Journal of Phylogenetics & Evolutionary Biology

Prosdocimi, J Phylogen Evolution Biol 2015, 3:2
DOI: 10.4172/2329-9002.1000152

Commentary Open Access

Teleological Thought in Gene Naming: The Quest for an Evolutionary-Based Taxonomy and Systematics for Genes

Francisco Prosdocimi*

Laboratório de Biologia Teórica e de Sistemas, Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Brazil

Abstract

The teleological though is a well-known misunderstanding of evolutionary biology and it is frequently understood as the vision that organs, species (and genes) have evolved to achieve specific ends. The theory of evolution as predisposed by Charles Darwin is based on differential reproduction of randomly produced variants with no direction. This commentary intend to look into molecular genetics and genomics trying to point out major evolutionary conceptual problems on naming genes according to their most likely, first-described function. Naming genes in accord to their function may lead to misunderstandings and finalistic thoughts assumptions that genes might have "evolved for" accomplishing that function. Besides, the idea that a gene would be responsible for a single and unique function in the cellular environment became over simplistic. We now accept that genes have multiple splicing variants, messenger RNAs can attach multiple molecules in the cytoplasm, proteins have multiple sites of recognition by the immune system (epitopes), interact with many other proteins and most genes present regulatory roles in the cell metabolism. Gene name databases and biological ontologies consist in a powerful repositoire of gene characteristics and will certainly help further studies of gene systematics and taxonomy. Genes should be considered free-living entities evolving by the processes of natural selection and self-organization that cannot foresee functions to evolve for. Here I discuss the intellectual teleological background behind naming genes based on their functions and suggest new approaches for gene naming and classification in the future.

Introduction

The thinking that organisms have evolved and modified their anatomical constitutions "to achieve specific aims" is probably one of the greatest misunderstandings related to biology and evolution. This kind of thought is known as the teleological thought.

Biological Darwinian evolution does not operate by causative, direct agents. The environment influences the genetic material just subtly by the force of natural selection. Random mutations will act in the DNA to change nucleotide bases that after translation may further generate modifications in amino acids, changing the 3D structure of the codified protein. Most non-synonymous mutations will slightly damage a given protein structure that has been working well along generations. Once in a while, those mutations will produce a molecular modification that will turn the individual harboring it fitter. The raise of fitness will make this individual capable to keep being alive and producing more offspring, making its genes more represented in further generations.

This molecular updated view of Darwinian theory has been challenged in the late 60's when it has been discovered that mutations that raise fitness consist in a vast minority and actually most genetic characteristics has a quasi-zero advantage in fitness. The classical paper by King and Jukes [1] was actually called "Non-Darwinian Evolution" and along with the paper of Motoo Kimura "Evolutionary Rate at the Molecular Level" [2] inaugurated a time of controversies. According to the neutralists, evidences have shown that randomness of mating and genetic drift play the major role in the evolution of complex organisms. According to them, natural selection would be an overestimated force in biological evolution.

Besides, in very few occasions a clear relationship between the genotype and the phenotype has been observed. These genes to character relationships have been difficult to be found and hard to be measured with precision. Therefore, modern geneticists seem to think that complex organisms evolve independent from the environment in many aspects.

On the other hand, since we live in a world on which facts can be traced to their specific and directed causes, the teleological thought is an intuitive idea that comes into mind when someone consider any topic of interest, such as the evolution of organisms - and also the evolution of genes. This finalist view has been extensively discussed by the most prominent philosophers of biology [3-7] although it has being applied mainly into the understanding on the evolution of species. What if we look into molecular genetics and contemporary genomics trying to see how teleology is conceptually buried in our thought?

Thought experiment: a teleological nomenclature for genes

At this point, it would be interesting to propose a thought experiment: let us suppose that someone would like to name a biological entity (a gene) according to a teleological perspective. Of course, it is wrong to do it. But it is just an experiment. So, let us suppose that a gene would have actually evolved to do something. And this something would be codifying a protein with a given enzymatic activity. Thus, let us name the gene with the enzymatic activity it has "evolved to" perform!

Although forcing the problem into a biased perspective, this is actually how most genes are named: based in the enzymatic activity they perform. For example, the genes of the first part of the glicolytic pathway are: Hexokinase, Phosphoglucose isomerase, Phosphofructokinase, Fructose-bisphosphate aldolase and Triosephosphate isomerase; all named after the catalytic reaction they perform even though many also play important regularory roles. Actually, the official entity for naming human genes, HUGO (Human Gene Nomenclature Committe) suggests that: "if possible, names should be based on function" [8]. It is clear that HUGO make an important job on defining unique names for genes, but it is time to go further in gene naming and systematics.

Even though evolutionists have always been ideologically fighting against the finalistic view of evolution; it seems to be incorporated in the intellectual background of scientific disciplines of biochemistry, genetics and molecular biology. It is reasonable to suppose that

***Corresponding author:** Francisco Prosdocimi, Laboratório de Biologia Teórica e de Sistemas, Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Brazil, Tel: +55 21 3938 6759; E-mail: prosdocimi@bioqmed.ufrj.br

Received April 05, 2015; **Accepted** May 05, 2015; **Published** May 12, 2015

Citation: Prosdocimi F (2015) Teleological Thought in Gene Naming: The Quest for an Evolutionary-Based Taxonomy and Systematics for Genes. J Phylogen Evolution Biol 3: 152. doi:10.4172/2329-9002.1000152

naming genes in accord to their putative function might lead to misunderstandings and finalistic thoughts that genes have "evolved for" accomplish a given function.

Actually, defenders of the intelligent design used the view that enzymes such as ATP synthase are highly complex molecular machines impossible to have evolved by natural selection together with teleological argumentation to disqualify the entire field of evolutionary biology [9]. Rephrasing old criticisms to evolution such as Paley's watchmaker argument, they suggest that multimeric, complex enzymes would require an intelligent maker.

Thus, a deeper understanding about the evolutionary nature of genes and how they have evolved from initial replicators must follow. Naming genes after the molecular reactions their protein perform is somehow accepting a finalistic view of evolution that has proven to be mistaken. In order to solve this problem we will need to study better gene ancestrality [6] and gene ressurection, for example [10-12]. Besides, a new tree-based system of gene taxonomy and systematics must arise.

Many discussions about the inappropriateness of gene naming in scientific journals have taken this relevant topic into account [13-17]. Moreover, a number of web-sites seem to refer into these gene-name problems as mere curiosities [18,19] .Maybe it is time to rethink that issue.

Gene naming known problems

Other challenges are related to the identification of common problems in gene naming. Seringhaus and collaborators (2008) classified gene names in four main classes:

(1) The ones presenting explicit meaning

(2) No explicit meaning

(3) Transferred naming system and

(4) Other problematic relationships.

Genes presenting explicit meaning are the ones on which their names reflect "in some intelligible way an underlying characteristic" of it. No-explicit meaning stands for genes named using funny or non-obvious characteristics as well as any series of non-meaningful characters defined by gene-predictor softwares. The transferred naming system is the case of arabidopsis genes superman and kryptonite on which one gene is named due to its relationship with another ones. The last of Seringhaus categories on which genes' names have shown to be "problematic". One example of this would be the case on which a gene has been named according to a given function that has further been proved as incorrect or misleading. Once more, the problem of associating gene names with their presently known function. Many researches have shown genes presenting more than one known function [20-25]. Besides, each protein has multiple sites of recognition by the immune system (epitopes), genes have multiple splicing variants and messenger RNAs can attach multiple molecules in the cytoplasm, making that a functional gene nomenclature system become over simplistic and naive. The dogma one enzyme-one reaction is broken: every gene is probably performing multiple functions in the cellular environment. The gene should be seen as an independent evolving entity and its codified protein has at least many immunological properties as an epitope besides its molecular functions. Why geneticis keep given gene names based on their "function"?

Thus, the problem with gene naming may be even more problematic even when they present a clear explicit meaning. Nevertheless, the most used documents prepared to help researchers to give name to genes, suggest exactly this approach and the Guidelines for Human Gene Nomenclature (1997) states the following in the section 1.3 about "Gene names": "1.3.1. Gene names should be brief and specific and should convey the character or the function of the gene" [26]. The problem is: the explicit meaning is almost always referring to the most likely and firstly described function of a gene. And this assumption may be wrong.

A critique to the idea of function in biology

Stuart Kauffman from the Santa Fe Institute is one the world leaders in the theory of chaos, in the study of complexity and self-organization .In a conference called "Evolution beyond Reductionism" (João Pessoa, Brazil, December 2014), he criticized the concept of function in biology. He asked to the audience: "What is the function of a screwdriver?" Although screwdrivers have been developed to screw, one could use them in an enormous multiplicity of purposes. This means that, even if an entity has been developed to do something, it can be used for many other issues.

In the case of genes, modern evolutionary theory suggests they did not evolve to any purpose at all. Genes should be understood as free-living molecular entities on which natural selection can act to shape it subtly.

The main argument to present is the idea that associating gene names to their functions can be seen as teleology. Scientific names of species are definitely not chosen according to the species' function. In fact, what would be the "function" of a given species? The process on which genes evolve is analogous to the process on which species evolve and it is so that their names should not be given based on their function as understood by biochemists. Such as species, genes exist as physical entities and they are subject to mutations, genetic drift and natural selection. The ones capable to cooperate better with others and their surrounding environment will survive and become represented in the next generations. Both genes and species execute a number of processes that may be called "functions", even if they do not "exists for" doing them. Genes and species exist only as evolutionary entities: they do not exist "to" accomplish anything. It would be really strange to call a cow by such a name like "milk producer" or "meat producer". If humans use organisms to some goal, it does not mean that organisms really exist for that and the same argument is valid when thinking about genes.

Description of the main characteristics of genes

However, even if a gene should not be named in accord to its function, it is clear that genes codify proteins that perform catalytic activities and these characteristics must be somehow attached to its representation. Then we finally come into some of the ideas originated mainly under the push of genomics sciences. First, the idea of creating a controlled vocabulary to describe the characteristics of gene [27] will certainly allow us to avoid the using of gene functions as their names.

Recapturing classical Darwinian theory, Thomas Huxley divides the knowledge about organisms in two main fields: morphology and physiology [28]. According to Huxley, morphology is the field that describes an organism in terms of their structural peculiarities while physiology represents the description of species regarding their environmental function. Besides the necessity of a common vocabulary to speak about gene names, there is the necessity to describe genes in regard of their morphology and physiology. The morphology of genes shall be understood as the characters underlying their reality in the physical world: such as their most common position in a given genome, their wild-type sequence of nucleotides and the related RNAs

and proteins they produce. The morphology of the genes should be described and named by their 3D-structure and it seems necessary to create ontologies such as a PDB-file format that would describe morphological characteristics of genes and proteins to be used in this complex characterization.

Considering physiology, a gene must be described in terms of which molecular function it performs, which biological processes it participates, which molecules it interacts as well as its relevance for known metabolic problems or associated diseases. Some other characters, such like the position in the cell on which genes' products have shown to be present (as physical entities) have helped researchers to infer their putative functions and make the link between the field of gene morphology and physiology.

Biological ontologies

Most of these described worries on the representation of genes in human understandable ways have been performed in the last years through a number of different ways. Many genome projects have been finished and genome annotated Genbank files [29] frequently present the position of genes in the genomes of sequenced species as well as their most likely RNA coding sequences and protein sequences in FASTA format [30]. Associated files produced during the time-consuming process of complete genome annotation are also available, linking genes into their molecular functions and the putative biological processes they have been involved [31-33]. Moreover, functional genomics data have developed bringing into light the role of previously unknown genes and suggesting new roles for known genes [34-37].

We might say that the problem in defining standards for gene representation have been solved recently, at least in great part, through the use of biological ontologies [38,39]. Some biological ontologies are used as common vocabularies defined to describe genes' physiological and morphological characteristics, such as gene ontology (GO) [27] and sequence ontology (SO) [40] respectively. GO is the most widely used ontology in molecular biology and it may be said that all the new sequenced genomes have their genes classified on each of the three main sub-ontologies. Thus, GO provides a controlled vocabulary of terms in regard of the molecular function, biological processes and the cellular components on which genes have been associated and a high number of articles have been published recently presenting further analyses based in this resource, corroborating its efficacy and utility for a number of applications. The SO is a newer resource and unhappily the new sequenced genomes still do not use this kind of annotation very often, although it has already been used successfully by the biggest genome resources on model organisms, such like the WormBase, FlyBase and the Mouse Genome Informatics group. Moreover, the usage of controlled vocabulary to describe sequence features has been proven to help annotation and database curation through the use of text-mining algorithms [41].

The advantage of using controlled vocabularies is enormous since it allows, for example, an easy and even automatic comparison a number of gene features between the genomes of a number of species.

Therefore, considering that (1) genes are entities evolving by no causative agent at all, (2) most genes have proven to present more than one function, and (3) we have already good ways to describe genes' features both morphologically and phisiologycally, what we probably need now is a restructuration on the way we give names to genes. Moreover, since genes evolve in similar ways to species (by random variation and selection), it has been already suggested that genes might be better described using binomial Linnaean names [42].

Modern ways to give names to genes

According to the HUGO Gene Nomenclature Committee (HGNC), the committee responsible to approve genes names in humans, the problems of nomenclature in human genetics were recognized as early as the 1960s and full guidelines for human gene nomenclature were presented at the Edinburgh Human Genome Meeting in 1979 [43]. Since then, groups of molecular biologists and bioinformaticians have been working to produce a controlled gene nomenclature database of terms containing approved symbols and names for each human gene [44]. New human genes shall be named following a number of rules described in the HGNC Guidelines for Human Gene Nomenclature which present in its very first page the following definition of a gene: "A gene is a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a gene may be characterized by sequence, transcription or homology." Therefore, even the concept of gene for modern gene taxonomists is primarily associated with its function. It was already demonstrated in here that reducing gene *raison-d'etre* to its function is incurring in teleology. Genes are free-living entities evolving by the processes of natural selection and self-organization that cannot foresee functions to evolve for.

Suggestions for a new gene nomenclature and classification

Since Carolus Linnaeus, organisms have been classified into groups due to the presence of some particular characteristics. Looking into organisms and their similarities and/or differences, naturalists have created criterions to group them together into a taxonomic class in a way that putatively represent their ancient evolution from a common ancestor. As already pointed out, it is time for an evolutionary approach on the origin of genes and the study of molecular Darwinism will be able to produce answers about the common ancestry relationships among all known genes. This knowledge about the origin and evolution of genes will also allow the study of many other interesting topics in biology. It will also be interesting to study mechanisms of "genetic speciation" (molecular cladogenesis) and how a duplicated gene evolves by accumulating mutations and diverging from its ancestors.

Of course, the whole molecular biology community must take such effort and the associations created to name genes: such as HUGO, must be convinced about the relevance of the present argumentation. It is maybe time for a new consortium on gene taxonomy and systematics.

Conclusion

Both (i) the aristotelic ladder of beings growing up from simple organisms "to" achieve the human complex brain, behavior and conscience and (ii) the lamarckian giraffes that have been longing their necks "to be able to" get the high leaves in trees are widely known examples of evolutionary mistakes associated to a finalistic view of evolution. Although Darwin's theories brought those thoughts to history, the way contemporary researchers name genes according to their molecular function is still contaminated by the teleological thought.

Such as species, genes do not evolve to perform any given function; genes accumulate random mutations and the ones that happen to cooperate in an integrated fashion, helping organisms' adaptation and reproduction are kept in the genome and passed on through generations.

Therefore, genes should not be named after their functions using names unrelated to their functions. Genes should be classified in groups related to their similarities and common ancestry, just like it has been done with species. It is possible that Linean taxonomy might be used

and, thus, genus, families, classes, orders and other taxonomic groups might be created to group together genes presenting some defined molecular characteristics that will attest for their lineage of common ancestry deriving from LUCA's genome or even earlier in time. We look forward to a time, hope not so far, on which evolutionary genes' names and evolutionary relationships will be better understood and classified.

References

1. King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164:788-98.
2. Kimura M (1968) Evolutionary rate at the molecular level. Nature 217: 624-626.
3. Huxley T (1870) Lay sermons, addresses and reviews. London.
4. Mayr E (1974) Teleological and teleonomic. A new analysis. Boston Studies in the Philosophy of Sciences 14: 91-117.
5. Mayr E (1988) Toward a new philosophy of biology. Cambridge, Massachusetts, Harvard University Press.
6. Mayr E (2004) What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline. Cambridge, UK, Cambridge University Press.
7. Ruse M (1973) The philosophy of biology. London, Hutchinson.
8. Wright MW (2014). A short guide to long non-coding RNA gene nomenclature. Hum Genomics 8:7.
9. Behe MJ (1996) Darwin's Black Box: The Biochemical Challenge to Evolution, Free Press, New York.
10. Cai W, Pei J, Grishin NV (2004) Reconstruction of ancestral protein sequences and its applications. BMC Evol Biol 4:33.
11. Thornton JW, Need E, Crews D (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science. 301:1714-1717.
12. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW (2012) Evolution of increased complexity in a molecular machine. Nature 481:360-364.
13. Vacek M (2001) A gene by any other name: whimsy and inspiration in the naming of genes. Am Sci November/December.
14. Petsko GA (2002) What's in a name? Genome Biol 3: COMMENT100.
15. Schwartz J (2006) 'Sonic Hedgehog' sounded funny, at first. The New York Times.
16. Seringhaus MR, Cayting PD, Gerstein MB (2008) Uncovering trends in gene naming. Genome Biol 9: 401.
17. Lacroix M (2009) Poor usage of HUGO standard gene nomenclature in breast cancer studies. Breast Cancer Res Treat 114:385-386.
18. Flynome (2001). Flynome, a database for Drosophila nomenclature.
19. Niku M (2001) Clever gene names.
20. Driessen AJ (2001) SecB, a molecular chaperone with two faces. Trends Microbiol 9: 193-196.
21. Zeng X, Kaplan S (2001) TspO as a modulator of the repressor/antirepressor (PpsR/AppA) regulatory system in Rhodobacter sphaeroides 2.4.1. J Bacteriol 18: 6355-6364.
22. Ebnet K, Suzuki A, Ohno S, Vestweber D (2004) Junctional adhesion molecules (JAMs): more molecules with dual functions? J Cell Sci 117:19-29.
23. Ryan RM, Vandenberg RJ (2005) A channel in a transporter. Clin Exp Pharmacol Physiol 32: 1-6.
24. Takagi M, Absalon MJ, McLure KG ,Kastan MB (2005) Regulation of p53 translation and induction after DNA damage by ribosomal protein L26 and nucleolin. Cell 123: 49-63.
25. Ni JQ, Liu LP, Hess D, Rietdorf J, Sun FL (2006) Drosophila ribosomal proteins are associated with linker histone H1 and suppress gene transcription. Genes Dev 20: 1959-1973.
26. White JA, McAlpine PJ, Antonarakis S, Cann H, Fraser K, et al. (1997) Guidelines for human gene nomenclature.
27. The Gene Ontology Consortium (2001) Creating the gene ontology resource: design and implementation. Genome Res 11: 1425-1433.
28. Huxley T (1893) Darwiniana. London, Mcmillan.
29. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. Nucleic Acids Res 36: 25-30.
30. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35(Database issue): D61-65.
31. Consortium CeS (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282: 2012-8.
32. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921.
33. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.
34. Burke C, Thomas T, Egan S, Kjelleberg S (2007) The use of functional genomics for the identification of a gene cluster encoding for the biosynthesis of an antifungal tambjamine in the marine bacterium Pseudoalteromonas tunicata. Environ Microbiol 9: 814-818.
35. Rezen T, Contreras JA, Rozman D (2007) Functional genomics approaches to studies of the cytochrome p450 superfamily. Drug Metab Rev 39: 389-399.
36. Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, et al (2008) Cancer proliferation gene discovery through functional genomics. Science 319: 620-624.
37. Wang X, Jia S, Geoffrey R, Alemzadeh R, Ghosh S, et al. (2008) Identification of a Molecular Signature in Human Type 1 Diabetes Mellitus Using Serum and Functional Genomics. J Immunol 180: 1929-1937.
38. Bada M, Hunter L (2007) Enrichment of OBO ontologies. J Biomed Inform 40: 300-315.
39. Smith B, Ashburner M, Rosse C, Bard J, Bug W, et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25: 1251-1255.
40. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, et al. (2005) The Sequence Ontology: a tool for the unification of genome annotations. Genome Biol 6: R44.
41. Vlachos A, Gasperin C, Lewin I, Briscoe T (2006) Bootstrapping the recognition and anaphoric linking of named entities in Drosophila articles. Pac Symp Biocomput: 100-111.
42. Seringhaus MR, Cayting PD, Gerstein MB (2008) "Uncovering trends in gene naming." Genome Biol 9: 401.
43. HUGO-Gene-Nomenclature-Committee (2007). About the HGNC. Cambridge. 2008.
44. Eyre TA, Ducluzeau F, Sneddon TP, Povey S, Bruford EA et al. (2006) "The HUGO Gene Nomenclature Database, 2006 updates." Nucleic Acids Res 34:319-321.