



# Non-Darwinian Molecular Biology

Alexander F. Palazzo\* and Nevraj S. Kejiou

Department of Biochemistry, University of Toronto, Toronto, ON, Canada

With the discovery of the double helical structure of DNA, a shift occurred in how biologists investigated questions surrounding cellular processes, such as protein synthesis. Instead of viewing biological activity through the lens of chemical reactions, this new field used biological information to gain a new profound view of how biological systems work. Molecular biologists asked new types of questions that would have been inconceivable to the older generation of researchers, such as how cellular machineries convert inherited biological information into functional molecules like proteins. This new focus on biological information also gave molecular biologists a way to link their findings to concepts developed by genetics and the modern synthesis. However, by the late 1960s this all changed. Elevated rates of mutation, unsustainable genetic loads, and high levels of variation in populations, challenged Darwinian evolution, a central tenant of the modern synthesis, where adaptation was the main driver of evolutionary change. Building on these findings, Motoo Kimura advanced the neutral theory of molecular evolution, which advocates that selection in multicellular eukaryotes is weak and that most genomic changes are neutral and due to random drift. This was further elaborated by Jack King and Thomas Jukes, in their paper “Non-Darwinian Evolution”, where they pointed out that the observed changes seen in proteins and the types of polymorphisms observed in populations only become understandable when we take into account biochemistry and Kimura’s new theory. Fifty years later, most molecular biologists remain unaware of these fundamental advances. Their adaptationist viewpoint fails to explain data collected from new powerful technologies which can detect exceedingly rare biochemical events. For example, high throughput sequencing routinely detects RNA transcripts being produced from almost the entire genome yet are present less than one copy per thousand cells and appear to lack any function. Molecular biologists must now reincorporate ideas from classical biochemistry and absorb modern concepts from molecular evolution, to craft a new lens through which they can evaluate the functionality of transcriptional units, and make sense of our messy, intricate, and complicated genome.

**Keywords:** neutral theory, evolution, junk DNA, junk RNA, robustness

## INTRODUCTION

We live in the post genomic era. Our ability to analyze nucleic acid sequences has increased many orders of magnitude in the past 2 decades, largely due to the advent of high throughput sequencing technology. This has allowed us to undertake large-scale molecular biology experiments unimaginable a few decades ago. For example, we are now able to isolate and sequence long non-coding RNAs that are present at a few copies per thousand cells. Our newfound ability to analyze

## OPEN ACCESS

### Edited by:

Maritza Jaramillo,  
Université du Québec, Canada

### Reviewed by:

Claire Benard,  
Université du Québec à Montréal,  
Canada

Jian-Rong Yang,  
Sun Yat-sen University, China

### \*Correspondence:

Alexander F. Palazzo  
alex.palazzo@utoronto.ca

### Specialty section:

This article was submitted to  
RNA,  
a section of the journal  
Frontiers in Genetics

**Received:** 07 December 2021

**Accepted:** 24 January 2022

**Published:** 16 February 2022

### Citation:

Palazzo AF and Kejiou NS (2022) Non-Darwinian Molecular Biology.  
*Front. Genet.* 13:831068.  
doi: 10.3389/fgene.2022.831068

DNA and RNA sequences has been exploited by large consortia, such as the ENCODE project, that have allocated significant resources to catalogue every biochemical event associated with the entire genome (ENCODE Project Consortium et al., 2007; ENCODE Project Consortium et al., 2012). Despite these developments, the molecular biology community seems adrift. This is exemplified by many individuals within the community who discover that a segment of non-coding DNA has some function, then falsely extrapolate their finding to the genome at large and erroneously conclude that junk DNA is in fact a vast repository of dark matter (“matter”, here being understood as a collection of poorly defined functional entities that have yet to be elucidated). In terms of advancing large scale theories about genome organization, or the contribution of lncRNAs to cellular homeostasis and organismal development, new insights are few and far between. To paraphrase Sydney Brenner, many in the field are drowning in a sea of data and starving for knowledge (Brenner, 2003).

## HOW DID WE GET TO THIS POINT?

From the 1850s to the 1950s, biochemistry had made spectacular advances in understanding cellular life. Despite this, the biochemical and molecular processes that explained heredity remained mysterious. It wasn't until 1953, with the elucidation of the structure of DNA, that a coherent molecular basis for heredity could be formulated. DNA contained sequence information where the series of nucleic acid bases along one polymer chain specifies the sequence of nucleic acid bases along a second polymer chain (Watson and Crick, 1953). Biological information was digital, each quantum being a nucleotide (C, A, T or G). This sequence in nucleic acids could then specify the order of other polymers, such as proteins (Crick, 1958).

Before the double helix, the biochemistry of life was investigated as a series of reactions. Afterwards, life was seen as being imbued with molecules that could carry information. With this shift, the budding field of molecular biology described these new molecules and their properties using terminology (e.g., sequence, transcription, translation, codon, frame) that could have been lifted from computer science and information theory, two disciplines that were also developing at the same time.

The degree to which the emphasis had shifted from investigating cellular activity through the lens of enzymology to that of information is nicely illustrated by a quote from Sydney Brenner (about the problem of co-linearity between DNA and protein sequence) (Wolpert, 1994):

At that time, the biochemists of the world were preoccupied by where you get the energy to make the proteins, and we had to spend weeks, months, saying “don't worry about energy, energy will look out for itself. The important thing is how to get everything in the correct order. How do you get everything specified in this order, that is, the genetic code.” I think that this is such an important and fundamental divergence from anything else in

Biology, that it is a total discontinuity, at least this is the way I've seen it. And it has of course constrained quite a lot of later developments in biology. And of course it crystalizes the types of problems you have to solve in a clear-cut way, because now they do not remain, sort of, vague problems that you can ask sort of rhetorical problems about. But you can actually sit down and ask, “If I had a gene and I could do the fine structure, and if I had a protein that I could sequence, then I could show whether the gene was co-linear.”

In the following decades molecular biologists fleshed out the mechanistic details of how information was duplicated and how it specified the synthesis of functional molecules such as proteins.

## THE MOLECULAR EVOLUTION REVOLUTION

The molecular basis of heredity not only changed our view of cellular and organismal activity, but also altered our understanding of evolution. In the first half of the 20th century, the theoretical basis of how mutations spread within a population were developed by population geneticists such as J.B.S Haldane, Sewall Wright and Ronald Fisher. By the 1950s, it was widely accepted that evolution occurred through natural selection at the genetic level. Given that the molecular basis for inheritance could now be linked to the sequence of bases in DNA, and this could be further extrapolated to the sequence of amino acids in proteins, it was unsurprising that some molecular biologists started to tackle evolutionary questions in the mid-1960s.

In 1962, Emile Zuckerkandl, under the supervision of Linus Pauling, found that the number of amino acid differences between the hemoglobin proteins isolated from any two species, correlated with how long ago they had diverged from a common ancestor (as per paleontological estimates)—a concept later dubbed the “molecular clock” (Zuckerkandl and Pauling, 1962; Margoliash, 1963; Zuckerkandl and Pauling, 1965). This established that species relatedness could be determined at the molecular level, and that protein paralogues evolve at a constant rate in different lineages. Around the same time, Jack Hubby and Richard Lewontin, used gel electrophoresis to demonstrate that different individuals within a species had a surprisingly high amount of variability in any given protein (Hubby and Lewontin, 1966; Lewontin and Hubby, 1966). The variability was far greater than what most population genetics models had predicted. Others had noted that many of these substitutions led to mostly neutral changes in proteins (Freese, 1962; Margoliash, 1963). All of these observations led to a crisis in evolutionary biology.

This crisis was further deepened by Motoo Kimura's 1968 publication “Evolutionary Rate at the Molecular Level” (Kimura, 1968). The amount of protein variation between and within species, estimated by the new molecular biology techniques, inferred such a high substitution rate that if they all consisted of alleles that were under selection, the cost of replacing these alleles (in “genetic deaths”) would be intolerable. Instead, Kimura

proposed that the majority of these mutations must be neutral. Furthermore, Kimura demonstrated that slightly deleterious or beneficial mutations behave like neutral mutations, provided that the absolute value of their selection coefficient was smaller than the inverse of the effective population size.

$$|s| < 1/N_e$$

In this equation, the selection coefficient ( $s$ ) is the average decrease or surplus in offspring in the mutant compared to a wildtype individual, and the effective population size ( $N_e$ ) is the number of breeding individuals that randomly mate in a species. In practice,  $N_e$  is much lower than the actual population size as it is dependent on a number of different parameters (Lynch, 2007). In humans, the effective population size has been estimated to be about 10,000 which is a bit lower than the typical number for most other animals (Charlesworth, 2009). If a mutation falls within the parameters of the above equation, it does not clear what Michael Lynch called “the drift barrier” (Lynch et al., 2011; Sung et al., 2012), and its fixation will be dictated by neutral evolution.

A year after Kimura’s paper, evolutionary biologist Jack King and Thomas Jukes independently established a similar conclusion (King and Jukes, 1969); however, many of their arguments relied on the biochemistry of proteins and the types of mutations seen in nature. Amino acid substitutions between species and allele variants within species occurred in regions that were less important (for example away from the active site of enzymes). These changes were typically conservative (swapping amino acids of similar biochemical nature), and rarely altered the activity of a protein. Moreover, the substitution rate was lower in genes that play fundamental roles in cell physiology. In the next few decades, important advances in the neutral theory were further developed by Kimura and his colleague Tomoko Ohta (Kimura and Ohta, 1974; Ohta, 2003). All of this data and analysis argued that most evolution was due to random neutral genetic drift at the molecular level.

At the time, natural selection was believed to be the main driver of evolution and the new neutral theory caused considerable controversy within the field—creating what is commonly referred to as the neutralist-selectionist debate (Nei, 2005). Eventually, DNA sequencing data validated predictions made by neutral theory, namely that the rate of fixation is higher in parts of the genome with low functional constraint; such as pseudogenes and introns (Kimura, 1991). It became increasingly difficult to justify why natural selection would maintain high rates of mutations in these genomic regions. Today, when researchers claim that a portion of a protein is important because it is conserved, they are in fact using arguments (whether they know it or not) that originated in the neutral theory of molecular evolution (Hughes, 2007).

Of course, mutations are not simply nucleotide or amino acid substitutions. INDELS, small and large, would be subject to the forces of neutral and near-neutral evolution. Geneticist Susumu Ohno, who investigated gene duplication and its evolutionary consequences, proposed that newly duplicated genes either gained new functions (neofunctionalization) or lost function entirely, becoming pseudogenes (Ohno, 1972a). Ohno also

recognized that the genome contained other non-functional entities, such as short repeats and that as much as 90% of the genome was likely non-functional (Ohno, 1972b). He called this “junk DNA”, a term that had been previously used colloquially in molecular biology circles (Graur, 2013).

To add fuel to the fire, in the 1970s the contribution of neutral evolution to phenotypic change was explored (Lande, 1976; Nei, 2013), although some of these ideas had been previously advanced by Sewall Wright (Wright, 1932). The idea that natural selection may not be the only mechanism for phenotypic evolution was further popularized by Lewontin and Stephen J. Gould. They warned those who studied phenotypic evolution that many of their explanations were “just-so stories” and that other factors may explain how any given phenotypic trait evolves (Gould and Lewontin, 1979). This includes not only neutral evolution, but also developmental constraints, pleiotropy and historical contingencies. Even today, the degree to which neutral evolution impacts phenotypic changes is poorly understood (Zhang, 2018).

The neutralist-selectionist debate is still ongoing. Some population geneticists in the selectionist camp have advocated that most mutations can reach fixation because they are linked to nearby alleles that are under positive selection, so called evolution by draft (Gillespie, 2000; Kern and Hahn, 2018). However, these arguments do not invalidate the idea that most mutations *by themselves* are nearly neutral, the main thesis of the neutralist camp. Indeed, as pointed out by several commentators, rampant draft, if anything, lowers the effective population size and would thus further erode natural selection’s ability to weed out slightly deleterious mutations (Jensen et al., 2019). Other factors, such as fluctuations in selection coefficients between generations can further diminish the power of natural selection to differentiate between slightly deleterious or beneficial from neutral mutations (Nei, 2013). Although many details about molecular evolution are still debated, the main tenants of the neutral theory have been confirmed by numerous studies. Predictions made by the neutral theory match closely what is seen in countless analyses of single nucleotide polymorphism (SNP) distributions and comparative genomic surveys (Lynch, 2007; Koonin, 2011; Nei, 2013; Graur et al., 2016).

## MOLECULAR BIOLOGY: DROWNING IN A SEA OF DATA AND STARVING FOR KNOWLEDGE

Although neutral theory has made important intellectual advances, and was fueled by advances in molecular biology, many life scientists are either unaware of these developments or have failed to assimilate these new concepts. As the spiritual descendants of Watson and Crick, most modern molecular biologists assume that DNA *is* information. This is implicit when they use the shorthand of “sequence” for a short stretch of DNA. And for the decades following the elucidation of the double-helical structure, there was little reason to deny this view given the furious pace at which researchers were discovering molecular machines that would copy, read, and translate the

code-script. Despite this, there were signs that the human genome may contain vast tracts of non-functional DNA that could harbor neutral mutations. First, several researchers, using concepts from population genetics, estimated that the human genome contained about 30,000 genes with an upper limit of 40,000 (Muller, 1966; King and Jukes, 1969; Ohno, 1972a; Hatje et al., 2019). Since the average protein size was known (about 50 kDa) this meant that most of the genome had to be non-coding. At the same time, it was observed that eukaryotic genomes contained a large fraction of repeated sequence (Britten and Kohne, 1968), and that much of this was composed of transposable elements that were unlikely serving any role for the organism but instead were the product of selfish DNA (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Then with the discovery of splicing, it was realized that much of the genome was transcribed into introns that were discarded soon after they were synthesized (Berget et al., 1977; Chow et al., 1977; Gilbert, 1978; Berk, 2016). Finally, during the completion of the human genome project, it was revealed that nearly 98% of the genome was non-coding. While many researchers were surprised (Hatje et al., 2019), this observation was entirely consistent with earlier findings and gene estimates.

One of the problems with accepting that the human genome consists mostly of junk DNA, is that additional functional DNA elements, beyond protein-coding genes, were widely known since the 1960s (although many researchers continue to state, erroneously, that *all* non-coding DNA was once thought to be junk). Leaving aside the fact that nearly a quarter of the genome consists of intronic sequence, it has been long known that genomes contain promoters (Ippen et al., 1968), enhancers (Banerji et al., 1983; Gillies et al., 1983; Mercola et al., 1983), silencers (Abraham et al., 1983), origins of replication (Beach et al., 1980), telomeres (Blackburn and Gall, 1978), centromeres (Clarke and Carbon, 1980), and a wide array of genes transcribed into various non-coding RNAs, including rRNA (Roberts, 1958; Morrow et al., 1974), tRNAs (Holley et al., 1965), snRNAs (Yang et al., 1981), and other non-coding RNAs (Stark et al., 1978; Walter and Blobel, 1982; Brannan et al., 1990; Brockdorff et al., 1992). Although many of the founders of molecular biology were quite open to the idea of junk DNA, many of the molecular biologists that came afterwards had hyper-adaptationist tendencies. With the discovery of every new functional non-coding RNA, these researchers would extrapolate, erroneously, this finding to all cryptic transcription and conclude that it is all functional. With the discovery that a given transposable element was co-opted for a functional purpose, these molecular biologists would conclude that all transposable elements have some hidden function. The rejection of junk DNA, in the mind of these researchers, led to the idea that the genome consisted of mostly “dark matter”.

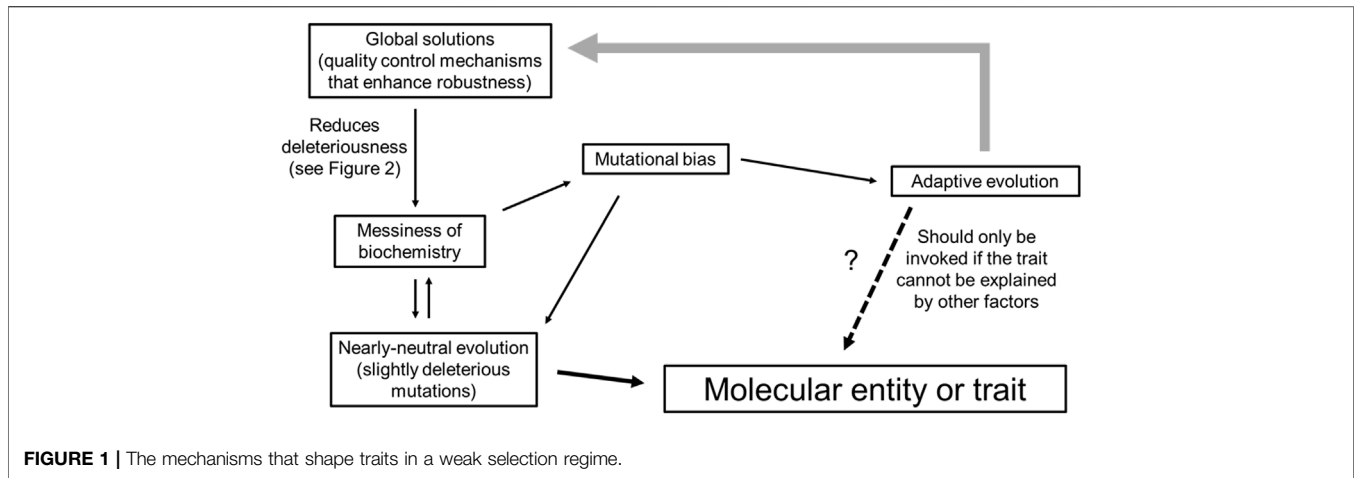
The pinnacle of this line of reasoning came with the Encyclopedia of DNA Elements, or ENCODE project, whose goal was to map all of the biochemical activities in the genome (ENCODE Project Consortium et al., 2007; ENCODE Project Consortium et al., 2012). Although the initial ENCODE publications, which documented all the biochemically active sites in 1% of the human genome, interpreted the data very conservatively, the second round of publications, which

examined the entire genome, famously did not. The most publicized “conclusion” from this work was that at least 80% of the genome had some function (ENCODE Project Consortium et al., 2012), again implying that junk DNA did not exist (Pennisi, 2012). Embedded in the collection of ENCODE papers was a compilation of all the messy reactions that we now have come to accept as commonplace in eukaryotic genomes. Focusing on the RNA results, it was seen that most of the genome was transcribed, albeit at some low level, that transcripts often began and ended at a wide variety of locations with many of these originating from intergenic regions, and that much of the transcriptional patterns were cell-type specific (Djebali et al., 2012). Similar findings were seen for other biochemical events that could be spatially mapped back to the genome (Spitz and Furlong, 2012; Thurman et al., 2012). This includes promoter-like regions, which appeared to outnumber protein-coding genes by an order of magnitude, and transcription factor binding sites, which were present throughout the genome, irrespective of whether they were near any genes. In response to these unsupported conclusions, other researchers pointed out that the new ENCODE results were entirely consistent with a junky, messy genome that is predicted by our modern understanding of molecular evolution (Eddy, 2012; Doolittle, 2013; Eddy, 2013; Graur et al., 2013; Niu and Jiang, 2013; Doolittle et al., 2014; Palazzo and Gregory, 2014; Ponting, 2017).

The problem is that when many molecular biologists are confronted with the idea that much of the genome, especially junk DNA, may be fixed by drift, they counter that natural selection would not tolerate waste or useless junk. They then use a few counterexamples, for example the presence of a functional non-coding RNA, as evidence against junk DNA. These conclusions are, of course, an over-extrapolation. Part of the problem is that by equating DNA to information, molecular biologists implicitly assume that all DNA is functional. Making the situation worse, many molecular biologists incorrectly equate natural selection with evolution. Biological function, however, implies that the trait, or element in question is currently under selective constraint (Doolittle et al., 2014; Linquist et al., 2020), and as described above, less than 10% of the human genome is under selection (Lindblad-Toh et al., 2011; Ward and Kellis, 2012; Palazzo and Gregory, 2014; Rands et al., 2014; Graur, 2017).

To move forward, the molecular biology community must not only gain a deeper appreciation of how evolution works on the molecular level, but also reincorporate concepts from biochemistry. Indeed, King and Jukes, two of the instigators of the molecular evolution revolution recognized this (King and Jukes, 1969). Their critique of pan-adaptationist thinking was partially made on biochemical grounds. They noted how observed changes in proteins, either between paralogs in different organisms or due to polymorphisms within an organism, were mostly due to conservative substitutions between chemically similar amino acids in parts of the protein which could tolerate such changes. For these reasons they argued that much of evolution was non-Darwinian.

In the next few sections, we will cover some of the modern ideas and concepts from both biochemistry and molecular



evolution, and try to build a more modern conceptual framework for molecular biology that is non-Darwinian in outlook. First, we discuss how cellular biochemistry is imperfect, by virtue of the chemical properties of macromolecules, and how genetic drift hinders their improvement and further exacerbates their messiness, especially in eukaryotes. Next, we consider how eukaryotic systems evolve robust quality control systems (“global solutions”) which have a two-fold effect of ameliorating numerous entities that experience slightly deleterious mutations (“local problems”), while also permitting the proliferation of additional cellular sloppiness. Then, we discuss mutational bias which drastically affects the path of genome evolution. Only by incorporating concepts of biochemical sloppiness, global solutions and mutational bias into neutral evolution, we can propose a null hypothesis for adaptive evolutionary claims (**Figure 1**). Lastly, we discuss how neutral evolutionary processes can increase the complexity of an organism which may in turn permit the evolution of novel adaptive traits.

## CELLS ARE SLOPPY

The bench scientist has an intuitive notion that biochemistry is imperfect. Many restriction enzymes that supposedly cut only one sequence also have weak activity on others (Wei et al., 2008). RNA binding domains that supposedly recognize one unique RNA motif, will bind almost any other RNA molecule when tested in isolation (Jankowsky and Harris, 2015), with apparent affinity for one motif only readily observed when they are presented with a mixture of RNAs (Ray et al., 2009). Of course, these results are due to the varying affinities a given RNA binding domain has for different RNA sequences. And these imperfections extend far beyond the test tube and the cellular environment. Clinicians and drug developers actively rely on the sloppiness of enzymes to interact with compounds never encountered in the natural world. Small molecules that are never seen in nature are nevertheless acted on by a cadre of cellular enzymes, such as in cytochrome p450 family members

(Atkins, 2020). Even in the absence of drugs, enzymes still engage in a variety of adventitious interactions with non-canonical protein partners or endogenous metabolites (Khersonsky and Tawfik, 2010; Tawfik, 2020). This is in part a consequence of the nearly-neutral theory of molecular evolution, that the underlying mutation which allows enzyme promiscuity is not sufficiently deleterious to be selected against. As a result, enzymes accumulate slightly deleterious mutations and as a result protein activities and interactomes become even more messy and error-prone (Levy et al., 2009). This is also due to basic principles of chemistry, that non-canonical reactions will occur, and the degree that these happen, as opposed to the most favorable reaction, is dictated by the Boltzmann factor between the two competing states:

$$\frac{P_1}{P_2} = e^{(E_1 - E_2)/kT}$$

In this equation, the ratio of the probabilities of two competing states ( $P_1$  and  $P_2$ ) are dictated by the energies associated with each state ( $E_1$  and  $E_2$ ), the Boltzmann constant ( $k$ ) and the temperature ( $T$ ). The greater the energy differential is, the more skewed the ratio will be. However, the probability of the unfavored state can never be zero as this would require an infinite energy differential. Although the Boltzmann factor is used to describe the population of two states at equilibrium, it can also be used to describe competition between two reactions, with  $E_1$  and  $E_2$  representing the activation energy of the competing reactions, although there are some differences (Sawato et al., 2019). Using Boltzmann distributions, one can use energetic differentials between tautomeric states of nucleic acid bases and their propensity to base pair, to predict some aspects of nucleotide misincorporation during DNA replication (Kimsey et al., 2015; Kimsey et al., 2018).

Focusing in on gene expression, messiness can be seen at every step. Nucleotide misincorporation during transcription and amino acid misincorporation during translation, which can be as high as  $10^{-3}$  per codon (Kramer and Farabaugh, 2007), will produce proteins that differ from their canonical sequence that further contribute to biochemical sloppiness (Drummond and Wilke, 2009). Indeed, with error rates this high, a protein such as

RanBP2/Nup358, which contains 3,224 amino acids, will have on average one amino acid misincorporation error for every molecule in a human cell. This messiness is not only true for simple molecular interactions and reactions, but for more complicated processes. Many random pieces of DNA can activate transcription (Reinke et al., 2008; Gerber et al., 2013; White et al., 2013; Gosselin et al., 2016) producing spurious transcripts from intergenic regions (Struhl, 2007; Palazzo and Lee, 2015; Garland and Jensen, 2020; Palazzo and Koonin, 2020). Similarly, RNA transcripts are often mis-processed (Saudemont et al., 2017; Xu and Zhang, 2021). And even when they are not, mRNAs are frequently translated using the wrong start site and fail to terminate translation at the canonical stop codon (Kosinski and Masel, 2020; Xu and Zhang, 2020).

Of course, the ratio between canonical and non-canonical biological reactions could be increased by the expenditure of additional energy, thus increasing the difference between  $E_1$  and  $E_2$ . For example, misincorporation of nucleotides is corrected by DNA repair enzymes and proof-reading machinery, both of which expend energy. However, if these non-canonical states are rare and only mildly deleterious, then there may not be enough benefit for the organism to invest in cellular machinery and energy which act to further drive the equilibrium toward the canonical state. Even if there is some benefit to pushing the equilibrium away from these deleterious states, there is diminishing returns as more and more energy is expended. However, probably the biggest obstacle, is the selective maintenance of genes that encode the additional machinery. Examples of this type of machinery are DNA repair enzymes and domains in polymerases responsible for proof-reading activity. The benefit of these genes, in terms of selection ( $s$ ), must clear the drift barrier in order for natural selection to maintain them. It is for this very reason that the DNA mutation rate is strongly influenced by the effective population size (Lynch et al., 2016).

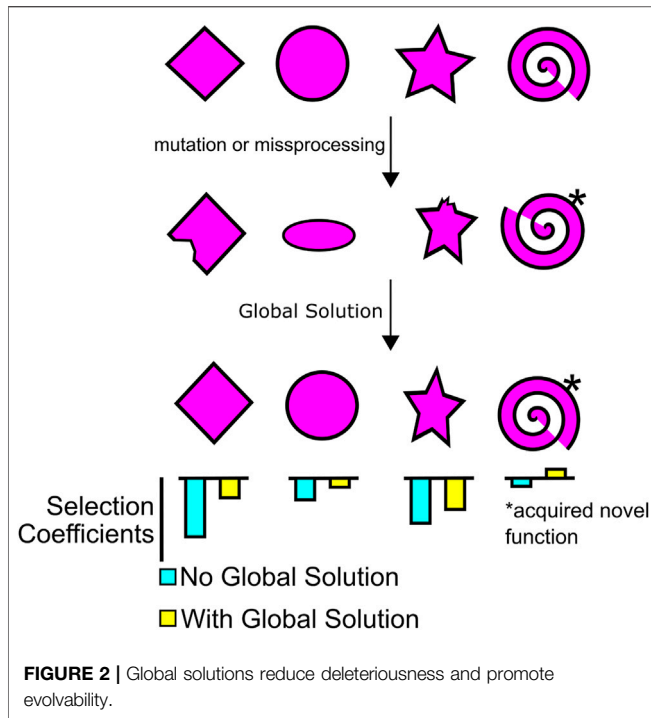
As long as the deleteriousness of these suboptimal activities remains below the critical threshold required for purging selection, they will not be eliminated and instead be subjected to evolution by drift (Lynch, 2007; Sung et al., 2012). With this in mind, it becomes obvious how cryptic transcription factor binding events or exotic RNA species can exist while serving no benefit (Palazzo and Gregory, 2014; Palazzo and Lee, 2015). Even tissue-specific expression will be subjected to unique noise given that each cell type expresses a unique set of transcription factors that activate distinct cryptic transcription start sites scattered throughout the genome (Levy et al., 2009; Palazzo and Lee, 2015). With the advent of powerful instruments that can identify rare and short-lived molecules, it should come as no surprise that we can document these non-adaptive entities. There is no doubt that further increasing the sensitivity of our detection instruments will reveal new biological insight, but it will also reveal an overwhelmingly large amount of cellular sloppiness.

## GLOBAL SOLUTIONS TO LOCAL PROBLEMS

If our cells are teeming with non-optimal reactions, mistake-riddled molecules, and non-adaptive processes, each being

slightly deleterious, but not enough to be eliminated by natural selection, how do cells, and by extension multicellular organisms, manage to survive? This situation is exacerbated if the genome is constantly absorbing slightly deleterious mutations that increase messiness. One can consider each of these mutations as a local problem with an associated cost to the fitness of the organism. When we consider that natural selection operates on each individual mutation, the cost of a single mutation (or “local problem”) may not be sufficient for its elimination. Instead, it appears that eukaryotic cells have what is known as “global solutions” to these local problems (Rajon and Masel, 2011; Rajon and Masel, 2013; Koonin, 2016). The concept of global solutions shares many aspects with the idea of “buffering” systems advanced by Marc Kirschner and John Gerhart (Gerhart and Kirschner, 1997; Kirschner and Gerhart, 1998; Kirschner and Gerhart, 2005). It also shares many features with robustness advanced by other theorists (Félix and Wagner, 2008; Masel and Siegal, 2009). Global solutions are robust cellular mechanisms that maintain homeostasis. Often, these act to buffer not only environmental changes but also genetic changes. A good example would be chaperones, which not only prevent protein misfolding at high temperatures, but also promote the folding of proteins that have acquired destabilizing mutations. In the absence of sufficient chaperone activity, the effects of many such mildly deleterious mutations are revealed (Rutherford and Lindquist, 1998; Queitsch et al., 2002).

Other global solutions that increase the robustness of cells include RNA quality control mechanisms, such as non-sense mediated decay (Khajavi et al., 2006; Lykke-Andersen and Jensen, 2015), and the RNA export machinery, which retains mis-processed mRNAs and spurious transcripts in the nucleoplasm thus preventing their translation into potentially toxic proteins (Palazzo and Akef, 2012; Palazzo and Lee, 2015; Palazzo and Lee, 2018). Indeed, the evolution of the nucleus likely permitted the extensive mRNA processing that is characteristic of eukaryotic cells (López-García and Moreira, 2006; Martin and Koonin, 2006) and acts as a global solution to reduce the deleteriousness of countless RNA mis-processing events (Koonin, 2016) and the potential toxic effects of junk RNA (Palazzo and Lee, 2015; Palazzo and Lee, 2018). Besides non-sense mediated decay and mRNA export, the eukaryotic gene expression pathway is filled with distinct machineries that are coupled to each other allowing them to act as global solutions to errors and sub-optimal products (Maniatis and Reed, 2002; Warnecke and Hurst, 2011; Palazzo and Akef, 2012). For example, the splicing machinery directly deposits nuclear export factors on spliced mRNAs (Zhou et al., 2000; Luo et al., 2001; Masuda et al., 2005). For this reason, the splicing machinery is said to be coupled to the mRNA nuclear export machinery, and as a result spliced mRNA is more efficiently exported than mRNAs produced from intronless genes (Luo and Reed, 1999; Valencia et al., 2008). Since most protein-coding genes are well spliced, while spurious transcripts are not, the former but not the later are well exported (Palazzo and Lee, 2015). The gene expression pathway contains many other coupling reactions, and all of these promote the processing, nuclear export and translation of RNAs that contain features that are over-represented in protein-coding genes (“mRNA identity”



features), while promoting the nuclear retention and degradation of spurious transcripts that lack these features (Palazzo and Akef, 2012; Palazzo and Lee, 2018; Palazzo and Kang, 2021). Overall, these systems act as global solutions and increase the robustness of the gene expression machinery.

Another important global solution is the piwi-associated RNA (piRNA) system which represses transposon activity in the germline of most eukaryotes (Czech et al., 2018). Instead of eliminating every transposon, the piRNA system can be mobilized to suppress the deleteriousness of most transposable elements that have an RNA intermediate in their life cycle. Indeed, it is believed that the piRNA, small interference RNA (siRNA), and micro RNA (miRNA) systems all evolved early in the evolution of eukaryotes to globally inhibit transposable elements and viruses (Shabalina and Koonin, 2008).

The ultimate effect of these global solutions/buffering systems is to simultaneously elevate the selection coefficient of numerous mutations that share a common problem (Figure 2). Some of these mutations will have their selection coefficient values raised from slightly negative toward zero, and in some cases from near-zero to positive. For example, a novel mutation may allow an enzyme to perform an additional beneficial activity but also destabilize it. The proteostatic system may blunt the deleteriousness of any unfolded protein, while further promoting the beneficial activity. Non-sense mediated decay and coupling reactions in the gene expression system may eliminate misprocessed transcripts, while permitting some expression of an mRNA containing a poorly spliced novel exon. In this way, global solutions “encourage” the development of new innovations. Using similar arguments, Kirschner and Gerhart advocated that buffering systems facilitate phenotypic variation and promote “evolvability”

(Gerhart and Kirschner, 1997; Kirschner and Gerhart, 1998; Kirschner and Gerhart, 2005). In a similar vein, many systems are likely present that increase the plasticity and phenotypic heterogeneity of organisms, and these appear to increase robustness (Rego et al., 2017; Mo et al., 2021).

## MUTATIONAL BIAS

Although many evolutionary models assume that mutation supplies a random assortment of variants for natural selection to work on, mutation is not random, but is biased and this has major effects on genomic composition and the direction that evolution takes (Sueoka, 1962; Stoltzfus and Yampolsky, 2009; Nei, 2013). For example, in most organisms, transversion mutations (e.g., G to A mutations) are more common than transition mutations (e.g., G to C mutations). When amino acid differences between the proteomes of two closely related species are tabulated, the observed changes can be predicted by models that integrate mutation bias and the biochemical similarity of the swapped amino acids (Stoltzfus and Yampolsky, 2009). Ultimately, the nucleotide content of most genomes is in equilibrium, meaning that the nucleic acid composition is dictated by the interconversion rate between each of the bases (Sueoka, 1962).

Although the overall genomic content is dictated by global biases, mutational biases can also vary along the genome and likely dictate local nucleotide composition. For example, nucleosome occupancy can suppress cytosine deamination (Chen et al., 2012), and DNA curvature can alter local mutation rates (Duan et al., 2018). In addition, collisions between RNA polymerase and replication forks cause certain types of damage that are more prevalent at the start of genes (Aguilera, 2002; Lin and Pasero, 2012; Lin and Pasero, 2017). In transcription bubbles, the coding strand of transcribed genes is periodically single stranded and is thus more susceptible to certain types of DNA damage than the template strand. This likely leads to biased mutations in the coding strand of genes that are transcribed in the germ line, ultimately altering the frequency of all four nucleotides, and making these differ between the coding and template strand (Polak and Arndt, 2008; Hodgkinson and Eyre-Walker, 2011). These mutational biases may explain certain features of protein-coding genes, such as the elevated GC-content found at their 5' ends (Palazzo and Kang, 2021).

Repair can also be biased. For example, transcribed genes undergo transcription coupled DNA repair which acts only on the template strand (Fousteri and Mullenders, 2008). DNA repair enzymes are more efficiently recruited to chromatin with certain histone modifications (Li et al., 2013; Supek and Lehner, 2017; Huang et al., 2018; Huang and Li, 2018). These biases likely explain why transcribed and intergenic DNA differ in their nucleotide content (Palazzo and Kang, 2021) and why highly expressed genes may have lower mutation rates than intergenic regions (Monroe et al., 2022). DNA repair polymerases also tend to have low fidelity, for example misincorporating G into A sites (McCulloch and Kunkel, 2008; Goodman and Woodgate, 2013).

Thus, the repair of certain types of DNA damage may generate secondary mutations, at a lower rate, that have their own biases.

Bias also happens at the level of INDELS. In most organisms, small deletions are more common than small insertions (Kuo and Ochman, 2009; Sung et al., 2016). This bias may explain why prokaryotic genomes tend to have very little intergenic sequence (Mira et al., 2001; Moran et al., 2009). Tandem repeats also tend to be easily lost in prokaryotes due to recombination enzymes and this may further limit the growth of their genomes. In contrast, eukaryotes are parasitized by transposable elements which insert themselves into the genomes of their host, significantly increasing the rate of large insertions, and thus driving up genome size (Fedoroff, 2012; Palazzo and Gregory, 2014). Since the deleterious effects of these insertional mutations (i.e., local problems) are buffered by the eukaryotic gene expression machinery and the piRNA system (i.e., global solutions), this results in bloated genomes common to most eukaryotic lineages.

Other biases exist as well. DNA recombination during meiosis and some DNA repair pathways involve strand invasion between homologous chromosomes. Recombination between paternal and maternal chromosomes that differ due to the presence of heterozygous SNPs results in the formation of DNA with single nucleotide mismatches, which are then corrected in a biased manner to favor G:C base pairs over A:T base pairs. This process, called GC-biased gene conversion is quite prevalent in metazoans (Duret and Galtier, 2009). It has been estimated that each generation the average human genome gains roughly 100 *de novo* single nucleotide mutations (Chintalapati and Moorjani, 2020), and experiences 13 GC-biased gene conversion events (Paudel et al., 2020). When one considers that at most only 10% of all mutations occur in functional parts of the genome (Lindblad-Toh et al., 2011; Ward and Kellis, 2012; Rands et al., 2014; Graur, 2017), GC-biased gene conversion has at least as much influence over genomic content as selection, and this has been observed in the human genome (Pouyet et al., 2018). Ultimately, biased mutation (and biased repair) drastically shapes genomic features, directing the paths of both adaptive evolution and random genetic drift (Figure 1).

## THE IMPORTANCE OF THE NULL HYPOTHESIS

A key strength for the neutral theory of molecular evolution is that it provides molecular biologist with a null model for testing evolutionary hypotheses. Populations are perpetually evolving given the constant change in allele frequency and mutation rates. Therefore, it is important to understand whether these changes are due to stochastic versus adaptive processes. To invoke adaptive forces as the main driver of a molecular trait, one must demonstrate that the trait could not have arisen solely due to drift and biased mutations. In addition, organisms that have invested in global solutions that blunt the deleteriousness effects of messiness, will tolerate mutations that create a certain degree of messiness. This messiness will accumulate as long as it does not pose a burden above the drift barrier. To invoke adaptive forces as the main driver for the existence of some entity or some

process, one must demonstrate that the trait is not simply a product of a messy organism (Figure 1).

To see how an understanding of modern evolutionary thinking can shed light into biology, we will briefly revisit the molecular clock and discuss our current understanding of this phenomenon. Although the rate of evolutionary change is constant for a given protein-coding gene in different lineages, it has been documented that different protein-coding genes have different clock rates. Over 2 decades ago, it was demonstrated that protein conservation in yeast species (i.e., how slow the clock runs for a given protein) correlates with its expression level (Pál et al., 2001). In contrast, conservation level had little correlation with other features, such as whether the protein was essential for viability, or any other measure of importance (Wang and Zhang, 2009).

How does one make sense of this observation? Adaptive models would presume that most changes are due to positive selection. The null model is that most change is due to neutral, or slightly deleterious mutations. If we assume that most mutations are nearly-neutral (the null hypothesis), why would these accumulate in lowly expressed genes? One possible explanation is that highly expressed genes cannot easily accommodate slightly deleterious mutations that marginally destabilize the folding of their encoded protein, as this would impose a greater burden on the protein homeostasis and folding machinery (Drummond et al., 2005). In contrast, the slight misfolding of lowly expressed proteins is more tolerable as it imposes less of a burden on protein homeostasis. Highly expressed proteins also contribute to a greater number of intermolecular interactions (in terms of the *types of interactors* and total *number of interactions*). As a result, small changes in these proteins would either disrupt a greater number of functional interactions, or promote a greater number of misinteractions (Levy et al., 2012; Yang et al., 2012; Mehlhoff et al., 2020). Global solutions can be used to blunt small changes, but are not as effective against big changes. As such, highly expressed genes are under greater evolutionary constraint and are more optimized in terms of their structural stability and/or interactions than lowly expressed genes, which can more easily rely on global solutions to fold properly. Although this “E-R anticorrelation” (Expression, Rate of evolution) is seen in all life forms (Zhang and Yang, 2015), and has somewhat been experimentally validated in yeast (Wu et al., 2022), it is likely that the reasons for this may differ in mammalian cells where constraints on highly expressed mRNAs, in terms of RNA folding, translation, and other aspects of RNA biology, may be more significant than in yeast or prokaryotes (Liao et al., 2006). Despite this, it is widely thought that the E-R anticorrelation is due to different levels of constraint and that any change is due to a greater tolerance for deleterious alterations in certain genes or their encoded proteins.

Concepts from molecular evolution, and the use of null models, have been used to understand the evolution of a wide array of entities and biochemical processes. This includes the evolution of 5'UTRs (Lynch et al., 2005), introns (Hong et al., 2006), RNA modifications (Liu and Zhang, 2018; Jiang and Zhang, 2019) and RNA processing (Melamud and Moul, 2009).



2009; Pickrell et al., 2010; Saudemont et al., 2017; Xu and Zhang, 2021). In all these studies, many of the observed phenomena have features that are consistent with the null hypothesis—that they are largely shaped by neutral evolution. When these produce undesirable products that are nevertheless present in organisms, their effects are small enough that they do not clear the drift barrier and are largely blunted by global solutions.

## WHY TOLERANCE FOR SLOPPINESS CAN ACT AS A CATALYST FOR THE EVOLUTION OF COMPLEXITY

Although the nature of complexity is widely disputed, many observers have sought to define it in terms of a network or a “system”. The complexity of a system is correlated with the number of parts and the number of interactions between these parts (Hinegardner and Engelberg, 1983; McShea, 2000). It has been naively assumed by some that biological complexity is a direct product of natural selection (Ekstig, 2015). In other cases, it has been pointed out that when one starts off with a simple state (in this case a primitive ancestor) that evolves (i.e., has the capacity to change), then it is almost a certainty that some of its progeny will drift towards a more complex state, the so called “Drunkard’s Walk” model (Gould, 1996). Other commentators have noted that if an evolving organism is built of similar components, be it duplicated genes or similar groups of cells that form body parts, each individual component over evolutionary time will accumulate unique changes (either adaptive or neutral) that will eventually cause each part to drift away from its copies and cause the system as a whole to become more complex (McShea and Brandon, 2010). This effect, the so called “zero force evolutionary law”, does not require selection *per se*. Although the drifting apart of these components, may explain how a many-component system becomes complicated over time, it does not explain how new parts are generated. Furthermore, these drift models do not explain one of the main dichotomies that we see in life on earth: that by most definitions, eukaryotes are more complicated than prokaryotes.

What could explain the striking difference in the relative simplicity of prokaryotes and complexity of eukaryotes? One major difference appears to be how these two forms of life evolve. Put simply, prokaryotes experience high levels of selection pressure, while eukaryotes do not (Lynch, 2007). In organisms where selection pressure is extreme, superfluous activity is wasteful and effectively eliminated by natural selection (Lynch, 2006). In organisms with a low effective population size, these extra features are not purged but instead are allowed to persist for extended evolutionary time and their deleterious effects are instead buffered by global solutions. Even within eukaryotes, lower effective population sizes correlate with longer and/or more numerous intergenic regions, introns, cryptic transcriptional start sites and other non-functional genomic entities (Lynch, 2007). Within this surplus of non-functional activity lies the raw substrates for the evolution of new components. To rephrase this idea in terms used by other evolutionary theorists, a sloppy cellular environment full of

junk contains many available substrates that can be tinkered with (Jacob, 1977; Jacob, 2001) and eventually exapted (Gould and Vrba, 1982) to form new functional parts. More recently, the detailed process of exaptation has been investigated, and surprisingly, the evolution of junk to functional entities often involves processes that do not rely on positive selection, but rather on neutral evolution.

## CONSTRUCTIVE NEUTRAL EVOLUTION

As discussed above, non-functional entities are inefficiently eliminated in organisms that evolve under weak selection regimes. In the case of junk RNA, their persistence allows them to explore sequence space over considerable evolutionary time, allowing them to potentially acquire additional activities, or what is known as “excess capacity” (Stoltzfus, 1999). In some cases, the excess capacity overlaps the activity of some other functional entity. When situations like this arise, the activity of both new and existing parts, will mutationally decay to the point where either one activity is eliminated, or both activities become essential for organismal fitness. In this second scenario, there is an “accidental dependency” on the newly created entity (Stoltzfus, 1999). This is an example of the larger phenomenon, known as constructive neutral evolution (Stoltzfus, 1999; Stoltzfus, 2012). Note, that at no point was a new activity shaped by positive selection. Rather, the excess capacity was accidental and often a byproduct of proteins and RNA that have messy non-optimal activities. Furthermore, the eventual retention of the new part was due to mutational decay of the old part. At the end of this process the organism does not experience an increase in fitness. It however gains a new functional part and hence an increase in its complexity. An example of this is the evolution of junk RNA into functional long non-coding RNAs (Palazzo and Lee, 2018; Palazzo and Koonin, 2020). Eventually, these new functional entities may be further co-opted to generate entirely new evolutionary innovations and used to explore novel evolutionary trajectories. Organisms that experience strong selection regimes tend to eliminate messiness and as a result do not have as many raw substrates to fuel constructive neutral evolution. Organisms that experience weak selection regimes do not eliminate messiness. Instead, these new non-functional entities are allowed to explore sequence space for extended periods of time, increasing the chance that they accidentally evolve a new excess capacity.

As described above, an increase in complexity is not only due to the creation of new parts, but also through the establishment of new interactions between existing parts. Again, the rewiring of interaction networks is enhanced by constructive neutral evolution. For example, proteins that have an established function may acquire initial mutations that give it additional properties (i.e., excess capacity). As we pointed out earlier, enzymes and proteins are inherently messy and can promote many non-optimal reactions that may not be initially selected for. In some cases, these new activities can simply be the generation of a novel binding site between two unrelated proteins. Initially, these additional binding events may serve no function, but may

allow the two proteins to structurally stabilize one another through non-specific chaperoning. These initial fortuitous mutations then allow for additional slightly deleterious mutations that structurally destabilize both proteins. The deleteriousness of these secondary mutations is of course blunted by the fact that the two proteins act as chaperones for each other. However, now the association of the two proteins, which originally was non-functional, is now required for their structural stability (i.e., accidental dependency). Like a molecular ratchet, the protein-binding interface becomes more and more essential as each complex member acquires more and more of these secondary destabilizing mutations. This type of constructive neutral evolution explains how cellular machineries, such as the ribosome and spliceosome, have gained components over evolutionary time (Stoltzfus, 1999; Gray et al., 2010; Lukeš et al., 2011; Stoltzfus, 2012; Linquist et al., 2020). This model has been experimentally validated in other multi-protein complexes and predicts the long-term evolution of amino-acid composition seen in many proteins that are part of large complexes (Hochberg et al., 2020).

## SUMMARY

Most molecular biologists use an antiquated model of how evolution shapes biological processes leading them to an

## REFERENCES

- Abraham, J., Feldman, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S., Broach, J. R., et al. (1983). Sites Required for Position-Effect Regulation of Mating-type Information in Yeast. *Cold Spring Harbor Symposia Quantitative Biol.* 47, 989–998. doi:10.1101/SQB.1983.047.01.113
- Aguilera, A. (2002). The Connection between Transcription and Genomic Instability. *EMBO J.* 21, 195–201. doi:10.1093/emboj/21.3.195
- Allan Drummond, D., and Wilke, C. O. (2009). The Evolutionary Consequences of Erroneous Protein Synthesis. *Nat. Rev. Genet.* 10, 715–724. doi:10.1038/nrg2662
- Atkins, W. M. (2020). Mechanisms of Promiscuity Among Drug Metabolizing Enzymes and Drug Transporters. *FEBS J.* 287, 1306–1322. doi:10.1111/febs.15116
- Banerji, J., Olson, L., and Schaffner, W. (1983). A Lymphocyte-specific Cellular Enhancer Is Located Downstream of the Joining Region in Immunoglobulin Heavy Chain Genes. *Cell* 33, 729–740. doi:10.1016/0092-8674(83)90015-6
- Beach, D., Piper, M., and Shall, S. (1980). Isolation of Chromosomal Origins of Replication in Yeast. *Nature* 284, 185–187. doi:10.1038/284185a0
- Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced Segments at the 5' Terminus of Adenovirus 2 Late mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 74, 3171–3175. doi:10.1073/pnas.74.8.3171
- Berk, A. J. (2016). Discovery of RNA Splicing and Genes in Pieces. *Proc. Natl. Acad. Sci. USA* 113, 801–805. doi:10.1073/pnas.1525084113
- ENCODE Project ConsortiumBernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., and Snyder, M. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* 489, 57–74. doi:10.1038/nature11247
- ENCODE Project ConsortiumBirney, E., Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., et al. (2007). Identification and Analysis of Functional Elements in 1% of the Human Genome by the ENCODE Pilot Project. *Nature* 447, 799–816. doi:10.1038/nature05874
- Blackburn, E. H., and Gall, J. G. (1978). A Tandemly Repeated Sequence at the Termini of the Extrachromosomal Ribosomal RNA Genes in Tetrahymena. *J. Mol. Biol.* 120, 33–53. doi:10.1016/0022-2836(78)90294-2

unrealistic hyper-adaptationalist view. A prime example of this is the interpretation of the ENCODE project results. Ultimately, this ultra-Darwinian mindset perpetuates the notion that the genome, and life itself, is like a Swiss watch—ornate, and complicated, with every part hand crafted for a specific purpose. This view is completely compatible with the idea that genome is pure information. However, this view is based on ignorance of developments in molecular evolution. It also ignores principles of biochemistry, that predict suboptimal reactions and widespread promiscuity. A more modern view of the eukaryotic cell, shaped by drift-dominated evolution, is a messy junk-filled entity, full of Rube-Goldberg contraptions that were hobbled together by non-adaptive forces. With this new vantage point, certain aspects of eukaryotic biology become clarified, including the evolution of complexity.

## AUTHOR CONTRIBUTIONS

The text was written by AP and NK.

## ACKNOWLEDGMENTS

We would like to thank E. Koonin, and L. Moran for useful feedback on our manuscript.

- Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The Product of the H19 Gene May Function as an RNA. *Mol. Cell. Biol.* 10, 28–36. doi:10.1128/mcb.10.1.28-36.1990
- Brenner, S. (2003). NOBEL LECTURE: Nature's Gift to Science. *Biosci. Rep.* 23, 225–237. doi:10.1023/B:BIRE.0000019186.48208.f3
- Britten, R. J., and Kohne, D. E. (1968). Repeated Sequences in DNA. *Science* 161, 529–540. doi:10.1126/science.161.3841.529
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., et al. (1992). The Product of the Mouse Xist Gene Is a 15 Kb Inactive X-specific Transcript Containing No Conserved ORF and Located in the Nucleus. *Cell* 71, 515–526. doi:10.1016/0092-8674(92)90519-i
- Charlesworth, B. (2009). Effective Population Size and Patterns of Molecular Evolution and Variation. *Nat. Rev. Genet.* 10, 195–205. doi:10.1038/nrg2526
- Chen, X., Chen, Z., Chen, H., Su, Z., Yang, J., Lin, F., et al. (2012). Nucleosomes Suppress Spontaneous Mutations Base-Specifically in Eukaryotes. *Science* 335, 1235–1238. doi:10.1126/science.1217580
- Chintalapati, M., and Moorjani, P. (2020). Evolution of the Mutation Rate across Primates. *Curr. Opin. Genet. Develop.* 62, 58–64. doi:10.1016/j.gde.2020.05.028
- Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An Amazing Sequence Arrangement at the 5' Ends of Adenovirus 2 Messenger RNA. *Cell* 12, 1–8. doi:10.1016/0092-8674(77)90180-5
- Clarke, L., and Carbon, J. (1980). Isolation of a Yeast Centromere and Construction of Functional Small Circular Chromosomes. *Nature* 287, 504–509. doi:10.1038/287504a0
- Crick, F. H. (1958). On Protein Synthesis. *Symp. Soc. Exp. Biol.* 12, 138–163.
- Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E. L., Fabry, M. H., Kneuss, E., et al. (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu. Rev. Genet.* 52, 131–157. doi:10.1146/annurev-genet-120417-031441
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of Transcription in Human Cells. *Nature* 489, 101–108. doi:10.1038/nature11233
- Doolittle, W. F., Brunet, T. D. P., Linquist, S., and Gregory, T. R. (2014). Distinguishing between “Function” and “Effect” in Genome Biology. *Genome Biol. Evol.* 6, 1234–1237. doi:10.1093/gbe/evu098

- Doolittle, W. F. (2013). Is Junk DNA Bunk? A Critique of ENCODE. *Proc. Natl. Acad. Sci.* 110, 5294–5300. doi:10.1073/pnas.1221376110
- Doolittle, W. F., and Sapienza, C. (1980). Selfish Genes, the Phenotype Paradigm and Genome Evolution. *Nature* 284, 601–603. doi:10.1038/284601a0
- Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. (2005). Why Highly Expressed Proteins Evolve Slowly. *Proc. Natl. Acad. Sci.* 102, 14338–14343. doi:10.1073/pnas.0504070102
- Duan, C., Huan, Q., Chen, X., Wu, S., Carey, L. B., He, X., et al. (2018). Reduced Intrinsic DNA Curvature Leads to Increased Mutation Rate. *Genome Biol.* 19, 132. doi:10.1186/s13059-018-1525-y
- Duret, L., and Galtier, N. (2009). Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annu. Rev. Genom. Hum. Genet.* 10, 285–311. doi:10.1146/annurev-genom-082908-150001
- Eddy, S. R. (2012). The C-Value Paradox, Junk DNA and ENCODE. *Curr. Biol.* 22, R898–R899. doi:10.1016/j.cub.2012.10.002
- Eddy, S. R. (2013). The ENCODE Project: Missteps Overshadowing a success. *Curr. Biol.* 23, R259–R261. doi:10.1016/j.cub.2013.03.023
- Ekstig, B. (2015). Complexity, Natural Selection and the Evolution of Life and Humans. *Found. Sci.* 20, 175–187. doi:10.1007/s10699-014-9358-y
- Fedoroff, N. V. (2012). Transposable Elements, Epigenetics, and Genome Evolution. *Science* 338, 758–767. doi:10.1126/science.338.6108.758
- Félix, M.-A., and Wagner, A. (2008). Robustness and Evolution: Concepts, Insights and Challenges from a Developmental Model System. *Heredity* 100, 132–140. doi:10.1038/sj.hdy.6800915
- Fousteri, M., and Mullenders, L. H. (2008). Transcription-coupled Nucleotide Excision Repair in Mammalian Cells: Molecular Mechanisms and Biological Effects. *Cell Res* 18, 73–84. doi:10.1038/cr.2008.6
- Freese, E. (1962). On the Evolution of the Base Composition of DNA. *J. Theor. Biol.* 3, 82–101. doi:10.1016/S0022-5193(62)80005-8
- Garland, W., and Jensen, T. H. (2020). Nuclear Sorting of RNA. *WIREs RNA* 11, e1572. doi:10.1002/wrna.1572
- Gerber, A., Esnault, C., Aubert, G., Treisman, R., Pralong, F., and Schibler, U. (2013). Blood-Borne Circadian Signal Stimulates Daily Oscillations in Actin Dynamics and SRF Activity. *Cell* 152, 492–503. doi:10.1016/j.cell.2012.12.027
- Gerhart, J., and Kirschner, M. (1997). *Cells, Embryos and Evolution*. 1st edition. Malden, Mass: Wiley, 656.
- Gilbert, W. (1978). Why Genes in Pieces. *Nature* 271, 501. doi:10.1038/271501a0
- Gillespie, J. H. (2000). Genetic Drift in an Infinite Population: The Pseudohitchhiking Model. *Genetics* 155, 909–919. doi:10.1093/genetics/155.2.909
- Gillies, S. D., Morrison, S. L., Oi, V. T., and Tonegawa, S. (1983). A Tissue-specific Transcription Enhancer Element Is Located in the Major Intron of a Rearranged Immunoglobulin Heavy Chain Gene. *Cell* 33, 717–728. doi:10.1016/0092-8674(83)90014-4
- Goodman, M. F., and Woodgate, R. (2013). Translesion DNA Polymerases. *Cold Spring Harbor Perspect. Biol.* 5, a010363. doi:10.1101/cshperspect.a010363
- Gosselin, P., Rando, G., Fleury-Olela, F., and Schibler, U. (2016). Unbiased Identification of Signal-Activated Transcription Factors by Barcoded Synthetic Tandem Repeat Promoter Screening (BC-STAR-PROM). *Genes Dev.* 30, 1895–1907. doi:10.1101/gad.284828.116
- Gould, S. J., and Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: a Critique of the Adaptationist Programme. *Proc. R. Soc. Lond. B Biol. Sci.* 205, 581–598. doi:10.1098/rspb.1979.0086
- Gould, S. J. (1996). *Full House: The Spread of Excellence from Plato to Darwin*. 1st ed. New York: Harmony Books.
- Gould, S. J., and Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology* 8, 4–15. doi:10.1017/S0094837300004310
- Graur, D. (2017). An Upper Limit on the Functional Fraction of the Human Genome. *Genome Biol. Evol.* 9, 1880–1885. doi:10.1093/gbe/evx121
- Graur, D., Sater, A. K., and Cooper, T. F. (2016). *Molecular and Genome Evolution*. Sunderland, Massachusetts: Sinauer Associates.
- Graur, D. (2013). *The Origin of Junk DNA: A Historical Whodunnit*. Judge Starling. Available at: <https://judgestarling.tumblr.com/post/64504735261/the-origin-of-junk-dna-a-historical-whodunnit>. (Accessed February 4, 2022).
- Graur, D., Zheng, Y., Price, N., Azevedo, R. B. R., 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 Biol. Evol.* 5, 578–590. doi:10.1093/gbe/evt028
- Gray, M. W., Lukeš, J., Archibald, J. M., Keeling, P. J., and Doolittle, W. F. (2010). Irremediable Complexity. *Science* 330, 920–921. doi:10.1126/science.1198594
- Hatje, K., Mühlhausen, S., Simm, D., and Kollmar, M. (2019). The Protein-Coding Human Genome: Annotating High-Hanging Fruits. *BioEssays* 41, 1900066. doi:10.1002/bies.201900066
- Hinegardner, R., and Engelberg, J. (1983). Biological Complexity. *J. Theor. Biol.* 104, 7–20. doi:10.1016/0022-5193(83)90398-3
- Hochberg, G. K. A., Liu, Y., Marklund, E. G., Metzger, B. P. H., Laganowsky, A., and Thornton, J. W. (2020). A Hydrophobic Ratchet Entrenches Molecular Complexes. *Nature* 588, 503–508. doi:10.1038/s41586-020-3021-2
- Hodgkinson, A., and Eyre-Walker, A. (2011). Variation in the Mutation Rate across Mammalian Genomes. *Nat. Rev. Genet.* 12, 756–766. doi:10.1038/nrg3098
- Holley, R. W., Appar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., et al. (1965). STRUCTURE OF A RIBONUCLEIC ACID. *Science* 147, 1462–1465. doi:10.1126/science.147.3664.1462
- Hong, X., Scofield, D. G., and Lynch, M. (2006). Intron Size, Abundance, and Distribution within Untranslated Regions of Genes. *Mol. Biol. Evol.* 23, 2392–2404. doi:10.1093/molbev/msl111
- Huang, Y., Gu, L., and Li, G.-M. (2018). H3K36me3-mediated Mismatch Repair Preferentially Protects Actively Transcribed Genes from Mutation. *J. Biol. Chem.* 293, 7811–7823. doi:10.1074/jbc.RA118.002839
- Huang, Y., and Li, G.-M. (2018). DNA Mismatch Repair Preferentially Safeguards Actively Transcribed Genes. *DNA Repair* 71, 82–86. doi:10.1016/j.dnarep.2018.08.010
- Hubby, J. L., and Lewontin, R. C. (1966). A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. I. The Number of Alleles at Different Loci in DROSOPHILA PSEUDOBOBSCURA. *Genetics* 54, 577–594. doi:10.1093/genetics/54.2.577
- Hughes, A. L. (2007). Looking for Darwin in All the Wrong Places: the Misguided Quest for Positive Selection at the Nucleotide Sequence Level. *Heredity* 99, 364–373. doi:10.1038/sj.hdy.6801031
- Ippen, K., Miller, J. H., Scaife, J., and Beckwith, J. (1968). New Controlling Element in the Lac Operon of *E. coli*. *Nature* 217, 825–827. doi:10.1038/217825a0
- Jacob, F. (2001). Complexity and Tinkering. *Ann. N. Y. Acad. Sci.* 929, 71–73. doi:10.1111/j.1749-6632.2001.tb05708.x
- Jacob, F. (1977). Evolution and Tinkering. *Science* 196, 1161–1166. doi:10.1126/science.860134
- Jankowsky, E., and Harris, M. E. (2015). Specificity and Nonspecificity in RNA-Protein Interactions. *Nat. Rev. Mol. Cell Biol.* 16, 533–544. doi:10.1038/nrm4032
- Jensen, J. D., Payseur, B. A., Stephan, W., Aquadro, C. F., Lynch, M., Charlesworth, D., et al. (2019). The Importance of the Neutral Theory in 1968 and 50 Years on: A Response to Kern and Hahn 2018. *Evolution* 73, 111–114. doi:10.1111/evo.13650
- Jiang, D., and Zhang, J. (2019). The Preponderance of Nonsynonymous A-To-I RNA Editing in Coleoids Is Nonadaptive. *Nat. Commun.* 10, 5411. doi:10.1038/s41467-019-13275-2
- Kern, A. D., and Hahn, M. W. (2018). The Neutral Theory in Light of Natural Selection. *Mol. Biol. Evol.* 35, 1366–1371. doi:10.1093/molbev/msy092
- Khajavi, M., Inoue, K., and Lupski, J. R. (2006). Nonsense-mediated mRNA Decay Modulates Clinical Outcome of Genetic Disease. *Eur. J. Hum. Genet.* 14, 1074–1081. doi:10.1038/sj.ejhg.5201649
- Kimsey, I. J., Petzold, K., Sathyamoorthy, B., Stein, Z. W., and Al-Hashimi, H. M. (2015). Visualizing Transient Watson-crick-like Mispairs in DNA and RNA Duplexes. *Nature* 519, 315–320. doi:10.1038/nature14227
- Kimsey, I. J., Szymanski, E. S., Zahurancik, W. J., Shakya, A., Xue, Y., Chu, C.-C., et al. (2018). Dynamic Basis for dGdT Misincorporation via Tautomerization and Ionization. *Nature* 554, 195–201. doi:10.1038/nature25487
- Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature* 217, 624–626. doi:10.1038/217624a0
- Kimura, M., and Ohta, T. (1974). On Some Principles Governing Molecular Evolution. *Proc. Natl. Acad. Sci.* 71, 2848–2852. doi:10.1073/pnas.71.7.2848
- Kimura, M. (1991). The Neutral Theory of Molecular Evolution: a Review of Recent Evidence. *Jpn. J. Genet.* 66, 367–386. doi:10.1266/jjg.66.367
- King, J. L., and Jukes, T. H. (1969). Non-Darwinian Evolution. *Science* 164, 788–798. doi:10.1126/science.164.3881.788
- Kirschner, M., and Gerhart, J. (1998). Evolvability. *Proc. Natl. Acad. Sci.* 95, 8420–8427. doi:10.1073/pnas.95.15.8420

- Kirschner, M. W., and Gerhart, J. C. (2005). *The Plausibility of Life: Resolving Darwin's Dilemma*. 1st edition. New Haven, London: Yale University Press.
- Koonin, E. (2011). *The Logic of Chance: The Nature and Origin of Biological Evolution*. Upper Saddle River N.J.: Pearson Education.
- Koonin, E. V. (2016). Splendor and Misery of Adaptation, or the Importance of Neutral Null for Understanding Evolution. *BMC Biol.* 14, 114. doi:10.1186/s12915-016-0338-2
- Kosinski, L. J., and Masel, J. (2020). Readthrough Errors Purge Deleterious Cryptic Sequences, Facilitating the Birth of Coding Sequences. *Mol. Biol. Evol.* 37, 1761–1774. doi:10.1093/molbev/msaa046
- Kramer, E. B., and Farabaugh, P. J. (2007). The Frequency of Translational Misreading Errors in *E. coli* Is Largely Determined by tRNA Competition. *RNA* 13, 87–96. doi:10.1261/rna.294907
- Kuo, C.-H., and Ochman, H. (2009). Deletional Bias across the Three Domains of Life. *Genome Biol. Evol.* 1, 145–152. doi:10.1093/gbe/evp016
- Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* 30, 314–334. doi:10.2307/240770310.1111/j.1558-5646.1976.tb00911.x
- Levy, E. D., De, S., and Teichmann, S. A. (2012). Cellular Crowding Imposes Global Constraints on the Chemistry and Evolution of Proteomes. *Proc. Natl. Acad. Sci.* 109, 20461–20466. doi:10.1073/pnas.1209312109
- Levy, E. D., Landry, C. R., and Michnick, S. W. (2009). How Perfect Can Protein Interactomes Be. *Sci. Signal.* 2, pe11. doi:10.1126/scisignal.260pe11
- Lewontin, R. C., and Hubby, J. L. (1966). A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of *DROSOPHILA PSEUDOOSCURA*. *Genetics* 54, 595–609. doi:10.1093/genetics/54.2.595
- Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., et al. (2013). The Histone Mark H3K36me3 Regulates Human DNA Mismatch Repair through its Interaction with MutSa. *Cell* 153, 590–600. doi:10.1016/j.cell.2013.03.025
- Liao, B.-Y., Scott, N. M., and Zhang, J. (2006). Impacts of Gene Essentiality, Expression Pattern, and Gene Compactness on the Evolutionary Rate of Mammalian Proteins. *Mol. Biol. Evol.* 23, 2072–2080. doi:10.1093/molbev/msl076
- Lin, Y.-L., and Pasero, P. (2012). Interference between DNA Replication and Transcription as a Cause of Genomic Instability. *Curr. Genomics* 13, 65–73. doi:10.2174/138920212799034767
- Lin, Y.-L., and Pasero, P. (2017). Transcription-Replication Conflicts: Orientation Matters. *Cell* 170, 603–604. doi:10.1016/j.cell.2017.07.040
- Lindblad-Toh, K., Garber, M., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., et al. (2011). A High-Resolution Map of Human Evolutionary Constraint Using 29 Mammals. *Nature* 478, 476–482. doi:10.1038/nature10530
- Linguist, S., Doolittle, W. F., and Palazzo, A. F. (2020). Getting clear about the F-word in Genomics. *PLoS Genet.* 16, e1008702. doi:10.1371/journal.pgen.1008702
- Liu, Z., and Zhang, J. (2018). Most m6A RNA Modifications in Protein-Coding Regions Are Evolutionarily Unconserved and Likely Nonfunctional. *Mol. Biol. Evol.* 35, 666–675. doi:10.1093/molbev/msx320
- López-García, P., and Moreira, D. (2006). Selective Forces for the Origin of the Eukaryotic Nucleus. *Bioessays* 28, 525–533. doi:10.1002/bies.20413
- Lukeš, J., Archibald, J. M., Keeling, P. J., Doolittle, W. F., and Gray, M. W. (2011). How a Neutral Evolutionary Ratchet Can Build Cellular Complexity. *IUBMB Life* 63, 528–537. doi:10.1002/iub.489
- Luo, M. J., and Reed, R. (1999). Splicing Is Required for Rapid and Efficient mRNA export in Metazoans. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14937–14942. doi:10.1073/pnas.96.26.14937
- Luo, M. L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., et al. (2001). Pre-mRNA Splicing and mRNA export Linked by Direct Interactions between UAP56 and Aly. *Nature* 413, 644–647. doi:10.1038/35098106
- Lykke-Andersen, S., and Jensen, T. H. (2015). Nonsense-mediated mRNA Decay: an Intricate Machinery that Shapes Transcriptomes. *Nat. Rev. Mol. Cell Biol.* 16, 665–677. doi:10.1038/nrm4063
- Lynch, M., Ackerman, M. S., Gout, J.-F., Long, H., Sung, W., Thomas, W. K., et al. (2016). Genetic Drift, Selection and the Evolution of the Mutation Rate. *Nat. Rev. Genet.* 17, 704. doi:10.1038/nrg.2016.104
- Lynch, M., Bobay, L.-M., Catania, F., Gout, J.-F., and Rho, M. (2011). The Repatterning of Eukaryotic Genomes by Random Genetic Drift. *Annu. Rev. Genomics Hum. Genet.* 12, 347–366. doi:10.1146/annurev-genom-082410-101412
- Lynch, M., Scofield, D. G., and Hong, X. (2005). The Evolution of Transcription-Initiation Sites. *Mol. Biol. Evol.* 22, 1137–1146. doi:10.1093/molbev/msi100
- Lynch, M. (2006). Streamlining and Simplification of Microbial Genome Architecture. *Annu. Rev. Microbiol.* 60, 327–349. doi:10.1146/annurev.micro.60.080805.142300
- Lynch, M. (2007). *The Origins of Genome Architecture*. Sunderland Mass: Sinauer Associates.
- Maniatis, T., and Reed, R. (2002). An Extensive Network of Coupling Among Gene Expression Machines. *Nature* 416, 499–506. doi:10.1038/416499a
- Margoliash, E. (1963). PRIMARY STRUCTURE AND EVOLUTION OF CYTOCHROME C. *Proc. Natl. Acad. Sci. U S A.* 50, 672–679. doi:10.1073/pnas.50.4.672
- Martin, W., and Koonin, E. V. (2006). Introns and the Origin of Nucleus-Cytosol Compartmentalization. *Nature* 440, 41–45. doi:10.1038/nature04531
- Masel, J., and Siegal, M. L. (2009). Robustness: Mechanisms and Consequences. *Trends Genet.* 25, 395–403. doi:10.1016/j.tig.2009.07.005
- Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the Human TREX Complex to mRNA during Splicing. *Genes Dev.* 19, 1512–1517. doi:10.1101/gad.1302205
- McCulloch, S. D., and Kunkel, T. A. (2008). The Fidelity of DNA Synthesis by Eukaryotic Replicative and Translesion Synthesis Polymerases. *Cel Res* 18, 148–161. doi:10.1038/cr.2008.4
- McShea, D. W., and Brandon, R. N. (2010). *Biology's First Law: The Tendency for Diversity and Complexity to Increase in Evolutionary Systems*. Chicago, London: University of Chicago Press.
- McShea, D. W. (2000). Functional Complexity in Organisms: Parts as Proxies. *Biol. Philos.* 15, 641–668. doi:10.1023/A:1006695908715
- Mehlhoff, J. D., Stearns, F. W., Rohm, D., Wang, B., Tsou, E.-Y., Dutta, N., et al. (2020). Collateral Fitness Effects of Mutations. *Proc. Natl. Acad. Sci. U S A.* 117, 11597–11607. doi:10.1073/pnas.1918680117
- Melamud, E., and Moul, J. (2009). Stochastic Noise in Splicing Machinery. *Nucleic Acids Res.* 37, 4873–4886. doi:10.1093/nar/gkp471
- Mercola, M., Wang, X. F., Olsen, J., and Calame, K. (1983). Transcriptional Enhancer Elements in the Mouse Immunoglobulin Heavy Chain Locus. *Science* 221, 663–665. doi:10.1126/science.6306772
- Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional Bias and the Evolution of Bacterial Genomes. *Trends Genet.* 17, 589–596. doi:10.1016/s0168-9525(01)02447-7
- Mo, N., Zhang, X., Shi, W., Yu, G., Chen, X., and Yang, J.-R. (2021). Bidirectional Genetic Control of Phenotypic Heterogeneity and its Implication for Cancer Drug Resistance. *Mol. Biol. Evol.* 38, 1874–1887. doi:10.1093/molbev/msaa332
- Monroe, J. G., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., et al. (2022). Mutation Bias Reflects Natural Selection in *Arabidopsis thaliana*. *Nature* 602, 101–105. doi:10.1038/s41586-021-04269-6
- Moran, N. A., McLaughlin, H. J., and Sorek, R. (2009). The Dynamics and Time Scale of Ongoing Genomic Erosion in Symbiotic Bacteria. *Science* 323, 379–382. doi:10.1126/science.1167140
- Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M., and Helling, R. B. (1974). Replication and Transcription of Eukaryotic DNA in *Escherichia Coli*. *Proc. Natl. Acad. Sci. U S A.* 71, 1743–1747. doi:10.1073/pnas.71.5.1743
- Muller, H. J. (1966). The Gene Material as the Initiator and the Organizing Basis of Life. *The Am. Naturalist* 100, 493–517. doi:10.1086/282445
- Nei, M. (2013). *Mutation-Driven Evolution*. Oxford: Oxford University Press.
- Nei, M. (2005). Selectionism and Neutralism in Molecular Evolution. *Mol. Biol. Evol.* 22, 2318–2342. doi:10.1093/molbev/msi242
- Niu, D.-K., and Jiang, L. (2013). Can ENCODE Tell Us How Much Junk DNA We Carry in Our Genome. *Biochem. Biophys. Res. Commun.* 430, 1340–1343. doi:10.1016/j.bbrc.2012.12.074
- Ohno, S. (1972a). So Much “Junk” DNA in Our Genome. *Brookhaven Symp. Biol.* 23, 366–370.
- Ohno, S. (1972b). An Argument for the Genetic Simplicity of Man and Other Mammals. *J. Hum. Evol.* 1, 651–662. doi:10.1016/0047-2484(72)90011-5
- Ohta, T. (2003). Origin of the Neutral and Nearly Neutral Theories of Evolution. *J. Biosci.* 28, 371–377. doi:10.1007/BF02705113

- Orgel, L. E., and Crick, F. H. (1980). Selfish DNA: the Ultimate Parasite. *Nature* 284, 604–607. doi:10.1038/284604a0
- Pál, C., Papp, B., and Hurst, L. D. (2001). Highly Expressed Genes in Yeast Evolve Slowly. *Genetics* 158, 927–931.
- Palazzo, A. F., and Akef, A. (2012). Nuclear export as a Key Arbiter of “mRNA Identity” in Eukaryotes. *Biochim. Biophys. Acta* 1819, 566–577. doi:10.1016/j.bbagr.2011.12.012
- Palazzo, A. F., and Gregory, T. R. (2014). The Case for Junk DNA. *Plos Genet.* 10, e1004351. doi:10.1371/journal.pgen.1004351
- Palazzo, A. F., and Kang, Y. M. (2021). GC-content Biases in Protein-Coding Genes Act as an “mRNA Identity” Feature for Nuclear export. *BioEssays* 43, e2000197. doi:10.1002/bies.202000197
- Palazzo, A. F., and Koonin, E. V. (2020). Functional Long Non-coding RNAs Evolve from Junk Transcripts. *Cell* 183, 1–12. doi:10.1016/j.cell.2020.09.047
- Palazzo, A. F., and Lee, E. S. (2015). Non-coding RNA: what Is Functional and what Is Junk. *Front. Genet.* 6, 2. doi:10.3389/fgene.2015.00002
- Palazzo, A. F., and Lee, E. S. (2018). Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. *Front. Genet.* 9, 440. doi:10.3389/fgene.2018.00440
- Paudel, R., Fedorova, L., and Fedorov, A. (2020). Adapting Biased Gene Conversion Theory to Account for Intensive GC-Content Deterioration in the Human Genome by Novel Mutations. *PLOS ONE* 15, e0232167. doi:10.1371/journal.pone.0232167
- Pennisi, E. (2012). Genomics. ENCODE Project Writes Eulogy for Junk DNA. *Science* 337, 11591161. doi:10.1126/science.337.6099.1159
- Pickrell, J. K., Pai, A. A., Gilad, Y., and Pritchard, J. K. (2010). Noisy Splicing Drives mRNA Isoform Diversity in Human Cells. *Plos Genet.* 6, e1001236. doi:10.1371/journal.pgen.1001236
- Polak, P., and Arndt, P. F. (2008). Transcription Induces Strand-specific Mutations at the 5' End of Human Genes. *Genome Res.* 18, 1216–1223. doi:10.1101/gr.076570.108
- Ponting, C. (2017). Biological Function in the Twilight Zone of Sequence Conservation. *BMC Biol.* 15, 71. doi:10.1186/s12915-017-0411-5
- Pouyet, F., Aeschbacher, S., Thiéry, A., and Excoffier, L. (2018). Background Selection and Biased Gene Conversion Affect More Than 95% of the Human Genome and Bias Demographic Inferences. *Elife* 7, e36317. doi:10.7554/eLife.36317
- Queitsch, C., Sangster, T. A., and Lindquist, S. (2002). Hsp90 as a Capacitor of Phenotypic Variation. *Nature* 417, 618–624. doi:10.1038/nature749
- Rajon, E., and Masel, J. (2013). Compensatory Evolution and the Origins of Innovations. *Genetics* 193, 1209–1220. doi:10.1534/genetics.112.148627
- Rajon, E., and Masel, J. (2011). Evolution of Molecular Error Rates and the Consequences for Evolvability. *PNAS* 108, 1082–1087. doi:10.1073/pnas.1012918108
- Rands, C. M., Meader, S., Ponting, C. P., and Lunter, G. (2014). 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLOS Genet.* 10, e1004525. doi:10.1371/journal.pgen.1004525
- Ray, D., Kazan, H., Chan, E. T., Castillo, L. P., Chaudhry, S., Talukder, S., et al. (2009). Rapid and Systematic Analysis of the RNA Recognition Specificities of RNA-Binding Proteins. *Nat. Biotechnol.* 27, 667–670. doi:10.1038/nbt.1550
- Rego, E. H., Audette, R. E., and Rubin, E. J. (2017). Deletion of a Mycobacterial Divisome Factor Collapses Single-Cell Phenotypic Heterogeneity. *Nature* 546, 153–157. doi:10.1038/nature22361
- Reinke, H., Saini, C., Fleury-Olela, F., Dibner, C., Benjamin, I. J., and Schibler, U. (2008). Differential Display of DNA-Binding Proteins Reveals Heat-Shock Factor 1 as a Circadian Transcription Factor. *Genes Dev.* 22, 331–345. doi:10.1101/gad.453808
- Roberts, R. (1958). *Microsomal Particles and Protein Synthesis*. New York, London, Paris and Los Angeles: Pergamon Press.
- Rutherford, S. L., and Lindquist, S. (1998). Hsp90 as a Capacitor for Morphological Evolution. *Nature* 396, 336–342. doi:10.1038/24550
- Saudemont, B., Popa, A., Parmley, J. L., Rocher, V., Blugeon, C., Necseulea, A., et al. (2017). The Fitness Cost of Mis-Splicing Is the Main Determinant of Alternative Splicing Patterns. *Genome Biol.* 18, 208. doi:10.1186/s13059-017-1344-6
- Sawato, T., Saito, N., and Yamaguchi, M. (2019). Chemical Systems Involving Two Competitive Self-Catalytic Reactions. *ACS Omega* 4, 5879–5899. doi:10.1021/acsomega.9b00133
- Shabalina, S. A., and Koonin, E. V. (2008). Origins and Evolution of Eukaryotic RNA Interference. *Trends Ecol. Evol.* 23, 578–587. doi:10.1016/j.tree.2008.06.005
- Spitz, F., and Furlong, E. E. M. (2012). Transcription Factors: from Enhancer Binding to Developmental Control. *Nat. Rev. Genet.* 13, 613–626. doi:10.1038/nrg3207
- Stark, B. C., Kole, R., Bowman, E. J., and Altman, S. (1978). Ribonuclease P: an Enzyme with an Essential RNA Component. *Proc. Natl. Acad. Sci. U.S.A.* 75, 3717–3721. doi:10.1073/pnas.75.8.3717
- Stoltzfus, A. (2012). Constructive Neutral Evolution: Exploring Evolutionary Theory's Curious Disconnect. *Biol. Direct.* 7, 35. doi:10.1186/1745-6150-7-35
- Stoltzfus, A. (1999). On the Possibility of Constructive Neutral Evolution. *J. Mol. Evol.* 49, 169–181. doi:10.1007/pl00006540
- Stoltzfus, A., and Yampolsky, L. Y. (2009). Climbing Mount Probable: Mutation as a Cause of Nonrandomness in Evolution. *J. Hered.* 100, 637–647. doi:10.1093/jhered/esp048
- Struhl, K. (2007). Transcriptional Noise and the Fidelity of Initiation by RNA Polymerase II. *Nat. Struct. Mol. Biol.* 14, 103–105. doi:10.1038/nsmb0207-103
- Sueoka, N. (1962). ON THE GENETIC BASIS OF VARIATION AND HETEROGENEITY OF DNA BASE COMPOSITION. *Proc. Natl. Acad. Sci. U S A.* 48, 582–592. doi:10.1073/pnas.48.4.582
- Sung, W., Ackerman, M. S., Dillon, M. M., Platt, T. G., Fuqua, C., Cooper, V. S., et al. (2016). Evolution of the Insertion-Deletion Mutation Rate across the Tree of Life. *G3: Genes, Genomes, Genet.* 6, 2583–2591. doi:10.1534/g3.116.030890
- Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., and Lynch, M. (2012). Drift-barrier Hypothesis and Mutation-Rate Evolution. *PNAS* 109, 18488–18492. doi:10.1073/pnas.1216223109
- Supek, F., and Lehner, B. (2017). Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes. *Cell* 170, 534–547. e23. doi:10.1016/j.cell.2017.07.003
- Tawfik, D. S. (2020). Enzyme Promiscuity and Evolution in Light of Cellular Metabolism. *FEBS J.* 287, 1260–1261. doi:10.1111/febs.15296
- Tawfik, O. K. A. D. S., and Tawfik, D. S. (2010). Enzyme Promiscuity: a Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* 79, 471–505. doi:10.1146/annurev-biochem-030409-143718
- Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., et al. (2012). The Accessible Chromatin Landscape of the Human Genome. *Nature* 489, 75–82. doi:10.1038/nature11232
- Valencia, P., Dias, A. P., and Reed, R. (2008). Splicing Promotes Rapid and Efficient mRNA export in Mammalian Cells. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3386–3391. doi:10.1073/pnas.0800250105
- Walter, P., and Blobel, G. (1982). Signal Recognition Particle Contains a 7S RNA Essential for Protein Translocation across the Endoplasmic Reticulum. *Nature* 299, 691–698. doi:10.1038/299691a0
- Wang, Z., and Zhang, J. (2009). Why Is the Correlation between Gene Importance and Gene Evolutionary Rate So Weak. *Plos Genet.* 5, e1000329. doi:10.1371/journal.pgen.1000329
- Ward, L. D., and Kellis, M. (2012). Evidence of Abundant Purifying Selection in Humans for Recently Acquired Regulatory Functions. *Science* 337, 1675–1678. doi:10.1126/science.1225057
- Warnecke, T., and Hurst, L. D. (2011). Error Prevention and Mitigation as Forces in the Evolution of Genes and Genomes. *Nat. Rev. Genet.* 12, 875–881. doi:10.1038/nrg3092
- Watson, J. D., and Crick, F. H. (1953). Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* 171, 964–967. doi:10.1038/171964b0
- Wei, H., Therrien, C., Blanchard, A., Guan, S., and Zhu, Z. (2008). The Fidelity Index Provides a Systematic Quantitation of star Activity of DNA Restriction Endonucleases. *Nucleic Acids Res.* 36, e50. doi:10.1093/nar/gkn182
- White, M. A., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2013). Massively Parallel *In Vivo* Enhancer Assay Reveals that Highly Local Features Determine the Cis-Regulatory Function of ChIP-Seq Peaks. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11952–11957. doi:10.1073/pnas.1307449110
- Wolpert, L. (1994). *Understanding the DNA Model*. Web of Stories. Available at: <https://www.webofstories.com/play/sydney.brenner/59>. (Accessed February 4, 2022).
- Wright, S. (1932). The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution. *Proc. sixth Int. congress Genet.* 1, 356–366.

- Wu, Z., Cai, X., Zhang, X., Liu, Y., Tian, G., Yang, J.-R., et al. (2022). Expression Level Is a Major Modifier of the Fitness Landscape of a Protein Coding Gene. *Nat. Ecol. Evol.* 6, 103–115. doi:10.1038/s41559-021-01578-x
- Xu, C., and Zhang, J. (2020). Mammalian Alternative Translation Initiation Is Mostly Nonadaptive. *Mol. Biol. Evol.* 37, 2015–2028. doi:10.1093/molbev/msaa063
- Xu, C., and Zhang, J. (2021). Mammalian Circular RNAs Result Largely from Splicing Errors. *Cell Rep* 36, 109439. doi:10.1016/j.celrep.2021.109439
- Yang, J.-R., Liao, B.-Y., Zhuang, S.-M., and Zhang, J. (2012). Protein Misinteraction Avoidance Causes Highly Expressed Proteins to Evolve Slowly. *Proc. Natl. Acad. Sci. U S A.* 109, E831–E840. doi:10.1073/pnas.1117408109
- Yang, V. W., Lerner, M. R., Steitz, J. A., and Flint, S. J. (1981). A Small Nuclear Ribonucleoprotein Is Required for Splicing of Adenoviral Early RNA Sequences. *Proc. Natl. Acad. Sci. U.S.A.* 78, 1371–1375. doi:10.1073/pnas.78.3.1371
- Zhang, J. (2018). Neutral Theory and Phenotypic Evolution. *Mol. Biol. Evol.* 35, 1327–1331. doi:10.1093/molbev/msy065
- Zhang, J., and Yang, J.-R. (2015). Determinants of the Rate of Protein Sequence Evolution. *Nat. Rev. Genet.* 16, 409–420. doi:10.1038/nrg3950
- Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The Protein Aly Links Pre-messenger-RNA Splicing to Nuclear export in Metazoans. *Nature* 407, 401–405. doi:10.1038/35030160
- Zuckermandl, E., and Pauling, L. (1962). “Molecular Disease, Evolution, and Genetic Heterogeneity,” in *Horizons in Biochemistry* (New York: Academic Press), 189–225.
- Zuckermandl, E., and Pauling, L. (1965). Molecules as Documents of Evolutionary History. *J. Theor. Biol.* 8, 357–366. doi:10.1016/0022-5193(65)90083-4

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Palazzo and Kejiou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.