Transposons for Philosophers[1]

Stefan Linquist

Although philosophers of biology have taken considerable interest in the nature of genetic information and in historical conceptions of the gene, there has been almost no philosophical discussion about transposable elements. For example, the recent book *Genetics and Philosophy: An Introduction* (Griffiths and Stotz 2013) contains less than a page of discussion about this topic. The lack of attention is surprising, in part, because transposable elements are responsible for producing the vast majority of DNA in eukaryotic genomes- at least 60% in humans. Aside of their biological importance, there are a number of controversies surrounding their discovery and functional interpretation that ought to be of interest to philosophers. In this paper I briefly review four philosophical controversies in the ongoing scientific study of transposons. These controversies are presented in roughly chronological order. In each case, I highlight reasons why philosophers should care about transposable elements.

The first controversy surrounds Barbara McClintock's discovery and functional interpretation of mobile DNA. McClintock is often portrayed as a victim of sexist and other institutional biases which allegedly prevented her peers from appreciating her insights. I review some of the arguments surrounding this claim while also presenting her hypotheses about the functional importance of TEs for the organism. I go on to suggest that in fact the scientific community was justified in rejecting her functional hypotheses, partly because McClintock violated norms governing the communication of novel scientific ideas.

The second section on selfish DNA presents an alternative perspective on the evolution of transposons, viewing them not as functional components of the organism but instead as genomic parasites. I briefly review the rise of this perspective, and offer three reasons why (I think) selfish DNA theory has recently fallen out of favour in recent years. This section will touch on a number of issues surrounding the evolution of transposons at multiple levels and the influence of adaptationist thinking in genomics.

Section 3 on the ENCODE debacle reviews a recent controversy over the appropriate use of function concepts in genomics. I briefly present an objection to the use of the causal role function and then explore some of the epistemic challenges surrounding the identification of genomic functions at multiple levels.

[1] This draft was written for the 2018 POBAM summer school and remains a work in progress. Please do not cite without author's permission.

The fourth and final section discusses some philosophical issues surrounding genome-level ecology. Proponents of this approach argue that transposable elements can only be properly understood if they are regarded as organism-like entities inhabiting a cellular environment. This "ecological" perspective will purportedly allow researchers to apply concepts and models from the fields of population and community ecology to explore their intra-cellular dynamics. Philosophers should care about this research program for at least two reasons. First, it raises the questions of what it means to take an ecological approach to any subject matter. Second, it offers an opportunity for us to reflect on how we assess whether some model or framework has explanatory value.

I have relatively little to say in what follows about the molecular biology of transposons. However, it might be helpful to present at least a few basic details. Transposable elements (TEs) or "jumping genes" are strands of DNA capable of moving from one chromosomal location to another, often replicating in the process. This is achieved using one of two strategies. Class 1 or "retrotransposons" employ a copy-and-paste strategy. These elements code for their own enzymes which identify the element, transcribe it into RNA, facilitate its translation into DNA, and then reinsert the replicated element back into a chromosome. Class 2 or "DNA transposons" employ a cut-and-paste strategy. They code for an enzyme that splices the chromosome, excises the element, and reinserts it into a new location.

Transposable elements can cause various sorts of mutations. This occurs most obviously when they insert into a protein coding gene or a regulatory region that provides some important phenotypic function for the organism. They can also cause major chromosomal rearrangements. These deleterious effects have led to the evolution of cellular mechanisms that silence TEs or prevent their insertion. Hence, in eukaryotic genomes most TEs have been rendered incapable of jumping. However, some families of transposon retain this capacity and undergo bursts of activity. This sometimes happens in response to environmental cues, prompting some researchers to view transposons as "genomic architects" which are potentially maintained by selection for this role. There are also cases where TEs have been functionally appropriated or "domesticated" to serve some host-beneficial function. As I discuss, these examples further suggest, to some, that it is mistake to view TEs in general as parasitic. In recent years there has been a revival of the idea that TEs should instead be regarded as functionally important components of the organismal genome.

1. The McClintock story.

Transposable elements were discovered in the 1940s by Barbara McClintock, a young researcher investigating the mechanisms of gene regulation in the corn plant *Zea mays*. This species of "Indian corn" is known for rows of multiply coloured kernels in the mature plant. The heritability of these patterns made pigmentation a useful diagnostic trait for early geneticists. After exposing corn plants to radiation, certain mutations can be identified as changes to the previous generation's pigmentation pattern. Linkage mapping also provided information about where on the chromosomes certain pigment genes were located. In addition, McClintock was skilled in the use of x-ray crystallography to observe gross morphological changes in chromosome structure.

Using this system, McClintock was able to track the movement of genetic elements through their telltale phenotypic effects. She noticed that certain pigment genes would become active at the same time that others fell silent. To her, it looked as if a single regulatory factor, previously responsible for supressing one gene, had jumped to a new location. Further exposure to radiation would sometimes cause this phenomenon to recur at new locations, causing previously suppressed genes to again be expressed while others are dampened. In other cases, McClintock noticed a "dosage-dependent" effect on gene expression. The movement of one element would sometimes cause only partial suppression of a target gene, resulting in a "variegated" pattern of pigmentation. After more radiation a second suppressor would sometimes move into position, causing the pigment gene to switch off completely. Again, these effects were reversible with additional radiation or when the suppressors jumped automatically to new locations. Using x-ray crystallography, McClintock was able to confirm that some of these modifications to gene expression corresponded with distinctive changes in chromosome structure. Eventually, it became apparent not only that the genome is composed of various mobile elements, but also that the translocation of certain elements can have a predictable, "switch-like" effects on gene expression.

McClintock proposed that these "controlling elements" (her name for transposons) are parts of the normal biological mechanism for regulating gene expression. From the outset of her career McClintock had been interested in the problem of cellular differentiation. That is, in the mechanism by which cells specialized into distinctive tissues through the regulation of protein

production. Her controlling elements were thought to play a key role in the process. The basic idea was controlling elements (somehow) move to the appropriate position depending on the stage of development and the cell lineage, where they impose "editorial" changes to the genome.

The fact that controlling elements were responsive to environmental cues inspired a second functional hypothesis. Later in her career, McClintock noticed that other environmental stressors besides radiation could likewise cause elements to jump. She hypothesized that such modifications could be beneficial to the organism, especially in the face of changing environments where previous patterns of expression might be less advantageous. Specifically, McClintock proposed that organisms respond to environmental challenges by inducing mutations that are likely to have some adaptive benefit for the organism or its offspring (McClintock 1984).

McClintock can thus be credited with two functional hypotheses. Her regulatory hypothesis proposed that the movement of genetic elements, from one locus to another, was a part of the normal ontogenetic mechanism for regulating gene expression. Her "genome shock" hypothesis, as it is called, held that natural selection favours the accumulation of transposons in order to make organisms more evolvable.

The unpopularity of McClintock's regulatory hypothesis is historically well documented. In her biography *A Feeling for the Organism*, Evelyn Fox Keller (1983) offers several explanations for the scientific community's negative reception of this idea. Part of the problem, Keller claims, is that few people were able to grasp the relevant details of her proposed mechanism. McClintock's writing is notoriously cryptic and she would often invent new theoretical terms without carefully defining or exemplifying them. Reportedly, McClintock also harboured a prickly social demeanor, which might have discouraged others from following up with her in person. Keller also points out that gender is likely to have played a role in discouraging some of her contemporaries from taking McClintock seriously as a theoretician.

In addition to these interpersonal factors, Keller points to an institutional explanation for the unpopularity of McClintock's regulatory hypothesis. The prevailing view of the gene was T.H. Morgan "beads on a string" model. This concept represented the gene above all as a static locus on the chromosome. Also influential was Beadle and Tatum's "one gene, one enzyme" hypothesis, which viewed genes as units of enzymatic production, not as units that engage directly in protein regulation. According to Keller, the rise of Crick's "central dogma" solidified the case against McClintock's regulatory hypothesis. She maintains that the idea of ontogenetic

information having a unidirectional flow (from DNA to RNA to protein) is incompatible with the idea that transposons regulate gene expression[2].

It is noteworthy that Keller, over the course of her career, has established herself as a leading critic of 20th Century gene concepts. She has written numerous books and articles arguing that the central dogma, among other related ideas, is being gradually overturned with the progressive march of molecular biology and genomics. On her account, the traditional concept of the gene is being abandoned in favour of a "reactive" model of the genome (Keller 2014). Among those who subscribe to this idea of a pending paradigm-shift in genomics, McClintock is often heralded as a patron saint (Shapiro 1992, Mattick 2009).

Interestingly, the one consideration not seriously entertained by Keller is that McClintock's views about gene regulation simply lacked adequate empirical support. As historian Nathaniel Comfort (1995, 1999, 2001) points out, there were good reasons for thinking that the effects she observed on pigment genes were anomalous.

> The plants in which she observed this behavior were sick. Their ancestors had been bombarded with X-rays and selected for their likelihood of undergoing spontaneous chromosomal rearrangements. Many died before reaching maturity; others were pale and fragile, with wrinkled leaves and shriveled kernels.

Indeed, the number of token transposition events McClintock recorded in her irradiated plants was much greater than anyone had observed in the wild. This might have suggested to some that the processes she observed were abnormal. To the contrary, McClintock insisted that, because of the switch-like regularity of these effects, both in location and in developmental timing, she must have been observing a normal ontogenetic mechanism. Comfort describes this inference as, "a great speculative leap, from mutant behaviour to normal activity" (1999, p. 139).

Between 1944 and 1953, McClintock identified several different types of mobile element in maize, each time insisting that they were part of a normal regulatory mechanism. Other researchers likewise found evidence of gene mobility, first in maize and later in bacteria, but they stopped short of embracing McClintock's functional interpretation. In the words of her close friend and fellow researcher, Waclaw Szybalski,

> While I was convinced that she had good evidence that things moved, that this has any significance for development I was not convinced, because she did not present anything to me which would convince me at that time or even later (Comfort, 1999, p 146).

[2] For a relevant discussion, see Rosenberg (2006), "Is epigenetic inheritance a counterexample to the central dogma?" History and Philosophy of the Lie Sciences 28:549-565.

By 1960, Comfort explains that her regulatory hypothesis was conclusively rejected in favour of Jacob and Monod's (1960) operon model, which provided an alternative mechanism for the enzymatic control of gene expression. This model was not only well corroborated, it was relatively easy to understand, and fit more coherently with received views about gene structure function.

As I have suggested, there is disagreement among historians not only about McClintock's scientific status, but also about the nature and extent of the biases working against her. On Keller's version of the story, McClintock was unjustifiably neglected by the scientific community. This assessment is defended, in part, by Keller's contention that McClintock's image of a dynamic genome was prescient. In other words, the fact that (according to Keller) contemporary genomics is now questioning the central dogma and other "static" views of the genome implies that McClintock was the victim of an unjustified bias. On the other hand, Comfort suggests that McClintock remained highly regarded as a scientist, in particular for her discovery of transposition, but that her regulatory hypothesis was deservedly rejected.

Let me now suggest one reason why I think that philosophers of biology should care about the McClintock story. Contemporary philosophers of science recognize that the success or failure of a theory cannot be explained entirely in terms of adherence to some logical procedure or 'scientific method" (Woodward and Goodstein, 1996). To a certain extent the success of science depends on the integrity of its social institutions and norms. One such institution is the tendency to assign notoriety to a scientist in proportion to the novelty of her discoveries (Koertge 1991). My understanding of Keller is that she thinks that this reward system broke down in the case of McClintock. Her regulatory hypothesis failed to receive due respect and attention from fellow scientists, Keller claims, because her scientific colleagues were blinded by various theoretical and gender biases.

But this example also helps us recognize other, somewhat subtler institutional norms surrounding the scientific acquisition of knowledge. For example, it is generally regarded as bad practice for a researcher to present her theory in a form that is inscrutable or unnecessarily opaque (Sperber 2010). McClintock violated this norm on several fronts. Not only is her writing dense, but she often invented new theoretical terms without carefully defining or illustrating them. This is not an acceptable mode of presentation for ideas that are controversial or novel. It is also noteworthy that many of McClintock's key publications are reports to scientific societies,

not peer reviewed articles. And in some of her more important publications, McClintock took the liberty of departing from the standard form of a scientific paper.

Let me explain why I take this latter point about form to be significant. Much has been written about the shortcomings of the standard form of a scientific paper. For example, the artificial separation of sections into "methods," "results," and "discussion" generates a misleading chronology of scientific practice. However, one of the benefits of this format is to facilitate communication. A reader knows where to look for specific pieces of information. By departing from this standard form, while at the same time introducing novel and controversial ideas, McClintock was again thumbing her nose at received norms surrounding scientific communication.

It is perhaps no surprise that Keller is apologetic in discussing these quirks. McClintock's cryptic writing style is attributed to her sharp intellect and her distinctively intimate relationship with her model organism. The obscure publication venues is explained in terms of the reporting requirements for her research grants. And the departure from the conventional form is explained as an effort to communicate with an audience who seemed reluctant to accept her ideas. In addition, the fact that McClintock was a woman is presented as a reason why some might have been less than patient in attempting to understand her work. If people had trouble following McClintock's writing, Keller suggests, it was as much their fault as it was hers.

However, I think that McClintock's violation of certain norms surrounding the communication of scientific ideas arguably justified her colleagues' rejection of her regulatory hypothesis. It would have been reasonable for her colleagues to assume that her model was not sufficiently developed to merit serious attention, given the various obscurities surrounding its presentation.

The point I am attempting to illustrate is that there is valuable philosophical work to be done in applying ideas from the social epistemology of science to historical episodes, such as the McClintock story. It is common in certain circles to see McClintock as a victim of prevailing scientific biases that were allegedly unjustified. The important lesson for proponents of this view is that successful science depends on more than just the empirical adequacy of a novel idea. The acceptability of a scientific theory can, and perhaps also should depend on a researcher's conduct within a community of scholars.

2. The rise of Selfish DNA

Throughout the 1960s and 70s, evidence of transposable elements continued to be found in a variety of model organisms. Although McClintock's regulatory hypothesis had been discredited, her evolvability hypothesis seems to have been more favourably received, at least among molecular biologists. It is important to recognize that the prevailing theoretical orientation in molecular biology (arguably, to this day) has been organism-level Panglossianism (Eddy 2012, Brunet and Doolittle 2015). Whenever some new category of genetic element is discovered, the default assumption is always that it serves some beneficial function for the organism.

This attitude is apparent, for example, in historical discussions of "junk" DNA. Sasumu Ohno coined this term in 1972 in the context of a fairly specific problem. It was widely assumed that selection tends to purge variation rather quickly from most genomes. At the same time, acquisition of a novel genetic function often requires the gradual accumulation of mutations at multiple loci. This raised the question of how do novel genetic functions evolve if the necessary components are always being eliminated.

Ohno proposed a specific mechanism in which a gene is first duplicated, thus avoiding the loss of original function. The redundant gene is then free to acquire mutations, occasionally resulting in a novel function. Like an old bicycle accumulating rust in the attic, Ohno's junk DNA was gradually losing functionality, except in the rare case when a new beneficial sequence would appear.

The reaction to Ohno's hypothesis reveals the extent of adaptationist thinking at the time. One of the earliest presentations of his theory was at The Brookhaven Symposium on Biology, in 1972, and a transcript of the discussion following his talk has been published (Ohno 1973). Perhaps the first thing to note is that every commentator was quick to propose some organism-level adaptive benefit for Ohno's junk DNA. Many of these candidate functions involved some kind of structural role, for example, as a genetic spacer. Ohno himself encouraged this interpretation:

> The chance of acquiring a new [coding] function by unrestricted accumulation of mutations, however, should be as small as that of an isolated population emerging

> triumphant as a new species. Degeneracy is the more likely fate. The creation of every new gene must have been accompanied by many other redundant copies joining the ranks of solent DNA base sequences, and these silent DNA base sequences many now be serving a useful but negative function of spacing those which have succeeded." (1972, 169).

His proposal that junk DNA might have a "negative" (i.e. noncoding) function illustrates that even Ohno was reluctant to accept that large amounts of DNA could persist without an organism-level fitness benefit. Indeed, in the commentary following Ohno's presentation, this opinion was explicitly stated by serval leading geneticists. Here is an excerpt of their discussion:

> W. Hennig "From all what is known so far we can conclude that probably the Nucleotide sequence as such does not matter. Furthermore the actual amount of simple sequence DNA (within some limits) seems not to be important. Since this kind of DNA is there one has to correlate it with some kind of function. "
>
> E.B. Ford "I think it just wouldn't be there unless it would do it [function as a spacer]. Something was a functional reason of some kind for it.
>
> W.S. Yunis This is what I emphasized earlier, that this DNA must have a functional value since nothing is known so widespread and so universal in nature that has proven useless.

At this stage it is important to insert a terminological note. When these researchers described junk DNA as "functional," they were referring to some unknown selected effect at the level of the organism. In other cases, "functional" refers even more narrowly to selected effects on the organism that are sequence specific. That is, to evolved functional roles in the organism that depend on the order of nucleotides in a genetic sequence. These remain the standard usages of "functional" in contemporary genomics. However, as I will explain, there has been at least one noteworthy departure from this practice.

Despite the pervasiveness of organism-level adaptationism in the early 1970s, by the end of the decade two emerging lines of thinking would lay the groundwork for an important challenge. The first was Dawkins' gene-centric perspective. Although Dawkins (1976) notoriously revelled in adaptationist thinking, he encouraged us to think about what it would mean for a genetic element to be truly selfish. Of course, as some of Dawkins' critics have

pointed out, he did not follow through with his own suggestion. Dawkins's "selfish" genes always promoted their own replication through the phenotype of some organism and in concert with an entire genome of cooperative replicators. Dawkins failed to provide an adequate origin story for how independently replicating genes might come together into a single genome (Okasha 2005). Without appealing to multi-level selection, it is unclear how Dawkins proposes to solve the free rider problem at the genetic level.

With respect to our understanding of TEs, Dawkins' perspective was nonetheless valuable for encouraging researchers to think of selection acting below the level of the organism. Then, in 1979, Gould and Lewontin published their "Spandrels" paper which challenged biologists to justify their reliance on adaptationist thinking more generally. Together, these two lines of thought provided the backdrop for an alternative approach to the evolution of transposons, and genetic elements more broadly.

In 1980 two articles appeared back to back in the journal Nature, offering a novel perspective on the function of transposable elements. Ford Doolittle and Carmen Sapienza (1980) criticized what they called the "phenotype paradigm" in molecular biology. This term refers to the popular tendency to assume that all genetic elements are selected exclusively for their contribution to the organismal phenotype. Doolittle and Sapienza argued that this assumption is unwarranted in the case of mobile DNA. They pointed out that transposable elements were not only capable of jumping to new locations, they also replicate in the process, potentially generating multiple copies per cell cycle. The upshot is that TEs are capable of multiplying more rapidly than most other components of the genome. Doolittle and Sapienza identified some of the implications of for popular ideas about the function of mobile elements:

> What we propose here is there are classes of DNA for which a 'different kind of explanation' may well be required. Natural selection does not operate on DNA only through organismal phenotype. Cells themselves are environments in which DNA sequences can replicate, mutate and so evolve… Furthermore, if it can be shown that a given gene (region of DNA) or class of genes (regions) as evolved a strategy which increases its probability of survival within cells, then no additional (phenotypic) ecplanation for its origin or continued existence is required (1980, 601).

Leslie Orgel and Francis Crick (1980) presented a similar perspective in their companion paper. They described transposons as "genomic parasites" that enjoy quasi-independent existence from the "host" organism. One of their central arguments revolved around the C-Value Paradox. It has long been recognized that species exhibit dramatic size differences in their genomes despite no marked difference in phenotypic complexity. For example, they pointed out that lilies and salamanders have 20 times more DNA than is found in the human genome (Orgel and Crick 1980). The best explanation for these size differences, they argued, was the differential success of transposable elements among species, independent of any beneficial effects on the host.

Although Selfish DNA theory reportedly encountered some initial resistance (Doolittle 1982), the idea that transposons qualify as a kind of genomic parasite has become one of the dominant explanations for their persistence and diversity (Kidwell and Lisch 2001). However, recent years have seen growing opposition to this perspective. Even McClintock's evolvability hypothesis is enjoying something of a renaissance. Currently, the field of genomics is in a state of entrenched disagreement about whether most non-coding DNA is likely to have an organismal function. My sense is that there are at least three reasons for this resurgence in popularity of the phenotype paradigm.

1) **The discovery of transposon "domestication."** Gould and Vrba (1982) first coined the term "exaptation" for the acquisition of secondary functions in traits. One of their hypothetical examples was the possible cooption or "domestication" of TEs or TE-derived sequences into the normal molecular machinery of the host cell. This phenomenon was first iddentified by Miller et al. (1992) in *Drosophila Guanche,* where one sub-family of P-elements appears to have gone to fixation in the host population due to its tendency to repress the transposition of other, active *P* elements. However, the most famous example of transposon domestication involves the production of telomeres in *Drosophila melagonaster* (Zhang and Rong 2005). Telomeres are the terminal ends of chromosomes that serve a number of important functions related to chromosome integrity. Most species encode an enzyme called telomerase which is responsible for generating telomeres. However, in one species of Drosophila this gene has degenerated. Instead, the production of telomeres in this species is executed by a transposable element that self-replicates and preferentially inserts into the ends of chromosomes. This element is "domesticated" in the sense that it has taken over this important function while also avoiding insertion into regions containing host genes,

where they are more likely to have harmful effects. This example of domestication is often cited by authors arguing for the likelihood of other, yet to be discovered organismal functions for TEs. In other words, the fact that TEs have been domesticated on a few occasions is used as an argument to suggest that this phenomenon is widespread.

2) **Stress induction is "switch-like."** TEs have been shown to become active in response to a variety of stressors including temperature fluctuations, UV light, and cocaine exposure. In some species of plant, activation occurs only in the areas where stress is applied. Hence, TE activation is conditionally activated and localized to certain tissues. To some, this "switch-like" behaviour is suggestive of an evolved response (Capy et al. 2000). A key assumption is that environmental stressors function as a signal to the organismal genome of a changing environment. In static environments most mutations reduce fitness. However, arguably, in a changing environment an increase mutation rate is less harmful or even beneficial. On this view, an enhanced mutation rate allows an organism to effectively "roll the dice" and improve its chances of hitting upon a novel beneficial phenotype.
3) **So much non-coding DNA**. As sequencing technology becomes increasingly efficient and affordable, a growing number of species are having their entire genomes sequenced. An emerging pattern is that conventional genes represent only a small fraction of most genomes. By far, the majority of DNA in eukaryotic genomes consists of either active transposable elements or their deactivated remnants. For example, roughly 60% of the human genome consists of TEs, most of which are now silent. A whopping 10% of our genomes consist of just one family of TEs (called alu) which is specific to primates. By contrast, the entire complement of conventional genes in the human genome is less than 4%. For some researchers, it seems unlikely that non-coding DNA would persist in such high proportions unless it served some organism-beneficial function.

By now, I hope that it is becoming clear why philosophers should care about transposable elements. Each of these augments raise interesting theoretical questions about the nature of selection and the criteria for identifying functions at different levels. For example, in the case of conventional parasites, it has been suggested that decreased virulence and eventual mutualism is an expected evolutionary outcome (Poulin 2007). It is an interesting theoretical question whether this expectation holds also genomic parasites.

The general topic of evolvability has received philosophical attention in recent years (e.g. Sterelny, 2007). It would be interesting to consider whether there are any conditions under which TEs could be maintained as a source of novel phenotypes. Also, does the fact that they can be conditionally activated in response to stress support this hypothesis?

In a recent publication by Tyler Brunet and Ford Doolittle (2015) the evolvability hypothesis is dismissed as teleological nonsense. Instead, these authors propose that TEs are under selection at two other levels. They are selected within genomes, as the selfish DNA hypothesis proposes, for their capacity to self-replicate. At the same time, Brunet and Doolittle propose that TEs are favoured at a macroevolutionary scale, by selection acting among species. TE activity can result in large-scale modifications to the genome, potentially causing reproductive isolation. Thus, populations harbouring larger proportions of active TEs are expected to speciate more rapidly than genomes with fewer mobile elements. Brunet and Doolittle propose that speciation is more than just an incidental byproduct of transposons. Rather, they argue that selection acting at the species level could explain the persistence of TEs and their widespread distribution among taxa. Brunet and Doolittle (2015) present a number of arguments in defense of this claim and their paper merits close attention.

It might come as a surprise that, even in this day and age, the discovery of so much non-coding DNA is often taken to suggest an organism-level function. After all, the existence of bloated genomes (e.g. in lilies and salamanders) was cited as one of the main arguments in defense of the selfish DNA hypothesis (Orgel and Crick 1980). To my knowledge, no compelling theoretical argument has been developed to suggest that this hypothesis is misguided. To the contrary, population genetic models lend support the Selfish DNA interpretation. For example, Charlesworth (1985) has shown that it would be extremely difficult for selection acting at the organism-level to purge the genome of transposable elements, even on the assumption that they tend to be mildly deleterious. As far as I can tell, many molecular biologists working on transposons are simply unaware of the theoretical literature surrounding their evolution. This partly explains our next episode in the history of transposon biology: The announcement by the ENCODE consortium that 80% of the human genome is functional. As I shall now explain, this claim raises yet another set of philosophical issues.

3. The ENCODE debacle

The Encyclopedia of DNA Elements Consortium (ENCODE) is a large international research collaboration that set out to identify all functional elements encoded in the human genome. This project is sometimes described as picking up where the Human Genome Project left off. More than 400 scientists have collaborated in this project, at a cost of over $200 million. Over the course of roughly a decade this project has generated various insights into the composition and structure of the human genome. However, the most widely publicized results were announced in September 2012, with the simultaneous publication of 30 research articles across a number of journals. Nearly all of the extensive media coverage surrounding this event focused on a single finding: "These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions" (ENCODE 2012, p. 57). The backlash from members of the genomics community was swift and harsh. ENCODE researchers were accused of exaggerating the significance of their claims, in part, by playing fast and loose with the conventional understanding of "function" (Doolittle 2013, Eddy 2013, Graur et al. 2013, Elliot et al. 2014). This episode has focused attention on the various meanings and uses of function concepts in genomics – a debate that philosophers are well positioned to weigh in on.

There have been at least two different diagnoses of what when wrong in ENCODE's use of function concepts. One view holds that they relied on a causal role (CR) notion of function which has no proper place in the field of genomics. Critics who defend this claim point to the permissiveness of the CR function concept. As philosophers have long recognized, the fact that CR functions are defined relative to the interests of an investigator can result in some rather counter-intuitive functional ascriptions. For example, a worn sparkplug has the function of causing an engine to misfire, provided that our aim is to functionally analyze the automobile's poor performance. Such considerations seem to lie behind the statement that, "the causal role concept of function can lead to bizarre outcomes in the biological sciences. As a result, most biologists use the selected effects concept of function." (Graur et al. 2013, p. 579).

I confess to being somewhat sympathetic to this objection. Philosophers tend to acknowledge the permissiveness of CR functions, but then assume that scientists will not be motivated to abuse this concept by positing obscure or unconventional functions. However, my sense is that philosophers have failed to appreciate the various non-epistemic incentives for labelling genetic elements (in some sense) "functional." One such motive is to grab headlines.

This appears to be part of the explanation for ENCODE's claim to have identified a function for 80% of the human genome. In the face of having to justify the massive expenditure on a program that has (arguably) delivered underwhelming results, the pressure was on to announce an earthshattering finding. Another, more common incentive is to exaggerate the significance of a research finding. The phrase "long dismissed as junk" has become a popular refrain for many genomics researchers who find evidence of a potential organismal benefit for some non-coding element. The availability of more than one function concept, one of which (CR functions) is much less epistemically onerous than the other (SE functions) is bound to foster confusion in this kind of environment.

Despite this reservation, I have on two occasions defended the cautious use of the CR function concept in the field of genomics (Elliot et al. 2014, Doolittle et al. 2014). Let me now offer an alternative suggestion for what went wrong in the ENCODE debacle.

Arguably, the problem with ENCODE stems from the particular proxies that they used to identify CR-functions. ENCODE borrowed proxies that had been developed for the detection of functions in the production or regulation of "normal" protein-coding DNA. They overlooked the fact that transposable elements exhibit many of the earmarks of protein coding or regulatory regions, despite lacking an organism-level function. For example, transposons are transcribed into RNA. They possess enzymatic binding sites. They are sometimes located close to regulatory regions. All of these properties potentially contribute to their selfish replication. Yet, somewhat astonishingly, ENCODE researchers took any one of these criteria as a sufficient condition for identifying a genetic element as "functional" for the organism. In Elliot et al. (2014) we draw the following analogy:

> Imagine an individual who wishes to use a metal detector to find valuables on a beach. First, he visits a jewelry store in order to establish the machine's ability to detect gold and silver. Satisfied, he begins scanning the beach. Occasionally, he hits on an old nail or a bottle cap, which causes the metal detector to light up and emit a sound. Technically, triggering the metal detector could be considered a CR function of these pieces of discarded metal. However, this is not the same thing as locating treasures, and it would be false to assume that every hit with the detector was identifying something useful just because this was the case in the jewelry store. Yet, this appears to be what ENCODE has done by employing assays that

> are normally used to find unambiguously functional elements (e.g., genes) and then considering any positive result elsewhere in the genome to be an indication of "biochemical function." Some of the hits identified by ENCODE may indeed be gold, but most could be bottle caps (p. 17).

All of this raises the question, what sorts of evidence can be used to distinguish element-level from organism-level functions. This question is deceptively simple. Let me now briefly review three standard proxies for organism level functionality, and show why they are inconclusive.

Sequence conservation. Within the genomics literature, evidence of a conserved sequence is sometimes taken to suggest an organism-level function. This inference is based on two related assumptions. First, it is generally assumed that, over extended periods of time, transposons tend to become deactivated because of their deleterious mutagenic effects. At the same time, if a sequence is beneficial to the organism, then "purifying" selection is thought to prevent it from accumulating mutations. Therefore, if a sequence is highly conserved, then it is probably functional for the organism. Graur et al (2013) identify conservation as the gold standard for inferring organismal function.

However, this line of reasoning suffers from a few shortcomings. One problem is that it is possible for selection acting just at the element-level to favour a specific sequence, simply because it facilitates efficient transposition. The first assumption, that over time mobile elements will be deactivated, is likewise questionable. A second problem is that this criterion focuses on functions associated with sequence specificity. If some TEs are selected for a structural role (e.g. as spacers) then there is no expectation that they will maintain a specific sequence. Hence, sequence specificity is neither necessary nor sufficient for organism-level functionality.

Conditional activation. As we have seen, the responsiveness of TEs to stress and related "switch-like" behaviour is sometimes taken to suggest an organism-level function. However, an alternative hypothesis is that stress tends to compromise the ability of the host organism to suppress TE activity. When TEs are activated in response to stress, this could be because the host organism has a weakened capacity to control this parasite. A better understanding of how cells suppress TE activity should help to identify the kind of evidence that would be required to decide among these hypotheses.

Proximity to protein coding regions. In general, regulatory elements tend to be positioned close to the genes that they control. So, if a transposon is located in a regulatory region or close to a gene, it often assumed that it likewise serves some regulatory function. However, an alternative explanation is that these regulatory regions provide 'safe zones' where an element can effectively avoid suppression by the host. Transposons are typically suppressed using a process of methylation, which involves the attachment of a methyl group to the DNA strand, thus preventing transcription. It might turn out that this is a sloppy mechanism –that methylation does not allow for precise targeting. In this case, it might be impossible to silence a TE that is embedded close to a gene because doing so would have harmful spillover.

In this section, I hope to have touched on a few additional reasons why philosophers should care about transposable elements. Philosophers of biology routinely distinguish between CR and SE functions, and it is generally assumed that both concepts have a place in contemporary biology. However, I have argued that this simple assessment ignores some important aspects of scientific practice. One such consideration is that researchers are often motivated by non-epistemic considerations to exaggerate the significance of their findings. Especially in the case of large and expensive research projects, there is pressure to deliver surprising results, and the identification of a non-coding region as SE functional is often regarded as such an achievement. This can provide an incentive for researchers to equivocate in their ascription of functions, perhaps even without realizing it. A second consideration is that it is easy for philosophers to overlook the epistemic challenges associated with identifying SE functions, especially when they can emerge at more than one level. At the same time, philosophers have long debated the question of pluralism about levels of selection. Much of this debate turns on the question of how to identify causal processes at multiple levels (Okasha 2016). Perhaps the debate over levels of selection in the transposon literature can provide some guidance on this front.

4. Genome-level ecology

Up to this point I have discussed two popular perspectives in the field of transposon biology which are occasionally at odds over the mechanisms of TE evolution and the level(s) at which they tend to be functional. The phenotype paradigm defaults to an organism-level interpretation of their function. The selfish DNA perspective begins with the assumption that

they coevolve in a parasitic relationship with their hosts. In recent years, a third approach has begun to emerge in the study of transposable elements. Genome-level ecology purports to offer a novel perspective for investigating transposon dynamics. In broad outline, this view regards transposons as akin to species which inhabit a local, intra-cellular environment. The aim is then borrow models and concepts from the discipline of ecology in order to explain transposon dynamics.

Without going into the details of this approach (see Linquist et al. 2013, 2015), let me jump to the question of why it is relevant to philosophers. I see at least two important points of contact. The first concerns the relationship between ecological and evolutionary modes of explanation in general. The second addresses our strategy for deciding whether some novel hypothesis or theory has explanatory power.

I suspect that if most philosophers of biology were asked to come up with a paradigmatic example of an evolutionary explanation, they would cite a case of natural selection acting on some trait: peppered moths, finch beaks, and guppy tails are well worn examples. Most of us take for granted the empirical challenges associated with developing such detailed explanations. It is a deceptively simple achievement to show how population-level phenotypic variation changes across generations in response to particular ecological factors. Often, it is much simpler to treat evolutionary and ecological processes independently from one another. Hence, ecologists often default to an idealized view of organisms, treating them as static functional types interacting with particular features of the environment. At the same time, evolutionary biologists often ignore the specifics of ecological relationships, focusing instead on how previous features of the population, such as effective population size, impact its evolution. Ecology is not ignored in principle on this approach. But instead of investigating the roles of particular ecological factors in selection, they are typically collapsed into a single variable –the selection coefficient.

My sense is that philosophers tend to view these simplifications as placeholders for a more complete or fleshed out explanation. The idea is that both ecological and evolutionary factors should be taken into account whenever possible. This attitude has become increasingly popular with the realization that selection can potentially occur over relatively short time scales (Schoener 2011).

Upon reflection, this preference for detailed eco-evo explanations relies on a questionable assumption. Most philosophers accept that a good mechanistic explanation should identify only a

subset of the causal factors contributing to a target phenomenon. An explanation that represents all of the "gory details" is usually less informative than one which identifies the central causal tendencies or, in some cases, the most causally salient difference makers. This realization alone should make us suspicious of the claim that both ecological and evolutionary factors must be taken into account whenever possible. Instead, we require a strategy for determining when it is sensible to adopt either an ecological or an evolutionary approach to some subject matter.

This issue becomes salient with the recent extension of ecological thinking into a variety of new domains. It is becoming increasingly popular to take an "ecological" approach to micro-organisms (O'Malley), cancer cells (Thomas et al. 2012), and transposable elements (Brookfield 2005). Ww might ask, what is distinctive about an *ecologica*l approach in all of these cases? And how, if at all, does it differ from an evolutionary mode of explanation?

A potential answer to these questions has recently been put forward in the context of geneome-level ecology (Linquist et al 2013). The basic idea is that ecological and evolutionary modes of explanation, as I suggested earlier, involve alternative strategies for simplifying a complex process. A purely ecological explanation regards the focal entity as a fixed type, unchanging over time,that engages in various functional relationships to its environment. A purely evolutionary explanation treats the focal entity as a variable population capable of changing over time. Previous states of the population explain later states. However, specific functional relationships between the population and its environment are ignored for the sake of simplicity. It is also possible to adopt a combined eco-evo approach, which attempts to track the effect of particular ecological factors on a variable population. But this is obviously more epistemically demanding.

How can we determine, given a particular phenomenon or pattern in nature, whether to adopt a purely ecological, a purely evolutionary, or a combined eco-evo approach? Our proposal (ibid) is to approach this question by partitioning the variance among the two types of factor. On this view, one begins by identifying candidate ecological and evolutionary variables that are likely to be relevant for some focal entity. In the case of transposable elements, relevant ecological factors are aspects of the host genome which are likely to impact their rates of replication or insertion. These might include the availability of transcription enzymes, the density of host genes, or the spatial distribution of non-coding regions. A potentially relevant evolutionary factor is the relative population sizes of different TE families in previous

generations. It then becomes a simple exercise of partitioning the variance between these two types of factor to determine the extent to which they each potentially explain a given pattern of variation in TE diversity or abundance (see Linquist et al 2013 for an example application).

To be clear, I am not proposing that the covariation between some factor, or combination of factors, itself constitutes an *explanation* of the relevant pattern. This approach is better understood as a litmus test for identifying whether some mode of explanation (evolutionary, ecological, or a combination of the two) is likely to pay dividends. To illustrate, suppose that we discover for a range of salamander genomes that the diversity of transposable element families covaries with the size of contiguous regions of non-coding DNA. (I am imagining something akin to a species/area relationship at the intra-genomic level). At the same time, suppose that there is only a weak correlation between the phylogenetic relatedness of these transposon communities and their diversity across host genomes. It would then be sensible to pursue a purely ecological approach to explaining this pattern of diversity. Of course, it would remain an open question which specific ecological mechanism explains the pattern. Transposon diversity might be driven by a diversity of local (genomic) "habitats", by increased competition among TEs in larger non-coding region, or some other factor.

Let me close by acknowledging a potential worry. It might be argued that this strategy inherits the various problems associated with the use of ANOVA to detect causal strength (Sober 1988, Northcott 2008). One limitation is that explanatory significance can only be assessed for a "population" of systems, not for token systems (Sober 1988). It makes no sense, for example, to ask whether that evolution is more relevant than ecology in explaining the diversity of TEs in the human genome. Perhaps a more significant worry is that an assessment of explanatory relevance is contingent on the particular set of systems being evaluated. For example, it might turn out that ecological factors are relevant for explaining TE diversity in primates, but not for explaining TE diversity in vertebrates. I suspect that part of this worry can be alleviated by the appropriate choice of systems. It would be unacceptable, for example, to base an assessment of explanatory relevance for the entire primate lineage on a sample of genomes that included only humans, baboons, and tarsiers. However, provided that samples are appropriate, I see nothing in principle mistaken about the idea that explanatory relevance is context sensitive. In ecology, it is a familiar idea that different processes are more or less relevant depending on the scale of the analysis.

Conclusion

I hope it has become apparent that the study of transposons is a neglected area of philosophical research. It is especially surprising that philosophers working on issues in genomics, biological function, levels of selection, and adaptationism have overlooked this philosophically rich subject. The history of transposon biology is also a useful case study for examining the factors influencing theory change in biology. Hopefully philosophers of biology will start to invest more time in philosophy of transposon biology in the near future.

References

Brookfield, J.F. (2005), "The ecology of the genome – mobile DNA elements and their hosts. Nature Reviews genomics 6: 128-136.

Brunet, T. D.P. Doolittle, W.F. (2015), "Multilevel selection theory and the evolutionary functions of transposable elements. Genome Biology and Evolution 7: 2245-2457.

Capy, P. Gasperi, G. Biemont, C. Bazini, C. (2000), "Stress and transposable elements: co-evolution or useful parasites?' Heredity 85: 101-106.

Comfort, N.C. (2001), The Tangled Field: Barbara McClintock's Search for the Patterns of Genetic Control. Harvard University Press: Cambridge Mass

Comfort, N.C. (1995), "Two genes, no enzyme: A second look at Barabara McClintock and the 1951 Cold Spring Harbor Symposium. Genetics 140: 1161-1166.

Comfort, N.C. (1999), "The real point is control: The reception of Barbara McClintock'ss controlling elements." The Journal of the History of Biology. 32: 133-162.

Doolittle, W. F. and Sapienza, C. (1980), "Selfish genes, the phenotype paradigm and genome evolution." Nature 284: 601-604.

Doolittle, W. F. (1982), "Selfish DNA after fourteen months." In G.A. Dover and R.B. Flavell (Eds) Genome Evolution. Academic Press: New York. P. 3-28.

Doolittle, W.F. (2013), "Is junk DNA bunk? A critique of ENCODE. Proceedings of the National Academy of Science USA.

Eddy, S. (2013). "The C-value paradox, junk DNA, and ENCODE. Current Biology 22: R898.

Elliott, T.A., S. Linquist and T.R. Gregory (2014). Conceptual and empirical challenges of ascribing functions to transposable elements. *The American Naturalist, 184 (1)*, 14-24.

Graur, D. Zheng, Y. Price, N. Azevedo, RBR. Zufall, R.A. Elhaik, E. (2013), "On the immortality of television sets: "Function" in the human genome according o the evolution-free gospel of ENCODE. Genome Biology and Evolution. 5: 578-590.

Griffiths, P.E. and Stotz, K. (2013). Genetics and Philosophy: An Introduction. Cambridge University Press.

Keller, E.F. (1983) A Feeling for the Organism. Freeman: New York.

Kidwell, M.G. and Lisch D.R. (2001), "Transposable elements, parasitic DNA, and genome evolution." *Evolution* 55: 1-24.

Linquist, S. B. Saylor, K. Cottenie, T.A. Elliott, S.C. Kremer & T.R. Gregory (2013**), "**Distinguishing ecological from evolutionary approaches to genetic elements**."** *Biological Reviews*, 88(3): 573-584

Linquist, S. Cottenie, K. Elliott, T.A. Saylor, B. Kremer, S.C. and Gregory, T.R. (2015). Applying ecological models to communities of genetic elements: the case of neutral theory. *Molecular Ecology*, 24: 3232-3242

Koertge, N. (1991) The function of credit in Hull's evolutionary model of science. Proceedings of the Philosophy of Science Associatiion 2:237-244.

Mattick, J.S. (2009), "Deconstructing the dogma." Annals of the New York Academy of Sciences. 1178:

McClintock, B. (1984), "The significance of responses of the genome to challenge." Science, 226, 792-801.

Northcott, R. (2008), "Can ANOVA measure causal strength?" The Quarterly Review of Biology, 83: 47-55.

Ohno, S. (1972), "So much 'junk' DNA in our genome. In Evolution of Genetic Systems. HH, Smith, SJ Price AH Sparow FW Studer and JD Yourno (Eds). Gordon and Breach: New York.

Ohno, S. (1973), "Evolutional reasons for having so much junk DNA." In Modern Aspects of Cytogenetics: Constitutive Heterochromatin in Man. In RA Pfeiffer (Ed FK Schattaner Verlag: Stuttgart, Germany. Pp 169-173. .)

Okasha, S. (2005). "Multilevel selection and the major transitions in evolution." Philosophy of Science 72:1013-1025.

Okasha, S. (2016), "The relation between kin and multilevel selection: an approach using causal graphs. British Journal for the Philosophy of Science, 67: 435-470.

Orgel, L. Crick, F. "Selfish DNA: the ultimate parasite." Bature 284: 604-607.

Poulin, R. (2007), "Are there laws in parasite ecology?" Parasitology, 134: 763-776.

Schoener, T.W. (2011), "The newest synthesis: Understanding the interplay between evolutionary and ecological dynamics." Science, New Series 331: 426-429.

Shapiro, J.A. (1992), Barbara McClintock, 1902-1992. Bioessays 14: 791.

Sober, E. (1988), "Apportioning causal responsibility." Journal of Philosophy, 85:303-318.

Sperber, D. (2010), "the guru effect." Review of Philosophy and Psychology, 1: 583-592.

Sterelny, K. (2007), "What is evolvability?" In Handbook of the Philosophy of Science. M Mathen and C. Stephens (Eds). Elesvier.

Thomas, S. et al (2012), "Applying ecologicl and evolutionary theory to cancer: a long and wionding road." Evolutionary Applications, 1-10.

Woodwaard, J. & Goodstein, D. (1996). Conduct, misconduct, and the structure of science. American Scientist 84: 479-490.