# THE PHYLOGENY OF tRNA MOLECULES AND THE ORIGIN OF THE GENETIC CODE

MASSIMO DI GIULIO

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

(Received September 23, 1993)

**Abstract.** The evolutionary relationships between transfer RNA (tRNA) molecules are analyzed by parsimony algorithms. The position of the topologies expected on the basis of the hypotheses made to explain the origin of the genetic code, on the frequency distribution of all the possible tree topologies of the evolutionary relationships between tRNAs seems to lead to the following conclusion: The hypothesis (Wong, J. T., Proc. Natl. Acad. Sci. USA, 1975, 72: 1909–1912) that sees the genetic code as a map of the biosynthetic relationships between amino acids seems to occupy a statistically significant position on these frequency distributions, thus reflecting a significant part of the tRNA phylogeny.

## Introduction

The transfer RNA (tRNA) molecules must have been present in several of the events that led to the development of life on earth and, in particular, they must have witnessed the mechanism that led to the structuring of the genetic code. These molecules could therefore be a precious means for investigating the early events that led to the stabilization of the organization of the genetic code.

The coevolution hypothesis of the genetic code (Wong, 1975) is in fact based on a mechanism involving the ancestors of tRNAs. This hypothesis identifies in a tRNA-like molecule the device through which precursor amino acids are supposed to have conceded part of their codon domain to product amino acids metabolically derived from them. This concession mechanism is supposed to have defined the organization of the genetic code (Wong, 1975; Wong, 1981). Thus, if the current tRNA molecules still remember the mechanism that led to the definition of the genetic code then, according to the coevolution hypothesis, the tRNAs of the amino acids in a precursor-product relationship should be more similar to one another than the tRNAs specific for non biosynthetically correlated amino acids (Wong, 1975; Wong, 1988; Szathmary and Zintzaras, 1992; Di Giulio, 1992). Moreover, if the physicochemical properties of amino acids were very important in the structuring of the genetic code (Woese *et al.*, 1966; Epstein, 1966; Goldberg and Wittes, 1966; Allf-Steinberger, 1969; Jungck, 1978; Weber and Lacey, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a; Di Giulio, 1989b; Haig and Hurst, 1991; Szathmary and Zintzaras, 1992; Lacey *et al.*, 1992) it is to be expected that the tRNAs of similar amino acids should be more correlated than tRNAs specific for amino acids with very different physicochemical properties (Woese, 1967; Szathmary and Zintzaras, 1992; Di Giulio, 1992).

Analyses of the evolutionary relationships between tRNA molecules have esta-

*Origins of Life and Evolution of the Biosphere* **24**: 425–434, 1994.

blished that these molecules have undergone a considerable divergence (Holmquist *et al.*, 1973; Eigen *et al.*, 1989). Nevertheless, at least two works (Fitch and Upper, 1987; Szathmary and Zintzaras, 1992) report data that suggests that the tRNA molecules still contain phylogenetic information. Therefore, in the present paper, parsimony algorithms will be used to analyze whether or not the sequences of tRNAs still contain traces of the mechanism that led to the definition of the organization of the genetic code.

## Materials and Methods

The tRNA sequences or tRNA genes and their alignment have been taken from Sprinzl *et al.* (1989). The estimation of the ancestral sequence of the tRNA specific for a certain amino acid has been made using 19 tRNA sequences of Asp, 12 of Asn, 23 of Glu, 22 of Gln, 19 of Thr, 15 of Ile, 13 of Met, 18 of Val, 19 of Leu, 24 of Phe, 18 of Tyr, 14 of Ser, 18 of Trp, 24 of His and 14 of Lys. The Appendix shows the code name (Sprinzl *et al.*, 1989) of the sequences used in this analysis. Of the approximately 76 nucleotides forming a tRNA only 51 have been used because the other positions are practically invariant (Eigen *et al.*, 1989). These 51 nucleotide positions are given explicitly in the legend of Figure 1.

The polar requirement values of the twenty amino acids have been taken from Woese *et al.* (1966).

Three programs from the PHYLIP package (Felsenstein, 1991) have been used: (i) DNAPARS, which estimates the specific sequence of a certain node in a phylogenetic tree by means of parsimony rules; (ii) CONSENSE, which defines a single consensus topology starting from a set of tree topologies; (iii) UPGMA, which constructs a tree topology starting from a distances matrix. This program has been used as a simple clustering procedure.

The following options from the PAUP package (Swofford, 1993) have been used: (i) *Alltrees*, which uses parsimony rules to analyze all the possible tree topologies for a set of sequences and provides a frequency distribution between the length of a specific topology in nucleotide substitutions and the number of topologies with that given length; (ii) *Lengths* and *Fit Measures*, which make it possible to calculate the length of a specific tree topology in nucleotide substitutions.

A final method has also been used. Fisher (1950) showed that the quantity $-2\ln P_i$ (where $P_i$ is the probability deriving from the i-th significance test) follows an $\chi^2$ distribution with 2 degrees of freedom. Thus, the quantity $-2\sum_{i=1}^{k} \ln P_i$ follows an $\chi^2$ distribution with 2K degrees of freedom. This method makes it possible to join the probabilities deriving from independent significance tests in a single value.

```
tRNA-Asp   GCCCTGGAGTGTTNTKCATACGGSCCGGCCGAGA-GCGDTCGGCCGGGGCG
tRNA-Asn   GCCKCCGAGCTCTTTGAGCGNTCGGACCGADTGGTGBARTTVCCGGBGGCG
tRNA-Glu   GCCCYGTGGTCTT-CGGATATCKCCCGGMGGCGA-CVGATCBGTCGGGGHA
tRNA-Gln   DGTBCCRGGTGTT-TDCACNCAGGAATCCTKCGA-CGAATTCGYGGGACHA
tRNA-Thr   GCCTCTGGGCTCTNTGAGCGCCTGAATCAGGAGGTGTGATCRCCDGGGGCT
tRNA-Ile   GGGCCAKAGCTCTTTGAGCGCGCGGACSGCGAGGTRGSDTSCTMTGGCCCA
tRNA-Met   GSCGGGGGGCTCTTAGAGCGCWGGWATCCKGAGGTNCRACCGNYCCCGSCA
tRNA-Val   GGGTTNGRGTCTTNTTGRCATCTCCYGGAGGAGGTGGCACGNCCNAACCCA
tRNA-Leu   GCGGGGRGGCCGCNTAGGCGCCGGRTTCCGGTGTCGTGATCACYCCCCGCA
tRNA-Phe   GCCGSGAAGCTCTTGGAGCRBCVGAATCBGVNTGTCCCWTGGGTCGCGGCA
tRNA-Tyr   CCGGCGWAGCTCTTTGAGCGGCGGAATCCGCAGGTGCTATGGCWCGCCGGA
tRNA-Ser   GGCGRGAGGCCGT-YAGGCGTBAGAATCTVATGTSGBGATCGCTCYCGBCG
tRNA-Trp   GGGGCCGGGCBCTNTGVGCGTCTGAATCAGAAGGTGGGWTCTCCGGSCCCA
tRNA-His   GCCGABGAGBSTTNWASVCWCAGSAMTSCTGCRA-SCGTTCGGTGTTGGCC
tRNA-Lys   GVSCCGBAGCTCTCTGAGCACCTGRATCAGGGGGTGYGATCRBVBGGSBCG
```

Fig. 1. This shows the ancestral sequences of the tRNA genes built using the DNAPARS program. The nucleotide positions in order are as follows: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 16, 17, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50, 51, 59, 60, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72 and 73. These numbers identify, in accordance with the official nomenclature (see, for example, Sprinzl *et al.*, 1989) the nucleotide positions. All the abbreviations used are standard. See text for further information.

## Results

### 1. The building of the ancestral sequence of tRNAs specific for a given amino acid

In a preliminary analysis using 51 nucleotide positions of tRNA sequences (see Legend of Figure 1) from archaebacteria, eubacteria, eukaryotes, chloroplasts and mitochondria and analyzing the sequences with the programs DNAPARS and CONSENSE (Felsenstein, 1991), certain conclusions were reached. (i) The mitochondria tRNA sequences, having undergone several base insertion and/or deletion events and with positions of dubious alignment, reflect this in a disordered behavior of these sequences on the tree topologies. Therefore, the mitochondria tRNA sequences have been eliminated from the analysis. However, this does not mean that the mitochondria sequences do not posses phylogenetic information. (ii) It was noted that in several cases the tRNA sequences from archaebacteria are closer to those of eukaryotes. In other words, the data seems to favour the topology: ((archaebacteria,eukaryotes),eubacteria,chloroplasts). However, this cannot be confirmed with confidence by the present data. Nevertheless, in the following analysis it has been assumed that the tRNA sequences of archaebacteria separated later than those of eukaryotes and earlier than those of eubacteria and chloroplasts. This assumption is supported by previous analyses (Fitch and Upper, 1987).

Starting from a set of tRNA sequences specific for a given amino acid and belonging to archaebacteria, eukaryotes, eubacteria and chloroplasts, it was assumed that the final tree had thus to have the topology: ((archaebacteria, eukaryotes), eubacteria,chloroplasts). The analysis thus consisted of starting from 20–30 tRNA sequences and building a set of more parsimonious tree topologies using the DNAPARS program. In building the concensus tree from this set of trees, the CONSENSE program made it possible to check whether the consensus tree thus obtained had the required topology. If not, the sequences that did not fit the required characteristics were eliminated from the analysis. The DNAPARS program was then used to build another set of more parsimonious tree topologies. The CONSENSE program established whether this new consensus topology had the necessary requisites, and so forth. Analysis was interrupted when the consensus tree topology of the set of more parsimonious tree topologies fitted the imposed tree topology.

At this point the sequence of the node that joined the sequences of archaebacteria to those of eukaryotes was determined (Fitch and Upper, 1987) by the set of more parsimonious trees, and was assumed as the ancestral sequence specific for that given amino acid. If the node in question varied in the set of more parsimonious tree topologies, the consensus sequence was determined and assumed as the specific ancestral sequenced.

The result of this analysis, using the sequences shown in the Appendix for the final analysis, is shown in Figure 1.

## 2. The tree topologies associated to the various hypotheses made to explain the origin of the genetic code

The biosynthetic relationships between amino acids have been transformed into binary trees by introducing a simple metric. Wong's Figure 1 (1975), which shows these relationships, assumed that every arrow indicating a transformation between amino acids had a distance of 1 unit. The pairs of precursor-product amino acids Asp-Asn and Glu-Gln were assigned with a distance of 0.5 unit. This is the consequence of the considerations that Wong (1975) makes on the biosynthetic relationships between these pairs of amino acids. The distance between Val, Leu, Phe and Tyr and all the other amino acids has been assumed to be 10 units because these amino acids are not biosynthetically correlated to the other amino acids (Wong, 1975). The distances between Leu-Phe, Leu-Tyr, Val-Phe and Val-Tyr have been attributed with a value of 5 units since another map of the biosynthetic relationships between amino acids (Taylor and Coates, 1989; Figure 1) shows that these amino acids might be biosynthetically closer to one another than to the other amino acids used in the present analysis. The introduction of these distances makes it easy to count the steps separating any two amino acids and thus establish the distance between them. Various combinations of amino acids thus defined various distance matrices and these matrices represented the input for the UPGMA program (Felsenstein, 1991) producing the tree topologies shown in Table I. The twelve

TABLE I

This shows the tree topologies, expressed according to the New Hampshire notation (Swofford, 1993; user's manual pp. 143–146), that can be associated to the coevolution hypothesis. The probability is defined in the text. The abbreviations are standard. See text for further information

| Tree topology | Probability |
|---|---|
| ((((Asp, Asn), Lys), ((Thr, Met), Ile)), (Ser, Trp), (Val, Leu)) | 0.24 |
| ((((Asp, Asn), Lys), ((Thr, Met), Ile)), (Ser, Trp), (Phe, Tyr)) | 0.12 |
| ((((Asp, Asn), Lys), ((Thr, Met), Ile)), (Val, Leu), (Phe, Tyr)) | 0.11 |
| ((((Glu, Gln), His), ((Thr, Met), Ile)), (Ser, Trp), (Phe, Tyr)) | 0.0091 |
| ((((Glu, Gln), His), ((Thr, Met), Ile)), (Val, Leu), (Phe, Tyr)) | 0.015 |
| ((((Glu, Gln), His), ((Thr, Met), Ile)), (Val, Leu), (Ser, Trp)) | 0.064 |
| ((((Glu, Gln), His), ((Asp, Asn), Lys)), (Val, Leu), (Ser, Trp)) | 0.12 |
| ((((Glu, Gln), His), ((Asp, Asn), Lys)), (Phe, Tyr), (Ser, Trp)) | 0.014 |
| ((((Glu, Gln), His), ((Asp, Asn), Lys)), (Phe, Tyr), (Val, Leu)) | 0.044 |
| ((((Asp, Asn), Lys), ((Thr, Met), Ile)), ((Glu, Gln), His), (Phe, Tyr)) | 0.021 |
| ((((Asp, Asn), Lys), ((Thr, Met), Ile)), ((Glu, Gln), His), (Ser, Trp)) | 0.069 |
| ((((Asp, Asn), Lys), ((Thr, Met), Ile)), ((Glu, Gln), His), (Val, Leu)) | 0.062 |

topologies in Table I were compared with Wong's Figure 1 (1975). In no case was any discrepancy noted in the biosynthetic relationships between amino acids (Wong, 1975; Figure 1) and the tree topologies expressing these relationships (Table I). These tree topologies (Table I) were thus assumed as the topologies expected on the basis of the coevolution hypothesis (Wong, 1975).

There are some indications tending to favour the hypothesis that polarity was important in defining amino acid allocations in the genetic code (Woese *et al.*, 1966; Di Giulio, 1989b; Haig and Hurst, 1991). It can thus be reasonably assumed that the property that is best reflected in the genetic code is polarity. I have thus assumed for the distance between two amino acids the absolute value of the differences between their polarity values. From these distances it was possible to build matrices that were used as input data for the UPGMA program. The tree topologies thus obtained are shown in Table II. These represent the topologies expected on the basis of the physicochemical hypothesis (Sonneborn, 1965; Woese *et al.*, 1966; Lacey and Mullins, 1983).

Finally, as regards the tree topologies associated to the ambiguity reduction hypothesis (Fitch, 1966; Fitch and Upper, 1987), these were determined through a single codon differentiation pattern (Fitch and Upper, 1987). This pattern involves the columns of the genetic code for which there are indications of the fact that the physicochemical properties of amino acids are indeed correlated with these (Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a). Thus, starting from the undifferentiated codon NNN and by performing ambiguity reduction (Fitch and Upper, 1987; Di Giulio, 1992) and always giving preference to the nucleotide in the second codon position (code columns), the tree topologies reported in Table III

TABLE II

This shows the tree topologies, expressed according to the New Hampshire notation (Swofford, 1993; user's manual pp. 143–146), that can be associated to the physicochemical hypothesis. The probability is defined in the text. The abbreviations are standard. See text for further information

| Tree topology | Probability |
|---|---|
| (((((Asn, Lys), Asp), (Thr, Ser)), Val), (Ile, Leu), (Met, Trp)) | 0.55 |
| ((((Asn, Lys), Asp), (Ser, Thr)), ((Met, Trp), Tyr), (Ile, Phe)) | 0.73 |
| (((((Asn, Lys), Asp), Thr), ((Ile, Leu), Phe)), Val, (Met, Tyr)) | 0.76 |
| (((((Ser, Thr), (Gln, His)), Glu), (Ile, Phe)), Tyr, (Met, Trp)) | 0.40 |
| (((((Met, Tyr), Val), ((Ile, Leu), Phe)), Thr), Glu, (Gln, His)) | 0.072 |
| (((((Ser, Thr), (Gln, His)), Glu), Val), (Ile, Leu), (Met, Trp)) | 0.37 |
| (((((Gln, His), Ser), (Asn, Lys)), (Asp, Glu)), Val, (Trp, Leu)) | 0.028 |
| (((((Gln, His), Ser), (Asn, Lys)), (Asp, Glu)), Tyr, (Trp, Phe)) | 0.031 |
| ((((Gln, His), (Asn, Lys)), (Asp, Glu), (Val, Tyr), (Leu, Phe)) | 0.032 |
| (((((Gln, His), (Asn, Lys)), (Asp, Glu)), Thr), (Tyr, Met), (Ile, Phe)) | 0.014 |
| (((((Gln, His), (Ser, Thr)), (Asn, Lys)), (Asp, Glu)), Ile, (Met, Trp)) | 0.053 |
| (((((Asn, Lys), (Gln, His)), (Asp, Glu)), Thr), (Met, Val), (Ile, Leu)) | 0.062 |

TABLE III

This shows the tree topologies, expressed according to the New Hampshire notation (Swofford, 1993; user's manual pp. 143–146), that can be associated to the ambiguity reduction hypothesis. The probability is defined in the text. The abbreviations are standard. See text for further information

| Tree topology | Probability |
|---|---|
| ((((((Asn, Lys), Asp), Trp), (Ser, Leu)), Thr), Val, (Ile, Met)) | 0.18 |
| (((((Asn, Lys), Asp), (Tyr, Trp)), (Ser, Phe)), Thr, (Ile, Met)) | 0.61 |
| ((((((Asn, Lys), Asp), Tyr), (Leu, Phe)), Thr), Val, (Ile, Met)) | 0.85 |
| ((((((His, Gln), Tyr), Trp), Glu), (Phe, Ser)), Thr, (Ile, Met)) | 0.32 |
| ((((((Gln, His), Tyr), Glu), (Phe, Leu)), Thr), Val (Ile, Met)) | 0.44 |
| ((((((Gln, His), Trp), Glu), (Ser, Leu)), Thr), Val, (Ile, Met)) | 0.11 |
| ((((Gln, His), Trp), ((Asn, Lys), (Asp, Glu))), Val, (Ser, Leu)) | 0.0062 |
| (((((Asn, Lys), (Asp, Glu)), (Ser, Phe)), Trp), Tyr, (Gln, His)) | 0.043 |
| ((((Asn, Lys), (Asp, Glu)), ((Gln, His), Tyr)), Val, (Phe, Leu)) | 0.14 |
| (((((Asp, Glu), (Asn, Lys)), ((Gln, His), Tyr)), Phe), Thr, (Ile, Met)) | 0.030 |
| (((((Asp, Glu), (Asn, Lys)), ((Gln, His), Trp)), Ser), Thr, (Ile, Met)) | 0.053 |
| ((((((Asp, Glu), (Asn, Lys)), (Gln, His)), Leu), Thr), Val, (Ile, Met)) | 0.082 |

are obtained. These were considered as the main topologies expected on the basis of the ambiguity reduction hypothesis (Fitch, 1966; Fitch and Upper, 1987).

## 3. Probabilities associated to the various hypotheses made to explain the origin of the genetic code

The *Alltrees* options of the PAUP package of programs (Swofford, 1993) makes it possible to examine all the possible tree topologies from a set of sequences and to build a frequency distribution between tree length and the number of trees with

a certain length. In the analysis, I have used 10 and 11 sequences (Tables I, II, and III) for a total number of examined trees of 2,027,025 and 34,459,425 respectively. This frequency distribution was built for each of the sequence combinations (see, for instance, Table I). Then by using the *Lengths* and *Fit Measures* option (Swofford, 1993), the length (in nucleotide substitutions) was evaluated for all the tree topologies reported in Tables I, II, and III. I then calculated the probability with which a specific tree topology is obtained, by seeing at which point in the distribution a given tree length is located. In other words, by summing all the topologies with a length equal to or less than the given one and dividing this number by that of all the possible tree topologies of 10 or 11 sequences. Thus the method reported by Fitch and Upper (1987) was used. These probabilities are shown in Tables I, II, and III alongside their respective topology. [If we analyze the topologies that are equally expected on the basis of coevolution hypothesis, in which the position of Ile is exchanged for that of Met, five of the nine probabilities in which Ile and Met are involved (Table I) show lower values than the ones indicated in Table I. Whereas, the remaining four probability values have the same value as the one indicated in Table I. Therefore, if we consider these topologies, which are equally probable, the analysis is even more favourable towards the coevolution hypothesis (see Discussion)].

The various combinations of sequences used to determine the probabilities (Tables I, II, or III) should give statistical information regarding the significance that would be obtained in the ideal (infeasible) analysis in which all the sequences shown in Figure 1 were simultaneously used. As can be seen (Tables I, II, and III), the probabilities are three by three dependent, in that groups of three of these probabilities deriving from topologies that differ only in two sequences seem to be of the same order of magnitude and it is, thus, the mean of these three values that gives a measure of how the various topologies are statistically significant. However, it can be noted that between groups of topologies that differ in at least three sequences the probabilities are mainly independent and Fisher's method (1950) can thus be used. By calculating the mean of the probabilities in a group of three topologies that differ only in two sequences and using the resulting probability values in the Fisher test (1950), the following results are obtained. For the coevolution hypothesis (Table I) a value of $\chi^2 = 22.4$ ($P<0.01$, df=8) is obtained; for the physicochemical hypothesis (Table II) a value of $\chi^2 = 16.6$ ($0.02<P<0.05$, df=8) is obtained; and finally for the ambiguity reduction hypothesis (Table III) a value of $\chi^2 = 15.0$ ($0.05<P<0.10$, df=8) is obtained.

## Discussion

In general, the results presented in this paper (Tables I, II, and III; see Results) seem to confirm previous analyses (Fitch and Upper, 1987; Szathmary and Zintzaras, 1992) which established that the tRNA sequences still contain sufficient phylogenetic information. In particular, the probabilities (Table I) that can be associated to

the coevolution hypothesis (Wong, 1975) are significant, apart from the value 0.24, at significance levels of about 10% (5 out of 12 are statistically significant at least at the 5% level). This, together with the value of $\chi^2 = 22.4$ ($P<0.01$, df=8; see Results), which joins the probabilities in a single value, are a clear indication that the mechanism (Wong, 1975; Wong, 1981; Di Giulio, 1993) proposed by the coevolution hypothesis might have been the main cause of the origin of genetic code organization.

The probabilities that can be associated to the physicochemical hypothesis (Sonneborn, 1965; Woese *et al.*, 1966; Lacey and Mullins, 1983) seem to correlate with tRNA phylogeny (Table II; $\chi^2 = 16.6$, $0.02<P<0.05$, df=8). This seems to confirm other types of analysis that linked the physicochemical properties of amino acids to the organization of the genetic code (Woese *et al.*, 1966; Epstein, 1966; Goldberg and Wittes, 1966; Allf-Steinberger, 1969; Jungck, 1978; Weber and Lacey, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a; Di Giulio, 1989b; Haig and Hurst, 1991; Szathmary and Zintzaras, 1992; Lacey *et al.*, 1992). However, the observed relationship between polarity distances and tRNA phylogeny (Table II; see Results) seems to be partly due to the fact that the physicochemical properties of amino acids are conserved between the amino acids in precursor-product relationship (Di Giulio, 1991; Di Giulio, 1992) more than to the hypothesis that assumes that similar amino acids effectively used similar tRNAs. This can also be partly seen in the probability values in Tables I and II; the correlation coefficient between these values is, in fact, close to significance ($r = +0.54$, $0.05<P<0.10$, $n = 12$, $df = 10$). In this sense, therefore, the coevolution and the physicochemical hypotheses would lead to predictions that are partly overlapping and thus not easy to differentiate (Di Giulio, 1991). Szathmary and Zintzaras (1992) did not, however, find a significant correlation between the biosynthetic distances between amino acids and the polarity distances. It is thus also possible that the observed correlation between the phylogeny of tRNAs and the polarity distances, on one hand, and between the phylogeny of tRNAs and the precursor-product relationships, on the other, may be due to two independent mechanisms that led to the definition of the organization of the genetic code. However, if we compare the coevolution and the physicochemical hypotheses on the basis of these probability values (Tables I and II), we get $\chi^2$ values that are clearly in favour of the coevolution mechanism ($\chi^2 = 72.5$, $P<<0.001$, $df = 24$ versus $\chi^2 = 52.4$, $P\simeq0.001$, $df = 24$). This need not imply that the amino acid properties were not important in structuring the genetic code but, rather, that they probably acted together with the coevolutionary transfer of codons from the precursor to the product amino acid (Wong, 1988).

As regards the ambiguity reduction hytpothesis (Fitch, 1966; Fitch and Upper, 1987), although this presents a highly significant probability value (Table III), it is the hypothesis that least seems to reflect the phylogeny of tRNAs (Table III, $\chi^2 = 15.0$, $0.05<P<0.10$, df=8). This does not seem to confirm, at least for the codon differentiation pattern used here, Fitch and Upper's conclusion (1987) that the tRNA phylogeny reflects the mechanism of the ambiguity reduction hypothesis.

The conclusion that seems to emerge from this paper is that the mechanism (Wong, 1975; Wong, 1981; Di Giulio, 1993) proposed by the coevolution hypothesis and which seems to have defined the organization of the genetic code, is still detectable in the phylogeny of tRNA molecules. The traces of this mechanism seem to have also been found in an equivalent analysis conducted on the trees of aminoacyl-tRNA synthetases (Di Giulio, 1992).

## Acknowledgements

I would like to thank one reviewer for his discussion and Prof. Antonio Lazcano for encouragement. I would also like to thank Marco Valenzi for computer help.

## References

Allf-Steinberger, C.: 1969, *Proc. Natl. Acad. Sci. USA* **64**, 584–591.
Di Giulio, M.: 1989a, *J. Mol. Evol.* **29**, 191–201.
Di Giulio, M.: 1989b, *J. Mol. Evol.* **29**, 288–293.
Di Giulio, M.: 1991, *Z. Naturforsch.* **46C**, 305–312.
Di Giulio, M.: 1992, *Origins Life Evol. Biosphere* **22**, 309–319.
Di Giulio, M.: 1993, *J. Mol. Evol.* **37**, 5–10.
Eigen, M., Lindemann, B. F., Tietze, M., Winkler-Oswatitsch, R., Dress, A., and von Haeseler, A.: 1989, *Science* **244**, 673–679.
Epstein, C. J.: 1966, *Nature* **210**, 25–28.
Felsenstein, J.: 1991, PHYLIP(Phylogeny Inference Package) version 3.4, University of Washington, Seattle.
Fisher, R. A.: 1950, in Statistical Methods for Research Workers (Oliver and Boyd, Edinburgh and London) 11th ed., p. 99.
Fitch, W. M.: 1966, *J. Mol. Biol.* **16**, 1–12.
Fitch, W. M. and Upper, K.: 1987, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 759–767.
Goldberg, A. L. and Wittes, R. E.: 1966, *Science* **153**, 420–424.
Haig, D. and Hurst, L. D.: 1991, *J. Mol. Evol.* **33**, 412–417.
Holmquist, R., Jukes, T. H., and Pangburn, S.: 1973, *J. Mol. Biol.* **78**, 91–116.
Jungck, J. R.: 1978, *J. Mol. Evol.* **11**, 211–224.
Lacey, J. C. Jr. and Mullins, D. W. Jr.: 1983, *Origins Life* **13**, 3–42.
Lacey, J. C. Jr., Wickramasingle, N. S. M. D., and Cook, G. W.: 1992, *Origins Life Evol. Biosphere* **22**, 243–275.
Sjostrom, M. and Wold, S.: 1985, *J. Mol. Evol.* **22**, 272–277.
Sonnerborn, T. M.: 1965, In: Evolving genes and proteins Bryson V. & Vogel H. J. (eds.), Academic Press, New York, pp. 377–397.
Sprinzl, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R.: 1989, *Nucl. Acids. Res.* **17**, r1–r156.
Swofford, D. L.: 1993, PAUP, Phylogenetic Analysis Using Parsimony, version 3.1.1, Laboratory of Molecular Systematics, Smithsonian Institution.
Szathmary, E. and Zintzaras, E.: 1992, *J. Mol. Evol.* **35**, 185–189.
Taylor, F. J. R. and Coates, D.: 1989, *BioSystems* **22**, 177–187.
Weber, A. L. and Lacey, J. C. Jr.: 1978, *J. Mol. Evol.* **11**, 199–210.
Woese, C. R.: 1967, In: The genetic code. Harper & Row, New York, pp. 179–195.
Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M., and Saxinger, W. C.: 1966, *Cold Spring Harbor Symp. Quant. Biol.* **31**, 723–736.
Wolfenden, R. V., Cullis, P. M., and Southgate, C. C. B.: 1979, *Science* **206**, 575–577.
Wong, J. T.: 1975, *Proc. Natl. Acad. Sci. USA* **73**, 1909–1912.
Wong, J. T.: 1981, *Trends Biochem. Sci.* **6**, 33–36.
Wong, J. T.: 1988, *Microbiol. Sci.* **5**, 174–181.

## Appendix

The following sequences of tRNAs of genes of tRNAs are from Sprinzl *et al.* (1989); The letter identifies the amino acid and the number the organism (in some cases the anticodon is specified):

K145, K148, K200, K304, K310, K335, K355, K360, K700, K770, K780, K870, K951, K955; H110, H120, H145, H235, H236, H250, H255, H258, H259, H310, H320, H333, H335, H340, H348, H355, H365, H570, H575, H780, H950, H640, H980, H995; W115, W120, W200, W201, W250, W304, W310, W315, W320, W335, W342, W355, W555, W570, W700, W870, W625, W970; S110, S121, S160, S203, S206, S236, S310, S313, S320, S355, S566, S570, S575, S590; Y145, Y148, Y200, Y235, Y250, Y260, Y230, Y304, Y310, Y315, Y335, Y355, Y365, Y375, Y615, Y655, Y625, Y640; F120, F160, F203, F236, F250, F280, F304, F310, F320, F335, F375, F350, F365, F560, F575, F590, F780, F830, F955, F620, F625, F630, F970, F650; L160, L190, L191, L235, L250, L255, L259, L304, L311, L320, L376, L378. L560, L645, L700, L780, L830, L955, L996; V110, V111, V112, V120, V235, V250, V252, V310, V327, V335, V355, V360, V365, V570, V780, V781, V830, V995; M180, M250, M310, M313, M315, M320, M324, M327, M335, M365, M635, M780, M625; I120, I121, I145, I235, I236, I320, I327, I335, I366, I570, I590, I780, I950, I580, I640; T145, T146, T148, T200, T203, T252, T253, T260, T304, T311, T313, T320, T335, T336, T365, T366, T780, T995, T970; Q110, Q120, Q145, Q200, Q235, Q250UUG, Q251CUG, Q250CUG, Q251UUG, Q304, Q310, Q335, Q313, Q530, Q995, Q530CUA, Q531, Q532, Q950, Q951, Q955, Q970; E120, E121, E145, E295, E304, E310, E315, E335, E355, E365, E375, E570, E571, E575, E770, E780, E781, E782, E950, E955, E956, E957, E995; N110, N120, N235, N236, N320, N335, N355, N370, N780, N830, N995, N996; D120, D145, D148, D203, D206, D304, D310, D315, D335, D355, D365, D570, D700, D780, D950, D955, D590, D830, D956.