The Genomic Challenge to Adaptationism

Sahotra Sarkar

ABSTRACT

Since the late 1990s, the characterization of complete DNA sequences for a large and taxonomically diverse set of species has continued to gain in speed and accuracy. Sequence analyses have indicated a strikingly baroque structure for most eukaryotic genomes, with multiple repeats of DNA sequences and with very little of the DNA specifying proteins. Much of the DNA in these genomes has no known function. These results have generated strong interest in the factors that govern the evolution of genome architecture. While adaptationist 'just so' stories have been offered (as typically occurs in every area of biology), recent theoretical analyses based on mathematical population genetics strongly suggest that non-adaptive processes dominate genome architecture evolution. This article critically synthesizes and develops these arguments, explicating a core argument along with several variants. It provides a critical assessment of the evidence that supports these arguments' premises. It also analyses adaptationist responses to these arguments and notes potential problems with the core argument. These theoretical analyses continue the molecular reinterpretation of evolution initiated by the neutral theory in 1968. The article ends by noting that some of these arguments can also be extended to evolution at higher levels of organization which raises questions about adaptationism in general. This remains a puzzle because there is probably little reason to doubt that many organismic features are genuine adaptations.

1 Introduction

Ever since Darwin and Wallace, natural selection has often been regarded as a major, if not the only, mechanism of evolutionary change. In what follows, this is the view that will be construed as 'adaptationism', though several nuances of that term will be discussed later (Section 2). Throughout the twentieth century, this adaptationist interpretation of evolution was also routinely challenged. In the first decades, for instance, de Vries ([1901], [1903]) emphasized mutations with large effects, while Hagedoorn and Hagedoorn-Vorstheuvel la Brand ([1921]) emphasized chance (Sarkar [2004]). A much more serious challenge to adaptationism began in the late 1960s after the emergence of molecular biology. Motivated by Haldane's ([1957]) argument for a cost to selection (due to the elimination of less fit individuals), Kimura ([1968]) argued that selection could not maintain the high levels of molecular polymorphism that had recently been recorded; rather, according to him, these variants must be neutral. Drawing on Kimura's calculations, King and Jukes ([1969]) went rhetorically further to announce the advent of a 'non-Darwinian' model of evolution.

The neutral theory was systematically criticized. Adaptationists (or 'selectionists', as they usually called themselves) reinterpreted the data, for instance, by invoking models with randomly fluctuating selection that mimicked the results of neutral models (Gillespie [1991]). Independent of these arguments, adaptationism was famously criticized by Gould and Lewontin ([1979]) at all phenotypic levels as consisting of 'just so' stories unsupported by credible evidence. There were many adaptationist responses to this argument, perhaps most famously by Mayr ([1983]), who argued that the long history of adaptationism in evolutionary biology had obviously been a success.

Gould and Lewontin's article initiated a controversy that continues today (Nielsen [2009]). What has changed is the context, with the availability (since 2000) of full genome sequences for an increasing number of species that have permitted tests of selection at increasingly higher levels of precision. Full-sequence data have also revealed complex architectures for eukaryotic genomes. Relevant complexities go well beyond the expectations of the 1990s (Sarkar [2006]), even though discovery of 'junk DNA', 'split genes', and other such genomic features in the late 1970s had alerted biologists to the complexity of eukaryotic genomes compared to prokaryotes (see Section 3). What has emerged in the genomic era is a dynamic model of the genome with a large role for mobile genetic elements (more accurately, mobile 'DNA

elements', because most of these elements are not associated with genes) in many lineages, including humans.

The purpose of this article is to argue that these developments in genomics present a new challenge to an adaptationist interpretation of evolution, at least at the level of the genome.[1] This challenge deserves attention—and scrutiny—because recent claims from the ENCODE project (Encode Project Consortium [2012]) have generated controversy over adaptation and function in the genome (Eddy [2012], [2013]; Graur *et al.* [2013]; Niu and Jiang [2013]). The purpose of that project was to catalogue all 'functional elements' in the human genome. Trouble arose because ENCODE's definition of 'function' allegedly removed reference to natural selection (and, consequently, to adaptation). On the basis of that definition, the ENCODE investigators argued that less than twenty percent of the genome consisted of 'junk DNA' as opposed to the textbook figure of more than ninety percent. According to them, more than eighty percent of the human genomic DNA was 'functional'. These claims have provoked explicit philosophical discussion of the proper definition of 'function' within the scientific literature (Eddy [2013]; Graur *et al.* [2013]). That issue remains unresolved at present. However, the arguments of this article support the claim that, if function is linked to adaptation, the figure claimed by ENCODE is exaggerated.

Even before the advent of genomics, challenges to the received view of evolution posed by the new developments in eukaryotic genetics (which were a prelude to genomics) were brought to philosophical attention in a remarkable piece by Doolittle ([1985]), but ignored in the philosophical literature (with (Ruse [1988]) apparently being the only exception). In recent work in evolutionary genomics, the challenge to adaptationism has been extended and forcefully urged in the biological literature by Lynch ([2007a], [2007b]; Lynch and Conery [2003]) and Koonin ([2009], [2012]), among others (for example, Maeso *et al.* ([2012]) and (Stoltzfus [2012])). This article aims to show that these new developments also deserve sustained philosophical attention because they fundamentally challenge how evolution should be viewed.

The developments in genomics referred to earlier extend the molecular reinterpretation of evolution initiated by the neutral theory. As in the case of the neutral theory, the arguments rely fundamentally on deploying mathematical results from population genetics at the level of DNA but, beyond the earlier analyses, the arguments below also draw heavily on physical properties of DNA that facilitate evolutionary changes in genomes. This is not to suggest that the physical mechanisms operating at the genomic level are all

---

[1] For a different set of arguments that also call for the re-evaluation of generally unsupported (at least, not fully supported) adaptationist claims using genomic techniques, see (Barrett and Hoekstra [2011]), who provide a list of cases in which traits/genes have been dubbed adaptive without adequate support.

well-understood. It is possible that there is a range of molecular mechanisms acting at the genomic level that have complex relations to possible adaptive dynamics at higher levels of organization (for example, the organismic level) and that these dynamics may affect the production of variation at the genomic level. However, this issue will be left for further analysis on another occasion; not enough is known about such mechanisms for them to warrant philosophical analysis at present.

Section 2 will make some preliminary observations about how adaptationism has been construed in the literature. It will note that the issue that will be at stake in this article is the question of whether selection is relevant irrespective of whether optimization is achieved. Thus what will be criticized is a very weak form of adaptationism, *ipso facto* excluding stronger forms. Section 3 will review some of the features of eukaryotic genomes that pose problems for adaptationism, including features discovered during the early period of eukaryotic genetics (Section 3.1) and the more recent findings of genomics (Section 3.2). Section 4 will build the case against adaptationism. To set the stage, it will begin by noting examples of just so adaptationist stories about genome architecture (Section 4.1). Next, it will present the core argument against adaptationism (Section 4.2); this formulation synthesizes several arguments present in the reviews by Lynch ([2007b]) and Koonin ([2012]), though the precise formulation given here is new. The evidence in favour of the soundness of this argument will be reviewed. Next, three variants of the core argument and the evidence in support of their premises will be discussed (Section 4.3). The first two are implicit in Lynch's (for example, [2007b]) work; the third is new. Finally, some examples of putatively non-adaptive features of genome architecture will be briefly described (Section 4.4).

Section 5 turns to a range of adaptationist responses and will assess them critically. It will argue that the most compelling one is a denial of one of the empirical premises of the core argument (namely, that there is a negative correlation between genome size and population size). Finally, Section 6 will return to the task of putting these arguments in their philosophical context. It will also note that the core argument is applicable at any level of organization—consequently it potentially challenges adaptationism at levels higher than that of genome architecture. The article thus ends with a puzzle: how to reconcile the core argument with likely adaptationist evolution at these levels, in particular, at the organismic level. This puzzle is left unresolved.

2 Preliminaries: Senses of Adaptationism

The term 'adaptation' can be used to refer to a process (of adaptation), to a state of affairs (for instance, a state of adaptation to some environment), or to an entity (that is, the biological feature that is an adaptation). Little confusion

typically results from this ambiguity since the context makes clear which use is relevant. The first of these uses is associated with what has been called 'empirical adaptationism': the 'view that natural selection is ubiquitous, free from constraints, and provides a sufficient explanation for the evolution of most traits, which are "locally" optimal, that is, the observed trait is superior to any alternative that does not require "redefining" the organism' (Orzack and Forber [2010]). The other two are associated with what has similarly been called 'explanatory adaptationism': 'the view that explaining traits as adaptation resulting from natural selection is the central goal of evolutionary biology' (Orzack and Forber [2010]). Finally, 'methodological adaptationism' has also been distinguished as 'the view that looking first for adaptation via natural selection is the most efficient approach when trying to understand the evolution of any given trait' (Orzack and Forber [2010]), though it is far from clear that the distinction between explanatory and methodological reductionism is of much salience. (The first strongly suggests—if not requires—the latter.)

Neither explanatory nor methodological adaptationism will be a concern of this article since they seem to have few, if any, proponents in genomics—the well-recognized complexities of genome sequences (which will be discussed in Section 3) typically preclude such a strong commitment to the dominance of natural selection. Rather, the focus will be on empirical adaptationism. Now, the definition of empirical adaptationism given above has two components that may not be compatible with each other in many circumstances: (i) the operation of natural selection, and (ii) the optimality of the end product. The trouble is that—except for the simplest cases of selection (simplest in the sense that the genetic basis for a trait is simple)—it is mathematically trivial to show that natural selection does not lead to an equilibrium that is a (local) maximum of the mean fitness of a population (Moran [1964]; Sarkar [2014]). Adaptationists have typically argued that such situations can be reinterpreted as one of constrained optimization, that is, optimization subject to constraints that are imposed by the structure of the genome (Orzack and Forber [2010]).

Lewens ([2009]), who offered a different taxonomy of adaptationism, subdivided empirical adaptationism into three more fine-grained categories: panselectionism, 'good-designism', and gradualism. It is unclear why the last of these (which only requires selection to operate slowly and step-by-small-step) is a category of adaptationism at all; it will be ignored here. However, the first of these corresponds to component (i) and the second of these corresponds, roughly, to component (ii).[2]

---

[2] Lewens ([2009]) distinguishes between constrained optimization and good design; however, the latter concept is left unexplicated.

This analysis will not rely on any optimality criterion. In what follows, empirical adaptationism will be taken to require only that the operation of natural selection is paramount and constitutes a sufficient explanation of a trait, that is, it will correspond to what Lewens ([2009]) called pan-selectionism. This choice is standard in recent discussions of evolution at the genomic level (for example, Barrett and Hoekstra [2011]); the term 'adaptationism' will be preferred to 'pan-selectionism' to maintain continuity with this literature. This means that the critique of adaptationism presented here is more in the spirit of the neutralist and nearly neutralist theories rather than that of Gould and Lewontin ([1979]), who required more than natural selection for adaptation (following Lewontin [1978]). In other words, from the perspective of this article, Lewontin ([1974]) was an advocate rather than a critic of adaptationism because he sided with the selectionists in the neutralism–selectionism debate. Thus, this choice makes the present critique logically stronger than that of Gould and Lewontin ([1979]) in the sense that it would accept as an adaptation any feature that is sufficiently explained by natural selection whether or not it constitutes a local optimum (see also, Lewontin [1978]). The point is that even this weak form of empirical adaptationism is challenged by the findings of recent genomics.

3 Genome Architecture

This section will summarize the problems and puzzles posed by eukaryotic genome architecture that have emerged over the past three decades. The focus is on eukaryotes because of the emergence of structural and behavioural complexity in them, especially at the macroscopic level, which has been of biological interest since before Darwin and Wallace.

Classical genetics conceived of the eukaryotic genome as paired linear sets of loci at each of which alleles (versions of genes) were specified.[3] Each of these sets corresponded to a chromosome. It was implicitly expected (presumably on adaptationist grounds), but with no empirical basis, that each position on the chromosome specified a gene that, in turn, specified a protein; otherwise there would be a potential for irrelevant waste in evolution.[4] This was referred

---

[3]   For expository simplicity, this discussion is limited to diploids and ignores sex chromosomes. Nothing conceptually new is introduced by incorporating these complexities. For more detail, see (Sarkar [1998]).

[4]   In the article that explicitly introduced the C-value paradox (see below for further discussion of this paradox), Thomas ([1971], p. 251) claimed that non-functional DNA 'offends the principle of parsimony'. The implicit adaptationism requires natural selection to achieve parsimony (the meaning of which remains unspecified). Similarly, after the demonstration of the existence of large segments of non-coding DNA, in a penetrating discussion of the C-value paradox, Moore ([1984], p. 425) put it thus: 'we might expect an economical use of DNA, such that most of it would code for protein (as it does in prokaryotes)'. Here the implicit adaptationism is that which requires economy in DNA use.

to as a 'beads-on-a-string' model (Dunn [1964]). However, the advent of the operon model for gene regulation in prokaryotes in the 1960s suggested that parts of the DNA sequence did not specify proteins but played regulatory roles. This did not pose a problem for adaptationism since these parts of DNA sequences still had a function for which they could have been selected.

By the late 1960s, it was also known that repeated DNA sequences were ubiquitous in eukaryotic genomes (Britten and Kohne [1968]), suggesting a possible regulatory role for such units (Britten and Davidson [1969], [1971]), though the evidence for such a role was non-existent. Moreover, starting with McClintock's ([1950], [1951]) work in the 1940s, it was also known that at least some eukaryotic genomes contained mobile DNA elements which, too, were hypothesized to play a regulatory role. Meanwhile, it also became clear that whole-genome duplication (ploidy change) was associated with some major taxonomic transitions in evolution. In particular, Ohno ([1970]) argued that both genome and tandem gene duplications were major mechanisms of evolution.

By 1971, biologists were aware of at least three aspects of eukaryotic genomes that could not easily be given an adaptationist story. These comprised what was dubbed the 'C-value paradox' with the C-value being the amount of DNA in a (haploid) genome of a germinal cell (Thomas [1971]): (i) closely related eukaryotic species had different DNA amounts in their genome (which, the C-value for a species, was long known to be a constant for that species) (p. 247); (ii) there was no good correlation between the C-value and the morphological complexity of a species (p. 24); (iii) eukaryotes seemed to contain much more DNA than required for the specification of their proteins (pp. 250–1). (For subsequent theoretical understanding of the C-value paradox, see (Gregory [2001], [2005]).)

3.1 Surprises of early eukaryotic genetics

Thus, there was some indication by 1970 that eukaryotic genomes would exhibit levels of complexity not seen in prokaryotes. Nevertheless, the demonstration in the late 1970s that much of eukaryotic DNA had no role in specifying proteins, and not even any discernible regulatory role, was unexpected.[5] Not only were large segments of DNA not involved in specifying proteins, non-coding sequences were found 'within genes', that is, within segments of DNA that specified a single amino-acid sequence (Berget *et al.* [1977];

---

[5] This is perhaps an understatement. Watson *et al.* ([1983], p. 91) were quoted earlier on the 'unexpected complexity of the eukaryotic genome'. Gilbert ([1978], p. 501) put it as follows: 'Our picture of the organization of genes in higher organisms has recently undergone a revolution'; and Crick ([1979], p. 270): 'There can be no denying that the discovery of splicing has given our ideas a good shake'. No adequate history of these developments is available; see (Sharp [2005]) for a partial history.

Chow *et al.* [1977]). These non-coding sequences were dubbed 'introns' by Gilbert ([1978]) with the coding parts comprising 'exons'. After an RNA transcript was produced from DNA in the nucleus, introns were 'spliced' out before translation at the ribosome in the cytoplasm. An added complexity was that most introns required enzymes for their removal but some did not. Moreover, splicing was not unique: 'alternative splicing' involved the production of more than one messenger RNA (*m*RNA) transcript from the same precursor RNA (and, therefore, from the transcribed DNA sequence). Alternative splicing raised the logical possibility of overlapping genes. These had already been observed in viruses in the mid-1970s; eukaryotic examples followed soon afterwards (Normark *et al.* [1983]). Splicing was found not to be restricted to *m*RNA but also occurred in transfer RNA (*t*RNA) and ribosomal RNA (*r*RNA) (Crick [1979]).

It soon became apparent that non-coding sequences, including introns and regions between genes, constituted most of the genome for all eukaryotic species that were studied. In 1978, Gilbert ([1978]) estimated introns to comprise five to ten times the size of exons in the genome. For most eukaryotes this turned out to be an underestimate. By 1977, it was known that genes often occurred in families, and that non-coding regions between genes included 'pseudogenes' or inactive variants of active genes (Jacq *et al.* [1977]). Repeated DNA sequences, already identified by Britten and Kohne ([1968]), turned out to be ubiquitous (Jelinek *et al.* [1980]). A welcome consequence of these developments was a resolution of the C-value paradox using the presence of non-coding DNA to explain the otherwise paradoxical patterns of variation (Lewin [1980]; Gregory [2001]).

More anomalies were discovered in the 1980s in the form of RNA editing, that is, modification of *m*RNA after splicing (Koslowsky [2004]). Editing processes observed included insertions (and, later, deletions) of codons at the ends of *m*RNA transcripts and in their interior. By 1990, observed editing processes included modification of nucleotides (Schuster *et al.* [1990]; Gualberto *et al.* [1990]). One consequence of these developments was that the relationship between gene and protein became indeterminate.

The discovery of RNA editing added a level of complexity to the control of gene expression. Further complexity was recognized in the 1990s through the discovery of RNA 'interference': RNA transcripts affecting the translation of *m*RNA (Guo and Kemphues [1995]; Rocheleau *et al.* [1997]). Meanwhile, alternatives to the standard genetic code began to be recorded from the 1980s (Caron [1990]). For the context of this article, the most significant development was the extent to which mobile DNA elements were found to be ubiquitous in eukaryotic genomes. More than any other feature, this led to the reconceptualization of genomes as dynamic entities rather than 'beads-on-a-string', what Shapiro ([1995]) dubbed a 'fluid genome'. By 1985, it was

clear that there were two types of mobile DNA elements, those based on a mechanism that included an intermediate RNA stage, and those that did not; the former were dubbed 'retrotransposons' (Boeke *et al*. [1985]). Without complete genome sequences what remained unclear was the extent to which genomes were composed of mobile DNA elements.

3.2 Genome structure, post-2001

By the late 1980s, it was clear that a theoretical understanding of the baroque architecture of eukaryotic genome was not immediately forthcoming. It was one of the factors that motivated the desire for full genome sequences, in particular the Human Genome Project (HGP).[6] The complex political and scientific history of the HGP is not of concern here (see, for example, Cook-Deegan [1994] and McElheny [2010]). By 2001, when the draft sequence of the human genome was published (IHGSC [2001]), besides thirty-nine bacterial species, the genomes of the yeast (*Saccharomyces cerevisae*), the nematode (*Caenorhabditis elegans*), and the fruit-fly (*Drosophila melanogaster*) had already been sequenced. Since then, eukaryotic full genome sequences continue to be reported at a steady rate. The largest eukaryotic genome recorded so far seems to be that of an endemic monocotyledon from Japan, *Paris japonica*, which has ~150,000 Mbp (million base pairs; Pellicer *et al*. [2010]). While this genome is yet to be fully sequenced, the smallest recorded nuclear genome, that of the intracellular parasite, *Encephalitozoon intestinalis*, has recently been sequenced and found to be approximately 2.3 Mbp (Corradi *et al*. [2010]). This variation in genome size will be relevant to the arguments of Section 4.

In 2001, the biggest surprise from the completed human genome sequence was the low number of genes.[7] In the 1990s, while Gilbert ([1992]) put 300,000 as the upper limit of the possible number, most estimates ranged between 60,000 and 140,000, with the 1990 plan for the HGP embracing an estimate of 100,000 (Fields *et al*. [1994]). Instead, the completed sequence suggested about 30,000–40,000 genes (IHGSC [2001]). Since then, this estimate has decreased to 20,000–25,000, with more recent estimates of around 22,500 (Pertea and Salzburg [2010]). The same estimate holds for the mouse *Mus musculus*, and is not much more than the 21,200 estimate for *C. elegans*; *D. melanogaster* has 16,000. Meanwhile, the mustard weed (*Arabidopsis thaliana*) has 25,000 estimated genes but rice (*Oryza sativa*) has as many as 60,200. The pufferfish (*Fugu rubripes*) has 38,000 genes.

---

[6] This point was repeatedly made in the early 1990s by some proponents of the HGP such as Gilbert (for example, [1990], [1991]). See Tauber and Sarkar ([1992], [1993]) for a contemporary analysis.

[7] This discussion is restricted to genes that specified amino acid sequences. All these gene numbers are predictions and must be viewed with caution; they may be incorrect by as much as twenty percent (Lynch [2007b]).

The paradoxical lack of correlation between perceived complexity and gene number has been called the 'G-value paradox' (Hahn and Wray [2002]). The number of genes is also not correlated with genome size. The original report on the sequence (IHGSC [2001]) noted that the human 'proteome' or protein set is much larger (and, in that sense, more complex) than that of invertebrates. This puzzle is resolved by the higher prevalence of alternative splicing in humans. According to recent estimates, more than half of the human genes are subject to alternative splicing, with an average of 2.6 transcript variants per gene; in contrast, only 20% of the genes are alternatively spliced in *C. elegans* and *D. melanogaster*, with an average of 1.3 transcript variants per gene (Lynch [2007b], p. 50).

There were other surprises in the complete human sequence of 2001. The original report claimed that there had been horizontal gene transfer of hundreds of bacterial genes into the human genome; however, this high estimate did not survive further analysis with more recent estimates being around 40 (Salzberg *et al.* [2001]; Kurland *et al.* [2003]; Keeling and Palmer [2008]). The distribution of human genes between the chromosomes and within them was highly uneven (compared to what was found for other species for which sufficient sequences were available at that time). Human genes tend to occur in clusters. Many more details have been added to the knowledge of the architecture of the human genome, and it does not appear that any important feature of the human genome is unique when compared to other eukaryotes. The human genome has about 4,000 pairs of duplicate genes and 5% consists of recently duplicated segments. Almost a third of the genes in the human genome appear to be 'orphans', that is, they have no homologue in any other well-characterized non-primate species. The human genome also has about 15,000 pseudogenes. In 2001, only about 2% of the human genome was estimated to specify amino acid sequences; since then that estimate has come down to 1% (Lynch [2007b], p. 43). The average exon length is 0.15 kB (kilobases); that for introns is 4.66 kB; thus, within each gene, the average intron to exon ratio is about 1:30. While reliable estimation of the amount of regulatory DNA is difficult for a variety of technical reasons, for humans, a minimal estimate is about 1.5 times that for DNA specifying proteins.

In this context, the most important result from 2001 was that almost 50% of the human genome consists of mobile DNA elements. There are about 100 mobile DNA genetic elements per protein-specifying gene. Among the mobile DNA, transposons form 2.8% of the human genome; retrotransposons form 41.8%. Retrotransposons consist of long interspersed elements at 20.4%, short interspersed elements at 13.1%, and long terminal repeat elements at 8.3%. Patterns in other species are equally peculiar. At one extreme is maize (*Zea mays*) in which 85% of the genome consists of mobile DNA elements; at the other extreme is the malarial parasite (*Plasmodium falciparum*), which seems

to have none; *A. thaliana* falls in between at 10% (Rebollo *et al.* [2012]). Mobile DNA elements are responsible for perhaps most large-scale structural changes in genomes including duplication (which is often involved in the genesis of novel genes).

4 The Case against Adaptationism

The baroque architecture of the human genome—and of most eukaryotic genomes—calls out for explanation. Given the long tradition of adaptationist thinking in evolutionary biology, it was perhaps inevitable that adaptationist just so stories proliferated in the wake of a recognition of the complexities of eukaryotic genome architecture. Section 4.1 will note a few of the more compelling just so stories and will begin the task of contrasting them to what happens when arguments are constrained to remain consistent with mathematical population genetics. Section 4.2 will develop the core argument against adaptationism and analyse the evidence in support of its premises. Three variants that modify one of the premises of the core argument are similarly treated in Section 4.3. Finally, some putative examples of non-adaptive features of eukaryotic genome architecture are described in Section 4.4.

4.1 Just so stories versus population genetics

There are a miscellany of relevant just so stories and the discussion here will be limited to some illustrative cases. What deserves emphasis are both their intuitive plausibility and the ease of their construction that Gould and Lewontin ([1979]) derided. For instance, both McClintock ([1950]) and Britten and Davidson ([1969]) assumed that repeated DNA segments had a regulatory role without evidence. The same story animates those today who invoke a regulatory function for the high diversity of small RNA fragments found in eukaryotic cells (for example, Fontdevila [2011]). Analysing splicing in 1979, Crick ([1979], p. 268) observed: 'It is impossible to think about splicing without asking what it is all for [...] how splicing arose in evolution?'. That it was already presumed in this formulation that an answer to the second question (how splicing arose in evolution) would involve answering the first (what splicing is for) betrays the adaptationist commitment that is being challenged in this article. Crick endorsed Gilbert's ([1978]) adaptationist 'exon shuffling' story (see below) for the occurrence of both introns and exons; he also noted the possibility that introns arose by specific DNA insertions into the genome (presumably due to standard physical and chemical factors) and 'splicing evolved as a defense by the cell against an insertion element it was harboring' (p. 269). But Crick presented no evidence.

What Crick was referring to was an earlier argument due to Gilbert ([1978]). When introns were discovered in the late 1970s, Gilbert ([1978]) offered two stories of their origin. Both were adaptationist: (i) Introns existed because they facilitated the speed of evolutionary change. Single point mutations (base changes), if they occurred at intron–exon boundaries, could lead to changes in proteins involving multiple amino acid residues (instead of a single one as would be induced by point mutations in exons). (ii) Introns facilitated exon shuffling, that is, the production of new proteins by bringing together different exons scattered through the genome. The absence of evidence did not prevent the latter story being widely promoted—among others, by Blake ([1978]), Darnell ([1978]), Doolittle ([1978]), and Tonegawa *et al.* ([1978]). (However, Doolittle ([1985]) took a more critical attitude.)

Adaptationist story-telling was not limited to just the existence of DNA repeats and introns. Two more examples will suffice here. Crick ([1979], p. 266) provided an adaptationist argument against the possibility of alternative splicing: 'Should a chromosomal gene arise whose transcript was processed to make more than one protein, I would expect that in the course of evolution the gene would be duplicated, one copy subsequently specializing on one of the proteins and the other copy on the other [. . .] one would expect multiple-choice genes to occur only rarely in the chromosomes of eukaryotes'. That this story did not survive the first full genome sequences serves as a reminder of the frailty of just so stories whenever they make precise predictions. Meanwhile, Normark *et al.* ([1983], pp. 499–500) offered an adaptationist story of the overlap of viral genes: 'these had evolved mainly to optimize the amount of genetic information that could be packaged in the phage head'.[8] This explanation obviously does not suffice for eukaryotes; so, in accord with the finest of adaptationist traditions, a new story was invented: 'an overlapping arrangement of genes can have important regulatory implications both at the level of expression and at the level of protein-protein interaction' ([1983], p. 500). No evidence was presented for either story.[9]

The salient point—and this is where Gould and Lewontin's ([1979]) critique is most relevant—is that these stories are no more than stories: they should not be embraced as a substitute for genuine theorizing. Moreover, as Lynch ([2007a], [2007b]) correctly emphasizes, intellectually respectable evolutionary theorizing must be based on population genetics theory, which forms the substantive core of the relevant evolutionary theory. As Lynch ([2007a], p. 8598) put it: 'the field of population genetics is now so well supported at the empirical level that the litmus test for any evolutionary hypothesis must be

---

[8] Crick ([1979], p. 266) tells essentially the same story: 'I adopt the attitude that in most cases this [the overlap of viral genes] is because viruses are short of DNA and, by various devices, their limited amount of DNA is made to code for more proteins than would otherwise be possible'.

[9] In fairness, it should be noted that the second was clearly intended as speculation.

consistency with fundamental population-genetic principles'. None of the molecular biologists whose views are being questioned in this section, especially those who attempted a theoretical understanding of molecular phenomena (for instance, Crick and Gilbert), explicitly deny Lynch's stricture. Nor does Fontdevila ([2011]) in an extended attempt to provide an adaptationist account of genome evolution.

What, exactly, does theoretical population genetics require? Recall from Section 1, though natural selection is a potentially major mechanism of evolution, drift may counter the effects of selection to be realized and may even lead to the fixation of less fit variants in a population (Haldane [1924], [1932]; Fisher [1930]; Wright [1931]). Even when a less fit variant does not get fixed, it may persist indefinitely in a population: natural selection may not be intense enough to eliminate it. The crucial determinant of the efficacy of natural selection is the population size, more accurately, the effective population size, N_e, about which more will be said below. The reason is straightforward: the smaller a population is, the more varied are the finite samples drawn from it. Thus, the smaller that N_e is, the stronger the effect of drift (Sarkar [2011a]); the inverse, $1/N_e$, is the relevant quantitative measure. This point is important because what is at stake in the core argument of this article is that N_e is small for most eukaryotes but large for most prokaryotes.

It should be emphasized that just so stories are also logically insufficient: to claim the possibility of adaptation, there must be some explicit empirically founded argument to show that, relative to N_e, the intensity of selection, s,[10] is large enough to allow the elimination of variants with lower fitness (as measured by s). (As will be seen below, what matters critically is the value of $|N_e s|$.) Philosophically, perhaps the most salutary aspect of the turn to population genetics in debates over adaptationism is that the mathematical theory of population genetics reduces the relevant debate to empirical questions that can be assessed on the basis of mathematical analysis and empirical data (and the attendant scientific controversies in the case of genomic architecture will be duly addressed below) rather than with plausibility of intuitions and the ingenuity of constructing the just so stories.

Much of theoretical population genetics was developed in the context of the received view of evolution (see Section 1). During the period in which these developments occurred (mainly the 1920s and 1930s), while genetic changes were recognized as being critical to evolution, not enough was known at the molecular level to characterize the variegated ways in which genomes are subject to alteration. Genetic changes were attributed to catch-all 'mutations', the term designating a black box that was yet to be opened. When that

---

[10] Here s represents the difference between the fitness of the two variants. Thus, $s = 0$ represents neutrality, if $s > 0$, the first variant is more fit than the second, and so on. For more detail, see any standard work on theoretical population genetics, for example, (Kimura [1983]).

situation changed, especially in the 1970s and 1980s, population-genetic models began to be constructed to incorporate other changes including, but not limited to, the proliferation of mobile DNA elements.

In this context, three points will be critical to the arguments of Sections 4.2 and 4.3. First, as alluded to earlier (Section 3.2), unravelling the sources and types of DNA variation has shown that the expansion and proliferation of DNA sequences is ubiquitous (Maeso *et al.* [2012]). Except in the case of most prokaryotes (and some small eukaryotes), which typically do not show such a proliferative proclivity, mobile DNA elements are implicated in this phenomenon. While many details are still missing, and a unifying model of DNA proliferation yet to be formulated, it appears clear that such expansion is driven by physical (including chemical) interactions.[11] This fact will play a central role in the core argument of Section 4.2 (and also in its variants in Section 4.3). Even if these elements subsequently assumed major functional roles, the origin of expanded genomes is due to physical processes in the same way that point mutations and recombination are due to physical interactions. All that may subsequently occur through co-option of the expanded DNA is that new functions may evolve and be implicated in the continued persistence of baroque genomes through natural selection. The arguments developed in Sections 4.2 and 4.3 will question this possibility.

Second, much of the baroque structure of the genome is almost certainly functionally detrimental because the larger a genome, the higher the likelihood of detrimental physical instability through physical changes (Lynch [2007b], Chapter 4). As early as 1983 it was realized that introns were a genetic liability that should be subject to negative selection. For instance, twenty-five percent of all mutations in globin genes that resulted in β-thalassemia in *Homo sapiens* arose from splicing errors (Treisman *et al.* [1983]). Similarly, most mobile DNA elements, which can harbour a variety of mutations, presumably have negative consequences. In the late 1980s, it was shown that the insertion of mobile DNA elements could result in disease (Kazazian *et al.* [1988]). Since then, evidence for maladaptiveness of mobile DNA element insertions has accumulated (Rebollo *et al.* [2012]). Indeed, such a deleterious effect may explain what has been called reductive genome evolution that is common to many lineages (Maeso *et al.* [2012]).

Third, the complexity of genomic changes does not challenge the point that N_e and s are the factors relevant to whether natural selection can eliminate

[11] Lynch ([2007a], [2007b], [2011]) calls all generation of genomic variation 'mutation' and many others have followed him here (for example, Maeso *et al.* [2012]). Such a terminological choice suggests that the mechanisms generating variation are far more unified than the evidence warrants. Lynch's terminology will not be adopted here, partly to underscore the fact that a unified account of variation is not available now, though it would be of great interest in generating a more complete account of genome evolution.

deleterious variants. If $(1/N_e) \gg |s|$, or equivalently, $|N_e s| \ll 1$, selection will be ineffective and evolution will be described by a nearly neutral theory (see Section 1; Ohta [1973], [1996], [2013]; Takahata [2001]). Since even $s \sim 0.1$ constitutes very strong selection, what is critical is the value of N_e. It should, therefore, come as no surprise that this has been the most prominent source of controversy (see Section 5). A few points about N_e are worth emphasis (Charlesworth [2002], [2009]; Charlesworth and Barton [2004]). Not only is N_e less than the number of individuals in the population (that is, N), it is typically much less than even the number of breeding individuals in a population. A variety of factors often lower N_e by several orders of magnitude: (i) If the population size changes, the long-term value of N_e is the harmonic mean of the values for each generation. If a population has recently expanded, $N_e \ll N$. (ii) Selection at loci linked to a given locus decreases the N_e value for that locus. This means that low levels of recombination may decrease N_e. (iii) Loci on sex chromosomes (in diploid populations) often have lower N_e than those on autosomal chromosomes. (iv) Most departures from random mating lower N_e. (v) Population substructure also leads to N_e being lower than N. This is not a complete inventory but it shows that, in almost all circumstances relevant to genome evolution, very probably $N_e \ll N$. Lynch ([2007a], p. 8600) provides some tentative estimates while emphasizing the many uncertainties. Rough estimates of $|N_e s|$ are $\sim 10^{-1}$ for prokaryotes; $\sim 10^{-2}$ for unicellular eukaryotes, invertebrates, and land plants; and $\sim 10^{-3}$ for vertebrates.

However, because the core argument below relies so heavily on this theoretical work, a caveat must be introduced. For historical populations it is impossible to produce precise estimates for N, N_e, or s. Consequently, the arguments below must rely on ordinal comparisons using ranges of estimates rather than on quantitative data. In this sense, for the time being, they still remain 'qualitative' without being merely 'verbal' (like the just so stories criticized earlier).

4.2 The core argument

The core argument developed here depends critically on the mathematical consequences of population genetics discussed at the end of Section 4.1. A version of it is implicitly formulated by Lynch ([2007a], [2007b]), but it is not explicitly formulated as it will be presented here; an even less explicit version is to be found in (Koonin [2012]). This argument has four premises:

P1: The physical properties of DNA and its cellular environment lead to increased genome size and its baroque structure.
P2: Genome size is negatively correlated with population size.
P3: Selection acts against larger genomes.

P4: Small population sizes prevent the elimination of features selected against unless selection is very strong.

_____

C: Genomes increase in size, diversity, and so on and persist even though selection acts against these features.

Thus, according to the core argument, Crick was in error when he claimed (though only in the context of introns): 'Even if it [a change in the genome] has already spread, it cannot spread indefinitely without having some advantage *since otherwise it would be deleted*' (Crick [1979], p. 268, emphasis added).

Lynch ([2011]) has correctly pointed out that, contrary to claims made by Pigliucci ([2007]) and Gregory and Witt ([2008]), the model of evolution that emerges from the core argument is not a neutral model. It assumes that changes in the genome are maladaptive—in Lynch's ([2011]) version it is a 'mutational-hazard' model. In this sense, it is essentially a nearly neutral model. Perhaps the single most telling piece of evidence in favour of this model is that in prokaryotes (and small eukaryotes), which have the largest N_e among all species, genomes have typically not expanded; presumably even weak negative selection suffices to maintain the compactness of these genomes (though other factors such as energetic consideration may have a role either directly or, more likely, by resulting in weak selection).

The critical issue is the status of the premises of the core argument. The most important of these premises is P4, which is the only one that incorporates an assumption about the dynamics of evolutionary change. The discussion of population genetics theory in Section 4.1 shows that P4 should be regarded as being beyond (reasonable) question. Some of the evidence in favour of premises P1 and P3 was also sketched in Section 4.1. In principle, premise P1 should be based on a detailed understanding of molecular mechanisms. Such an understanding is not available at present and it must be regarded as an empirical generalization derived from studies of changes in genome size and complexity in phylogenetic lineages.

Premise P3 is similarly an empirical generalization. There is one important class of exceptions. The evidence in favour of it (sketched in Section 4.1) that supported a 'mutational-hazard' model may not be applicable when genome expansion is due to ploidy change (whole-genome duplication). Such ploidy change is ubiquitous amongst plants and can also occur in bacteria. In these cases, the premises of the core argument are not all satisfied—and, as should then be no surprise, varied genome sizes occur irrespective of population size (see, also, Section 4.4).

Perhaps, the most relevant point in this context is that these premises (P1 and P3) are not the focus of criticism from adaptationists who would deny the

conclusion, C. What these criticisms focus on is the premise P2. It has been presumed as an empirical generalization by Lynch ([2007a], [2007b]). More will be said about its epistemic status in Section 4.3, where it will be replaced by other assumptions to generate three variants of the core argument. It will also be discussed in some detail as part of the adaptationist responses in Section 5.

4.3 Three variants of the core argument

This section will analyse three variants of the core argument generated by replacing premise P2 with alternatives. The first of these arguments, which will be called the 'body size' (BS) argument, replaces P2 with two other premises:

> P2.1: Genome size is positively correlated with body size.
> P2.2: Body size is negatively correlated with population size.

It should be clear that premise P2 is a logical consequence of premises P2.1 and P2.2 of the BS argument. The model on which the BS argument is based goes back to Lynch and Conery ([2003]); it is also implicitly invoked by Lynch ([2007b], p. 41). The ecological evidence for premise P2.2 is overwhelming. Moreover, going beyond correlations (though this is all that is required by the dynamical premise P4 to generate conclusion C), small population size is very likely a necessary consequence of large body size because of physiological and resource constraints. However, because small population size may result from factors other than large body size, the BS argument has a more limited scope than the core argument.

For the BS argument, the crucial issue is the status of premise P2.1. It seems to be contradicted by one of the considerations that led the formulation of the C-value paradox (recall Section 3): there is no correlation between genome size and organismic complexity, with size as a surrogate for complexity. However, this absence of correlation may be a result of focussing on outliers in each genome or body size class (Lynch [2007b], p. 32). Once all the data are included, there may well be the requisite correlation. A recent review by Dufresne and Jeffery ([2011]) reports a positive correlation between genome size and body size in several taxa including aphids, flies, mollusks, flatworks, and copepods. However, some taxa do not show such a correlation; these include oligochaete annelids and beetles. Mammals show a positive correlation at the levels of species and genera but not at higher taxonomic levels. Moreover, the data remain sparse. It deserves emphasis that the status of premise P2.1 is particularly salient for the debate on adaptationism. If it is correct, the BS argument is at least highly plausible and this plausibility makes the core argument (which has weaker premises) even more likely to be sound. In that case, the handful of studies that purport to deny premise P2 of the core

argument (namely, a negative correlation between genome and population sizes in some taxa—see the discussion of the adaptationist response in Section 5) lose some of their force and can be treated as exceptions, at least for the time being and until similar results are obtained from an exhaustive set of taxa. Finally, note that the evidence for premises P2.1 and P2.2 also constitutes evidence for premise P2 of the core argument.

The second variant argument supplements the BS argument with an additional premise:

> P2.3: Large body size is selected for during evolution.

This argument is only being considered here because it has been invoked in this context: Lynch ([2007b], p. 41) offers it because it has the advantage of specifying a mechanism for the increase of body size. However, this reticulation of the BS argument weakens the case against adaptationism since selection is given some role, though an indirect one, in the origin of genomic architectures. Additionally, it generates the empirical problem of finding evidence for selection for large body size. Whether there is any compelling evidence for this claim remains a matter of controversy. The focus in the rest of this article will remain on the BS argument itself without this addition.

The final argument to be considered replaces premise P.2.1 in the BS argument by:

> P2.1*: Larger body size results from larger genome size.

Premise P2.1* is intended to suggest that there is some mechanism that leads to or enables (and it is deliberately vague on this point in the absence of relevant evidence) the formation of larger bodies; it is neutral on whether there is any selection for body size. The point is that it does not require selection. Moreover, if premise P2.2 is also taken to incorporate the mechanism mentioned earlier, this argument (which will be called the 'genome size' argument) goes beyond correlations. But the empirical status of premise P2.1* remains to be explored. It is introduced here only because of its plausibility.

4.4 Examples: Non-adaptive features of the genome

The discussion of Sections 4.2 and 4.3 shows that there is ample, though not fully decisive, evidence in favour of all the premises of the core argument and only slightly less support for those of the BS argument. The only problematic premise is P2 or (P2.1 and P2.2), and its status will be explored again in Section 5. Meanwhile, the scope of the genomic challenge to adaptationism will be illustrated here using details of four genomic features that seem to have

non-adaptive explanations. These examples also show how the core argument can be deployed in individual cases:

(1) Genomes are streamlined in microbial species but bloated in multicellular lineages (Lynch [2006], [2007b]; Maeso *et al.* [2012]): As noted in Section 4.1, $|N_e s|$ is larger in microbial species than in multicellular lineages (and, among microbes, largest for prokaryotes). Consequently, selection is much more effective for the former than for the latter. Given that larger genomes have deleterious consequences, excess DNA appears to have been removed from the microbial genomes by selection (that is, through reductive genome evolution). A recent review also found recurrent reductive genome evolution in several eukaryotic lineages for which $|N_e s|$ is estimated to have been sufficiently large (Maeso *et al.* [2012]); thus, the streamlining of genomes is not limited to prokaryotic (or even microbial) species depending on whether the premises of the core argument are correct. This means that, while selection can explain the streamlining and simplification of microbial genomes, the baroque structure and expansion of the genomes of multicellular species requires a non-adaptive explanation. An alternative adaptationist hypothesis is that compactness of prokaryotic genomes is due to indirect selection for metabolic features; Lynch ([2006]) reviewed the evidence for this possibility and concludes that it is at best equivocal. Moreover, even this alternative hypothesis does not provide an adaptationist argument for the expansion of the other eukaryotic genomes.

(2) Local genome sequences are conserved but genome structure is not (Koonin [2009]): There is likely to be strong selection for those genome sequences that specify proteins (that is, for classical genes); sufficiently strong selection would ensure local sequence conservation even in populations with low N_e. No such constraint operates on genome structure. Even if structural changes are maladaptive, they could persist in the population. Given a random origin of these structural variations, the result would be their diversity, that is, non-conservation. These structural changes include the loss of operons in almost all eukaryotes (Lynch [2006]).

(3) Differential proliferation of mobile DNA elements in unicellular versus multicellular species (Lynch [2007b]): For the same reasons as in the first example, mobile DNA elements can proliferate more successfully in multicellular than in unicellular species because the former have lower N_e than the latter. This is a pattern seen across taxa.

(4) Variation in organelle genome architecture between animals and plants (Lynch [2007b]): Animal mitochondrial genomes are highly streamlined, while plant organelle genomes (including mitochondrial genomes) are extraordinarily bloated with DNA that does not specify proteins. The best estimates for N_e for the two groups are roughly equal, which means that drift cannot account for the observed difference. Instead, what explains the difference is that the rate of genome changes in animal organelles (for example, rates of mutation or ploidy change) is lower than that in plant by a factor of 100, which allows DNA segments to accumulate in the latter. In contrast, the mutation rate in animal mitochondria makes their population-genetic features similar to those of prokaryotes. (Recall what matters is $|N_e s|$ and that s is correlated with u, the per-nucleotide mutation rate (Lynch [2007a], p. 8599).) Thus, physical processes (mechanisms of genome change) explain this difference between animal and plant organelle genomes.

Perhaps what deserves most emphasis is the diversity of phenomena that are thus being subsumed under a single general picture of genome architecture evolution. More examples are discussed by Lynch ([2007a], [2007b]), Koonin ([2009], [2012]), and Maeso *et al.* ([2012]), from where these examples were drawn.

5 Adaptationist Responses

There have been many attempts to show that there are 'signatures' of selection in human and other genomes; these are typically based on departures of sequences from expectations from neutral models (for example, Harris [2013]). Since many of these attempts claim success, these analyses would seem to provide support for adaptationism—and present problems for the core argument of Section 4 (and its variants). However, because of what was already said regarding the second example of Section 4.4, these analyses are not relevant to the question of large-scale genome structure and architecture. Reports of selection at the level of sequences are thus compatible with the core argument. Moreover, as Barrett and Hoekstra ([2011]) have pointed out, many of the claims of adaptation based on sequence data suffer from a critical incompleteness: the fitness of the corresponding phenotypes is not independently assessed.

Adaptationists have, therefore, appropriately focussed on questioning premise P2 in various ways, and this section will describe and assess these responses. Note that the claim that selection is weak in most relevant circumstances is not questioned in the ongoing debate over its role. Rather, the issue

at stake is the value of N_e, which must be correlated with the genome size for premise P2. Therefore, the problem that must be broached is that of the reliable estimation of N_e, especially in the distant past for the lineages that have evolved into extant organisms. In contrast, reasonable estimates of past genome sizes may be inferred from known mechanisms of genomic change; there is compelling evidence of large historical genome size in most of the relevant lineages (Koonin [2012]). Thus, a negative correlation between N_e and genome size, if established with sufficient reliability, would do much to resolve the status of the core argument and, therefore, that of adaptationism.

These adaptationist objections have sometimes focussed only on N_e (rather than its correlation with genome size); for instance, Fontdevila ([2011], p. 13) has emphasized the uncertainties of such estimates by arguing (somewhat strangely) that past fluctuations and bottlenecks could have biased estimates downwards. While this particular objection is mathematically implausible (for reasons noted at the end of Section 4.1), these uncertainties do exist. However, the most pertinent adaptationist objection concerns P2 directly, that is, the status of the posited correlation between smaller population and larger genome size. The most systematic evidence for such a correlation was presented early by Lynch and Conery ([2003]) using estimates of $N_e u$ where u is the per-nucleotide mutation rate which is correlated with s. Lynch and Conery presented data from forty-three species across thirty taxa ranging from bacteria, angiosperms, fungi, and mammals that gave rise to a 66% correlation (using Pearson's coefficient). Daubin and Moran ([2004]) questioned the accuracy of their $N_e u$ estimates for bacteria; however, that does not bring Lynch and Conery's general conclusions into question.

Supporting evidence for the presumed correlation came from an analysis by Yi and Streelman ([2005]; Yi [2006]) of genomes of freshwater and marine ray-finned fish. Freshwater species have lower N_e than marine species and larger genome sizes. However, critics argued that the larger genome sizes may be due to ancient polyploidy (Gregory and Witt [2008]). Whitney *et al.* ([2010]) found no correlation between genome size and N_e for 205 plant species. Ai *et al.* ([2012]) found no correlation between N_e and genome size in a study of ten diploid Oryza species. However, none of these studies may be germane to the issues under debate because the negative correlation between genome size and N_e is not expected at the level of species within genera or even at the level of genera; rather, it is expected at higher taxonomic levels (Lynch [2007b]). Nevertheless, it remains a problem that the appropriate phylogenetic level is not specified in the core argument, raising the objection that the premises of the core argument (and its variants) are untestable (Fontdevila [2011]). This objection may be met by noting that, for each specific case, the level can be specified: it is that at which the relevant phenotypic variation occurs.

It seems clear that the relevant taxonomic level will have to much higher than that of genera. Nevertheless, much more work is necessary before this objection can be entirely dismissed.

Even more compelling are objections raised by Whitney and Garland ([2010]), who challenged the correlation reported by Lynch and Conery ([2003]) on the grounds that each species could not be treated as an independent data point (in the correlation coefficient computations) because of shared phylogenetic histories. Once these are taken into account, they argued, no statistically significant correlation remains. Lynch ([2011]) responded by pointing out that the genomic features being used may not have a shared evolutionary history among related taxa; the relevant common ancestry could be sufficiently distant to be irrelevant or the feature could have emerged through convergent evolution. Moreover, as a general point, if the relevant taxonomic level is higher than that of the genus, the effect of phylogenetic relatedness becomes progressively weaker. While Whitney *et al.* ([2011]) remained unconvinced and continued to argue for the relevance of phylogeny, disputants agree that the issues at stake can ultimately only be decided by an analysis of much larger sets of taxa.

Following Charlesworth and Barton ([2004]), Whitney and Garland ([2010]) and Whitney *et al.* ([2011]) also object that a correlation between N_e and genome size may arise because of a correlation between N_e and other organismic features such as body size, mating system, developmental rate, or metabolic rate. However, this objection rests on a logical fallacy: the core argument, and its premise P2, only assume a correlation irrespective of where it comes from. There is only one mitigated sense in which low N_e may 'explain' large genome size, namely, if there is a correlation between the two features, that correlation will facilitate further expansion of the genome. However, this possibility is independent of the question as to what initially induced the correlation. Moreover, the BS model explicitly connects both N_e and genome size to body size. Unlike Whitney and Garland's ([2010]) earlier objection from spurious correlations in the data sets (discussed in the last paragraph), this new objection does little to mitigate the genomic challenge to adaptationism. These arguments support a tentative conclusion that, short of definitive analyses verifying and extending the conclusions of Whitney and Garland ([2010]) to a much wider range of taxa, the genomic challenge to adaptationism remains unmet.

6 Concluding Remarks

The HGP may not have delivered on almost all of its medical promises (Sarkar [2001]; Hall [2010]). However, few of its original critics (for example, Sarkar and Tauber [1991]) would doubt that, by jump-starting genomics, it has

helped expand the frontiers of biology, including evolutionary biology, in unprecedented and unexpected ways. This has been clear since the 2001 publication of the draft human genome that brought into focus its bloated structure and an unexpected paucity of protein-specifying genes. The full sequencing of genomes of other species established that the human genome was not peculiar among eukaryotes—the many odd features of eukaryotic genomes ('odd' in the sense that their occurrence could not be easily explained) were noted earlier in Section 3.2. The challenge has become to explain how and why these features emerged and why they are distributed among taxa in the way that they are.

These genomic features were sufficiently odd that adaptationist just so stories, though hardly non-existent (see Section 4.1), have failed to gain much traction. The problem is that it is far easier to argue that the peculiar features of bloated eukaryotic genomes are maladaptive than that they are adaptive. However, claims of maladaptation must also be based on the strictures of population genetics theory. The aim of this article has been to show how this is done by drawing on and extending the analyses of Lynch (for example, [2007a], [2007b]) and Koonin (for example, [2009], [2012]), and putting them in their historical and philosophical context. Five aspects of these arguments deserve further emphasis.

First, the arguments discussed in Sections 4.2 and 4.3 fall squarely within the tradition that started with the neutral theory and continued with the nearly neutral theory of molecular evolution (see Section 1). As noted earlier, these arguments extend the molecular reinterpretation of evolution initiated by Kimura ([1968]) and King and Jukes ([1969]) that called into question the role of natural selection in evolution. Moreover, this extension goes beyond the question of 'mere' molecular composition that motivated the neutral and nearly neutral theories: it moves into the realm of complex structural traits at the genomic level.

Second, Lynch (for example, [2007a], [2007b]) sometimes suggests that non-adaptationist models should be regarded as null models for (classical) statistical hypothesis testing. This can be justified on the ground that a model with no selection is simpler than a model that posits selection (which is an additional assumption) and, therefore, more appropriate as a null model. However, here, the arguments in Section 4.2 have been cast to show that the core argument and its variants offer a better explanation of genome architecture than adaptationist alternatives because a low effective population size (N_e) makes selection ineffective. Thus, the claims defended in this article are stronger than the non-rejection of a null model. Additionally, these arguments do not assume the framework of (classical) statistical hypothesis testing. Therefore, they do not fall afoul of those approaches that reject that

framework, such as and most importantly, Bayesian inference, which is increasingly becoming the standard in many areas of biology.[12]

Third, in the absence of quantitative estimates of the intensity of selection (and recall that even N_e estimates, let alone those of s, are problematic—see Section 4.1), the arguments and inferences of this article remain qualitative (but not merely verbal—recall Section 4.1). In that sense, these arguments will require much further development to achieve the level of rigour traditionally associated with theoretical population genetics (Charlesworth [2008]). This point was also emphasized by Lynch ([2007b], Chapter 13).

Fourth, the arguments developed in Sections 4.2 and 4.3 are silent about the mechanisms of genome expansion and increased structural complexity. These details have to be filled out before we have a reasonably complete account of the evolution of genome architecture. This reservation will be developed further in the last paragraph of this article.

Fifth, the arguments of Sections 4.2 and 4.3 depend critically on mathematical analyses of models from population genetics. That these analyses support a conclusion of obvious, wide-ranging evolutionary significance, namely, the ineffectiveness of natural selection to prevent the maladaptive expansion of eukaryotic genomes, underscores the centrality of mathematical reasoning in evolutionary analysis. This centrality was famously challenged by Mayr ([1963]) and defended by Haldane ([1964]), who pointed out that Mayr had not followed the mathematical arguments of Fisher, Haldane, and Wright (Sarkar [2007b]). However, though many adaptationists follow Mayr in favouring verbal argument over mathematical analysis, this is not the case with the adaptationists whose objections were analysed in Section 5. In this case, the very nature of the exchanges testifies to the significance of mathematical analysis in evolutionary biology.

This article will end with a puzzle. Lynch ([2007a], [2007b]) has maintained that non-adaptive evolution of genome architecture may be compatible with adaptive evolution of other traits and may indeed facilitate it, for instance, by the future co-option of redundant or non-functional genomic segments. As noted in Section 1, given that the physical mechanisms operating at the genomic level are not well understood, it is quite likely that there are a range of molecular mechanisms and processes acting at the genomic level that may have complex relations to possible adaptive dynamics at higher levels of organization. This appears to support Lynch's claim of potential co-option of genomic resources. However, one worry is worth noting; it raises a puzzle

---

[12] Arguably, the arguments of this article are better incorporated into a likelihood framework insofar as they are concerned with the probability of the data given a hypothesis. Note that there is much in common between the likelihood framework and Bayesian inference—the latter combines the use of likelihoods with priors (Sarkar [2011b]). Developing this theme is beyond the scope of this article.

that cannot be resolved at present. No matter whether there is such a potential for the co-option of genomic segments to allow adaptive evolution, the postulated low N_e will equally prevent natural selection from being effective with respect to these traits as it did for genomic architecture: the problem of small effective population sizes will not go away. Implicitly using this fact, Lynch ([2007c]) has also proposed non-adaptive models of gene regulatory networks that already go beyond the level of genome architecture. However, Wagner ([2008]) has suggested that there can be consistency between neutrality and selectionism based on the properties of networks. The issue remains unresolved at present.

In any case, unless selection is strong enough to counteract the effects of sampling fluctuations (that is, drift), adaptive evolution of these other features also becomes unlikely. Yet, there is no good reason to doubt that a significant number of phenotypic features at the organismic level (and probably at higher levels of organization) are results of selection and, in many cases, satisfy the stronger optimality condition (which leads to a stronger sense of adaptation than the one used in this article—recall Section 2). Therefore, at the very least, if the core argument (or any of its variants) is even approximately sound (in the sense that its premises, including P2, are at least approximately correct), such evolution is unlikely to have occurred through ubiquitous weak selection. Resolving this puzzle remains a task for the future.

Acknowledgements

Departments of Integrative Biology and Philosophy
University of Texas at Austin
Austin, TX 78712, USA
sarkar@austin.utexas.edu

References

Ai, B., Wang, Z. -S. and Ge, S. [2012]: 'Genome Size Is Not Correlated with Effective Population Size in the *Oryza* Species', *Evolution*, **66**, pp. 3302–10.

Barrett, R. D. H. and Hoekstra, H. E. [2011]: 'Molecular Spandrels: Adaptation at the Genetic Level', *Nature Reviews Genetics*, **12**, pp. 767–80.

Berget, S. M., Moore, C. and Sharp, P. [1977]: 'Spliced Segments at the 5′ Terminus of Adenovirus 2 Late RNA', *Proceedings of the National Academy of Sciences*, **74**, pp. 3171–5.

Blake, C. C. F. [1978]: 'Do Genes-in-Pieces Imply Proteins-in-Pieces?', *Nature*, **273**, p. 267.

Boeke, J. D., Garfinkel, D. J., Styles, C. A. and Fina, G. R. [1985]: '*Ty* Elements Transpose through an RNA Intermediate', *Cell*, **40**, pp. 491–500.

Britten, R. J. and Davidson, E. H. [1969]: 'Gene Regulation for Higher Cells: A Theory', *Science*, **165**, pp. 349–57.

Britten, R. J. and Davidson, E. H. [1971]: 'Repetitive and Nonrepetitive DNA Sequences and a Speculation on the Origin of Evolutionary Novelty', *Quarterly Review of Biology*, **46**, pp. 111–38.

Britten, R. J. and Kohne, D. E. [1968]: 'Repeated Sequences in DNA', *Science*, **161**, pp. 529–40.

Caron, F. [1990]: 'Eucaryotic Codes', *Experientia*, **46**, pp. 1106–17.

Charlesworth, B. [2002]: 'Effective Population Size', *Current Biology*, **12**, pp. R716–7.

Charlesworth, B. [2008]: 'Book Review: The Origin of Genomes—Not by Natural Selection?', *Current Biology*, **18**, pp. R140–1.

Charlesworth, B. [2009]: 'Effective Population Size and Patterns of Molecular Evolution and Variation', *Nature Reviews Genetics*, **10**, pp. 195–205.

Charlesworth, B. and Barton, N. [2004]: 'Genome Size: Does Bigger Mean Worse?', *Current Biology*, **14**, pp. R233–5.

Chow, L. T., Gelinus, R. E., Broker, T. R. and Roberts, R. J. [1977]: 'An Amazing Sequence Arrangement of the 5′ Ends of Adenovirus 2 Messenger RNA', *Cell*, **12**, pp. 1–98.

Cook-Deegan, R. [1994]: *The Gene Wars: Science, Politics, and the Human Genome*, New York: W. W. Norton.

Corradi, N., Pombert, J.-F., Farinelli, L., Didier, E. S. and Keeling, P. K. [2010]: 'The Complete Sequence of the Smallest Known Nuclear Genome from the Microsporidian *Encephalitozoon intestinalis*', *Nature Communications*, **1**, p. 77, doi:10.1038/ncomms1082.

Crick, F. H. C. [1979]: 'Split Genes and RNA Splicing', *Science*, **204**, pp. 264–71.

Darnell, J. E. [1978]: 'Implications of RNA: RNA Splicing in Evolution of Eukaryotic Cells', *Science*, **202**, pp. 1257–60.

Daubin, V. and Moran, N. A. [2004]: 'Comment on "The Origins of Genome Complexity"', *Science*, **306**, p. 978a.

de Vries, H. [1901]: *Die Mutationstheorie. Versuche und Beobachtungen über die Entstehung von Arten im Pflanzenreich*, Band 1, Leipzig: Veit.

de Vries, H. [1903]: *Die Mutationstheorie. Versuche und Beobachtungen über die Entstehung von Arten im Pflanzenreich*, Band 2, Leipzig: Veit.

Doolittle, W. F. [1978]: 'Genes-in-Pieces: Were They Ever Together?', *Nature*, **272**, pp. 581–2.

Doolittle, W. F. [1985]: 'Some Broader Evolutionary Issues which Emerge from Contemporary Molecular Biological Data', in P. Asquith and P. Kitcher (*eds*), *PSA 1984: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, East Lansing, MI: Philosophy of Science Association, pp. 129–44.

Dufresne, F. and Jeffery, N. [2011]: 'A Guided Tour of Large Genome Size in Animals: What We Know and Where We Are Heading', *Chromosome Research*, **19**, pp. 925–38.

Dunn, L. C. [1964]: 'Old and New in Genetics', *Bulleting of the New York Academy of Medicine*, **40**, pp. 325–33.

Eddy, S. R. [2012]: 'The C-value Paradox, Junk DNA, and ENCODE', *Current Biology*, **22**, pp. R898–9.

Eddy, S. R. [2013]: 'The ENCODE Project: Missteps Overshadowing a Success', *Current Biology*, **23**, pp. R259–61.

ENCODE Project Consortium [2012]: 'An Integrated Encyclopedia of DNA Elements in the Human Genome', *Nature*, **489**, pp. 57–74.

Fields, C., Adams, M. D., White, O. and Venter, J. C. [1994]: 'How Many Genes in the Human Genome?', *Nature Genetics*, **7**, pp. 345–6.

Fisher, R. A. [1930]: *The Genetical Theory of Natural Selection*, Oxford: Clarendon Press.

Fontdevila, A. [2011]: *The Dynamic Genome: A Darwinian Approach*, Oxford: Oxford University Press.

Gilbert, W. [1978]: 'Why Genes in Pieces?', *Nature*, **271**, p. 501.

Gilbert, W. [1990]: 'Current State of the HGI', *Oral Presentation, Meeting on the Historical and Social Study of the Human Genome Initiative*, Department of the History of Science, Harvard University.

Gilbert, W. [1991]: 'The Scientific Origins of the Human Genome Initiative', *oral presentation, workshop on the Philosophy and History of Molecular Biology: New Perspectives*, Boston Center for the Philosophy of Science, Boston University.

Gilbert, W. [1992]: 'A Vision of the Grail', in D. J. Kevles and L. Hood (*eds*), *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, Cambridge, MA: Harvard University Press, pp. 83–97.

Gillespie, J. H. [1991]: *The Causes of Molecular Evolution*, New York: Oxford University Press.

Gould, S. J. and Lewontin, R. C. [1979]: 'The Spandrels of San Marco and the Panglossian Paradigm', *Proceedings of the Royal Society of London B*, **205**, pp. 581–98.

Graur, D., Zheng, Y., Price, N., Azevedo, R. B., Zufall, R. A. and Elhaik, E. [2013]: 'On the Immortality of Television Sets: "Function" in the Human Genome According to the Evolution-Free Gospel of ENCODE', *Genome Biology and Evolution*, **5**, pp. 578–90.

Gregory, T. R. [2001]: 'Coincidence, Coevolution, or Causation? DNA Content, Cell Size, and the C-value Enigma', *Biological Reviews*, **76**, pp. 65–101.

Gregory, T. R. [2005]: 'Synergy between Sequence and Size in Large-Scale Genomics', *Nature Reviews Genetics*, **6**, pp. 699–708.

Gregory, T. R. and Witt, J. D. [2008]: 'Population Size and Genome Size in Fishes: A Closer Look', *Genome*, **51**, pp. 309–13.

Gualberto, J. M., Weil, J. H. and Grienenberger, J. M. [1990]: 'Editing of the Wheat Cox III Transcript: Evidence for Twelve C to U and one U to C Conversions and for Sequence Similarities around Editing Sites', *Nucleic Acids Research*, **18**, pp. 3771–6.

Guo, S. and Kemphues, K. J. [1995]: '*par*-1, a Gene Required for Establishing Polarity in *C. elegans* Embryos, Encodes a Putative Ser/Thr, Encodes a Putative Ser/Thr Kinase That Is Asymmetrically Distributed', *Cell*, **81**, pp. 611–20.

Hagedoorn, A. L. and Hagedoorn-Vorstheuvel la Brand, A. C. [1921]: *The Relative Value of the Processes Causing Evolution*, The Hague: Martinus Nijhoff.

Hahn, M. W. and Wray, G. A. [2002]: 'The G-value Paradox', *Evolution and Development*, **4**, pp. 73–5.

Haldane, J. B. S. [1924]: 'A Mathematical Theory of Natural and Artificial Selection: Part I', *Transactions of the Cambridge Philosophical Society*, **23**, pp. 19–41.

Haldane, J. B. S. [1932]: *The Causes of Evolution*, London: Harper and Brothers.

Haldane, J. B. S. [1957]: 'The Cost of Natural Selection', *Journal of Genetics*, **55**, pp. 511–24.

Haldane, J. B. S. [1964]: 'A Defense of Beanbag Genetics', *Perspectives in Biology and Medicine*, **7**, pp. 343–59.

Hall, S. S. [2010]: 'Revolution Postponed', *Scientific American*, **303**, pp. 60–7.

Harris, E. E. [2013]: 'Gene Evolution and Human Adaptation', *eLS*, doi: 10.1002/9780470015902.a0020773.pub2.

International Human Genome Sequencing Consortium (IHGSC) [2001]: 'Initial Sequencing and Analysis of the Human Genome', *Nature*, **409**, pp. 860–921.

Jacq, C., Miller, J. R. and Brownlee, G. G. [1977]: 'A Pseudogene Structure in 5S DNA of *Xenopus laevis*', *Cell*, **12**, pp. 109–20.

Jelinek, W. R., Toomey, T. P., Leinwand, L., Duncan, C. H., Biro, P. A., Choudary, P. B., Weissman, S. M., Rubin, C. M., Houck, C. M., Deininger, P. L. and Schmid, C. W. [1980]: 'Ubiquitous, Interspersed Repeated DNA Sequences in Mammalian Genes', *Proceedings of the National Academy of Sciences*, **77**, pp. 1398–402.

Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. and Antonarakis, S. E. [1988]: 'Hemophilia A Resulting from *de novo* Insertion of L1 Sequences Represents a Novel Mechanism for Mutation in Man', *Nature*, **332**, pp. 164–66.

Keeling, P. J. and Palmer, J. D. [2008]: 'Horizontal Gene Transfer in Eukaryotic Evolution', *Nature Reviews Genetics*, **9**, pp. 605–18.

Kimura, M. [1968]: 'Evolutionary Rate at the Molecular Level', *Nature*, **217**, pp. 624–6.

Kimura, M. [1983]: *The Neutral Theory of Molecular Evolution*, Cambridge: Cambridge University Press.

King, J. L. and Jukes, T. H. [1969]: 'Non-Darwinian Evolution: Most Evolutionary Change in Proteins May Be due to Neutral Mutations and Genetic Drift', *Science*, **164**, pp. 788–98.

Koonin, E. V. [2009]: 'Darwinian Evolution in the Light of Genomics', *Nucleic Acids Research*, **37**, pp. 1011–34.

Koonin, E. V. [2012]: *The Logic of Chance: The Nature and Origin of Biological Evolution*, Upper Saddle River, NJ: Pearson Education.

Koslowsky, D. J. [2004]: 'A Historical Perspective on RNA Editing: How the Peculiar and Bizarre Became Mainstream', in J. M. Gott (*ed.*), *RNA Interference, Editing, and Modification: Methods and Protocols*, Totowa, NJ: Humana Press, pp. 161–97.

Kurland, C. G., Canback, B. and Berg, O. G. [2003]: 'Horizontal Gene Transfer: A Critical Review', *Proceedings of the National Academy of Sciences*, **100**, pp. 9658–62.

Lewens, T. [2009]: 'Seven Types of Adaptationism', *Biology and Philosophy*, **24**, pp. 161–82.

Lewin, B. [1980]: 'Structure and Evolution of Interrupted Genes', *Annals of the New York Academy of Sciences*, **354**, pp. 453–65.

Lewontin, R. C. [1974]: *The Genetic Basis of Evolutionary Change*, New York: Columbia University Press.

Lewontin, R. C. [1978]: 'Adaptation', *Scientific American*, **239**, pp. 212–30.

Lewontin, R. C. [2008]: 'Heredity and Heritability', in S. Sarkar and A. Plutynksi (*eds*), *The Blackwell Companion to the Philosophy of Biology*, Oxford: Blackwell, pp. 40–57.

Lynch, M. [2006]: 'Streamlining and Simplification of Microbial Genome Architecture', *Annual Review of Microbiology*, **60**, pp. 327–49.

Lynch, M. [2007a]: 'The Frailty of Adaptive Hypotheses for the Origins of Organismic Complexity', *Proceedings of the National Academy of Sciences*, **104**, pp. 8597–604.

Lynch, M. [2007b]: *The Origins of Genome Architecture*, Sunderland, MA: Sinauer.

Lynch, M. [2007c]: 'The Evolution of Genetic Networks by Non-adaptive Processes', *Nature Reviews Genetics*, **8**, pp. 803–13.

Lynch, M. [2011]: 'Statistical Inference on the Mechanisms of Genome Evolution', *PLoS Genetics*, **7**, p. e1001389, doi: 10.1371/journal.pgen.1001389.

Lynch, M. and Conery, J. S. [2003]: 'The Origins of Genome Complexity', *Science*, **302**, pp. 1401–4.

Maeso, I., Roy, S. W. and Irimia, M. [2012]: 'Widespread Recurrent Evolution of Genomic Features', *Genome Biology and Evolution*, **4**, pp. 486–500.

Mayr, E. [1963]: *Animal Species and Evolution*, Cambridge, MA: Harvard University Press.

Mayr, E. [1983]: 'How to Carry out the Adaptationist Program', *American Naturalist*, **121**, pp. 324–34.

McClintock, B. [1950]: 'The Origin and Behavior of Mutable Loci in Maize', *Proceedings of the National Academy of Sciences*, **36**, pp. 344–55.

McClintock, B. [1951]: 'Chromosome Organization and Genic Expression', *Cold Spring Harbor Symposia on Quantitative Biology*, **16**, pp. 13–47.

McElheny, V. K. [2010]: *Drawing the Map of Life: Inside the Human Genome Project*, New York: Basic Books.

Moore, G. P. [1984]: 'The C-value Paradox', *BioScience*, **34**, pp. 425–9.

Moran, P. A. P. [1964]: 'On the Nonexistence of Adaptive Topographies', *Annals of Human Genetics*, **27**, pp. 283–93.

Nielsen, R. [2009]: 'Adaptationism: 30 Years after Gould and Lewontin', *Evolution*, **63**, pp. 2487–90.

Niu, D. K. and Jiang, L. [2013]: 'Can ENCODE Tell Us How Much Junk We Carry in Our Genome?', *Biochemical and Biophysical Research Communications*, **430**, pp. 1340–3.

Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, B., Lindberg, F. P. and Olsson, O. [1983]: 'Overlapping Genes', *Annual Review of Genetics*, **17**, pp. 499–525.

Ohno, S. [1970]: *Evolution by Gene Duplication*, Berlin: Springer.

Ohta, T. [1973]: 'Slightly Deleterious Mutant Substitutions in Evolution', *Nature*, **246**, pp. 96–8.

Ohta, T. [1996]: 'Development of Neutral and Nearly Neutral Theories', *Theoretical Population Biology*, **49**, pp. 128–42.

Ohta, T. [2013]: 'Molecular Evolution: Nearly Neutral Theory', *eLS*; doi:10.1038/npg.els.0001801.

Orzack, S. H. and Forber, P. [2010]: 'Adaptationism', in E. N. Zalta (*ed.*), *Stanford Encyclopedia of Philosophy*, <plato.stanford.edu/archives/fall2010/entries/adaptationism/>.

Pellicer, J., Fay, M. F. and Leitch, I. J. [2010]: 'The Largest Eukaryotic Genome of Them All?', *Botanical Journal of the Linnaean Society*, **164**, pp. 1–15.

Pertea, M. and Salzberg, S. L. [2010]: 'Between a Chicken and a Grape: Estimating the Number of Human Genes', *Genome Biology*, **11**, p. 206.

Pigliucci, M. [2007]: 'Postgenomic Musings', *Science*, **317**, pp. 1172–3.

Rebollo, R., Romanish, M. T. and Mager, D. L. [2012]: 'Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes', *Annual Review of Genetics*, **46**, pp. 21–42.

Rocheleau, C. E., Downs, W. D., Lin, R., Wittmann, C., Bei, Y., Cha, Y.-H., Ali, M., Priess, J. R. and Mello, C. C. [1997]: 'Wnt Signaling and an APC-Related Gene Specify Endoderm in Early *C. elegans* Embryos', *Cell*, **90**, pp. 707–16.

Ruse, M. [1988]: *Philosophy of Biology Today*, Binghamton, NY: State University of New York Press.

Salzberg, S. L., White, O., Peterson, J. and Eisen, J. A. [2001]: 'Microbial Genes in Human Genome: Lateral Transfer or Gene Loss?', *Science*, **292**, pp. 1903–6.

Sarkar, S. [2014]: 'Formal Darwinism: Some Questions', *Biology and Philosophy*, **29**, pp. 249–57.

Sarkar, S. [1998]: *Genetics and Reductionism*, New York: Cambridge University Press.

Sarkar, S. [2001]: 'Reductionism in Genetics and the Human Genome Project', in R. Singh, C. Krimbas, D. B. Paul and J. Beatty (*eds*), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, Volume 2, New York: Cambridge University Press, pp. 23–252.

Sarkar, S. [2004]: 'Evolutionary Theory in the 1920s: The Nature of the Synthesis', *Philosophy of Science*, **71**, pp. 1215–26.

Sarkar, S. [2006]: 'From Genes as Determinants to DNA as Resource: Historical Notes on Development and Genetics', in E. Neumann-Held and C. Rehmann-Sutter (*eds*), *Genes in Development: Re-Reading the Molecular Paradigm*, Durham: Duke University Press, pp. 77–95.

Sarkar, S. [2007a]: *Doubting Darwin? Creationist Designs on Evolution*, Oxford: Blackwell.

Sarkar, S. [2007b]: 'Haldane and the Emergence of Modern Evolutionary Theory', in M. Matthen and C. Stephens (*eds*), *Philosophy of Biology*, New York: Elsevier, pp. 49–86.

Sarkar, S. and Tauber, A. [1991]: 'Fallacious Claims for HGP', *Nature*, **353**, p. 691.

Sarkar, S. [2011a]: 'Drift and the Causes of Evolution', in P. K. Illari, F. Russo and J. Williamson (*eds*), *Causality in the Sciences*, Oxford: Oxford University Press, pp. 445–69.

Sarkar, S. [2011b]: 'Sober on Intelligent Design', *Philosophy and Phenomenological Research*, **83**, pp. 683–91.

Schuster, W., Hiesel, R., Wissinger, B. and Brennicke, A. [1990]: 'RNA editing in the Cytochrome b Locus of the Higher Plant *Oenothera berteriana* Includes a U-to-C Transition', *Molecular and Cellular Biology*, **10**, pp. 2428–31.

Shapiro, J. A. [1995]: 'The Discovery and Significance of Mobile Genetic Elements', in D. J. Sherratt (*ed.*), *Mobile Genetic Elements*, Oxford: IRL Press, pp. 1–17.

Sharp, P. A. [2005]: 'The Discovery of Split Genes and RNA Splicing', *Trends in Biochemical Sciences*, **30**, pp. 279–81.

Stoltzfus, A. [2012]: 'Constructive Neutral Evolution: Exploring Evolutionary Theory's Curious Disconnect', *Biology Direct*, **7**, p. 35.

Takahata, N. [2001]: 'Molecular Evolution: Neutral Theory', *eLS*, doi:10.1038/npg.els.0001800.

Tauber, A. I. and Sarkar, S. [1992]: 'The Human Genome Project: Has Blind Reductionism Gone Too Far?', *Perspectives in Biology and Medicine*, **35**, pp. 220–35.

Tauber, A. I. and Sarkar, S. [1993]: 'The Ideology of the Human Genome Project', *Journal of the Royal Society of Medicine*, **86**, pp. 537–540.

Thomas, C. A. [1971]: 'The Genetic Organization of Chromosomes', *Annual Reviews of Genetics*, **5**, pp. 237–56.

Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. and Gilbert, W. [1978]: 'Sequence of a Mouse Germ-line Gene for a Variable Region of an Immunoglobulin Light Chain', *Proceedings of the National Academy of Sciences*, **75**, pp. 1485–9.

Treisman, R., Orkin, S. H. and Maniatis, T. [1983]: 'Specific Transcription and RNA Splicing Defects in Five Cloned β-thalassemia Genes', *Nature*, **302**, pp. 591–6.

Wagner, A. [2008]: 'Neutralism and Selectionism: A Network-based Reconciliation', *Nature Reviews Genetics*, **9**, pp. 965–74.

Watson, J. D., Tooze, J. and Kurtz, D. T. [1983]: *Recombinant DNA: A Short Course*, New York: W. H. Freeman.

Whitney, K. D., Baack, E. J., Hamrick, J. L., Godt, M. J. W., Barringer, B. C., Bennett, M. D., Eckert, C. G., Goodwillie, C., Leitch, I. J. and Ross-Ibarra, J. [2010]: 'A Role for Nonadaptive Processes in Plant Genome Size Evolution?', *Evolution*, **64**, pp. 2097–109.

Whitney, K. D., Boussau, B., Baack, E. J. and Garland, T. [2011]: 'Drift and Genome Complexity Revisited', *PLoS Genetics*, **7**, e1002092, doi:0.1371/journal.pgen.1002092.

Whitney, K. D. and Garland, T. [2010]: 'Did Genetic Drift Drive Increases in Genome Complexity?', *PLoS Genetics*, **6**, e1001080, doi:10.1371/journal.pgen.1001080.

Wright, S. [1931]: 'Evolution in Mendelian Populations', *Genetics*, **16**, pp. 97–159.

Yi, S. V. [2006]: 'Non-adaptive Evolution of Genome Complexity', *BioEssays*, **28**, pp. 979–82.

Yi, S. V. and Streelman, J. T. [2005]: 'Genome Size is Negatively Correlated with Effective Population Size in Ray-finned Fish', *Trends in Genetics*, **21**, pp. 643–6.