THE PHYLOGENY OF tRNAs SEEMS TO CONFIRM THE PREDICTIONS OF THE COEVOLUTION THEORY OF 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 August 5, 1994)

Abstract. An extensive analysis of the evolutionary relationships existing between transfer RNAs, performed using parsimony algorithms, is presented. After building up an estimate of the tRNA ancestral sequences, these sequences are then compared using certain methods. The results seem to suggest that the coevolution hypothesis (Wong, J.T., 1975, Proc. Natl. Acad. Sci. USA 72, 1909–1912) that sees the genetic code as a map of the biosynthetic relationships between amino acids is further supported by these results, as compared to the hypotheses that see the physicochemical properties of amino acids as the main adaptative theme that led to the structuring of the genetic code.

1. Introduction

The hypotheses proposed to explain the origin of genetic code organization are highly diversified as far as the selective pressures on which they are based are concerned, even if they may lead to equivalent predictions. Sonneborn (1965) suggests that the allocations of amino acids in the genetic code were determined by a selective pressure tending to reduce the deleterious effects of mutations. Whereas, Woese *et al.* (1966) propose that the selective pressure tended to reduce the translation errors of the primitive genetic message. The ambiguity reduction hypothesis (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987) is not essentially very different from these hypotheses. This hypothesis suggests that groups of similar codons were initially assigned to chemically correlated groups of amino acids and that the genetic code thus underwent a process tending to reduce this ambiguity (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987). Lacey and Mullins (1983), on the other hand, discuss the correlation between the properties of amino acids and the properties of the anticodonic nucleotides, thereby proposing the anticodon hypothesis (Lacey *et al.*, 1992). All the above hypotheses predict that the physicochemical properties of amino acids must be related to the organization of the genetic code.

Other hypotheses predict the existence of a stereochemical relationship between the amino acid and the corresponding codons (or anticodons). According to these hypotheses, such interactions lay the foundations for the molecular bases of the genetic code. For instance, Shimizu (1982) proposed that a complex formed of four bases, three of which belonging to the anticodon, has a lock and key relationship with the corresponding amino acid. There are a number of models based on

Origins of Life and Evolution of the Biosphere **25:** 549–564, 1995.

these stereochemical relationships (for example, see: Balasubramanian *et al.*, 1980; Hendry *et al.*, 1981). Crick (1968) refuted the existence of any stereochemical relationship between the amino acid and the corresponding codons (or anticodons) and proposed the frozen accident theory. His evolutionary scheme (Crick, 1968) predicts that, starting from a small number of amino acids, the genetic code expanded its vocabulary and was subsequently frozen at a certain stage of development, although it nevertheless ensured that similar amino acids had codons that are in some way correlated. Finally, a strictly evolutionary hypothesis was introduced by Wong (1975). This hypothesis (Wong, 1975) suggests that the structure of the codon system is primarily an imprint of the prebiotic pathways of amino acid formation. Consequently, the origin of the genetic code could be elucidated on the basis of the precursor-product relationships between amino acids defined through their biosynthesis (Wong, 1975).

The molecules that could contain vestiges of the mechanism that led to the structuring of the genetic code are aminoacyl-tRNA synthetases and transfer RNAs (tRNAs). Of these two classes of molecules, the tRNAs should better reflect the mechanism that defined the genetic code as they establish the fundamental link between the language of nucleic acids and that of proteins. Furthermore, in light of the RNA world hypothesis (Joyce, 1989), tRNA-like molecules might have been primarily involved in several of the events that gave rise to the origin of life on earth (Weiner and Maizels, 1987) and they might have mediated the first contacts between these two languages.

Phylogenetic analyses of the evolutionary relationships between tRNAs have nevertheless shown that these molecules have undergone a considerable divergence (Holmquist *et al.*, 1973; Eigen *et al.*, 1989). However, there are some indications (Cedergren *et al.*, 1980; Fitch and Upper, 1987; Szathmary and Zintzaras, 1992; Di Giulio, 1994) that suggest that tRNAs still contain detectable phylogenetic information. The present paper presents an extensive analysis of the evolutionary relationships existing between tRNAs, performed using parsimony algorithms. This analysis will enable a comparison to be made between the various hypotheses proposed to explain the origin of the genetic code organization.

2. Materials and Methods

The tRNA sequences or the tRNA genes and their alignment have been taken from the data base of Sprinzl *et al.* (1991). The appendix shows the code names (Sprinzl *et al.*, 1991) of all the sequences used in the present analysis. Of the 99 nucleotide sites reported by Sprinzl *et al.* (1991) that form a tRNA, 96 have been used, eliminating the common sequence CCA at the 3′ end from the analysis. Furthermore, the tRNA sequences containing several modified nucleotides have been transformed into DNA sequences through the modified nucleotide table reported by Sprinzl *et*

al. (1991), i.e. substituting the modified nucleotide symbol with the corresponding main nucleotide.

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

The UPGMA program comes from the packet of programs in PHYLIP (Felsenstein, 1991). This program was used with a simple clustering procedure.

The following options contained in PAUP (Swofford, 1993) were used: (i) for the estimation of the ancestral sequences of tRNAs obtained through use of parsimony rules, a *Heuristic Search* was used with branch swapping given by tree bisection-reconnection (=TBR); other options in effect were: (a) simple sequence addition, (b) MULPARS and (c) steepest descent; (ii) the sequences of the relevant nodes have been obtained through use of the option: possible character-state assignments to internal nodes; the character-state optimization has been obtained through accelerated transformation (ACCTRAN); (iii) all the possible tree topologies for a set of sequences have been analyzed through *Exhaustive Search.* The *Alltrees* search constructs a frequency distribution between the length, in nucleotide substitutions, of a specific tree topology and the number of all the possible topologies with that given length; (iv) *Lengths and Fit Measures* calculates the length, in nucleotide substitutions, of a specific tree topology; (v) *Random Trees* makes it possible, starting from a set of sequences, to estimate the mean length, in nucleotide substitutions, of a tree population and the relative standard deviation, by generating a sample of random trees.

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

3. Results

3.1. The construction of tRNA ancestral sequences

The tRNA ancestral sequence specific for a given amino acid has been determined using tRNA sequences coming from archaebacteria, eukaryotes, eubacteria and chloroplasts. A total of 1129 sequences were used and the number varies from a minimum of 23 for Cys to a maximum of 103 for Leu (see Appendix).

Starting from a set of tRNA sequences specific for a given amino acid, the following general topological constraint was enforced on tree topologies: ((archaebacteria, eukaryotes), eubacteria, chloroplasts). Then, by using the *Heuristic Search* technique (Swofford, 1993) and setting the options as referred in Materials and Methods above, a more parsimonious set of trees was produced. This set of trees

```
                1111122222222233333344444444555566666667777
     12345679123770023467890127890123456901793456789012 3
Asp  GSCCCSGATGTTNCTKCARCCTNCCACGSAGGAGAGCGGRCGKCSGGGSCG
Asn  TCCTCMGACTCTTTNGAGGGTCGGCAACCGATTGGGYAGATRCYKGGGGAG
Lys  GRGTTGGACTCTATTGAGRTCTGNCAATCAGWVGGGSGGGCGBCCAGCYCR
Thr  GCCCKTTACTCTNTNGAGACCTCCTAAGGAGGAGGGYGGASRYAATSGGCT
Ile  GGGCTATACTCTNTGGAGGCACSSCAASGGKGAGGCCKAAMGGATGGCCCA
Met  GGCGGGGACGCTTTNGMGGYYGGGCAACCCGGAGGGAGNACTCCCCCGCYA
Glu  GCCCCCTGTCTCCCTGGAACCGCCCACGGCGGTAAMCGGACGKAGGGGGCA
Gln  TGGSCYGAYGTTNNAWCRGWCGGGTGWYCCGTCAWCGAGATCGCRGSVCAG
Pro  SGGGATGGCGYTTCNKCGDYTKGMTGAKCMAGAGGSCRNAYGKCATCCCSA
Arg  GGGCCCGACYYTTTTGRGACTGGCCAAGCCAGGGRGGGGACNCCGGGSCCG
Ala  GGGSSTAACTCTTNTGAGGCTTCCTAAGGAAGAGGATCGGGSTTACSTCCA
Val  GGGTCCGAYTYTNTCKARGTCTCCNACGGAGAAGGGSCGAGGTCGGACCCA
Leu  GCGGGTGGCSGYTMAASGGCMRGACGATCYKGTNGGTGGRCTCCRCCCGCA
Phe  GCCGGGAACTCTTNNGAGRGWGGACRATCCWCGTGSSCAWGSSTCYCGGCA
Tyr  SSGCCGAACYSTTTARRGGGCGGACRATCCGCTKKGCTGAGGCTCGGCSSA
Gly  GCGSSKGATKTTTTTGMAGTGAGCTAAGCTCAMGASCGGTCGKCMSSCGCW
Ser  GGAGRGAGCCGCTTGAGGGCYGGTCAAACCRGTAGGGGNACTCTCYCTCCG
Cys  GSCGGCAGCCATCNNASGAGCGGACAATCCGTTATCCCNAGGGTGYCGCCT
Trp  GGGKCCGAYTCCCTTGARGMCKGTCAAWCMGTRKGGWGATYTCCGGSCCYR
His  GCSGAGGAYSWTNTNASRACARGATGCTCYTGCAWGCGGTCGKYCTYSGCC
```

Fig. 1. The ancestral sequences of the tRNA genes reconstructed through the use of parsimony rules are shown. Only the non-invariant positions and the phylogenetically informative sites are shown. The vertical numbers at the top of the figure identify the nucleotide positions in accordance with the official nomenclature (see, for example, Sprinzl *et al.* (1989)). The two positions with the number 17 indicate 17 and 17A in order and the positions with the number 20 indicate 20A and 20B. All the abbreviations used are standard (Nomenclature Committee of the International Union of Biochemistry, 1986). See text for further information.

has been reduced to a single consensus tree, the strict consensus tree, by means of the *Consensus Tree Options* (Swofford, 1993). The *Tree Description Options* (Swofford, 1993) are then used to define the possible character-state assignments to internal nodes of the strict consensus tree by using the accelerated transformation (ACCTRAN) as a method to optimize the character state. All the above manages to estimate the sequences of the two nodes which link, in the first case, the archaebacteria sequences to those of the eukaryotes and, in the second case, the eubacteria sequences to those of the chloroplasts. The sequences of these two nodes are then used to determine the tRNA ancestral sequence, defining it through the two simple rules given below. Let $\{A\}$ and $\{B\}$ be the sets of nucleotides that are found in homologous sites in the two sequences of the nodes being investigated. If $\{A\} \cap \{B\} = \varnothing$ ($\cap$ = intersection; $\varnothing$ = empty set) then the nucleotide site in the ancestral sequence will be given by $\{A\} \cup \{B\}$ ($\cup$ = union); whereas, if $\{A\} \cap \{B\} = \{C\}$ then the nucleotide site in the ancestral sequence will be given by $\{C\}$. In other words, the consensus sequence between the two nodal sequences

TABLE I

This shows the tree topologies, expressed using the New Hampshire notation (Swofford, 1993; user's manual pp. 143–146), which can be associated to the genetic code coevolution hypothesis. The probability is defined in the text. The probabilities in brackets refer to tree topologies (data not shown) constructed using the metric based on the number of enzymatic steps. All the abbreviations used are standard. See text for further information.

Tree topology	Probability	
((((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((Glu, Gln), Pro)), Leu), Ala, Val)	0.28	(0.69)
(((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((Glu, Gln), Pro)), Phe, Tyr)	0.093	(0.047)
((((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((Glu, Gln), Pro)), Trp), Ser, Cys)	0.066	(0.047)
(((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((Ala, Val), Leu)), Phe, Tyr)	0.11	(0.33)
((((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((Ala, Val), Leu)), Trp), Ser, Cys)	0.056	(0.15)
(((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((Ser, Cys), Trp)), Phe, Tyr)	0.019	(0.019)
(((((Asp, Asn), Lys), ((Glu, Gln), Pro)), ((Ala, Val), Leu)), Phe, Tyr)	0.096	(0.35)
((((((Asp, Asn), Lys), ((Glu, Gln), Pro)), ((Ala, Val), Leu)), Trp), Ser, Cys)	0.054	(0.35)
(((((Asp, Asn), Lys), ((Glu, Gln), Pro)), ((Ser, Cys), Trp)), Phe, Tyr)	0.050	(0.068)
(((((Asp, Asn), Lys), ((Ala, Val), Leu)), ((Ser, Cys), Trp)), Phe, Tyr)	0.029	(0.18)
(((((Thr, Ile), Met), ((Glu, Gln), Pro)), ((Ala, Val), Leu)), Phe, Tyr)	0.049	(0.049)
((((((Thr, Ile), Met), ((Glu, Gln), Pro)), ((Ala, Val), Leu)), Trp), Ser, Cys)	0.067	(0.067)
(((((Thr, Ile), Met), ((Glu, Gln), Pro)), ((Ser, Cys), Trp)), Phe, Tyr)	0.043	(0.043)
(((((Ala, Val), Leu), ((Ser, Cys), Trp)), ((Thr, Ile), Met)), Phe, Tyr)	0.035	(0.035)
(((((Ala, Val), Leu), ((Ser, Cys), Trp)), ((Glu, Gln), Pro)), Phe, Tyr)	0.038	(0.038)

is calculated. The sequence thus defined is expected to provide an estimation of the tRNA ancestral sequence specific for a given amino acid.

The result of this analysis is shown in Figure 1.

3.2. THE TREE TOPOLOGIES EXPECTED ON THE BASIS OF THE VARIOUS HYPOTHESES PROPOSED IN ORDER TO EXPLAIN THE ORIGIN OF THE GENETIC CODE

The biosynthetic relationships between amino acids have been transformed into binary trees on the basis of the following qualitative arguments. (i) First of all, Wong's Figure 1 (1975) and the relative considerations on the biosynthetic relationships between amino acids reported in that work define the general relationships between groups of amino acids; (ii) Taylor and Coates' Figure 1 (1989) has been used to decide, on the basis of the biosynthetic steps, the order that the amino acids in strict biosynthetic relationships have to assume in the tree topology; (iii) the structure of the genetic code has been used as a general guideline in order to solve certain ambiguities deriving, for example, from determining the relative positions of the groups of amino acids in the tree topologies. The three above points make

TABLE II

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

Tree topology	Probability
(((((Asp, Glu), ((Asn, Lys), Gln)), ((Thr, Pro), Ala)), (Ile, Leu)), Met, Val)	0.18
((((((Asn, Lys), Gln), (Asp, Glu)), (Thr, Pro)), (Ile, Phe)), Met, Tyr)	0.13
(((((Asp, Glu), ((Asn, Lys), Gln)), ((Thr, Pro), Ser)), (Ile, Cys)), Met, Trp)	0.12
((((((Asn, Lys), Asp), (Thr, Ala)), ((Ile, Leu), Phe)), Val), Met, Tyr)	0.56
((((((Asn, Lys), Asp), ((Thr, Ala), Ser)), ((Ile, Leu), Cys)), Val), Met, Trp)	0.46
((((((Asn, Lys), Asp), (Ser, Thr)), ((Ile, Phe), Cys)), Trp), Met, Tyr)	0.54
(((((Asp, Glu), ((Asn, Lys), Gln)), (Pro, Ala)), (Val, Tyr)), Phe, Leu)	0.073
((((((Asp, Glu), ((Asn, Lys), Gln)), ((Pro, Ala), Ser)), Val), Trp), Leu, Cys)	0.054
(((((Asp, Glu), ((Asn, Lys), Gln)), (Ser, Pro)), (Phe, Cys)), Tyr, Trp)	0.025
((((((Asn, Lys), Asp), (Ser, Ala)), ((Val, Tyr), Trp)), Cys), Leu, Phe)	0.57
(((((Tyr, Met), Val), ((Ile, Leu), Phe)), ((Thr, Pro), Ala)), Glu, Gln)	0.78
(((((((Thr, Pro), Ala), Ser), Gln), Glu), ((Ile, Leu), Cys)), Val), Met, Trp)	0.68
(((((((Thr, Pro), Ser), Gln), Glu), ((Ile, Phe), Cys)), Trp), Met, Tyr)	0.87
((((((Thr, Ala), Ser), (((Ile, Leu), Phe), Cys)), Val), Trp), Met, Tyr)	0.57
(((((((Pro, Ala), Ser), Gln), Glu), ((Val, Tyr), Trp)), Cys), Leu, Phe)	0.61

it easy to determine the tree topologies that can be associated to the genetic code coevolution hypothesis (Wong, 1975). These tree topologies are reported in Table I. A comparison between these topologies (Table I) and, for example, Wong's Figure 1 (1975) (see also Figure 1 in Taylor and Coates (1989)) immediately shows the straightforward criteria used to construct these topologies. In order to define the tree topologies of the coevolution hypothesis, I have also used a convenient metric based on the number of enzymatic steps of the biosynthetic pathways between amino acids (Szathmary and Zintzaras, 1992). The number of biosynthetic steps (Taylor and Coates, 1989, Figure 1) has been taken as the distance between any two amino acids. It is thus easy to define distance matrices between amino acids, and such matrices represent the input to the UPGMA program (Felsenstein, 1991) which has produced tree topologies (data not shown) analogous to those reported in Table I. It must be pointed out, however, that in numerous cases these topologies are very different from the predictions of the coevolution hypothesis (Wong, 1975).

There are a number of indications tending to favor the hypothesis that the polarity distances were extremely important in defining the allocations of the amino acids in the genetic code (Woese *et al.*, 1966; Di Giulio, 1989a; Haig and Hurst, 1991;

TABLE III

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

Tree topology	Probability
((((((Ala, Thr), Pro), (((Ile, Met), Val), Leu)), Gln), (Asn, Lys)), Asp, Glu)	0.23
((((((Ile, Met), Phe), (Thr, Pro)), (Gln, Tyr)), (Asn, Lys)), Asp, Glu)	0.22
(((((Asp, Glu), (Asn, Lys)), Gln), ((Ile, Met), ((Ser, Pro), Thr))), Cys, Trp)	0.20
(((((Asn, Lys), Asp), Tyr), (((Ile, Met), Val), (Phe, Leu))), Thr, Ala)	0.33
(((((((Ile, Met), Val), Leu), ((Thr, Ala), Ser)), (Cys, Trp)), Asp), Asn, Lys)	0.39
(((((((Ile, Met), Phe), (Ser, Thr)), (Cys, Trp)), Tyr), Asp), Asn, Lys)	0.19
(((((Asp, Glu), (Asn, Lys)), (Gln, Tyr)), ((Leu, Phe), Val)), Ala, Pro)	0.096
(((((((Ser, Pro), Ala), (Val, Leu)), (Trp, Cys)), Gln), (Asn, Lys)), Asp, Glu)	0.25
((((((Ser, Pro), Phe), (Cys, Trp)), (Gln, Tyr)), (Asn, Lys)), Asp, Glu)	0.092
(((((((Asn, Lys), Asp), Tyr), (Cys, Trp)), (Ala, Ser)), Val), Leu, Phe)	0.57
((((((Ile, Met), Val), (Leu, Phe)), ((Ala, Thr), Pro)), Glu), Gln, Tyr)	0.51
((((((Glu, Gln), (Cys, Trp)), ((Thr, Ala), (Ser, Pro))), Leu), Val), Ile, Met)	0.43
((((((Ile, Met), Phe), ((Ser, Pro), Thr)), (Cys, Trp)), Glu), Gln, Tyr)	0.77
((((((Cys, Trp), Tyr), ((Thr, Ala), Ser)), (Phe, Leu)), Val), Ile, Met)	0.23
((((((Tyr, Gln), Glu), (Cys, Trp)), ((Phe, Leu), Val)), Ala), Ser, Pro)	0.89

Goldman, 1993; Di Giulio *et al.*, 1994). I have therefore assumed, as the distance between two amino acids, the absolute value of the difference between their polarity values (Woese *et al.*, 1966). These distances make it easy to construct distance matrices, which represent the input to the UPGMA program (Felsenstein, 1991). The tree topologies thus constructed are reported in Table II. These represent the topologies expected on the basis of the physicochemical hypotheses (Sonneborn, 1965; Woese *et al.*, 1966; Lacey and Mullins, 1983).

If the genetic code underwent an ambiguity reduction process (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987), then the column arrangement of the amino acids in the genetic code must also have been involved in this process. In fact, there are some observations indicating that the physicochemical properties of amino acids are correlated to the columns of the genetic code (Nelsestuen, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989b). I have therefore constructed the tree topologies by reducing the ambiguity according to the columns of the genetic code. Thus, starting from the undifferentiated codon NNN, performing ambiguity reduction (Fitch and Upper, 1987; Di Giulio, 1992; Di Giulio, 1994), always giving precedence to the second codon position (columns of the genetic code) and always clustering purines with purines and pyrimidines with pyrimidines,

we obtain the tree topologies reported in Table III. (In these calculations Ser was assumed to be encoded only by the UCN codons.) These topologies are considered to be the main topologies expected on the basis of the ambiguity reduction hypothesis (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987).

3.3. THE STATISTICAL SIGNIFICANCE OF THE VARIOUS HYPOTHESES PROPOSED IN ORDER TO EXPLAIN THE ORIGIN OF THE GENETIC CODE

The biosynthetic relationships between amino acids (Wong, 1975; Taylor and Coates, 1989) identify six groups of amino acids that are in a clear precursor-product relationship or in a strict biosynthetic relationship. The main representatives (precursor) of these groups are Asp, Thr, Glu, Ala, Ser and Phe. I have therefore examined all the possible combinations of some members of these six groups taking them four at a time (see for example Table I). This analysis consisted of examining all the possible tree topologies with 11 or 12 sequences (Figure 1; see for example Table I) using the *Exhaustive Search* technique (Swofford, 1993), and thus generating the frequency distribution between the length, in nucleotide substitutions, of a specific tree topology and the number of all the possible topologies with that given length. I then located the point in this distribution at which a specific expected topology is found, by calculating the length (in nucleotide substitutions) of this topology using *Lengths and Fit Measures* (Swofford, 1993). Finally, I determined the probability of obtaining a specific tree topology by calculating the ratio between the sum of all the topologies having a length equal to or shorter than the expected one and the number of all the possible topologies of trees with 11 or 12 sequences. Thus, I have used the method reported by Fitch and Upper (1987). These probabilities are shown in Tables I, II and III beside their respective topology. (In Table I the probability values given in brackets refer to tree topologies (data not shown) built using the metric based on the number of enzymatic steps).

In an ideal analysis, all twenty tRNA ancestral sequences (Figure 1) should be simultaneously used. This is not possible with the method used so far here as the number of all the possible binary tree topologies for twenty sequences is simply huge. However, by means of the *Random Trees* option (Swofford, 1993), I have generated one million random trees with all twenty sequences (Figure 1) and I have estimated that the mean length (in nucleotide substitutions) of these trees is 303.33 and the standard deviation is 7.82. I then constructed the tree topologies expected on the basis of the three hypotheses in question using the methods already mentioned above. These topologies are shown in Table IV. I have calculated the length of these topologies (Table IV) which were found to be 290 nucleotide substitutions for the coevolution hypothesis, 292 for the physicochemical hypothesis and 291 for the ambiguity reduction hypothesis. Finally, I have used normal distribution to determine the probability of obtaining the observed lengths. For the coevolution hypothesis, I have obtained a length of 290 and the probability is, therefore, given by: $P(x \leq 290) = P(Z \leq (x\text{–m})/\text{s}) = P(Z \leq (290.5\text{–}303.33)/7.82) = P(Z \leq -1.6407)$

= 0.050. (If continuity correction is not used, the probability is lower.) Analogous calculations define the probabilities for the other two hypotheses and these are reported in Table IV.

4. Discussion

The hypotheses proposed in order to explain the origin of the genetic code imply that specific evolutionary relationships must exist between the tRNA molecules (Woese, 1967; Wong, 1975; Fitch and Upper, 1987; Szathmary and Zintzaras, 1992). The analysis herein presented has tested three of these hypotheses (Tables I, II and III). A comparison between these results (Tables I, II and III) seems to establish that the mechanism suggested by the coevolution hypothesis (Wong, 1975; Wong, 1981; Di Giulio, 1993) was the main factor determining the organization of the genetic code. In fact, the probabilities (Table I) show a general trend towards statistical significance with only two values out of fifteen above 10% and with seven significant values at least at the 5% level. This trend is confirmed if we unite the probability values (Table I) into a single value (Fisher, 1950) ($\chi^2 = 85.4$, $P < 0.001$, $n = 15$, df = 30), i.e. if we consider these values as independent observations, and this seems to be at least partly true (Di Giulio, 1994).

If we observe the genetic code form the point of view of the polarity distances, then we see that these distances must have played an important role in determining its organization (Woese *et al.*, 1966; Di Giulio, 1989a; Haig and Hurst, 1991; Goldman, 1993; Di Giulio *et al.*, 1994). It is thus curious that this correlation disappears when the polarity distances are compared with the phylogeny of tRNAs (Table II; $\chi^2 = 38.9$, $0.10 < P < 0.20$, $n = 15$, df = 30) although three probability values have a certain suggestion of significance (Table II). One possible interpretation of this is that the physicochemical properties of amino acids played an important role only in determining which codons were conceded from the precursor amino acid to the product (Wong, 1980) and this might have determined a certain optimization of the amino acid properties according to the columns of the genetic code (Nelsestuen, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989b; Taylor and Coates, 1989) and, more generally, of the overall organization of the genetic code (Woese *et al.*, 1966; Epstein, 1966; Goldberg and Wittes, 1966; Allf-Steinberger, 1969; Jungck, 1978; Weber and Lacey, 1978; Di Giulio, 1989a; Haig and Hurst, 1991; Lacey *et al.*, 1992; Goldman, 1993; Di Giulio *et al.*, 1994). However, the fundamental organization of the genetic code must clearly have developed through its rows (Dillon, 1973; Taylor and Coates, 1989) by means of the mechanism of codon concession from the precursor to the product amino acid, introduced by Wong (1975, 1981) and confirmed by the relationship between the phylogeny of tRNAs and the precursor-product relationships (Table I) (Di Giulio, 1994). This is shown even more clearly by the lack of correlation between the

TABLE IV

This shows the tree topologies, expressed using the New Hampshire notation (Swofford, 1993; user's manual pp. 143–146), which can be associated in the first case (1) to the genetic code coevolution hypothesis, in the second case (2) to the physicochemical hypothesis, and in the third case (3) to the ambiguity reduction hypothesis. The probability in brackets refers to tree topologies (data not shown) constructed using the metric based on the number of enzymatic steps. The probabilities have been calculated through normal distribution. All the abbreviations used are standard. See text for further information.

Tree topology	Probability	
1. (((((((Asp, Asn), Lys), ((Thr, Ile), Met)), ((((Glu, Gln), His), Pro), Arg)), ((Ala, Val), Leu)), ((Ser, Gly), (Trp, Cys))), Phe, Tyr)	0.050	(0.46)
2. (((((((Pro, Thr), Ala), (Ser, Gly)), ((((Trp, Met), Tyr), Val), (((Leu, Ile), Cys), Phe))), (Glu, Asp)), ((Gln, His), Arg)), Lys, Asn)	0.083	
3. (((((((((Ile, Met), Val), (Phe, Leu)), ((Ala, Thr), (Ser, Pro))), (((Cys, Trp), Arg), Gly)), ((Gln, His), Tyr)), (Lys, Asn)), Asp, Glu)	0.065	

phylogeny of tRNAs and the decomposition of the genetic code in a column-wise manner (Table III; $\chi^2 = 36.7$, $0.10 < P < 0.20$, $n = 15$, df = 30).

In the analysis (Table IV) that uses all twenty reconstructed tRNA ancestral sequences, the probabilities associated to the various hypotheses are of the same order of magnitude (Table IV) and therefore such as not to enable any discrimination between the various hypotheses. The quasi-statistical significance of the physicochemical hypothesis and the ambiguity reduction hypothesis (Table IV) could, however, imply that the evolutionary relationships between tRNAs actually reflect the polarity distances between amino acids at some times and the biosynthetic relationships between amino acids at others (Szathmary and Zintzaras, 1992). In my opinion, this interpretation is probably false both in light of the close relationship between the biosynthetic pathways of amino acids and the organization of the genetic code (Dillon, 1973; Wong, 1975; Miseta, 1989; Taylor and Coates, 1989) and in view of the fact that these authors (Szathmary and Zintzaras, 1992) used the metric based on the number of enzymatic steps that generally does not seem to adequately reflect the coevolution hypothesis (Tables I and IV). (As already said above various topologies constructed using this metric do not seem to give an adequate representation of the coevolution hypothesis as they do not accurately define the biosynthetic families of amino acids identified by this hypothesis (Wong, 1975) unlike the topologies shown (Tables I and IV). Therefore, the generally higher probabilities (Tables I and IV) associated to these topologies could simply reflect this fact.)

Recently Szathmary (1993) discussed the various hypotheses proposed to explain the origin of the organization of the genetic code. He suggests (Szathmary, 1993) that the stereochemical model C4N (Shimizu, 1982) must have been compatible to a large extent with the mechanism suggested by Wong (1975, 1981). Although the physicochemical properties of amino acids are linked to the organization of the genetic code, as already mentioned, I nevertheless find it difficult to see how a stereochemical model (Shimuzu, 1982) can have laid the molecular foundations for the genetic code and, at the same time, be compatible with the biosynthetic relationships between amino acids. This would imply that the stereochemical relationships between the amino acid and the anticodon have conditioned the biosynthetic choices between amino acids or that they must, at least, have been such as to enable the genetic code to be imprinted by the precursor-product relationships. It seems to me more natural to think that the stereochemical effects cannot have been so strong as to justify stereochemical models for the origin of the genetic code. Nevertheless, these stereochemical effects, along with the physicochemical ones, must have acted together with the coevolutionary transfer of codons from the precursor to the product (Wong, 1988). When the codon domain of a precursor was divided among several products (Wong, 1975) the subdivision arrangement that optimized the physicochemical advantages or the advantages deriving from the reduction of translation errors (Woese, 1965) must have been selected on alternative arrangements (Wong, 1988). It seems to me that only in this sense could the biosynthetic

relationships between amino acids and the physicochemical constraints have joined together.

In conclusion, this and a previous analysis (Di Giulio, 1994) have shown that the phylogeny of tRNAs reflects the biosynthetic relationships between amino acids more clearly than it reflects the properties of amino acids. Therefore, if future analyses make it possible to obtain a more reliable estimation of the tRNA ancestral sequences (possibly by incorporating the mitochondrial sequences of tRNAs in the analysis) and confirm these results, then the evolutionary origin of the genetic code might be easier to accept.

Acknowledgements

I would like to offer my sincerest thanks to Gianluigi Blasi and Marco Valenzi for their help.

Appendix

The following sequences of tRNAs or genes of tRNAs are from Sprinzl *et al.* (1991):

RD0500, DD0660, DD0680, DD0740, RD1140, RD1660, DD1140, DD1180, DD1230, DD1260, DD1500, DD1540, DD1570, DD1660, DD2440, DD2520, DD2600, DD2680, DD2700, DD2920, DD3200, DD3280, DD6220, DD6280, DD6320, DD6900, DD7560, DD7740, DD8100, DD9160, DD9161, RD6040, RD6280, RD7920, RD9160, RD9161, RD9220, RD9280; RN0380, RN0500, RN0620, DN0660, DN0680, RN1140, RN1660, DN1140, DN1180, DN1230, DN1350, DN1351, DN1500, DN1540, DN1541, DN1570, DN1660, DN2520, DN2600, DN2700, DN2720, DN2740, DN2920, DN3200, DN3320, DN6050, DN6051, DN6060, DN6160, DN6280, DN7100, DN7740, DN7920, DN9990, DN9991, RN6280, RN6940, RN9160, RN9280, RN9990, RN9991; RK0500, RK0501, DK0660, DK0680, DK0740, RK1140, RK1141, RK1540, RK1541, RK1660, DK1140, DK1141, DK1200, DK1220, DK1230, DK1231, DK1350, DK1540, DK1660, DK2000, RK2530, DK2520, DK2580, DK2600, DK2920, DK3200, DK3220, DK3230, DK3240, DK6050, DK6051, DK6052, DK6160, DK6161, DK6280, DK6281, DK6320, DK7560, DK7680, DK7740, DK7741, DK7920, DK8040, DK8100, DK8101, DK9160, DK9990, DK9991, RK6280, RK6281, RK6820, RK7740, RK7741, RK8101, RK9160, RK9161, RK9162, RK9220, RK9221, RK9222; RT0380, RT0500, RT0501, DT0660, DT0661, DT0680, DT0740, DT1140, DT1141, DT1180, DT1230, DT1540, DT1541, DT1542, DT1580, DT1581, RT1140, RT1141, RT1180, RT1540, RT1660, RT1661, DT1660, DT1661, DT1662, DT1663, DT1664, DT1820, DT1821, RT3280, DT2460, DT2520, DT2600, DT2601, DT2640, DT2680, DT2700, DT2701, DT2720, DT2920, DT2921, DT3200, DT3280, DT3281, DT3360, DT6050, DT6160, DT6161, DT6280, DT6281, DT7740, DT9990, DT9991, RT6280, RT6281, RT9280; RI0500, RI0501, RI0660, DI0680, DI1140, DI1141, DI1180, DI1230, DI1260, DI1540, DI1541, DI1542, DI1620, DI1660, DI1661, DI1820, DI1860, DI1900, DI2100, RI1140, RI1141, RI1180, RI1580, RI1660, RI1661, RI1662, DI2400, DI2410, DI2440,

DI2480, DI2520, DI2540, DI2550, DI2570, DI2580, DI2590, DI2600, DI2601, DI2620, DI2700, DI2701, DI2720, DI2760, DI2840, DI2920, DI2921, DI2922, DI3080, DI3220, DI3280, DI3281, RI2720, RI3280, RI3281, DI6280, DI6281, DI6320, DI7740, DI8100, RI6280, RI6360, RI6940, RI8100; RM0500, RM0900, DM0680, DM0900, DMO960, DM1140, DM1180, DM1230, DM1231, DM1260, DM1540, DM1541, DM1660, DM1750, RM1140, RM1540, RM1580, RM1660, DM2520, DM2600, DM2610, DM2640, DM2680, DM2700, DM2701, DM2720, DM2760, DM2840, DM2920, DM3280, RM2530, RM2560, RM3280, DM6160, DM6280, DM6900, DM7740, RM6280, RM6820, RM6940, RM8100, RM9220, RM9990; RE0500, RE0501, DE0660, DE0680, DE0700, DE1140, DE1180, DE1230, DE1340, DE1500, DE1540, DE1570, DE1660, RE1140, RE1660, RE1661, RE1662, DE2440, DE2500, DE2520, DE2600, DE2680, DE2700, DE2920, DE3200, DE3280, DE3360, RE2440, RE2640, RE2680, DE6160, DE6161, DE6280, DE6281, DE6320, DE6321, DE7680, DE7740, DE7741, DE7742, DE8100, DE9160, DE9161, DE9162, DE9990, DE9991, RE6280, RE6320, RE6780, RE6781, RE6940, RE7740, RE8100, RE9160, RE9990; RQ0380, RQ0500, DQ0660, RQ1140, RQ1660, RQ1661, DQ1140, DQ1200, DQ1230, DQ1340, DQ1341, DQ1540, DQ1660, DQ1661, RQ2640, DQ2520, DQ2600, DQ2700, DQ2920, DQ3220, DQ3240, DQ6050, DQ6051, DQ6060, DQ6160, DQ6280, DQ6281, DQ9990, DQ9991, DQ9992, RQ6080, RQ6081, RQ6082, RQ8100, RQ8101, RQ9160, RQ9280, RQ9990, RQ9991; RP0500, RP0501, RP0502, DP0660, DP0680, DP0740, DP1140, DP1180, DP1260, DP1360, DP1400, DP1500, DP1540, DP1560, DP1660, DP1661, DP1662, DP1700, DP1740, DP1780, RP1140, RP1180, RP1540, RP1700, RP1701, RP1702, DP2520, DP2601, DP2680, DP2700, DP2720, DP2920, DP3000, DP3200, RP3280, DP6280, DP6980, DP6981, DP7560, DP7740, DP8040, DP8041, DP8100, DP8101, DP9160, DP9161, DP9990, DP9991, RP6280, RP6360, RP6940; RR0380, RR0500, RR0501, RR0502, DR0660, DR1140, DR1141, DR1180, DR1181, DR1230, DR1260, DR1500, DR1540, DR1660, DR1661, DR1662, DR1663, DR1664, DR1700, DR1780, RR1140, RR1141, RR1540, RR1661, RR1662, RR1663, RR1664, DR2440, DR2480, DR2520, DR2540, DR2600, DR2601, DR2602, DR2680, DR2700, DR2701, DR2720, DR2740, DR2920, DR2921, DR3040, DR3200, DR3201, DR3280, DR3320, DR3321, RR2530, DR6050, DR6051, DR6052, DR6160, DR6161, DR6280, DR6281, DR6282, DR6320, DR6321, DR7560, DR7740, DR7741, RR6280, RR6281, RR6282, RR6820, RR8100, RR8101, RR9280, RR9281; DA0340, DA0380, DA0420, DA0580, DA0620, DA0660, DA0670, DA0680, DA0780, DA0940, DA0980, DA0981, RA0380, RA0500, RA0501, RA0502, DA1140, DA1180, DA1230, DA1260, DA1540, DA1541, DA1542, DA1543, DA1620, DA1660, DA1661, DA1820, DA1860, DA1900, DA2100, DA2240, RA1140, RA1180, RA1540, RA1660, RA1662, DA2400, DA2410, DA2440, DA2480, DA2520, DA2540, DA2570, DA2580, DA2590, DA2600, DA2620, DA2700, DA2720, DA2840, DA2920, DA3280, DA6160, DA6280, DA6281, DA6320, DA6740, DA7680, DA7681, DA7740, DA7920, DA8100, RA6280, RA6360, RA7680, RA7681, RA9990, RA9991; RV0380, RV0381, RV0382, RV0500, RV0501, DV0660, DV0860, RV1140, RV1180, RV1460, RV1540, RV1660, RV1661, RV1662, DV1140, DV1180, DV1230, DV1350, DV1351, DV1500, DV1540, DV1570, DV1660, DV1661, DV2520, DV2600, DV2601, DV2640, DV2700, DV2701, DV2720, DV2721, DV2840, DV2920, DV2921, DV3200, DV3240, DV3280, RV3280, DV6160, DV6161, DV6280, DV6281, DV6320, DV6740, DV7740, DV7741, DV7920, DV9990, DV9991, DV9992, DV9993, DV9994, DV9995, RV6280, RV6281, RV6282, RV6360, RV6940, RV7740, RV7741, RV7742, RV8100, RV9161, RV9220, RV9990; RL0500, RL0501, RL0502, RL0503, RL0504, DL0440, DL0660, DL0680, DL0860, DL0980, DL0981, RL1140,

RL1141, RL1142, RL1460, RL1540, RL1660, RL1661, RL1662, RL1700, RL2020, RL2100, RL2101, DL1140, DL1141, DL1200, DL1220, DL1230, DL1231, DL1232, DL1310, DL1540, DL1541, DL1542, DL1543, DL1544, DL1660, DL1661, DL1662, DL1663, DL1664, DL1700, DL1750, DL1780, DL1940, DL1980, DL2000, RL2840, RL2841, RL2842, RL3160, RL3161, RL3162, RL3280, DL2520, DL2521, DL2522, DL2600, DL2601, DL2602, DL2700, DL2701, DL2702, DL2720, DL2721, DL2740, DL2800, DL2920, DL2921, DL2922, DL3280, DL3360, DL3361, DL3400, DL6160, DL6200, DL6280, DL6281, DL6980, DL7560, DL7740, DL7741, DL7920, DL8100, DL9160, DL9161, DL9162, DL9990, DL9991, RL6280, RL6281, RL6282, RL6360, RL6980, RL6981, RL6982, RL6983, RL7070, RL7560, RL9280, RL9400, RL9401, RL9990; RF0500, DF0660, DF0860, DF1140, DF1180, DF1230, DF1260, DF1540, DF1541, DF1660, RF1140, RF1460, RF1540, RF1580, RF1660, RF2020, RF2060, DF2520, DF2600, DF2700, DF2720, DF2920, DF3360, RF2520, RF3160, RF3280, DF6200, DF6280, DF6281, DF6320, DF6740, DF7740, DF7920, DF9160, RF6040, RF6120, RF6200, RF6280, RF6281, RF6320, RF6780, RF6820, RF6860, RF6940, RF7020, RF7680, RF7681, RF7740, RF7920, RF8100, RF8101, RF8102, RF9220, RF9280, RF9281, RF9340, RF9990; RY0500, DY0660, DY0740, DY1140, DY1200, DY1540, DY1580, DY1660, DY1661, DY1820, RY1140, RY1460, RY1540, RY1541, RY1660, RY1661, DY2520, DY2600, DY2680, DY2700, DY2920, DY3200, DY3280, DY3360, RY2560, DY6050, DY6160, DY6280, DY6740, DY6741, DY6742, DY6743, DY7060, DY7200, DY7740, DY7920, DY7921, DY7922, DY9990, DY9991, RY6120, RY6280, RY6320, RY6360, RY6820, RY6821, RY6940, RY7060, RY7061, RY7740, RY9280, RY9990, RY9991; RG0380, RG0500, RG0501, RG0502, RG0503, RG0620, DG0860, DG0960, DG1140, DG1180, DG1200, DG1230, DG1350, DG1500, DG1540, DG1541, DG1542, DG1580, DG1581, DG1660, DG1661, DG1662, DG1820, DG2000, RG1140, RG1180, RG1310, RG1540, RG1660, RG1661, RG1662, RG1700, RG1701, DG2440, DG2520, DG2521, DG2600, DG2601, DG2640, DG2641, DG2680, DG2681, DG2700, DG2701, DG2920, DG2921, DG3200, RG2530, DG6280, DG7140, DG7180, DG7680, DG7740, DG7741, DG8100, DG9160, DG9161, DG9990, DG9991, RG6280, RG6281, RG6820, RG6940, RG7680, RG7681, RG9990, RG9991; RS0380, RS0500, RS0501, RS0502, DS0440, DS0680, DS0860, DS1140, DS1141, DS1180, DS1230, DS1231, DS1250, DS1260, DS1500, DS1520, DS1540, DS1541, DS1542, DS1570, DS1660, DS1661, DS1663, DS1664, RS1140, RS1141, RS1180, RS1540, RS1541, RS1542, RS1660, RS1661, RS1662, RS1663, RS1664, DS2480, DS2520, DS2521, DS2600, DS2601, DS2602, DS2640, DS2680, DS2700, DS2701, DS2702, DS2720, DS2721, DS2722, DS2920, DS2921, DS2922, DS3200, DS3240, DS3280, DS3281, DS6060, DS6160, DS6161, DS6162, DS6240, DS6241, DS6280, DS6281, DS6282, DS6283, DS6284, DS6320, DS6321, DS6322, DS6740, DS6741, DS6742, DS6743, DS6744, DS6745, DS7240, DS7740, DS7741, DS7800, DS9280, DS9990, DS9991, DS9992, DS9993, RS6280, RS6281, RS6282, RS6940, RS7040, RS7740, RS7741, RS7742, RS9160, RS9161, RS9162, RS9282, RS9991; RC0500, DC0380, DC0500, DC1140, DC1230, DC1260, DC1350, DC1540, DC1660, RC1140, RC1660, DC2440, DC2520, DC2600, DC2680, DC2700, DC2720, DC2920, DC3280, DC6280, DC8100, DC8101, RC6280; RW0500, DW0460, DW0500, DW1140, DW1141, DW1230, DW1250, DW1251, DW1540, DW1660, RW1140, RW1141, RW1250, RW1251, RW1540, RW1660, DW2440, DW2520, DW2600, DW2680, DW2700, DW2720, DW2920, DW3000, DW3200, RW3160, RW3280, DW6160, DW6161, DW6280, DW6740, DW6741, DW7560, DW8040, RW6280, RW6820, RW8040, RW9280; RH0380, RH0500, DH0660, DH0680, DH1140, DH1230, DH1540, DH1541, DH1660, DH1700, DH1740, DH1780, RH1140, RH1660, RH1700, DH2520, DH2600, DH2700, DH2720,

DH2880, DH2920, DH2960, DH3020, DH3120, DH3200, DH3230, DH3240, DH3280, DH3360, DH6160, DH6280, DH6320, DH7740, DH8100, RH6280, RH6281, RH6940, RH7740, RH9460, RH9990.

References

Allf-Steinberger, C.: 1969, *Proc. Natl. Acad. Sci. USA* **64**, 584–591.

Balasubramanian, R., Seetharamulu, P. and Raghunathan, G.: 1980, *Origins Life Evol. Biosphere* **10**, 15–30.

Cedergren, R.J., LaRue, B., Sankoff, D., Lapalme, G. and Grosjean, H.: 1980, *Proc. Natl. Acad. Sci. USA* **77**, 2791–2795.

Crick, F.H.C.: 1968, *J. Mol. Biol.* **38**, 376–379.

Di Giulio, M.: 1989a, *J. Mol. Evol.* **29**, 288–293.

Di Giulio, M.: 1989b, *J. Mol. Evol.* **29**, 191–201.

Di Giulio, M.: 1992, *Origins Life Evol. Biosphere* **22**, 309–319.

Di Giulio, M.: 1993, *J. Mol. Evol.* **37**, 5–10.

Di Giulio, M.: 1994, *Origins Life Evol. Biosphere* **24**, 425–434.

Di Giulio, M., Capobianco, M.R. and Medugno, M.: 1994, *J. Theor. Biol.* **168**, 43–51.

Dillon, L.S.: 1973, *Bot. Rev.* **39**, 301–345.

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.

Goldman, N.: 1993, *J. Mol. Evol.* **37** , 662–664.

Haig, D. and Hurst, L.D.: 1991, *J. Mol. Evol.* **33**, 412–417.

Hendry, L.B., Bransome, E.D.Jr., Hutson, M.S. and Campbell, L.K.: 1981, *Proc. Natl. Acad. Sci. USA* **78**: 7440–7444.

Holmquist, R., Jukes, T.H. and Pangburn, S.: 1973, *J. Mol. Biol.* **78**, 91–116.

Joyce, G.F.: 1989, *Nature* **338**, 217–224.

Jungck, J.R.: 1978, *J. Mol. Evol.* **11**, 211–224.

Lacey, J.C. Jr. and Mullins, D.W. Jr.: 1983, *Origins Life Evol. Biosphere* **13**, 3–42.

Lacey, J.C. Jr., Wickramasingle, N.S.M.D. and Cook, G.W.: 1992, *Origins Life Evol. Biosphere* **22**, 243–275.

Miseta, A.: 1989, *Phys. Chem. Physics Med. NMR* **21**, 237–242.

Nelsestuen, G.L.: 1978, *J. Mol. Evol.* **11**, 109–120.

Nomenclature Committee of the International Union of Biochemistry: 1986, *Mol. Biol. Evol.* **3**, 99–110.

Shimizu, M.: 1982, *J. Mol. Evol.* **18**, 297–303.

Sjostrom, M. and Wold, S.: 1985, *J. Mol. Evol.* **22**, 272–277.

Sonneborn, T.M.: 1965, Degeneracy of the genetic code: extent, nature and genetic implications. In: Bryson V & Vogel HJ eds. Evolving genes and proteins. 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.

Sprinzl, M., Dank, N., Nock, S. and Schon, A.: 1991, *Nucl. Acids Res.* **19** (supplement), 2127–2171.

Swofford, D.L.: 1993, PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1, Laboratory of Molecular Systematics, Smithsonian Institution.

Szathmary, E.: 1993, *Proc. Natl. Acad. Sci. USA* **90**, 9916–9920.

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.
Weiner, A.M. and Maizels, N.: 1987, *Proc. Natl. Acad. Sci. USA* **84**, 7383–7387.
Woese, C.R.: 1965, *Proc. Natl. Acad. Sci. USA* **54**, 1546–1552.
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* **72**, 1909–1912.
Wong, J.T.: 1980, *Proc. Natl. Acad. Sci. USA* **77**, 1083–1086.
Wong, J.T.: 1981, *Trend Biochem. Sci.* **6**, 33–36.
Wong, J.T.: 1988, *Microbiol. Sci.* **5**, 174–181.