

# Deep ancestry of orthologs and a theoretical, gradualist perspective for the formation of the LUCA's (Last Universal Common Ancestor) genome

Francisco Prosdocimi<sup>1\*</sup> and Sávio Torres de Farias<sup>2,3\*</sup>

<sup>1</sup> Laboratório de Biologia Teórica e de Sistemas, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

<sup>2</sup> Laboratório de Genética Evolutiva Paulo Leminsk, Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brasil.

<sup>3</sup> Departamento de Filosofia, Programa de Pós-Graduação em Filosofia, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brasil.

\* Correspondence: prosdocimi@bioqmed.ufrj.br; stfarias@yahoo.com.br

**Keywords:** ancestry of orthologs, ancestry of gene families, gene genealogy, FUCA, LUCA, Origins of Life, gradualism, evolutionary biology.

## Abstract

Genes and gene trees have been extensively used to study the evolutionary relationships among populations, species, families and higher systematic clades of organisms. This brought modern Biology into a sophisticated level of understanding about the evolutionary relationships and diversification patterns that happened along the entire history of organismal evolution in Earth. Genes however have not been placed in the center of questions when one aims to unravel the evolutionary history of genes themselves. Thus, we still ignore whether Insulin share a more recent common ancestor to Hexokinase or DNA polymerase. This brought modern Genetics into a very poor level of understanding about sister group relationships that happened along the entire evolutionary history of genes. Many conceptual challenges must be overcome to allow this broader comprehension about gene evolution. Here we aim to clear the intellectual path in order to provide a fertile research program that will help geneticists to understand the deep ancestry and sister group relationships among different gene families (or orthologs). We aim to propose methods to study gene formation starting from the establishment of the genetic code in pre-cellular organisms like the FUCA (First Universal Common Ancestor) until the formation of the highly complex genome of LUCA (Last UCA), that harbors hundreds of genes families working coordinated into a cellular organism. The deep understanding of ancestral relationships among orthologs will certainly inspire biotechnological and biomedical approaches and allow a deep understanding about how Darwinian molecular evolution operates inside cells and before the appearance of cellular organisms.

# 1. Introduction

*common  
ancestry of genes*

Genes and gene trees have been extensively used to study the evolutionary relationships among populations, species, genus, families and higher systematic clades of organisms. Here we propose the inverse thought and methodology. In order to reconstruct the deep evolutionary history and *understand the common ancestry of ortholog genes*, genes must be taken in the center of evolutionary questions. Under this inverted line of reasoning, we will use species as tools to allow a deeper understanding about the evolutionary systematics of ortholog genes and gene families.

*most recent common  
ancestor (MRCA)*

Although we know today that humans share a more recent common ancestor to cows than fishes, for example, we still ignore whether the gene *Insulin* share a more recent common ancestor to *Hexokinase* than to *DNA polymerase* genes. This means that the history of appearance and the intricate pattern of evolution and sister clade relationships between genes and gene families is largely unknown by contemporary researchers.

*from LUCA to the past*

Evolutionary systematics of genes often goes merely until the level of *gene families* and do not scales up in taxonomic hierarchy to allow a deep understanding about the ancient origin of genes in the past. Many conceptual challenges must be overcome to allow this broader comprehension about gene evolution. First, we need to go back in the history of life to get back into a time before cellular organisms existed. The very first cell is normally considered to be LUCA, the Last Universal Common Ancestor (Penny and Poole, 1999; Forterre and Philippe, 1999; Doolittle,

2000; Delaye *et al.*, 2005), better named as LUCeLLA (Last Universal Cellular Ancestor) by some (Nasir *et al.*, 2012). A recent top-down work trying to reconstruct the genome of this last cellular ancestor by looking into genes conserved in the three domains of life (Woese *et al.*, 1990) identified that these organisms possibly presented at least 355 gene families operating to produce a highly complex metabolism (Weiss *et al.*, 2016). We have previously named this epistemological challenge to get our understanding back from LUCA to consider a pre-cellular world as the first dogma for the emergence of biological systems (Prosdocimi *et al.*, 2018a; Prosdocimi *et al.*, 2018b).

*darwinism as  
anti-teleological*

Second, we need to understand that genes cannot be named by a single molecular function as suggested today by international committees like the HGNC Guidelines for Human Gene Nomenclature (Wain *et al.*, 2002; HGNC, 2018). This brings immediately into a teleological idea that genes have evolved to accomplish molecular functions. As anti-teleological by essence, Darwinism must be better integrated into Genomics. Genes are actually highly complex entities with multiple, moonlighting activities (Jeffery, 1999; Copley, 2012), and they must be understood under a wider perspective, named more properly and clustered in higher systematic clades.

*first genes were  
involved in Translation*

This comprehension needs that we go back into the origin of first genes when nucleic acids started to interact with proteins in mutualism (Vitas and Dobovišek, 2018). The First Universal Common Ancestor (FUCA) produced the first genes just after the maturation of an initial genetic code and the process of protein synthesis (Prosdocimi *et al.*, 2018c). These first FUCA genes were most likely encoding ribosomal proteins, primitive tRNA-aminoacyl transferases and other proteins that helped the Translation system to stabilize and get more chances of being maintained (Farias *et al.*, 2016).

*from the first genes  
to the LUCA's genome*

Starting from these putative first genes, a new method must be developed to understand the forward evolutionary process that allowed the production of the few hundred gene families encountered in LUCA genome (355 gene families according to Weiss *et al.*, 2016). The Darwinian evolution from FUCA to LUCA must be proposed based on specific mechanisms and considering gradualist assumptions. These studies will lead into a more profound understand about what is a gene; and we propose here analogies as *gene-as-species* and *gene-as-organ* to provide basis into a better understanding about the very nature of genes.

Ancestral genes reconstruction for each family of orthologs must be produced and evolutionarily classified in clusters and clades. To accomplish this challenge, it may be necessary to go beyond sequence-based approaches and use innovative methods that will take on account information such as 2D and 3D structures of RNAs and proteins.

*cells are complex  
molecular ecosystems*

The *gene-as-species* analogy also implicates that cells must be reinterpreted as highly complex nano-ecosystems of interacting molecules. Thus, ecological concepts must be inherited by molecular biology, such as mutualism, niche construction, competition, population dynamics, etc. The deep understanding of ancestral relationships among orthologs will certainly inspire biotechnological and biomedical approaches and allow a deep understanding about the cell environment and the Darwinian molecular evolution that operates inside cells.

## 2. Teleological Challenge

*finalist thought in  
natural philosophy*

The teleological thought is related to the study of the ends or purposes of something. In western philosophy the term is applied more specifically to the understanding about the final causes of the universe. The question whether universe is operating to achieve specific ends is an important ground to differentiate between (i) science and philosophy, from one side, and (ii) religious beliefs, from another. Greek thinkers were theists and Aristotle described end or purpose as the fourth cause; the others being matter, form and agent. According to the fourth cause, things just exist because there has been a reason (God) that brought it to existence. Inheriting the Aristotelian idea, the natural philosophy tradition saw the organisms existing in Earth as degenerations of ideal organisms existing in the mind of God. The *demiurge* was an artisan-god responsible to imagine perfect organisms in his creative, omnipotent mind. With that wonderful view, this human-like god then sculpted these perfect forms in clay and with His divine blow He brought them to life, to the real world.

*finalist thought in  
daily life*

In our current world of facts, however, a more daily use of teleology is fare and helpful. Each day, we wake up, make plans and go forward to accomplish them step-by-step. Reasoning and planning about our lives require the use of teleology and a Nietzschean willing to put them in action. Therefore, the search for causes and purposes guides our life in society. These currently daily-life facts and reasons make the teleological thought so hardwired in our neural network of thoughts and feelings that is extremely hard to think counter wise.

*finalist thought has  
no role in modern science*

Nevertheless, modern science cannot accept the teleological thought to drive their ends. Under the materialist, nihilist view of the contemporary scientific reasoning, everything that happened in the history of the universe is *contingent* and might not have been. The idea of contingency looks to the past and explains the present by events occurring in the past, not by directional and intentional God-like forces shaping things to be.

*finalist thought  
in biology*

In the history of Biology, the teleological thought has been very recently expurgated from Academia. Until about the year 1850, the most prominent biologists kept explaining the existence of beings by the desire of God. Eminent scientists and natural philosophers from the XVIII<sup>th</sup> century like Carolus Linnaeus (1707-1778) and the Comte de Buffon (1707-1788) believed in that explanation. Jean-Baptiste Lamarck (1744-1829) thought explicitly that organisms evolved in a way to modify their organs and structures to achieve specific goals and was probably one of the most enthusiasts of the teleological thought in Biology. Even Charles Darwin (1809-1882) was educated to accept this view, but something went wrong when he tried to rationalize what he has seen when the HMS Beagle navigated around the world.

*darwinism is  
anti-finalist par excellence*

In 1859, with the publication of "The Origin of Species" by Darwin, Biology underwent a deep modification on its roots and paved its way to become a modern science (Darwin, 1859). The Darwinian thought is anti-teleologic *par excellence*, abolishing any need for causes. *Natural selection* is that incredible force at the hard core of Darwinian theory that chooses the individuals that are already fitted. The selection operated by nature is based in the variability that already exists in any population and, together with the environment, performs a *posthoc* judgment. The individual was already fitted and has been selected because of its fitness.

*biology has nothing  
to do with perfection*

It is clear to understand why the Darwinian thought was extremely controversial at his time; as it keeps being controversial nowadays. Abolishing the needs of a God to explain the origins of biological organisms (and the origins of men), the interpretation of Darwinism also invited naturalists to question the fitness and adaptation of organisms face environments. Instead of focusing in the wonder about near-perfectly fitted organs and structures found in organisms, Darwinism brought to light the enormous complexity of biological systems. Ecological relations among organisms in nature happen under a very narrow equilibrium, near to the imperfection, disorder, randomness and chaos. Any modification on the fine tune of ecosystems can lead to death and extinction; as it has been happening along the entire history of biological evolution in Earth. Darwinian thought therefore turned Biology into mechanistic and materialistic, and it is going to become reductionist with the development of biochemistry, genetics and molecular biology.

### 3. The gene-for-function Challenge

*teleological explanation  
of biological functions*

Even if the whole purpose of beings cannot be traced to any higher forces after Darwin, some subtle uses of teleology can still be found in modern sub-disciplines of Biology such as anatomy, development biology, and genetics. This way, the teleological thought keeps being used to explain the origins and functions of organs; and also the origins and functions of genes and proteins as it was the *raison d'être* of these entities: but under a Darwinian perspective they were *not originated for* anything.



*function X*  
*biological role*

Biologists and biologist teachers/professors keep confusing (i) “why it has appeared” with (i’) “why it has been maintained”. The problem is so serious that Ernst Mayr (probably the most eminent German evolutionary theorist from the XX<sup>th</sup> century) suggested that we should change the term *function* for *biological role* (Mayr, 1992). This seems to be the case once the latter indicates that the gene or organ might perform also other roles and dilutes the feeling of having one single reason of existence.

*teleological function X*  
*history of development*

What needs to be clarified is the fact that function explains merely why the evolutionary process has maintained an organ or a gene. Function may have nothing to do with the primary causes about how it aroused for the first time. In this *gene-as-organ* analogy, in order to understand **why** a given organ and/or gene was originated, one needs to overlook its modern function and tries to study **how** this entity aroused, *i. e.*, **what** was its *history of development*. Also, the suggestion that an organ has a single function is clearly reductionistic; organs may have a main role, but they can often do much more than that.

*one gene-*  
*many enzymes*

The comprehension of gene/protein functions is getting broader over time. Beadle-and-Tatum’s idea of *one gene-one enzyme* was very important at their time (Beadle and Tatum, 1941), but it has been refuted since long with the broad understanding of alternative splicing, for example. Even if we cannot define precisely what a gene is, it is clear that it is complex entity that present multiple aspects, interacts in different forms with different compounds and have never evolved to accomplish any specific function in the teleological sense. Maybe it has achieved a partial function by chance and that function got better under selective pressure, but it has not been designed to anything.

*every protein*  
*shines at the moonlight*

In the last years, many proteins thought to have a single function were shown to present multiple functions, these are being described as *moonlighting proteins* (Henderson and Martin, 2011; Jia et al., 2013; Gancedo et al., 2016; Jeffery, 2018). It seems that genetic community is accepting better the fact that

moonlighting effect is actually the rule for genes and proteins. It is getting hard to find an enzyme that presents one single role in the cellular environment. It is coming the time when researchers that try to associate a gene/protein with one single role will sound extreme reductionist. Genes and proteins can be (i) expressed in different locations in the cell; they can be (ii) expressed in different tissues; can (iii) interact with different cofactors; they can sometimes (iv) form dimers, oligomers; (v) participate in the formation of higher complexes; they (vi) present multiple domains and sites for binding and interaction; (vii) they are target of alternative splicing and/or cleavage; (viii) can be controlled by micro-RNAs and; (ix) can act as epitopes to the immune system. Every gene encodes a moonlight protein.

Following that reason we should take on serious account the criticisms of Stuart Kauffman to the notion of function in Biology (Kauffman, 2011). He makes a good point when he asks: “*What is the function of the screwdriver?*” Even if humans use screwdrivers mainly to screw screws, it has multiple functions and can be used in multiple ways. The screwdriver example denotes clearly that even if something has been teleologically designed by an engineer to accomplish a specific function, it will most likely be used in multiple ways.

It maybe that a better entity to be compared to the gene should another sort of natural kind, such as a cow. We hope that most readers will agree that a gene is more similar to a cow than to a screwdriver. And then we ask: “*what is the function of a cow?*” For us, asking the function of a cow is as awkward as asking the function of a gene. Under this *gene-as-species analogy*, or the naturalist analogy, genes should be viewed as biological species-like entities that perform ecological functions inside a cell. This view also supports the understanding of Biology as a *fractal organized system* on which the gene is to the cell as the species is to the environment. Biology presents different levels and layers, but its logic operates under similar patterns of organization on each level.

*a broader understand  
of function*

*the function  
of a cow*

*the logic of  
biology is fractal*

Contrarily to these questions and reinforcing the longstanding idea that each gene has one single function, HUGO Gene Nomenclature Committee (HGNC), the entity responsible to propose guidelines for gene naming in humans, suggests: “if possible, names should be based on function”. Shall we ask back: “which of them?”

## 4. The natural history of genes

In order to understand the origin and diversification of genes we must look into the scenarios that lead to the initial organization of the biological systems. In the mostly accepted contemporary view, it has been proposed that the first informational molecule has been the RNA. Thus, these molecules capable of self-replication and catalysis have produced an initial pre-cellular metabolism under an RNA-world (Kruger *et al.*, 1992; Gilbert, 1986; Higgs and Lehman, 2015).

Also, the first genes have been produced as RNA molecules that have been duplicated to generate other genes. This process of early gene evolution is in accordance with the current process on which new genes appear nowadays. The origin of new genes is undergone by a process of duplication of an ancient gene and further modification of either one or both duplicates. Under these scenarios genes can subfunctionalize, neofunctionalize or extinguish (Lynch and Force, 2000; Massingham *et al.*, 2001). Genes can also gain or lose specific parts or domains.

What must be notice in those scenarios is that, independently of the path followed after duplication, all gene products have a common origin and, therefore, it can be possible to trace the

*new genes evolve by  
duplication and modification*

*the common ancestry of  
genes is a history to be told*

evolutionary history of gene lineage; at least in theoretical grounds. At this point, the process of cladogenesis of a new gene is similar to the path on which a population gives rise to a species. Thus, we fall again in the *gene-as-species analogy*, allowing us to understand and inherit the same concepts here. Thus, the same assumptions used to categorize species under taxonomic and systematics basis shall be used for genes. After the current theorization on these matters, the application of this knowledge alongside with the development of new methods will allow the production of the complete tree of sister clade relationships among ortholog genes, elucidating the gene tree of life in Earth and allowing geneticists to understand the ancestry of ortholog gene families.

## 5. The ancestry of Orthologs

*different phases of  
gene evolution*

Other consequences from the *gene-as-species analogy* consists in the recognition of genes as entities that evolve inside a cell that are made of DNA, RNA, and proteins that interact with the most different metabolites under a highly complex intracellular ecologic-like environment. In that sense, this analogy invites us to consider each gene as a different species; and we might go further to consider DNA, RNA and protein as different morph types along the maturation of an organism; such as egg, embryo and adult, for example. The role of a gene, let's say the Insulin in the cell, might be seen as analogous to the role of the cow in the ecosystem. The role of each individual cow is to be compared to the role of each individual insulin molecule in a cell.

*the ancestry of  
species is well known*

*the ancestry of  
genes is UNKNOWN*

And then we come to one of the most interesting corollaries of the gene-as-species analogy: with the advance of species' systematics, we know reasonably well what is the ancestry relationship of the cow to almost any other species. For example, it is very clear that cows are more related to humans than to fishes, as cow and humans are common members of the Class Mammalia. However, we have absolutely no idea about the ancestry relationship between, for instance, Insulin, Hexokinase, phosphofructokinase or DNA polymerase genes.

*a Linnean hierarchy  
to classify gene clades*

The gene-as-species analogy makes one immediately imagine a system of *hierarchical evolutionary classification* of genes' in trees; such as we recognize nowadays for species. We know reasonably well which species compose a genus, which genera compose a family, which families compose an order, which orders compose a class, which classes compose a phylum, which phylums compose a kingdom and which kingdoms compose a domain under the history of life in Earth. Evolutionary-based systematics (of species) allowed us to deeply comprehend the ancestry relationships among species and higher species' clades, enlightening our understanding about the evolutionary pattern of kinship among cellular organisms.

*a methodological challenge  
for deep gene ancestry*

However, this evolutionary hierarchical classification is completely lacking when we think in the evolutionary history of genes. Genes individuals and species are studied nowadays, but evolutionary investigations about the ancient history of gene ancestry are nearly absent in literature. In part, this happens due to a methodological problem, that is: as the information content present in nucleic acid molecules is limited, the very same nucleotidic base (on a specific position in a given gene) can mutate so many times that the these recurrent mutations (named homoplasy) avoid the knowledge about the ancestral state of the molecule, bringing noise into the evolutionary account of gene ancestry using sequence-based analyses. Therefore, DNA and RNA sequence data are considered nowadays as too noisy to allow the reconstruction go far in the

past. On the other hand, the use of other polymorphic features beyond sequences may allow a more trustworthy classification, such as (i) 2D and 3D structures of RNAs and proteins, (ii) presence of conserved motifs and signatures, (iii) binding of cofactors, (iv) codon usage, (v) presence of ancient codons (RNY) (Shepherd, 1981) and amino acids (Gly, Ala, Ser) and many other molecular characters. These features will need to be evaluated with precision to allow a *bona fide* construction of a reliable character-state matrix that will allow a somewhat precise classification of gene ancestry relationships.

*common ancestry  
among orthologs*

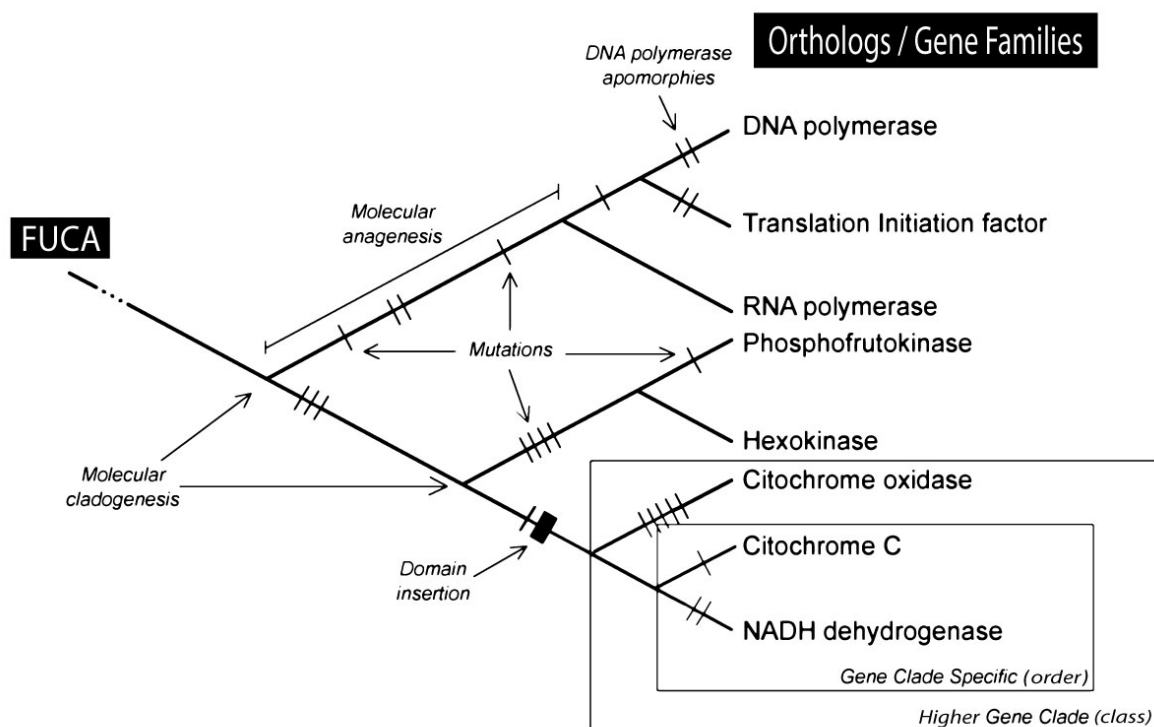
Contemporarily, normal science in evolutionary biology denies the use of different gene families in a single tree. According to the standard view, genes must be homologs (orthologs) to allow tree reconstruction. We understand that this rule represents a narrow view of gene evolution and whether we extend the time to the past, it becomes clear that different orthologs will also share a common gene ancestry.

Also, gene-tree reconstructions have been performed to answer questions about species' relationships. This is a paradigm that must be overcome to allow the understanding of deep evolutionary relationships amongst orthologs. Figure 1 suggest how new gene trees should be built under the current proposal.

*gene evolution:  
from FUCA to LUCA*

Changing the conceptual framework as we propose here will need new methods that make profit of using species from the most different taxa in order to understand *the evolutionary relationship among genes*. We want to know all the evolutionary order on which genes arose since FUCA until LUCA, *i. e.*, from the first to the last universal common ancestor. Weiss *et al.* (2016) present evidence that LUCA's genome should contain about 355 gene families. On the other hand, we conceptualized that FUCA probably produced the first genes that were most likely parts of the larger ribosomal subunit (Prosdocimi, José and Farias, 2018) and accessories to the initial formation of the protein synthesis apparatus and the genetic code. But what

happened with gene evolution from FUCA to LUCA? How the biochemical pathways of LUCA have been assembled? How stabilizing proteins that built the ancient ribosome and the genetic code were co-opted to produce metabolic pathways? This is still a story to be told.



**Figure 1:** An illustrative evolutionary tree to understand *the ancestry of orthologs*. Different orthologs (or gene families) should be OTUs (operational taxonomy units) in gene-centered trees. Boxes indicate the proposal of gene clades higher than families based on evolutionary classification. Understanding deep evolutionary relationships among genes is crucial to understand the origins of life and the emergence of biological systems. This knowledge can provide important clues to understand the origin of multimeric complexes and biochemical pathways in both pre-LUCA and post-LUCA scenarios, with possible applications for biotechnology and biomedical innovations.

*protein structure  
databases may store  
ancient relations*

Although it is nowadays missing a good method to understand the deep evolution of genes, there are a couple methods that can be already used with reasonable certainty. First, even sequence analyses can indicate more recent relationships between orthologs. With sequence-based approach we can understand the most recent histories of gene formation, mainly working with genes that were originated in recent species' clades. Protein structures can also help us to cluster gene families into older relationships of ascendance. Thus, services like CATH (Dawson et al., 2017) and SCOP (Murzin et al., 1995) have not been scrutinized as important sources of ancient gene ancestry relationships, but they clearly contain data to allow a reconstruction.

More ancient relationships can be possibly glimpsed using methods currently used to the production of ancestral genes. Farias et al. (2015) have shown that the production of ancient sequences of tRNA could reveal features to the formation of the first genes. Using different levels into the parameter of Complete Deletion to build gene ancestors under maximum likelihood approaches, the works of Farias' group are demonstrating that it is possible to rebuild the ancestral core of molecules to understand their evolution. The production of ancestor molecules must be done with precision, using appropriate models of nucleotidic evolution.

*protein evolution  
by accretion*

As expected, proteins possibly evolved by the risen of inefficient, error-prone catalytic sites. Adding new amino acids that formed new stabilizing layers through the *mechanism of accretion* has further protected these slightly catalytic peptides. Natural selection acting molecularly would further allow the formation of protected catalytic sites that achieved better chemical efficiency and stabilized back the whole system. This way, specialized enzymatic activities or sites for molecule binding may have been created.



## 6. Conclusions

*back to the emergence  
of biological systems*

In order to rebuild the ancient history of genes, we need to put the question about genes in focus when building evolutionary trees. To do that, we must understand that most genes are descendants of a first encoded gene that has been originated at the emergence of biological systems, when the genetic code started to be established in the pre-cellular organism named FUCA. Although genes can arise by different mechanisms, the most known form is by divergence after duplication.

At the time of FUCA, the first encoded genes were most likely some sort of proto-ribosomal proteins that stabilized the relationship between RNAs and oligopeptides encoded by them (Prosdocimi, José and Farias, 2018). It is likely that genes involved in the process of Protein Synthesis, such as tRNA-aminoacyl transferases and protein factors have been the first encoded genes, even if more recent and efficient proteins could have replaced some of them further.

This means that genes can be understood to be evolutionary linked in the past under a tree-shaped pattern. There is also the possibility that gene modules and motifs have been the actual agents of evolution. It is clear that shuffling of motifs and functional modules happened as genes evolved. These events make the past history of genes highly complex to be measured and evaluated. Nevertheless, a significant part of gene evolution was made of duplications of ancient genes that passed through process of mutational divergence and further sub/neo-functionalization. Whole biochemical pathways can be explained to evolve by sub-functionalization and specialization of specific activities of binding and catalysis. Starting with the

random formation of inefficient catalytic sites with enzymatic activity, the duplication of an initial gene that bound and modified some molecule very inefficiently may have evolved into two new genes. In one of the new copies, a more precise and specialized binding could evolve; while in the other, a catalytic site could also ameliorate after random mutations and natural selection. This seems to be the principle of sub/neo-functionalization and a number of examples do exist evidencing this mechanism as a common path for the evolution and engineering of genes (Hahn et al., 2009; Conant et al., 2014; Freeling et al., 2015).

Regarding gene nomenclature, our proposal here is that genes must have proper names or; alternatively, they should be classified in binomial types such as species and possibly they can be Latinized or Esperantized to produce Linnean-like names. The worst option, however, is keeping the gene name associated to a single function. This is an extreme reductionism and masks the complex nature of each gene, bringing teleological, non-darwinian thoughts to obscure our comprehension about the genetics. Most genes have moonlighting properties and they must be understood as complex as they are.

Cells must be understood as highly multifaceted environments on which different clades of genes interact under multi-level and complex systems biology. The gene can probably be interpreted as a multi-form entity that can be DNA, RNA and proteins. Under each morphotype, genes relate with the environment differently, interacting with different molecules and producing different outputs to the environment.

The knowledge about the stepwise transitions that make the initial genes from FUCA to produce the complex genome of LUCA (with 355 gene families) must be studied. Even if we do not have good methods today to get into this sort of knowledge, it is clear that the production of ancestral genes and the study of

RNA and protein structures will allow many gene clades to be constructed. Step by step, the whole scenario will be known. Maybe in a few years or dozens of years from now we will be able to understand well the deep ancestry relationships among orthologs. This will certainly make clearer how the biochemical pathways were built gene by gene. And then, intelligent design advocates will need to find other arguments than the therefore refuted irreducible complex (Behe, 1996).

**Conflicts of Interest:** The authors declare no conflict of interest.

## 7. Acknowledgements

We would like to thank FAPERJ (JCNE E-26/202.810/2015) for funding FP.

## 8. References

1. Beadle GW, Tatum EL. Genetic Control of Biochemical Reactions in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America*. 1941;27(11):499-506.
2. Behe, M. Darwin's black box: the biochemical challenge to evolution. Free Press (1996).
3. Conant GC, Birchler JA, Pires JC. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr Opin Plant Biol*. 2014 Jun;19:91-8.
4. Copley SD. Moonlighting is mainstream: paradigm adjustment required. *Bioessays*. 2012 Jul;34(7):578-88. doi: 10.1002/bies.201100191. Review. PubMed PMID: 22696112.
5. Darwin CR, 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray. [1st edition]
6. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, Orengo CA, Sillitoe I. CATH: an expanded resource to predict protein function

through structure and sequence. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D289-D295.

7. Delaye L, Becerra A, Lazcano A. The last common ancestor: what's in a name? *Orig Life Evol Biosph.* 2005, 35(6):537-54.

8. Farias ST, Rêgo TG, José MV. tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. *Life (Basel).* 2016 Mar 23;6(2). pii: E15. doi: 10.3390/life6020015.

9. Freeling M, Scanlon MJ, Fowler JE. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Curr Opin Genet Dev.* 2015 Dec;35:110-8.

10. Forterre P, Philippe H. The last universal common ancestor (LUCA), simple or complex? *Biol Bull.* 1999 Jun;196(3):373-5; discussion 375-7.

11. Gancedo C, Flores CL, Gancedo JM. The Expanding Landscape of Moonlighting Proteins in Yeasts. *Microbiol Mol Biol Rev.* 2016 Jul 27;80(3):765-77.

12. Gilbert W. The RNA world. *Nature* 1986, v.319, p. 618.

13. Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered.* 2009 Sep-Oct;100(5):605-17.

14. Henderson B, Martin A. Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. *Infect Immun.* 2011 Sep;79(9):3476-91.

15. Higgs PG, Lehman N. The RNA World: molecular cooperation at the origins of life. *Nat Rev Genet.* 2015 Jan;16(1):7-17. doi: 10.1038/nrg3841.

16. HGNC (HUGO gene nomenclature committee) website: <https://www.genenames.org/about/guidelines> accessed August, 15th (2018).

17. Jeffery CJ. Moonlighting proteins. *Trends Biochem Sci.* 1999 Jan;24(1):8-11. Review. PubMed PMID: 10087914.

18. Jeffery CJ. Protein moonlighting: what is it, and why is it important? *Philos Trans R Soc Lond B Biol Sci.* 2018 Jan 19;373(1738). pii: 20160523.

19. Jia B, Cheong GW, Zhang S. Multifunctional enzymes in archaea: promiscuity and moonlight. *Extremophiles.* 2013 Mar;17(2):193-203.

20. Kauffman S. There Are More Uses For A Screwdriver Than You Can Calculate. *Cosmos and Culture.* 2011 April;4(4) <https://www.npr.org/sections/13.7/2011/04/04/135113346/there-are-more-uses-for-a-screwdriver-than-you-can-calculate> (accessed on August, 15<sup>th</sup>, 2018).

21. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell.* 1982, 31(1):147-57.

22. Lynch M, Force A. The probability of duplicate gene preservation by subfunctionalization. *Genetics*. 2000 Jan;154(1):459-73.
23. Massingham T, Davies LJ, Liò P. Analysing gene function after duplication. *Bioessays*. 2001 Oct;23(10):873-6.
24. Mayr E. The Idea of Teleology. *J Hist Ideas*. 1992, 53(1): 117-135
25. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*. 1995 Apr 7;247(4):536-40.
26. Nasir A, Caetano-Anollés G. A phylogenomic data-driven exploration of viral origins and evolution. *Sci Adv*. 2015 Sep 25;1(8):e1500527.
27. Penny D and Poole A. The nature of the last universal common ancestor. *Curr Opin Genet Dev*. 1999, 9(6):672-7.
28. Prosdocimi F. Teleological Thought in Gene Naming: The Quest for an Evolutionary-Based Taxonomy and Systematics for Genes. *J Phylogen Evolution Biol* 2015, 3(2): 152.
29. Prosdocimi F, Jheeta S, Farias STD. Two Dogmas for the Emergence of Biological Systems: Cell Theory and Self-Replication. *Preprints* 2018a, 2018050445 (doi: 10.20944/preprints201805.0445.v1).
30. Prosdocimi F, Jheeta S, Farias STD. Conceptual challenges for the emergence of biological system: Cell Theory and Self-Replication. *Medical Hypotheses* 2018b, 119: 79-83.
31. Prosdocimi, F.; José, M.V.; Farias, S.T.D. Be Introduced to the First Universal Common Ancestor (FUCA): The Great-Grandmother of LUCA (Last Universal Common Ancestor). *Preprints* 2018c, 2018060035 (doi: 10.20944/preprints201806.0035.v1).
32. Shepherd JC. Periodic correlations in DNA sequences and evidence suggesting their evolutionary origin in a comma-less genetic code. *J Mol Evol*. 1981;17(2):94-102.
33. Vitas M, Dobovišek A. In the Beginning was a Mutualism - On the Origin of Translation. *Orig Life Evol Biosph*. 2018 Jun;48(2):223-243.
34. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. 2016, 25;1(9):16116.
35. Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S. Guidelines for human gene nomenclature. *Genomics*. 2002 Apr;79(4):464-70.
36. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 1990, 87(12):4576-9.