IUBMB *Life*, 56(3): 125–130, March 2004

Taylor & Francis
healthsciences

Hypothesis Paper

Symmetry Preservation in the Evolution of the Genetic Code

José Eduardo M. Hornos[1], Lígia Braggion[1], Marcio Magini[2] and Michael Forger[3]
[1]*Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, BR-13560-970 São Carlos, SP, Brazil*
[2]*Instituto de Pesquisa e Desenvolvimento, Universidade do Vale do Paraíba, Av. Shishima Hifumi, 2911, BR-12244-000 São José dos Campos, SP, Brazil*
[3]*Instituto de Matemática e Estatística, Universidade de São Paulo, Caixa Postal 66281, BR-05315-970 São Paulo, SP, Brazil*

Summary

The standard genetic code is found to exhibit an exact symmetry under a finite group of order 4 known in mathematics as the *Klein group*. The same symmetry is also present in almost all non-standard codes, mitochondrial as well as nuclear. Analysis of the phylogenetic tree for the evolution of the mitochondrial codes reveals that all changes along the main line of evolution preserve this symmetry, with a tendency towards symmetry enhancement. In the side branches of the evolutionary tree, the majority of changes also respect the symmetry. The few exceptional cases where it is broken correspond to reassignments that appear to be unstable or incomplete. Since the Klein group emerges naturally from the symplectic model for the prebiotic evolution that has led to the standard code, we interpret these results as lending support to the hypothesis that this symmetry has been selected during the evolution of the genetic code, not only before but also after establishment of the standard code.

IUBMB *Life*, 56: 125–130, 2004

Keywords Evolution; genetic code; symmetry.

The significance of the regularities observed in the standard genetic code and in non-standard codes has been a long-standing problem. In particular, the degeneracies in the codon to amino-acid assignments, as well as the changes that appear to have occurred in these assignments during evolution, have been the subject of much discussion. The most important deviations from the standard code, first found in 1979 in studies of the human cytochrome oxidase subunit II gene, are UGA and AUA coding for Trp and Met, rather than for Stop and Ile, respectively. Since then, many other examples of non-standard codes have been identified, burying definitively the hypothesis of a universal code and triggering an extensive debate about the evolution of the genetic code. A large amount of experimental data relevant to this question, together with phenomenological and theoretical considerations, has been assembled by Osawa, Jukes and collaborators (*1, 2*). In recent years, the number of articles on the problem has steadily increased and the pertinent literature has become so extensive that it has become almost impossible to compile a reasonably complete bibliography. As an illustration of the great variety of approaches to the subject, we may quote the discussion on the chronological order in which the twenty aminoacids have been incorporated into the code: there are presently about 40 models referring to various aspects of the early evolution of life on Earth that lead to predictions about this order and based on very different approaches; see Ref. (*3*) for an overview. Among them is the algebraic model for the evolution of the genetic code first proposed in 1993 (*4*) and further developed in subsequent years (*5–8*).

The present paper is devoted to the further development of the algebraic model. Our main goal is to lay the ground for making closer contact between the predictions of this model and experimentally accessible facts from evolutionary biology.

The basic idea underlying the algebraic model for the evolution of the genetic code is the concept of symmetry, viewed as the 'raison d'être' of the degeneracies observed in the genetic code. In order to avoid misunderstandings, we would like to emphasize from the very beginning that symmetry arguments are not intended nor able to replace biochemical considerations. Rather, their role in understanding the evolution of the genetic code is to postulate constraints on the possible pathways of evolution – constraints whose validity must be checked independently. For example, the by now generally accepted picture of a primordial evolution of the genetic code, accompanied and characterized by a stepwise inclusion of more and more amino acids into the machinery of protein synthesis, is completely consistent with the picture of stepwise symmetry breaking, starting out from a large

Received 21 July 2003; accepted 12 January 2004

Address correspondence to: José Eduardo Hornos, Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, BR-13560-970 São Carlos, SP, Brazil. E-mail: hornos@if.sc.usp.br

ISSN 1521-6543 print/ISSN 1521-6551 online © 2004 IUBMB
DOI: 10.1080/15216540410001687838

primordial symmetry which is broken in a sequence of steps until reaching a final state of strongly reduced symmetry. Of course, such a picture would be of little value if this process were to lead to a complete breaking of all symmetries, so it is crucial to realize that the genetic code observed in extant organisms does indeed exhibit an interesting and non-trivial residual symmetry.

A familiar example from physics may help to illustrate the kind of reasoning proposed here. Newton's universal law of gravitation is a central force law, which means that it exhibits rotational symmetry. This symmetry by itself implies conservation of angular momentum and constrains planetary orbits to be planar, but the specific dependence of the force on the distance between the planet and the sun cannot be predicted by this symmetry alone. Conversely, the observed fact that all planetary orbits are planar is a fingerprint of rotational symmetry and is experimentally accessible without complete knowledge of the force law.

We begin our exposition with a brief summary of the main points of the algebraic model; a detailed presentation of the subject, including the mathematical background, is available in the literature (*7*).

The starting point of the algebraic model is to consider codons as vectors, more specifically as basis vectors in an abstract 64-dimensional space: the space of codons. The notion of symmetry can be implemented by allowing a group of matrices to transform these vectors, in the same way that the ordinary rotation group transforms vectors in 3-dimensional physical space. The next step is to decompose the codon space by assembling the codons belonging to the same amino acids in mutually orthogonal subspaces. The mathematical procedure to implement this decomposition is called symmetry breaking and has been widely used in various fields of science. The process begins by selecting an ancestor group of matrices G and then choosing a descending sequence of subgroups $G_1 \supset G_2 \supset \ldots \supset G_n$. In each step of the process, the symmetry is lowered by requiring the codon vectors to be related only under the action of the next subgroup. In the beginning, all vectors can be connected by matrices belonging to G. After the first breaking, the codon space will be divided into subspaces formed by vectors that can be connected to each other by matrices belonging to G_1. Continuing in this way, one finally arrives at a decomposition of the codon space into 21 subspaces whose dimensions correspond to the degeneracies found in the genetic code. These subspaces are assigned to the amino acids and the stop codons. The search for the correct ancestor group and the adequate chain of subgroups is feasible due to the Cartan classification theorem, which organizes the simple continuous symmetries into four families and five exceptional groups. This search has been carried out, and the most natural ancestor group found is the symplectic group Sp(6), widely used in classical mechanics, together with the chain $Sp(6) \supset Sp(4) \times SU(2) \supset SU(2) \times SU(2) \times SU(2) \supset O(2) \times SU(2) \times U(1)$.

The decomposition of the codon space corresponding to this chain is shown in Fig. 1. In the first step, six independent groups of codons are generated, corresponding to six primitive amino acids. In the subsequent steps, degeneracy decreases until the standard code is reached. It should be noted that the symmetry breaking has been partially frozen in the last step and that no information derived from mitochondrial or other non-standard codes has been used in the search.

More insight into the role of symmetries in the evolution of the genetic code can be obtained by considering the so-called *weight diagram* of the codon representation of the group Sp(6). In this diagram, the codons can be arranged geometrically as points in a three-dimensional space, represented by small spheres, as shown in Fig. 2.

This diagram is composed of two familiar pieces: an external truncated octahedron with a smaller octahedron inside. In each vertex of the internal octahedron we see four codons, while in the center of each hexagonal face of the external truncated octahedron there are two codons. (The corresponding spheres in Fig. 2 are drawn separately

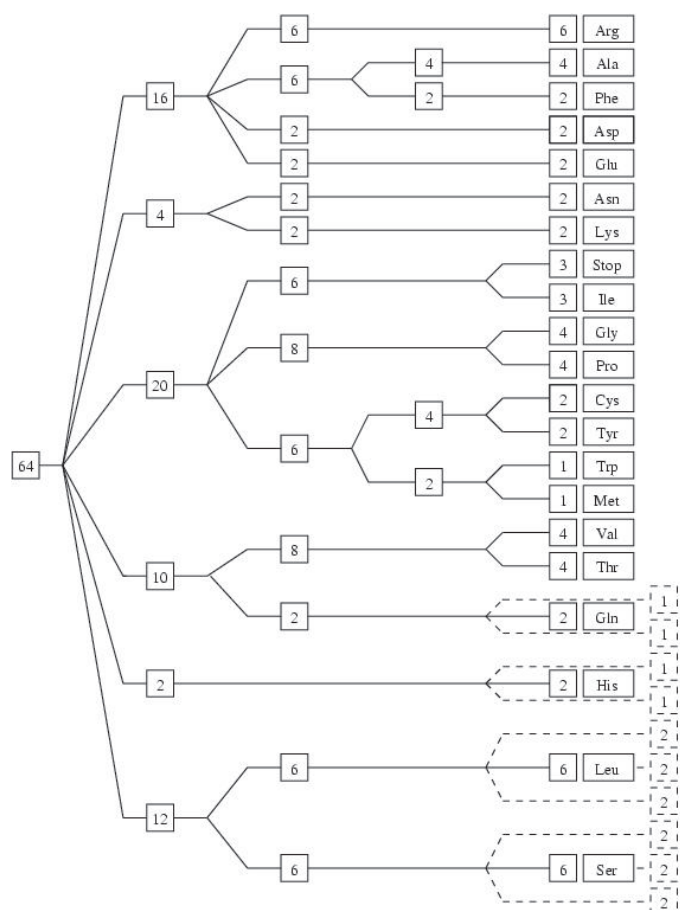

Figure 1. Tree of evolution for the standard genetic code in the Sp(6) model, with amino acid assignments as in Ref. (*7*).

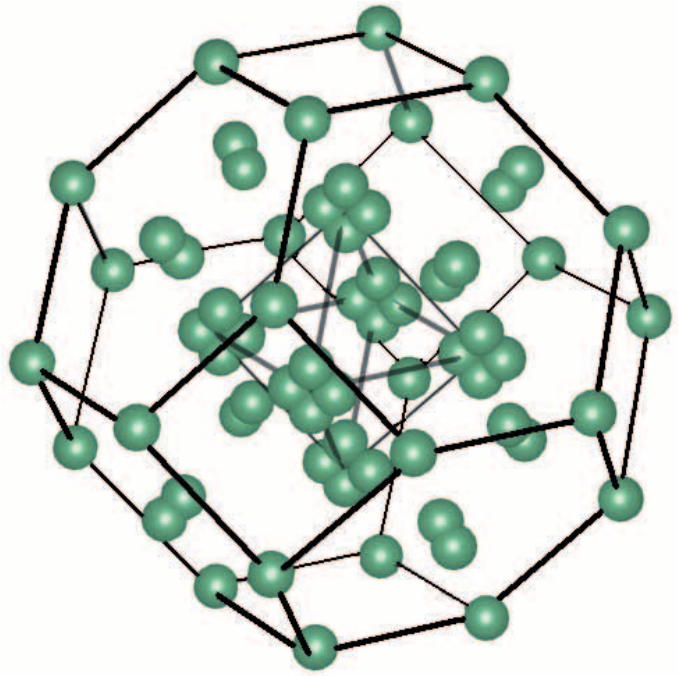

Figure 2. Weight diagram of the codon representation of Sp(6) before the first symmetry breaking.

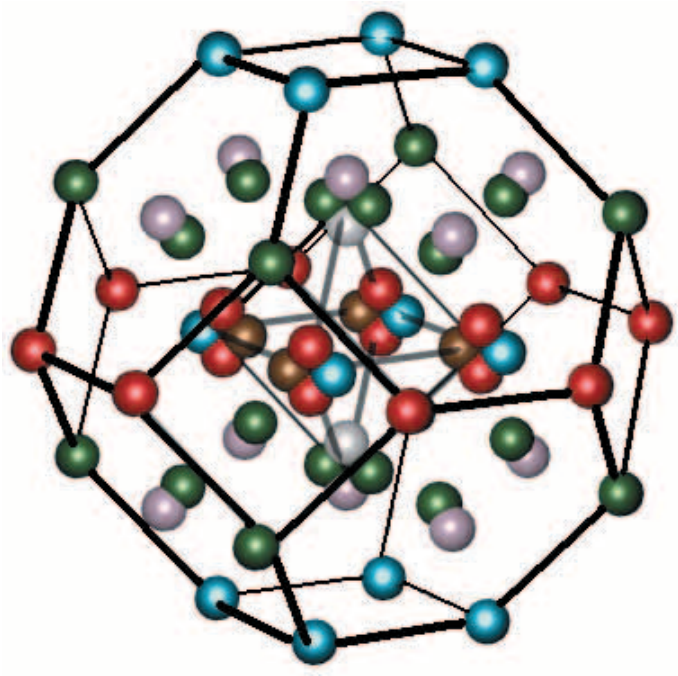

Figure 3. Weight diagram of the codon representation of Sp(6) after the first symmetry breaking.

but should be thought of as superimposed, with their centers located at the same points.) This degeneracy is not related with the code but is an intrinsic property of the representation. The invariance group of this object is familiar to crystallographers, chemists and solid-state physicists. It is the octahedral group O_h composed of 48 operations: (i) rotations by 90° around an axis joining opposite vertices of the internal octahedron, (ii) rotations by 120° around an axis joining the centers of opposite hexagonal faces of the external truncated octahedron; (iii) rotations by 180° around an axis connecting the middle points of opposite edges of the internal octahedron; (iv) inversion and (v) composition of inversion and the operations described before. This group is also the Weyl group of the symplectic group Sp(6).

Fig. 2 shows the situation before the beginning of the symmetry breaking, when there is no distinction between the codons. The first step in the evolution of the code is shown in Fig. 3. This primitive code has only six amino acids that are distinguished in the picture by different colors.

Continuing the process, the standard code is reached in the final step, where the 20 amino acids and the stop codons should be distinguished by 21 colors. Fig. 4 shows part of the resulting picture, namely the internal octahedron, including some amino acid assignments.

A complete illustration can be given by taking horizontal sections and specifying the amino acid and codon assignments as in Ref. (*7*): this is shown in Fig. 5.

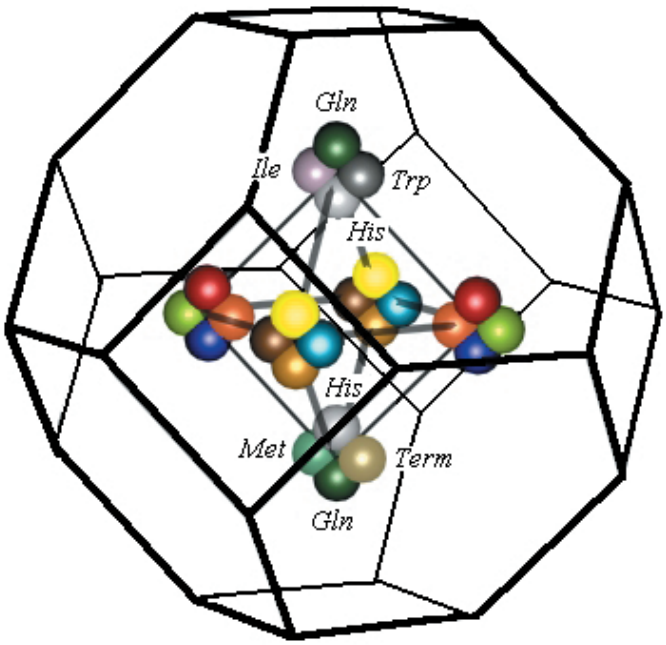

Figure 4. Weight diagram of the codon representation of Sp(6) after the last symmetry breaking: central octahedron.

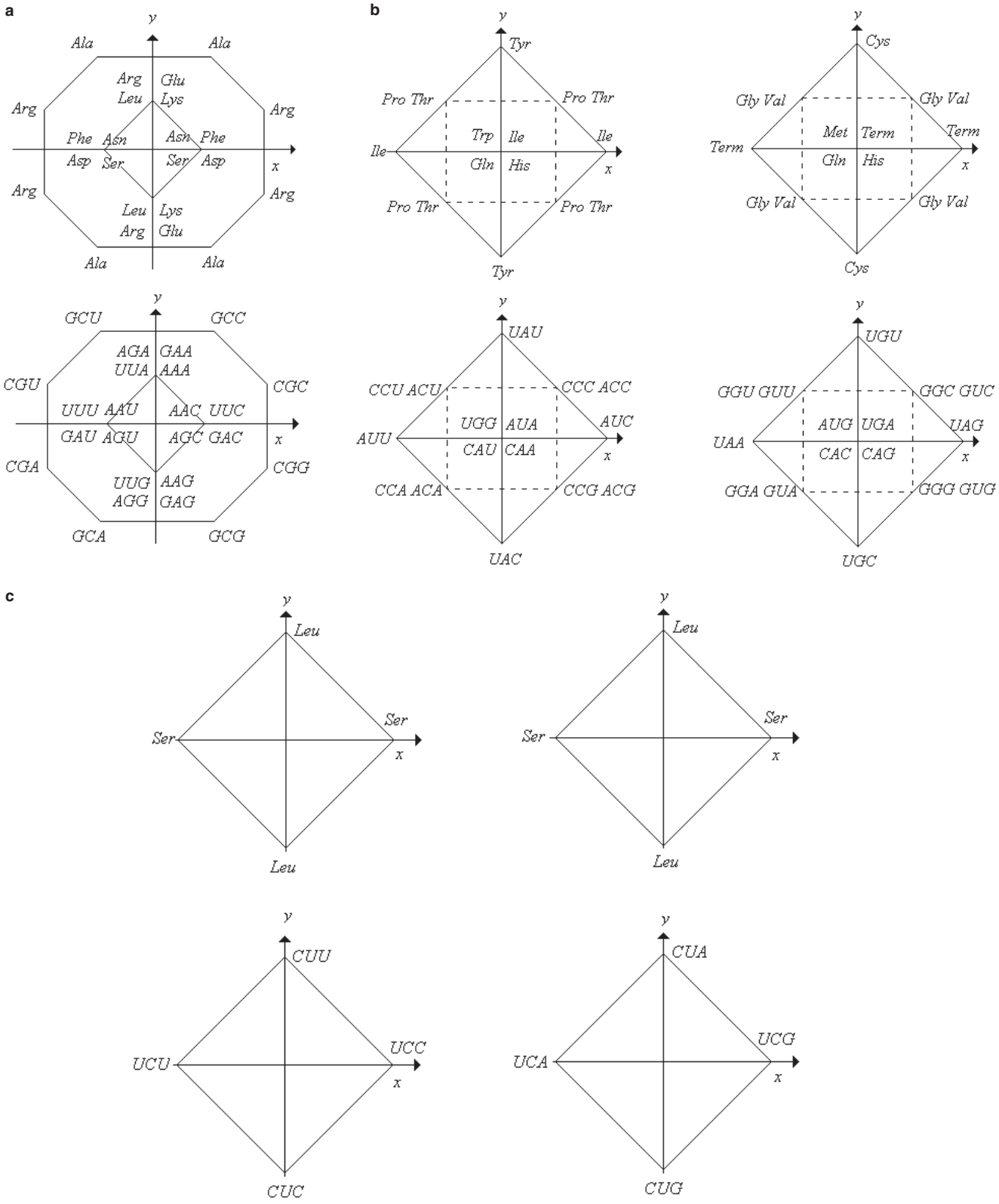

Figure 5. Amino acid and codon assignments for the standard code: **a.** plane $z = 0$. **b**. planes $z = +1$ (left) and $z = -1$ (right). **c**. planes $z = +2$ (left) and $z = -2$ (right).

The original octahedral symmetry has already been broken in the primitive code (Fig. 3), but there is a residual symmetry that survives up to the end of the symmetry breaking process. It is composed of four operations that are easily identified in Fig. 4 and Fig. 5; apart from the identity (or trivial transformation), they are: (i) the reflection in the xz-plane, (ii) the reflection in the yz-plane and (iii) the rotation by 180° around the z-axis. *This discrete group of four elements, known in mathematics as the Klein group, is an exact symmetry group of the standard genetic code*.

A more detailed picture of how the Klein group acts on codon space can be gained by inspection of Fig. 5, using that in the Cartesian coordinates employed above, a point (x,y,z) is taken to $(x,-y,z)$ under (i), to $(-x,y,z)$ under (ii) and to $(-x,-y,z)$ under (iii). It turns out that this action leaves the nucleotides in the first and second position of each codon unaltered while transforming the four possible nucleotides in the third position by a permutation that depends nontrivially on the first two. Thus following the general rule that symmetry expresses itself through degeneracies, we may assemble all codons that are forced to be synonymous by the Klein symmetry alone into boxes; this leads to the arrangement shown in Table 1.

Table 1
Synonymous codons in genetic codes with Klein symmetry

UUU UUC UUA UUG	UCU UCC UCA UCG	UAU UAC UAA UAG	UGU UGC UGA UGG
CUU CUC CUA CUG	CCU CCC CCA CCG	CAU CAC CAA CAG	CGU CGC CGA CGG
AUU AUC AUA AUG	ACU ACC ACA ACG	AAU AAC AAA AAG	AGU AGC AGA AGG
GUU GUC GUA GUG	GCU GCC GCA GCG	GAU GAC GAA GAG	GGU GGC GGA GGG

Any genetic code that is compatible with the Klein symmetry must assign the same amino acid to all codons belonging to the same box in Table 1. Obviously, the standard code satisfies this requirement.

Fig. 6 shows the phylogenetic tree for the evolution of mitochondria, reproduced from (*2*). The changes that occur along the main line of evolution are labeled as (1), (2), (3) and (7). (1) represents the reassignment of the standard Stop codon UGA to Trp and (3) indicates the reassignment of the standard Ile codon AUA to Met, whereas (2) and (7) stand for the reassignment of the pair of standard Arg codons AGA and AGG to Ser and to Stop, respectively. Inspection of Fig. 4 and Fig. 5 reveals that all these changes occur inside the internal octahedron and that each of them preserves the symmetry under the Klein group. Indeed, the changes (1) and (3) each affect only one codon lying on the z-axis, whereas the changes (2) and (7) affect two codons lying on the y-axis and forming a symmetric pair under reflection in the xz-plane. Moreover, when the two reassignments (1) and (3) occur together, they even enhance the symmetry of the internal octahedron, since an additional invariance under reflection in the xy-plane appears. This symmetry enhancement is however incomplete because the asymmetry under exchange of the $z = +1$ and $z = -1$ planes in the external truncated octahedron remains. (Note that the amino acid assignments in the $z = +2$ and $z = -2$ planes are invariant under this additional reflection.) We therefore conclude that the changes along the main line of evolution preserve the symmetry under the Klein group, with a tendency towards restoring symmetry under a larger group. This is in accordance with the proposal that the evolution of the mitochondrial codes can be viewed as retrogression from the standard code to an earlier code, with a higher degree of symmetry.

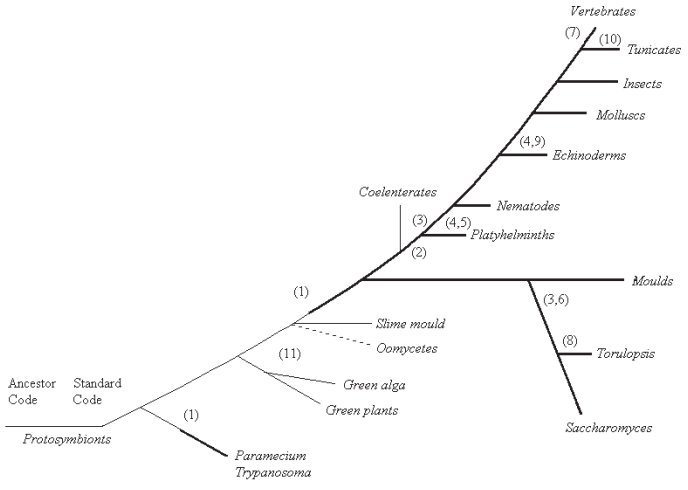

Figure 6. Phylogenetic tree for the evolution of mitochondrial codes (*2*).

In the side branches of the evolutionary tree shown in Fig. 6, we can identify three types of reassignments, the first one being labelled by (6), (9) and (10). (6) represents the conversion of the standard family box of Leu codons CUN, located along the y-axis in the $z = +2$ and $z = -2$ planes, to Thr, which occurs in yeast mitochondria (*Saccharomyces cerevisiae, Torulopsis glabrata*). (9) is the converse to (3), restoring the standard meaning of the previously reassigned codon AUA, from Met back to Ile, as found in the mitochondria of echinoderms. (10) indicates the reallocation of the pair of standard Arg codons AGA and AGG to Gly, which is similar to (7) and is typical for the mitochondria of tunicates. All these changes preserve the symmetry under the Klein group.

In the transition labeled by (4), which occurs in the mitochondria of platyhelminths and echinoderms, the standard Lys codon AAA is converted to Asn. Similarly, in the transition labelled by (5), which was also observed in the mitochondria of platyhelminths, the standard Stop codon UAA – not the standard Stop codon UGA subject to frequent reassignments – is converted to an amino acid, presumably Tyr. Both of these reassignments break the invariance under the full Klein group but are invariant under the 'chiral' symmetry generated by the reflection in the yz-plane.

Finally, (8) and (11) stand for the appearance of unassigned or nonsense codons: in case (8), also found in yeast mitochondria (*Torulopsis glabrata*), this happens to the family box of standard Arg codons CGN, whereas in case (11), found in green algae, it occurs for the three codons CGG, UGA and UAG. In the first case, symmetry under the Klein group is preserved, while in the second case, it is completely broken. However, it must be borne in mind that, according to the codon capture theory (*1*,*2*), such a reallocation represents a transient state on the way to a new assignment, so that in these cases, the question whether the symmetry is broken or not is probably not relevant.

An analogous analysis of the main non-standard nuclear codes, based on (*1*,*2*), leads to similar conclusions. The most important changes here are the reassignment of UGA from Stop to Trp already encountered before, occurring in certain eubacteria such as *Mycoplasma* and *Spiroplasma*, the reassignment of UAA and UAG from Stop to Gln, as found in various species of ciliated protozoans (*Tetrahymena thermophilia, Paramecium spp.*, *Oxytricha, Stylonichia*) and of *Acetabularia*, and finally the reassignment of UGA from Stop to Cys, encountered in other species of ciliates (*Euplotes*). All of these preserve the symmetry under the Klein group. The only known exception occurs for a group of yeast species (*Candida*), where the standard Leu codon CUG has been reassigned to Ser: again, this change breaks the invariance under the full Klein group but is invariant under the 'chiral' symmetry generated by the reflection in the yz-plane. Interestingly, this code variation is found to have been reversed in another group of yeast species that branched off later; moreover, both of these modifications are rather recent (estimated to have occurred 900 and 150 million years ago, respectively).

Summarizing our results, we have first of all identified an exact group of symmetries for the standard genetic code: the Klein group. It has emerged from the algebraic model for the evolution of the genetic code, based on the symplectic group Sp(6), by inspecting the corresponding weight diagram and Weyl group. As far as we know, this is the first exact symmetry observed in a fundamental biological process whose associated group is larger than the usual chiral group, and it explains part of the degeneracies observed in the standard genetic code. Moreover, this symmetry has a far larger class of universality than the standard code itself, being preserved in the main phylogenetic line and in most side branches of the evolution of mitochondria, as well as in most non-standard nuclear codes known to date. Finally, the small number and the characteristic features of the exceptions that have been found so far lend support to the picture that a reassignment violating the Klein symmetry is structurally unstable or at least much less stable than one compatible with the Klein symmetry and tends to be reverted or complemented by the same reassignment for its partner(s) under the Klein symmetry.

We conclude that the Klein symmetry in the genetic code has been selected by evolution and expect that it can be used to predict other changes to be detected through future experimental investigations.

ACKNOWLEDGEMENTS

This research was generously supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil.

REFERENCES

1. Osawa, S., Jukes, T. H., Watanabe, K., and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiological Rev.* **56**, 229–264.
2. Osawa, S. (1995) *Evolution of the Genetic Code*. Oxford University Press, Oxford.
3. Trifonov, E. N. (2000) Consensus temporal order of aminoacids and evolution of the triplet code. *Gene* **261**, 139–151.
4. Hornos, J. E. M., and Hornos, Y. M. M. (1993) Algebraic model for the evolution of the genetic code. *Phys. Rev. Lett.* **71**, 4401–4404.
5. Hornos, J. E. M., and Hornos, Y. M. M. (1994) A search for symmetries in the genetic code. *J. Biol. Phys.* **20**, 289–294.
6. Forger, M., Hornos, J. E. M., and Hornos, Y. M. M. (1997) Global aspects in the algebraic approach to the genetic code. *Phys. Rev. E* **56**, 7078–7082.
7. Hornos, J. E. M., Hornos, Y. M. M., and Forger, M. (1999) Symmetry and symmetry breaking: an algebraic approach to the genetic code. *Int. J. Mod. Phys. B* **13**, 2795–2885.
8. Antoneli, F., Braggion, L., Forger, M., and Hornos, J. E. M. (2003) Extending the search for symmetries in the genetic code. *Int. J. Mod. Phys. B* **17**, 3135–3204.