, D., Yokoyama, S., 1993. An operational RNA code for amino acids and possible relationship to genetic code. Proc. Natl. Acad. Sci. USA 90, 8763–8768.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., Steinberg, S., 1998. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 26, 148–153.

Steinberg, S., Cedergren, R., 1994. Structural compensation in atypical mitochondrial tRNAs. Nat. Struct. Biol. 1, 507–510.

Tanaka, T., Kikuchi, Y., 2001. Origin of cloverleaf of transfer RNA—the double-hairpin model: implication for the role of tRNA intron and the long extra loop. Viva Origino 29, 134–142.

Widmann, J., Di Giulio, M., Yarus, M., Knight, R., 2005. tRNA creation by hairpin duplication. J. Mol. Evol. 61, 524–530.