International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Continuity in Evolution from Codons to Amino Acids

Nisha Gohain

Abstract*: Continuity plays a crucial role in any discussion on evolution. The genotype-phenotype map focus crucial role in evolutionary biology. A nearness relation in the phenotype space allows distinguishing continuous evolution from discontinuous one. In this paper we have investigated continuity in evolution of amino acids. Uniform structures in the space of amino acids as well as in the set of codons are introduced. Here any codon is a genotype and the corresponding amino acid is the phenotype. We examined the continuity of the genotype-phenotype map with respect different uniform structures of codons.

***Keywords*: Amino acids, Continuity, Genotype space, Phenotype space, Uniformity.**

I. INTRODUCTION

Evolutionary biology is concerned with the determination of the organism's fitness, changes, adaptability in the nature. Genotype and phenotype are two important aspects of biological evolution. Understanding which phenotype is accessible from which genotype is fundamental for understanding the evolutionary process [1]. The genotype is the genetic makeup of an individual organism. And the phenotype is the actual appearance of an organism which describes the processes of development of the organism, that is, the phenotype is the observed characteristic of the organism.

Evolution is change of phenotypes with respect to time. Mathematically expressed, it is a function

$$\phi : R \to P$$

Where R is the time axis, say real line with a suitable topology and P is the phenotype set. This function is called evolutionary trajectory or evolutionary map.

The change in phenotype is actually caused by change in genotype. Thus with respect to time genotype changes and that change in genotype produce changes in phenotype which is observable. Hence the evolutionary map can be expressed as

composition of two maps:

$$g : R \to G \text{ and } f : G \to P \text{ that is } f \cdot g = \phi$$

Here G stands for the genotype set.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Nisha Gohain*, Directorate of Open and Distance Learning, Dibrugarh University, Assam, India. Email: gohainnisha@gmail.com

Many authors tried to model the genotype phenotype map, where they considered the RNA sequences as genotype and the RNA shapes as phenotypes. The RNA folding models becomes more realistic in genotype to phenotype map. Moreover these models are simple and computationally tractable [2]. In [3], Muller and Wagnar works on population genetics and states that the Neo-Darwinian framework is sufficient to explain characteristics of phenotypic evolution such as innovation. After that working on the same field, Fontana and Buss [4] argued that before selection can determine the fate of a new phenotype, that phenotype must first be produced, or accessed, by means of variational mechanisms.

Fontana and Schuster [1] and Cupal et al. [5] equipped the phenotype with topology based on probability of mutation. Stadler et al. [6] suggested that the set of phenotypes P with a pre-topology defined in terms of the GP map f: G→ P can describe the evolutionary mechanism.

II. MATERIALS AND METHOD

Now to study whether the evolution is continuous or not, we need to examine whether the changes from one phenotype to another is gradual or not. For that we need some notion of nearness or accessibility among phenotypes. As the evolution in phenotypes is mediated by change in genotypes, we also require a nearness relation in the set of genotypes. Fontana et al. ([1],[2],[7],[8]) considered evolution of RNA shapes where various nearness relations were discussed. In case of RNA evolution, the genotype is the RNA sequences and the phenotypes are the corresponding structures.

Our study will be concerned with amino acid evolution. Genetic code lies at the heart of this evolution. The genetic code is a composition of 64 codons. Each codon codes one of the 20 amino acids used in the synthesis of proteins. The amino acids are encoded by more than one codon. There are 64 codons.

Our present study is mainly a theoretical investigation of the pattern of amino acid evolution mediated by different kinds of mutation of codons

A. Topological Aspects and Continuity

In this section we discuss the basic concepts of topology and the notion of continuity.

Topology

A structured set is called a space. A set can be structured in different ways and one such method is by equipping it with a topology. Topology on a given set is a structure that enables us to discuss notions like convergence and continuity. The notion of topology is an abstraction of the concrete idea of distance (metric).

A topology on a set can be defined in many equivalent ways, namely in terms of open (closed) sets, interior operator, closure operator, neighbourhood system, cluster points, boundary operator, uniformity. The most frequently used definition of topology is in terms of open (closed) sets.

Definition: Let X be set. A topology on a set X is a non-empty collection T of subsets of X such that

- The empty set and whole set X belongs to T.
- Arbitrary union of members of T is in T (respectively finite intersection of members of T is in T).
- Finite intersection of members of T is in T (respectively arbitrary union of members of T is in T).

The members of T are called open sets (resp. closed sets).

Ali et al.[9] discussed continuity of evolution of amino acids by taking the neighbourhood structures based on vicinity. In this paper we attempted to make a similar study based on uniform structure. The idea of uniformity was introduced by A. Weil [10].

A uniform structure is induced by a family of relations which are called nearness relation or proximity relation. A relation on X is a subset, say U of $X \times X$. Here $(x, y) \in U$ can be interpreted as 'x is U-related to y ' or ' x is U-near to y '.

A relation need not be symmetric in the sense that it is not required that $(x, y) \in U$ implies $(y, x) \in U$. That is, ' x is U-near to y ' need not imply ' y is U-near to x '. However for our discussion the relations will be symmetric. Though nearness relation does not represent numerical distance, yet we can talk of degree of nearness (level of resolution). If U and are two nearness relations such that $U \subseteq U'$, then we say that any pair of points which are U-near are also U'-near. Then U expresses a greater degree of nearness in comparison to U'. If a pair of elements are both U-near and U'-near, they are also $U \cap U'$-near.

If U and U' are two nearness relations, then their composition is defined as

$$U \circ U' = \{(x, y) : \exists\ z : (x, z) \in U \text{ and } (z, y) \in U'\} \qquad (1)$$

Since $U, U' \subseteq U \circ U'$, elements of U and U' are near on a finer scale than elements of $U \circ U'$.

Elements z in equation (1) can be interpreted as in between x and y [6].

One obvious assumption for a nearness relation U is that $(x, x) \in U \quad \forall x \in X$. Such a relation is called reflexive. Mathematically it is expressed as $\Delta X \subseteq U$, where $\Delta X = \{(x, x) : x \in X\}$ is called the diagonal.

Following are some definitions regarding uniformity.

Definition: A uniformity on a set X is a non-empty collection $\mathfrak{A}$ of subsets of $X \times X$ satisfying the following properties

(i) $\Delta X \subset U$ for each $U \in \mathfrak{A}$

(ii) if $U \in \mathfrak{A}$ then $U^{-1} \in \mathfrak{A}$

(iii) if $U \in \mathfrak{A}$ then there exists $V \in \mathfrak{A}$ such that $V \circ V \subset U$.

(iv) if $U, V \in \mathfrak{A}$ then $U \cap V \in \mathfrak{A}$ and

(v) if $U \in \mathfrak{A}$ and $U \subset V \subset X \times X$ then $V \in \mathfrak{A}$

Members of $\mathfrak{A}$ are called entourages.

The set X along with the uniformity $\mathfrak{A}$ is called a uniform space. It is denoted by (X, $\mathfrak{A}$).

Axiom (i) states that the relations are reflexive. Axiom (ii) states that the relations are symmetric and axiom (iii) states that for each $U \in \mathfrak{A}$, there is a $V \in \mathfrak{A}$ such that whenever x is V-close to y and y is V-close to z, then x is U-close to z.

Definition: Let $(X, \mathfrak{A})$ be a uniform space. Then a family ß of $\mathfrak{A}$ is said to be a base for $\mathfrak{A}$ if every member of $\mathfrak{A}$ contains some member of ß; while a subfamily S of is said to be a sub-base for $\mathfrak{A}$ if the family of all finite intersections of members of S is a base for $\mathfrak{A}$.

Proposition: Let X be a set and $S \subset P(X \times X)$ be a family such that for every $U \subseteq X$ the following conditions hold:

i. $\Delta X \subset U$,

ii. U^{-1} contains a member of S, and

iii. There exists $V \in S$ such that $V \circ V \subset U$. Then there exists a unique uniformity U for which S is sub-base.

Throrem: For a uniform space $(X, \mathfrak{A})$ let $\mathfrak{T}_v$ be the family. Then $\mathfrak{T}_v$ is a topology on X. U[x] is defined to be the set

$$\{y \in X : \text{ there exists some } x \in X \text{ such that } (x, y) \in U\}.$$

Continuity

The notion of continuity which lies at the core of topology can be defined in a variety of equivalent ways. Roughly speaking, a map *f* from a space X to a space Y is continuous if small (gradual) change in X produces small (gradual) change in Y. The small (gradual) change will be determined by the topology on X. What may be gradual with respect to one topology may not be so with respect to another topology.

Let X and Y be two topological spaces.

A map *f:* $X \to Y$ is said to be continuous at $x \in X$, if

$$B \in \mathcal{M}(f(x)) \Rightarrow f^{-1}(B) \in \mathcal{N}(x),$$

where $\mathcal{N}$ and $\mathcal{M}$ denote the neighbourhood systems at x and *f(x)* respectively.

If *f* is continuous at each point of X, then we say $f : X \to Y$ is continuous.

Below we consider other equivalent definitions of continuity.

(1) $Cl(f^{-1}(B)) \subseteq f^{-1}(Cl(B)), \quad \forall B \in P(Y)$

(2) $f(Cl(A)) \subseteq Cl(f(A)), \quad \forall A \in P(X)$

(3) For every neighbourhood M of *f(x)*, ∃ neighbourhood N of x: $f(N) \subseteq M$.

(4) $\mathcal{M}(f(x)) \subseteq f(\mathcal{N}(x)) \quad \forall x \in X$, where $\mathcal{N}$ is the neighbourhood system at x and $\mathcal{M}$ is the neighbourhood system at *f(x)*.

Definition: A function $f : X \to Y$ between uniform spaces is called uniformly continuous if for every entourage V in Y there exists an entourage U in X such that for every (x_1, x_2) in U we have $(f(x_1), f(x_2))$ in V.

Similar to continuous functions between topological spaces, which preserve topological properties, are the uniform continuous functions between uniform spaces, which preserve uniform properties.

If we fixed one of the x_i, then the above definition gives continuity of the function at x_i.

III. TOPOLOGY IN CODON SET AND AMINO ACID SET

The 20 amino acids are coded by the 64 codons. We consider the 64 codons as genotype space and the 20 amino acids as the phenotype space. We have already mentioned that the understanding of which phenotype is accessible from which genotype is fundamental for understanding the evolutionary process. The notion of accessibility is studied using different topological structures [1, 2, 5-7, 8, 11-16].

Different forms of generalized topologies are used to study the Genotype-Phenotype Map Model. Here we make an attempt to study the model using the notion of uniformity.

An amino acid is coded by codons. Evolution of one amino acid from another is mediated by mutation in the corresponding codons.In this paper the continuity of the function mapping codons to amino acids is being investigated. To express continuity explicitly, we must define when one phenotype is close to another.

We have introduced uniform structure on the phenotype set (amino acids) using phenotype set (amino acids) using phenotypic proximity based on similarity of properties. The uniformity on the codon set is defined by considering relations between codons based on different types of mutations and the uniformity of the amino acid set is defined by considering three important properties of amino acids.

We consider the following three important properties of amino acids: hydrophillic- hydrophobic, polar-nonpolar and aliphatic-aromatic-neutral(neither aliphatic nor aromatic). These three properties give us three relations between amino acids.

R1: Two amino acids are related if both are hydrophilic or both hydrophobic.

R2: Two amino acids are related if both are polar or both non-polar.

R3: Two amino acids are related if both are aliphatic or aromatic or both neutral.

Let X represents the set of codons and Y represents the set of amino acids.

From the proposition, we observe that the three relations R1, R2and R3form a sub-base for some unique uniformity on Y.

The uniformity induced by the above relations show that two amino acids are close or near to each other if they share one of the properties mentioned above. Thus transformation of an amino acid into another is a continuous one if the transformed amino acid falls in the proximity of the original amino acid. However as mentioned already, the change in amino acid is mediated by mutation of the corresponding codons, the change in amino acid must be considered in the perspective of the mutation of codons.

First we introduce a uniformity defined by a proximity relation which is given by all possible one point mutations. With respect to this uniformity on the codon set and the uniformity already discussed on the amino acid set, it was observed that the GP-map is nowhere continuous.

Next we explore for some weaker versions of proximity relation for codons and investigate whether the GP-map is continuous with respect to the uniformities generated by these relations. Accordingly we have considered uniformities on the codon set based on proximity relation defined by different types of mutations.

Any three bases among the four DNA bases form a codon and the importance of base position is suggested by the error (accepted mutation) frequency found in the codons. The frequency of errors decreases from the third base to first base and then next to the second base. That is the second base is biologically most relevant and third base is least relevant base in the codon. The number of hydrogen bond and the chemical type of bases are two important factors involved in codon-anticodon interaction. Taking in mind the evolutionary importance of base positions, hydrogen bond number and the chemical type of bases, we have defined four uniformities based on different mutations.

Case 1: The uniformity is given by the following relations as sub-base.

R1: The mutation of second and/or third base preserve the chemical property (purine changes to purine and pyrimidine changes to pyrimidine).

R2:The mutation of first and/or third base preserve the chemical property (purine changes to purine and pyrimidine changes to pyrimidine).

R3:The mutation of first and/or second base preserve the chemical property (purine changes to purine and pyrimidine changes to pyrimidine).

Case 2: The uniformity is given by the following relations as sub-base.

R1: The mutation of first base preserve the chemical property.

R2: The mutation of second base: preserve the chemical property.

R3: The mutation of third base: preserve the chemical property.

Case 3: The uniformity is given by the following relations as sub-base.

R1: The mutations of first base either preserve the chemical property or have same number of hydrogen bonds.

R2: The mutations of second base either preserve the chemical property or have same number of hydrogen bonds.

R3: The mutations of third base either preserve the chemical property or have same number of hydrogen bonds.

Case 4: The uniformity is given by the following relations as sub-base.

R1: The mutation of second and/or third base changes to a non Watson-crick base.

R2: The mutation of first and/or third base changes to a non Watson-crick base and/or second base remains same.

R3: The mutation of first and/or second base changes to a non Watson-crick base.

IV. RESULTS AND DISCUSSION

Since, there is a canonical map between the set of codons and the set of amino acids, so with the help of this canonical map and the uniformity, we can check the continuity of the map.

From the above four cases we observe that In case 1, the GP map is discontinuous at all codons. In case 2, the GP map is continuous at the codons AUA, GUA, GUG and discontinuous at all the remaining codons.

In case 3, the GP map is continuous at the codons UAC, UAU, GCA, GCC, GCG, GCU, AGA, AGC, AGG, AGU, CGA, CGC, CGG, CGU, UGC, UGU, AUA, AUU, CUC, CUG, GUA, CUU, CUA, GUC, AUC, GUG and discontinuous at all the remaining codons.

In case 4, the GP map is continuous at the codons AAA, AAG, CAA, CAC, GAC, UAC, AUA, GAG, UAU, AUC, GAU, AUG, AUU, CUA, CUC, GUA, GUC, CUG, GUG, GAA, CUU, GUU, CAU, CAG, UUA, UUA, AAU, UUC, UUG, AAC, UUU and discontinuous at all the remaining codons.It is observed that the GP map is continuous at those codons which either code for strong hydrophilic (K, N, Q, H, E, D, Y) or strong hydrophobic (I, M, L, V, F) amino acids.

The 64 codons code the 20 amino acids. Evolution of one amino acid from another is mediated by mutation in the corresponding codons. A change in codons produces corresponding changes in the coded amino acids. So it is reasonable to check the continuity of the evolution of amino acids through mutationof corresponding codons. Different topological structures can be used to check the continuity. We used the notion of uniformity on the codon set as well as on the amino acid set. The uniformity is basically based on some relations and it reflects the corresponding evolutionary pattern of amino acids.

For example, if we consider case 3, the GP map is continuous at AUA. We observe that the codons proximate to AUA are CCA, GCC, CCC, CCU, CUC , GCA, , GCG, GGA, GGG, GGU, UGG, AUA, GGC, AUG, CCG, AUU, CUA, , GUA, GUC, GUG, GUU, UUA, CUG, AUC, GCU, UUC, UUG, CUU, UUU. The codon AUA codes for the amino acid Isoleucine (I) and its proximity codons codes for the amino acids P, A, G, W, M, L, V and F.

Therefore we can say that if the codon mutates to a codon under the uniformity,thenthe continuityof the GP map at AUA reflects the evolution of the amino acid I to P, A, G, I, M, L, V and F.

Next, we consider another codon says UUC, where the GP map is discontinuous:

The codon UUC code for the amino acid Phenylalanine (F). The amino acids which lie within the proximity of the amino acid F are P, A, G, W, I, M, L, V.

Here we noticed that the amino acids F and G are close but their coded codons UUC and GGA are not related under case 3.

V. CONCLUSION

In this paper we have studied how property is preserved in the evolution of amino acids, when evolution is generated by mutation of corresponding codons. One can adopt other framework for studying evolution of amino acids, but we have restricted ourselves to property based setting. We have seen that depending on the proximity relations defined for codon, the continuity of the GP-map changes. We have introduced uniform structure to discuss continuity of GP-map. Other suitable structures like, topology, pre-topology can also be introduced. The model we present here is theoretical and hypothetical and need to be validated by data/ experimental work so as to claim its resemblance to real life situations.

REFERENCES

1. . Fontana, P. Schuster, Continuity in Evolution: On the Nature of Transitions, Science 280, 1998, pp1451-1455.
2. W.Fontana, P. Schuster, Shaping Space: The possible and the Attainable in RNA Genotype-Phenotype Mappings, J. Theor. Biol. 194, 1998, pp 491-515.
3. G.B. Müller and G.P. Wagner, Novelty in evolution: restructuring the concept, Annual Review of Ecology and Systematics, 1991, 22, pp229-256.
4. W. Fontana and L.W. Buss: The arrival of the fittest: Toward a theory of biological organization, Bull. Math. Biol., 1994, 56: pp1-64.
5. J. Cupal, S. Kopp and P.F. Stadler: RNA shape space topology, Alife 6, 2000, pp 3-23.
6. P. Stadler, B. Stadler, Genotype phenotype maps. Biol Theory 3, 2006, pp268–279.
7. W. Fontana, T. Griesmacher, W. Schnabl, P.F. Stadler, P. Schuster, Statistics of landscapes based on free energies, replication and degradation rate constants of RNA secondary structures. Monatsh. Chem., 122, 1991, pp795-819.
8. W. Fontana, D.A.M Konings, P. Schuster, Statistics of RNA secondary structures. Biopolymers, 33, 1993, pp 1389-1404.
9. T. Ali and C.K. Phukan, Topology in Genetic Code Algebra, Mathematical Sciences International Research Journal, 2013, 2(2), pp179-182.
10. A. Weil, Sur les espaces à structure uniformeet sur la topologiegénérale, Act. Sci. Ind. 551, 1937.
11. S. Gnilka, On continuity in extended topologies. Ann. Soc. Math. Pol., Ser. I, Commentat.Math.37, 1997, pp99-108.
12. P. Schuster; W. Fontana, I.L. Hofacker, From sequences to shapes and back: a case study in RNA secondary structures. Proc. Biol. Sci., 1994, 255, pp279-284.
13. K.F. Lau, K.A. Dill, A lattice statistical mechanics model of the conformational and sequence spaces of proteins Macromolecules, 1989, 22, pp3986-3997.
14. F. Mynard, J.S. Gavin, Phenotype spaces, Department of Mathematical Sciences Technical report Series, Georgin Southern University, 2010.
15. G.D. Rose and R. Wolfenden, Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. Ann Rev. BiophysBiomolStruct. 1993, 22, pp381-415.
16. B. Stadler and P. Stadler, The topology of evolutionary biology, Modeling in Molecular BiologySpringerVerlag, 2004, pp. 267-286.

AUTHOR PROFILE

Dr. Nisha Gohain, Assistant Professor Directorate of Open and Distance Learning Dibrugarh University Assam, India. Email: gohainnisha@gmail.com

Reasearch Area: Mathematical Biology, Algebra, Topology.

Publications:

(i) Analysis of amino acids network based on distance matrix: Physica A, 452, 69–78, 2016.**[ISSN 0378-4371]**

(ii) Lattice Structure and Distance Matrix of Genetic Code, Journal of Biological Systems, 23(3), 485-504, 2015. **[ISSN 1793-6470]**

(iii) Reducing Redundancy of Codons Through Total Graph: American Journal of Bioinformatics (Science Publications), 4, doi : 10.3844/ajbsp.2015.1.6, 2015. **[ISSN 1948-9870]**

(iv) Graph Theoretic Approach to Analyze Amino Acid Network: International Journal of Advances in Applied Mathematics and Mechanics, 2(3), 31-37, 2015. **[ISSN 2347-2529]**

(v) Centrality Measures in Graph of Amino Acids: Mathematical Forum, XVI, 74-87, 2014. **[ISSN 0972-9852]**