An Evolutionary Classification of Genomic Function

Dan Graur*, Yichen Zheng, and Ricardo B.R. Azevedo

Department of Biology and Biochemistry, University of Houston

*Corresponding author: E-mail: dgraur@uh.edu.

Abstract

The pronouncements of the ENCODE Project Consortium regarding “junk DNA” exposed the need for an evolutionary classification of genomic elements according to their selected-effect function. In the classification scheme presented here, we divide the genome into “functional DNA,” that is, DNA sequences that have a selected-effect function, and “rubbish DNA,” that is, sequences that do not. Functional DNA is further subdivided into “literal DNA” and “indifferent DNA.” In literal DNA, the order of nucleotides is under selection; in indifferent DNA, only the presence or absence of the sequence is under selection. Rubbish DNA is further subdivided into “junk DNA” and “garbage DNA.” Junk DNA neither contributes to nor detracts from the fitness of the organism and, hence, evolves under selective neutrality. Garbage DNA, on the other hand, decreases the fitness of its carriers. Garbage DNA exists in the genome only because natural selection is neither omnipotent nor instantaneous. Each of these four functional categories can be 1) transcribed and translated, 2) transcribed but not translated, or 3) not transcribed. The affiliation of a DNA segment to a particular functional category may change during evolution: Functional DNA may become junk DNA, junk DNA may become garbage DNA, rubbish DNA may become functional DNA, and so on; however, determining the functionality or nonfunctionality of a genomic sequence must be based on its present status rather than on its potential to change (or not to change) in the future. Changes in functional affiliation are divided into pseudogenes, Lazarus DNA, zombie DNA, and Jekyll-to-Hyde DNA.

Key words: Functional DNA, literal DNA, indifferent DNA, rubbish DNA, junk DNA, garbage DNA, pseudogene, Lazarus DNA, zombie DNA, Jekyll-to-Hyde DNA.

Introduction

Genomic sequences are frequently categorized according to biochemical activity, regardless of whether or not such activity is biologically meaningful. Two erroneous equivalencies are particularly common. The first equivalency, usually espoused in the medical literature, erroneously equates “noncoding DNA”—that is, all regions in the genome that do not encode proteins—with “junk DNA”—that is, all regions in the genome that are neither functional nor deleterious (e.g., Krams and Bromberg 2013; Mehta et al. 2013). The second, more pernicious equivalency transmutes every biochemical activity into a function (e.g., ENCODE Project Consortium 2012; Sundaram et al. 2014; Kellis et al. 2014). Distinguishing between what a genomic element does (its causal-role activity) from why it exists (its selected-effect function) is a very important distinction in biology (Huneman 2013; Brunet and Doolittle 2014). Ignoring this distinction, and assuming that all genomic sites that exhibit a certain biochemical activity are functional, as was done by ENCODE Project Consortium (2012), is essentially equivalent to claiming that following a collision between a car and a pedestrian, a car’s hood would be ascribed the “function” of harming the pedestrian while the pedestrian would have the “function” of denting the car’s hood (Hurst 2013).

The ENCODE debate (Eddy 2012; Graur et al. 2013; Niu and Jiang 2013; Doolittle 2013; Palazzo and Gregory 2014) exposed the need for an evolutionary classification of genomic elements according to their selected-effect function. Such a classification is also needed to dispose of the widespread misconception according to which evolutionary processes can ever produce a genome that is wholly functional. Actually, evolution can only produce such a genome if and only if 1) the effective population size is enormous—infinite to be precise, 2) the deleterious effects of increasing genome size by even a single nucleotide are considerable, and 3) the generation time is very short. Not even in the commonest of bacterial species on Earth are these conditions met. In species with small effective population sizes and long generation time,
such as humans and perennial plants, a genome that is 100% functional is contrary to reason.

The Classification

Our classification scheme starts with the premise that all genomes are the products of natural evolutionary processes, rather than intelligent design and, hence, contain both functional and nonfunctional parts. “Function” in the context of this article is understood as selected-effect function (Millikan 1989; Neander 1991, 2002; Graur et al. 2013). That is, a sequence is functional if it is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

We first divide the genome into functional DNA and rubbish DNA (fig. 1). “Functional DNA” refers to any segment in the genome whose selected-effect function is that for which it was selected and/or by which it is maintained. Most functional sequences in the genome are maintained by purifying selection. Less frequently, functional sequences exhibit telltale signs of either positive or balancing selection. There are many methods for identifying functional genomic segments under various selective regimes (e.g., Nielsen 2005; Vitti et al. 2013). “Low-level noncoding RNA transcription” (e.g., Kellis et al. 2014), for example, is not sufficient to assign functionality. Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.

Functional DNA is further divided into “literal DNA” and “indifferent DNA.” In “literal DNA,” the order of nucleotides is maintained in the genome by natural selection because of its function. Furthermore, function is always defined in the present tense. In the absence of prophetic powers, one cannot use the potential for creating a new function as the basis for claiming that a certain genomic element is functional. For example, the fact that a handful of Alu elements have become functional cannot be taken as support for the hypothesis that all Alu elements are functional. The Aristotelian distinction between potentiality and actuality is crucial.
rasues that exist in the genome despite being actively selected against. The reason
that detrimental sequences are observable is that selection is neither omnipotent nor efficient. At any slice of evolutionary
time, segments of garbage DNA (on their way to becoming extinct) may be found in the genome. Garbage DNA is expected
to have a high turnover rate in evolution, but its disappearance from the genome is not instantaneous.

The affiliation of a DNA segment to a particular functional
category may change during evolution. Because there are
quite a lot of garbage DNA too. Junk DNA is expected
to persist in the genome for very long periods of evolution-
time; garbage DNA should be a more transient
phenomenon.

Changes in Functional Affiliation

The affiliation of a DNA segment to a particular functional
category may change during evolution. Because there are
four functional categories, there may be 12 possible such changes (fig. 2). Several such changes are known to occur
quite frequently. For example, junk DNA may become garbage DNA if the effective population size increases; the oppo-
site will occur if the effective population size decreases (Ohta 1973). Many of the 12 possible changes have been docu-
mented in the literature. Here, we suggest a nomenclature
for five such changes. Pseudogenes, for instance, represent a change in functional status from literal DNA to junk DNA,
whereas some diseases are caused by either a change from
functional DNA to garbage DNA (e.g., Chen et al. 2003) or
from junk DNA to garbage DNA (Cho and Brant 2011).
Rubbish DNA mutating to functional DNA may be referred
to as “Lazarus DNA,” so named after the second most famous
resurrected corpse in literature, Lazarus of Bethany
DNA may mutate to garbage DNA, in which case we suggest
the term “Hyde DNA” based on the fictional transformation
of a benevolent entity into a malicious one (Stevenson 1886).
Alternatively, junk DNA may become garbage DNA, for
which the term “zombie DNA” has been suggested (Kolata
2010).

Acknowledgments

R.B.R.A. was supported by a grant from the National Science
Foundation (1354952). D.G. wishes to express his awe and
reverence at Eric D. Green’s pioneering status as the first self-
awardee and self-awardee of NIH grants.

Literature Cited

Brosius J, Gould SJ. 1992. On “genomenclature”: a comprehensive (and
respectful) taxonomy for pseudogenes and other “junk DNA”. Proc
Brunet TDP, Doolittle WF. 2014. Getting “function” right. Proc Natl Acad
Sci U S A. 111:E3365.
Cavalier-Smith T. 1978. Nuclear volume control by nucleoskeletal DNA,
selection for cell volume and cell growth rate, and the solution of the
Chen YH, et al. 2003. KCNQ1 gain-of-function mutation in familial atrial
Cho JH, Brant SR. 2011. Recent insights into the genetics of inflammatory
Doolittle WF. 2013. Is junk DNA bunk? A critique of ENCODE. Proc Natl
Ehret CF, de Halle G. 1963. Origin, development, and maturation of
organelles and organelle systems of the cell surface in Paramecium. J
Ultrastruct Res. 23 (Supplement 6):1–42.


Hurst LD. 2013. Open questions: A logic (or lack thereof) of genome organization. BMC Biol. 11:58.


Stevenson RL. 1886., Strange case of Dr. Jekyll and Mr. Hyde. New York: Charles Scribner’s Sons.

