"Evolutionary Roles of Transposable Elements – The Science and the Philosophy"
University Club, Dalhousie University
October 20-21, 2018

ABSTRACTS

T. Ryan Gregory (University of Guelph)
Junk DNA, genome size, and the onion test

It has been known for ~70 years that the amount of DNA in the genomes of eukaryotes varies enormously among species and bears no relation to organismal complexity. Among animals alone, genomes vary more than 7,000-fold in size; across all eukaryotes, the range may exceed 200,000-fold. This extraordinary variability is the result of differential abundances of non-genic DNA, especially in the form of transposable elements. Genome size is correlated with key cellular parameters including cell size and cell division rate, which may lead to organism-level correlations with such features as body size, metabolic rate, and developmental rate. Patterns of genome size variability likely reflect differing evolutionary mechanisms and pressures, from those that add or subtract DNA within genomes to factors that constrain or enhance changes in genome size through organism-level effects. This seminar will provide an overview of the major patterns, correlates, and mechanisms of change in genome size among animals, and will explore the ways in which these are relevant in the debate over potential functions (or lack thereof) for non-genic DNA – as exemplified by the so-called "onion test".

Stefan Linquist (University of Guelph)
Four decades debating junk DNA and the Phenotype Paradigm is (somehow) alive and well

A standard trope in molecular biology and genomics says that most types of non-coding DNA have "long been dismissed as junk." The suggestion is that these regions were neglected by earlier researchers to the extent that hypotheses about their functional significance went untested or were just ignored. More subtly, this idea is sometimes used to bolster our expectation that particular non-coding regions will turn out to have some beneficial effect for the organism, especially when the molecular evidence is inconclusive. The idea, it seems, is that non-coding DNA ought to be regarded as functional for the organism until proven otherwise – thus compensating for the (alleged) tendency to discount its importance.

This paper presents a very different story about junk DNA. We show that the term has undergone at least three changes in meaning since its coinage in 1972. "Junk DNA" in the original mechanistic sense referred to duplicated genes with a small chance of becoming functionally repurposed. This evolved in the 1980s into a narrow sense of junk DNA, which referred to regions that lacked sequence-specific effects that are under selection at the level of the organism. This narrow-sense definition is entirely compatible

with the possibility that non-coding regions (1) possess selected-effect functions below or above the organism level; and (2) possess various selected-effect functions for the organism, but not associated with their sequence specificity. Indeed, many such hypotheses about the multi-level and non-coding functions of junk DNA have been entertained and tested since the early 1980s.

However, with the emergence of the Human Genome Project (HGP) in the 1990s, the term "junk DNA" took on a broader definition. Critics of the HGP appropriated this term to describe DNA that lacked any biological significance whatsoever. At the same time, these critics promoted the view that, since most of the human genome is "junk" (in the broad sense), it would be wasteful of limited resources to sequence it at great expense. This combination of ideas paved the way for contemporary researchers (such as the ENCODE Consortium, in 2012) to argue that non-coding DNA has long been considered "junk" in the broad sense, thus suggesting that it will be revealed to have at least some biological significance after all. Of course, the more accurate claim, that non-coding DNA has long been considered junk *in the narrow sense*, does not serve their rhetorical purposes nearly so well.

Two take-home messages emerge from this discussion. The first is that philosophers, funding agencies, and other vested parties should be skeptical of the claim that non-coding DNA has long been dismissed as functionally uninteresting. This suggestion first arose to discourage a certain type of genomic research, and is now used to bolster our faith in the alleged organismal benefits of non-coding DNA. Second, a striking limitation of these debates is that they have failed to move beyond the Phenotype Paradigm (Doolittle and Sapienza, 1980). The question of whether a sequence is "junk" (in any sense) has concerned its effects solely on the organism, while discouraging careful thought about functions at other levels.

Chris Ponting (University of Edinburgh)
90.8-92.9% of the human genome evolved neutrally

This is the estimate of the amount of the human genome that evolves neutrally with respect to insertions and deletions. It assumes nothing about where functional DNA lies, and yet finds that almost none lies within transposable element sequence. The model has not been challenged since it was proposed over a decade ago and it is important because it justifies the expectation that a particular stretch of DNA lacks function (as opposed to effect) until proven otherwise.

Paul Griffiths (University of Sydney)
Both adaptation and adaptivity are relevant to diagnosing function

The concept of adaptation is complemented by the concept of something being adaptive: what is now an adaptation must first have been adaptive. The 'selected effect' concept of function (also known as 'etiological' or 'proper' function) is dependent in the same way on a prior concept of function: the function for which trait was selected must be a function

that it first performed before being selected. It is widely assumed that this, second concept of function is merely the minimal, permissive concept of 'causal role' function – any effect which causally depends on a trait is a causal-role function of that trait. I have argued that the second concept of function is far more interesting. It bears *exactly* the same relation to 'selected effect function' that being adaptive does to adaptation. Cashing out this concept of function-as-adaptivity means cashing out the concept of being adaptive, that it, how the activity of a trait at some point in time affects evolutionary dynamics in a way that helps explain the future evolution of that trait. For this reason I have termed this concept 'evolutionary function'. Other philosophers of biology have argued that there is no such concept of function, because any concept of function must either collapse into selected effect function (adaptation) or into the minimal, permissive concept of causal-role function. I will rebut some of these arguments. My thesis in this talk is that the concept of evolutionary function is a valuable alternative to selected role function in understanding the function of genomic elements.

W. Ford Doolittle (Dalhousie University)
Selfish genes and selfish DNA: is there a difference?

Dawkins' widely-read book argued that a gene's phenotypic benefit to its bearers should be seen as nothing more than a mechanism by which its own spread and maintenance are ensured, a means but not an end. The 1980 "selfish DNA" papers went what their authors considered to be one step further, claiming that both the insertion sequences being investigated as endogenous mutagens in bacteria and repetitive DNA in eukaryotes are best understood as such "selfish" elements with no necessary individual phenotypic expression (hence not "selfish genes"), and countering claims that such elements, because they might someday prove useful, were retained "for the good of the species". The views are compatible but were differently motivated. I will agree with and elaborate on the conclusion of Agren (2016), that *"Selfish genetic elements have provided empirical ammunition in the disagreements between proponents of the genes-eye view and multi-level selectionists",* taking the position of the latter.

Justin Garson (Hunter College of the City University of New York)
Biological functions, the liberality problem, and transposable elements

Do transposable elements (TEs) have functions of their very own? In other words, do TEs have parts and processes that have the function of helping the TE replicate? Answering this question forces us to clarify the very concept of biological function. Many philosophers (including myself) accept the selected effects theory of function, which says, roughly, that in order for a trait to have a function it must have been selected for. The selected effects theory, however, is beset with well-known liberality problems, that is, it assigns functions too liberally. I argue that in order to solve the liberality problem, we must restrict functions to members of *populations*, that is, collections of individuals bound together by fitness-relevant interactions, such as competition, cooperation, and perhaps parasitism. Once we

accept this restriction, then in order to find out whether TEs have functions of their very own, we have to find out whether, and to what extent, TEs form populations in this sense. The scientific evidence to date is ambivalent and largely speculative. As a consequence, we don't yet know whether TEs, generally and for the most part, have functions of their very own.

Joyce Havstad (Oakland University)
Evolutionary thinking about critique of function talk

There is a decades-long scientific tradition of over-attributing function to genes (e.g., Dawkins 1976; ENCODE 2012). There is a parallel tradition of critiquing this practice (e.g., Doolittle and Sapienza 1980; Doolittle 2013; Elliott, Linquist, and Gregory 2014). Why does this cycle keep repeating itself? Why is function-talk so readily ascribed to genes, and why doesn't critique stick?

I posit the need for some evolutionary ways of thinking about science reform: thinking that could help us shape our recommendations in ways that are both popular *and* effective. Mechanistic thinking says that if we're representing ourselves as experts and making recommendations, then we should include in our recommendations something about how such recommendations might be implemented. Thinking about fitness implies that if we want those who adopt our recommendations to succeed, then we should consider their competition and environment—we should consider what their fitness will be in their environment, especially relative to others who may not be adopting our recommendations. And co-evolutionary thinking says that if we want our recommendations to have their intended effect(s), then we should theorize about the ways in which our recommendations might be misinterpreted, misapplied, adapted, co-opted, mimicked, and made obsolete.

In short, we should think about the potential evolutionary trajectories of our recommendations for scientific reform—and we should design those recommendations to withstand such evolutionary processes. After advocating for such methods, I consider how they apply to critical discussion of, for instance, ENCODE.

Guillaume Bourque (McGill University)
Impact of transposable elements on human gene regulatory networks

To help refine our understanding of human non-coding DNA, consortiums such as ENCODE, the NIH Roadmap and the International Human Epigenome Consortium (IHEC) have released more than 10,000 epigenomic maps across multiple cell types. What is needed for the functional interpretation of these datasets and for the accurate prediction of the impact of genetic variation, is a better understanding of the origin of the biochemical activity detected. Using open chromatin data sets from these consortiums in normal, embryonic, and cancer cells, we show that 44% of open chromatin regions are in transposable elements (TEs) and that this proportion reaches 63% for primate-specific regions. Moreover, although a number of studies have now demonstrated that TEs can

contribute to the remodeling of cis-regulatory networks, we argue that this is likely the exception rather than the rule. We contend that the unique selection forces acting on these TE-derived sequences will need to be carefully taken into account to partition the meaningful biochemical signal from the noise.

Ulrich Stegmann (University of Aberdeen)
On parity, genetic causation and coding

Like many other areas in the life sciences, genomics routinely employs informational and causal notions. The significance of these notions is contested among philosophers of biology, as is the nature of the phenomena they purport to pick out. My talk will sketch some of the issues and how I approach them. In particular, I focus on the alleged 'parity' between genetic and non-genetic causes, the claim that DNA can 'control' certain processes, and the genetic code.

Stephen Downes (University of Utah)
Does examining underlying alternate gene concepts help us understand the claims made by ENCODE and GWAS researchers?

Encyclopedia of DNA Elements (ENCODE) and Genome Wide Association Study (GWAS) researchers make analogous claims about the potency of areas of the genome traditionally understood as non-functional or non-coding. Further, GWAS and ENCODE researchers point to each other's results to reinforce their own. There is a difference in the language used to present the findings by each of these researchers. ENCODE researchers use functional and causal language in attributing a role to non-coding DNA in producing phenotypes, whereas GWAS researchers (mostly) use the language of the risk associated with certain single nucleotide polymorphisms (SNPs) for certain phenotypes (most commonly disease phenotypes). I propose that understanding these claims about the potency of regions of the genome can be enhanced by assessing the alternate underlying gene models relied on by each group of researchers. I explain and defend my use of alternate "gene models" as opposed to the more standard alternate gene concepts or gene definitions in this context. I also suggest that our continued commitment to the term "Genome" as opposed to "DNAome" or "Nucleotideome" is partly responsible for some of our current conceptual difficulties.

Alexander Palazzo (University of Toronto)
How nuclear retention and cytoplasmic export of RNAs reduces the deleteriousness of junk DNA

Eukaryotic cells are divided into two compartments: the nucleus where RNA is synthesized and processed, and the cytoplasm, where mRNA is translated into proteins. This subdivision allows for the extensive global surveillance of newly synthesized RNA, so

that junk RNA and misprocessed transcripts are retained in the nucleus and/or degraded. These quality control processes reduce the deleteriousness of junk RNA, and as a result diminishes the selection pressure to eliminate junk DNA. These quality control processes would also prevent the elimination of cryptic splicing signals and may explain why RNA splicing appears to be very error-prone in many eukaryotes. Ultimately, the preservation of junk RNA and misprocessed transcripts may fuel the evolution of lncRNAs and alternative splicing through non-adaptive processes.

David Haig (Harvard University)
Pax somatica

Transposable elements (TEs) and sedentary sequences of somatic cells have a common interest in survival and reproduction of the germ-track (*Keimbahn*). Somatic transposition has costs to individual fitness. Therefore, suppression of somatic transposition is usually an adaptation of transposable elements. In the germ-track, mutations to TEs that inhibit transposition are favored at every site a TE is located but moving to new sites selects for mutations that enhance the ability to transpose. Transposition scatters ready-made functional elements throughout the genome that are subject to natural selection.

Cedric Feschotte (Cornell University)
Experimental approaches to test the contribution of transposable elements to organismal function

Most eukaryotic genomes are replete with sequences derived from selfish genetic elements, such as transposable elements and endogenous viruses. Large-scale efforts to map the biochemical activity of mammalian genomes in a wide range of tissues and cell types have revealed that these elements commonly display chromatin marks or transcriptional factor binding events consistent with tissue/cell type-specific enhancer or promoter activity. But to what extent these activities contribute to the regulation of the host genome and to organismal fitness remain largely unknown. Novel approaches to manipulate the genome precisely and systematically, such as CRISPR-cas9 systems, offer an unprecedented opportunity to tackle these outstanding questions. In this talk, I will summarize a growing body of experimental results, produced by our group and others, suggesting that transposable elements exert pervasive effects on gene expression, which appear to be consequential for and intricately linked to organismal development and physiology. Interestingly, many of these activities are provided by elements that are not deeply conserved across species, but evolutionarily young, which point at the volatile nature of the sequences and molecular mechanisms underlying fundamental biological processes, including early embryonic development, innate immunity, and circadian clocks. These observations call for a nuanced definition of biological function.