Evolutionary honing in and mutational replacement: how long-term directed mutational responses to specific environmental pressures are possible

Adi Livnat[1,2*] Daniel Melamed,[1,2]

[1]Department of Evolutionary and Environmental Biology, University of Haifa, 3498838, Israel,
[2]Institute of Evolution, University of Haifa, 3498838, Israel
* To whom correspondence should be addressed; E-mail: alivnat@univ.haifa.ac.il

Abstract: Recent results have shown that the human malaria-resistant hemoglobin S mutation originates de novo more frequently in the gene and in the population where it is of adaptive significance, namely, in the hemoglobin subunit beta gene compared to the non-resistant but otherwise identical 20A→T mutation in the hemoglobin subunit delta gene, and in sub-Saharan Africans, who have been subject to intense malarial pressure for many generations, compared to northern Europeans, who have not. This finding raises a fundamental challenge to the traditional notion of accidental mutation. Here we address this finding with the replacement hypothesis, according to which pre-existing genetic interactions can lead directly and mechanistically to mutations that simplify and replace them. Thus, a gradual evolutionary process under selection can hone in on interactions of importance for the currently evolving adaptations, from which large-effect mutations follow that are relevant to these adaptations. We exemplify this hypothesis using multiple types of mutation, including gene fusion mutations, gene duplication mutations, A→G mutations in RNA-edited sites and transcription-associated mutations, and place it in the broader context of a system-level view of mutation origination called Interaction-based Evolution. Potential consequences include that similarity of mutation pressures may contribute to parallel evolution in genetically related species, that the evolution of genome organization may be driven by mutational mechanisms, that transposable element movements may also be explained by replacement, and that long-term directional mutational responses to specific environmental pressures are possible. Such responses need to be further tested by future studies in natural and artificial settings.

Keywords: Non random mutation; interaction-based evolution; hemoglobin S; directed mutation; parallelism; genome organization

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

1 Introduction

Mutation rates have long been measured as averages across genomic positions: across the whole genome, instances of a motif, the stretch of a gene, etc. (Kondrashov, 2003; Hodgkinson and Eyre-Walker, 2011; Rahbari et al., 2016; Carlson et al., 2018). Recently, we developed a method to measure mutation rates at the single mutation resolution and applied it to a 6 bp region in the human hemoglobin subunit beta (*HBB*) gene that contains the site of the malaria resistant HbS mutation and to the identical region in the hemoglobin subunit delta (*HBD*) gene in sperm samples from both sub-Saharan African and northern European donors (Melamed et al., 2022). Results showed that mutation rates at the single mutation resolution varied much more and in a different manner than expected from previous studies based on average mutation rates (cf. Harris, 2015; Harris and Pritchard, 2017). For example, mutation-specific rates varied substantially between the two genes and between the two populations, even though the mutations appeared on the identical local genetic sequence (Melamed et al., 2022). This suggests that much of the signal in mutation rate variation is in mutation-specific rates and thus had not been measurable before. Furthermore, the overall point mutation rate in *HBB* was significantly larger in Africans than in Europeans (Melamed et al., 2022), a difference at least two orders of magnitude larger than previously measured differences between human populations (Harris, 2015; Harris and Pritchard, 2017). Finally, a combination of statistically significant tests showed that the HbS mutation originates de novo more frequently in the *HBB* gene, where it provides resistance to malaria, compared to the non-resistant but otherwise identical mutation in *HBD*, and in sub-Saharan Africans, who have been experiencing intense malaria selection pressure for many generations, compared to northern Europeans, who have not (Melamed et al., 2022). Thus, it originates de novo more frequently where it is of adaptive significance.

These results emphasize the importance of revisiting modern synthetic concepts (Noble, 2008; Pigliucci and Müller, 2010; Noble et al., 2014; Noble, 2015). According to modifier theory, a random mutation that changes the average mutation rate, such as a mutation in a DNA-repair factor, can be fixed by natural selection, provided that it changes the average mutation rate across a large enough number of loci with which it remains linked for a long enough period

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

of time. In this manner, many potential mutations, each occurring with only a small probability, figure into its selective benefit (Leigh Jr, 1970; Feldman and Liberman, 1986; Altenberg et al., 2017). Accordingly, it was recently argued that the evolution of mutation rate modifiers reduced the average mutation rate in *Arabidobsis thaliana* in essential genes, which constitute 30% of the plant's genome (Monroe et al., 2022). However, because modifier theory cannot explain increases in the rates of specific mutations at specific base positions (Leigh Jr, 1970; Feldman and Liberman, 1986; Altenberg et al., 2017), it cannot address the patterns observed in the hemoglobin study. Furthermore, since *HBB* is an essential gene, only low mutation rates across the *HBB* gene would have been predicted based on the claim (Monroe et al., 2022) that essential genes are better protected from mutations.

Given all the above, how can the hemoglobin findings be interpreted? One possibility is that humans have a genomic fragility that by coincidence causes rapid origination of the malaria resistant HbS mutation both in the gene and in the population where it is of adaptive significance. However, this would leave anomalous data from the first measurement of mutation rates at the single mutation resolution, regarding what has been until now a quintessential textbook example of random mutation—the HbS mutation (Freeman and Herron, 1998; Hartl and Clark, 2007).

Another possibility is that the HbS mutation originates as a direct response to the immediate environment (Luria and Delbrück, 1943; Cairns et al., 1988), in accord with Lamarckism (Sarkar, 1990). However, given the apparent limitations of Lamarckism (Haig, 2007), we find it unlikely.

A third possibility is that the HbS mutation originates in a manner that is neither accidental nor Lamarckian and instead demonstrates a long-term directed mutational response to a specific environmental pressure (Livnat, 2013, 2017). However, if correct, then how could the genome "know" to generate a mutation more frequently where it is of adaptive significance? Our goal in this paper will be to propose an outline of an answer that is not susceptible to the difficulties of modifier theory, random mutation and Lamarckism.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

2 The used-fused hypothesis for nonrandom gene fusion mutations

We begin with an example that may at first seem unrelated, but later will be seen as directly relevant. The TRIM5 and cyclophilin A encoding genes fused at least twice independently by retroposition in at least two different simian lineages (Virgen et al., 2008; Nisole et al., 2004; Sayah et al., 2004; Liao et al., 2007; Brennan et al., 2008; Wilson et al., 2008; Newman et al., 2008), producing a fusion protein (TRIM5-CypA) that protects against certain lentiviruses, including HIV-1 (Nisole et al., 2004; Sayah et al., 2004). Compared to the independent recurrence of the same point mutation, it is very difficult to attribute the independent recurrence of such a gene fusion mutation to chance. In the latter, not one point mutation but multiple similar breakpoints fusing the same two loci must appear independently by chance twice, and mathematically, if the probability of the former is small, that of the latter is negligible.

However, genes that are used together are more likely to be transcribed at the same time and place in the nucleus—for example in co-expression domains (Soler-Oliva et al., 2017), topologically associated domains (Dixon et al., 2012) and transcription factories (Jackson et al., 1993; Edelman and Fraser, 2012; Papantonis and Cook, 2013) where DNA loops bring also remote coactive genes together. Therefore, to explain gene fusion, we argued that this presence of the two co-active genes at the same time and place, with the chromatin open at both loci due to transcription, enables various downstream mechanisms, such as reverse transcription of the RNA (perhaps aided by trans-splicing), transposable-element mediated translocation, recombination-based mechanisms, DNA breaks induced by the spatial chromosomal organization and active gene transcription, and other mechanisms to generate a gene fusion (Livnat, 2017; Bolotin et al., 2021). We furthermore hypothesized that because the genetic information that indicates that the two genes work together, such as shared cis elements and transcription factors that bind to them, is present in the DNA and accessible in the germline, this fusion effect applies not only to pairs of genes that serve germline functions but also to pairs that serve somatic functions (Livnat, 2017; Bolotin et al., 2021). Indeed, the germline-specific phenomenon of transcriptional promiscuity allows many somatic genes to be regularly transcribed in the germline (Kleene,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

2005; Melé et al., 2015; Xia et al., 2020) and thus to participate in mutational mechanisms involving interactions between genes (Livnat, 2013, 2017). Thus, we hypothesized that, in the course of evolution, genes that are used together repeatedly and persistently in a certain context are incomparably more likely than other genes to be fused together by a mutational mechanism (Livnat, 2017; Bolotin et al., 2021): it is genes that are used together that can be fused together—henceforth the "used-fused" effect. This hypothesis offered the first explanation for why there are recurrent gene fusions, and why they are common, both in evolution (Carvalho et al., 2010; Livnat, 2013) and in genetic disease and cancer (Li et al., 2008; Osborne, 2014). Recently, we found empirical support for this hypothesis, including its applicability to both germline and somatic genes (Bolotin et al., 2021).

Several consequences of this used-fused effect are directly relevant to the central question raised earlier regarding long-term functionally directed mutations. First, in contrast to modifier theory, where the average mutation rates across many loci are presumed to be affected in a blanket manner (Walsh and Lynch, 2018), the genetic interaction that leads to a fusion mutation via the used-fused effect increases the fusion probability specifically of the two genes that interact. Also in contrast to modifier theory, where the mutation rate across many loci can be attributed to a single controlling locus (Walsh and Lynch, 2018), the genetic information that causes the increase in the rate of a particular fusion mutation is complex, as it involves all of the information that makes these two genes interact tightly, such as promoters, enhancers, transcription factors and epigenetic marks of at least two loci.

Second, before the fusion, the two genes had to be activated separately and their products had to meet, whereas the fusion has chunked them together into a single unit that can be activated as one. Therefore, the outcome of the fusion can be seen as a local mutational simplification of gene regulation (Livnat, 2017; Bolotin et al., 2021).

Third, because gene fusion is a discrete event, it may have seemed as though it must arise by a sudden evolutionary "jump" where a sequence is randomly translocated from one context to another. However, according to the used-fused hypothesis, the genetic interaction that leads to a fusion mutation has already evolved and repeatedly occurred through the generations. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

other words, two genes first come to interact with each other tightly in the course of evolution through the gradual accumulation of multiple other heritable changes of smaller effect in various loci, which in turn leads to a situation where their fusion becomes likely to arise (Livnat, 2017; Bolotin et al., 2021). A gradual, long-term process of evolution of regulation can pave the way to a discrete mutational change (Livnat, 2013, 2017; Bolotin et al., 2021; Melamed et al., 2022).

Fourth, although it leads to local simplification, the used-fused effect does not lead to ever more simplicity and diminution of the genome. Fusion often comes together with or is preceded by gene duplication of the source copies, and therefore often leads not from two genes to one, but from two genes to three. In the long term, this increases the genetic vocabulary and the extent of interactions between genes, and thus it increases complexity (Livnat, 2017; Bolotin et al., 2021).

According to the above, the used-fused effect replaces a phenotype—an interaction between genes—with an innate, ready-made object—a gene (Livnat, 2017). In addition, the above demonstrates that mutation rates can respond in a mutation-specific manner to complex information in the genome that has been accumulated as a result of past evolution (Livnat, 2013, 2017); that a gradual evolutionary process can lead to a discrete mutational change (Livnat, 2013, 2017; Bolotin et al., 2021); that mutational phenomena can result in local genetic regulatory simplification (Livnat, 2017; Bolotin et al., 2021); and that, in the context of gene duplication, local simplification can result in a global increase in complexity (Livnat, 2017; Bolotin et al., 2021).

3 The replacement hypothesis is an extension of the used-fused hypothesis

The used-fused effect is one of multiple examples of a broader hypothesis, called "the replacement hypothesis," that an evolved, system-level phenomenon can lead directly and mechanistically to a new and specific mutation that replaces and simplifies this pre-evolved phenomenon (Melamed et al., 2022). We elaborate on this hypothesis below through examples.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

3.1 Gene duplication as mutational replacement

The mutational phenomena that enable whole gene duplication (e.g., non-allelic homologous recombination; Gu et al., 2008) are essential for long-term evolution. They can generate two functional gene copies from one, thus enabling the copies to specialize by the gradual accumulation of mutations under selection (Ohno, 1970). Here we hypothesize that these mutational phenomena could be even more useful in the long term than previously considered.

Several findings show that elevated transcription increases the chance of a gene duplication mutation (Hull et al., 2017; Wilson et al., 2015; Thomas and Rothstein, 1989; Aguilera and Gaillard, 2014; Hamperl et al., 2017), likely by local destabilization of the genome (Aguilera, 2002; Gómez-González and Aguilera, 2019) followed by downstream mechanisms involving RNA (Kaessmann et al., 2009) or DNA intermediates (Chicote et al., 2020; Durkin et al., 2012; Takahashi and Innan, 2020). A direct connection between transcription and gene duplication mutation rates was exemplified in yeast, where copy number amplification of the CUP1 gene results in adaptation to high copper concentration: by replacing the CUP1 promoter sequence with a galactose-inducible promoter, Hull et al. (2017) have shown that CUP1 transcription induces gene copy number amplification under galactose stimulation and in the absence of copper selection, suggesting that copy number amplification of the wild-type CUP1 can be caused mechanistically by transcription. The mechanism likely involves transcription-inducible formation, accumulation and reintegration of extrachromosomal circular DNA (eccDNA) (Hull et al., 2017).

How shall we interpret this mechanistic link between elevated transcription and gene duplication mutation? Examining it from the random mutation perspective, it is merely one phenomenon untied to others. Examining it from a Lamarckian perspective, it is limited in its ability, as it can facilitate evolutionary adaptation mainly in unicellular organisms in certain situations. However, Livnat (2017) has argued that heritable change in general is neither accidental nor Lamarckian: instead of only responding directly to an environmental cue, regulatory activity that has gradually evolved through the generations and has come to cause elevated transcription can increase the probability of a gene duplication mutation in the germline. Then,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

the gene duplication mutation replaces the previous regulatory activity with an innate ability to produce a large amount of the gene product (Livnat, 2017).

As in the case of the used-fused mechanism, it has been argued that this elevated-transcription–based gene duplication mechanism applies not only to genes that serve germline functions but also to genes that serve somatic functions, due to germline expression of somatic genes, without a Lamarckian transmission of information from the soma to the germline (Livnat, 2017). Evidence exists consistent with the operation of the focal mechanism in sperm (Vinckenbosch et al., 2006; Park et al., 2012; Mouakkad-Montoya et al., 2021). Also as in the case of gene fusion, this mechanism may facilitate long-term evolution: where high quantities of the gene product are needed, it could alleviate the pressure on gene expression, and where usage of a gene by multiple processes leads to its high expression, duplication and specialization of it into two or more paralogs could facilitate evolution by resolving pleiotropy (Livnat, 2017).

Because this gene duplication mechanism replaces pre-existing regulatory activity with an innate ability to produce a large amount of the product, it is an example of the replacement hypothesis; its action in general is not an immediate and direct response to the environment, and it facilitates both unicellular and multicellular organism evolution.

3.2 Mutational replacement of RNA editing

Following the common posttranscriptional regulatory mechanism carried out by members of the adenosine deaminase acting on RNA family (ADARs)—A-to-I RNA editing—the inosine (I) is usually recognized as guanine (G) during mRNA translation (Bass, 2002). Comparing RNA A-to-I edited sites to nonedited sites in one species, the former are more likely to correspond to sites in other species where A→G DNA substitutions occurred (Grauso et al., 2002; Ohlson et al., 2007; Tian et al., 2008; Xu and Zhang, 2014; Chen et al., 2014), and are also more likely to exhibit A/G DNA polymorphisms in the same species (Popitsch et al., 2020; Danecek et al., 2012). In *Drosophila melanogaster*, for example, A/G DNA polymorphisms are approximately twice more common in A-to-I edited than non-edited sites, and the polymorphisms almost always (98%) feature G rather than C or T (Popitsch et al., 2020). Interpreting these data from the random mutation perspective, some authors suggested that A-to-I editing is a rescuing mecha-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

nism from past, deleterious G→A substitutions (Pinto et al., 2014); others suggested that both A-to-I RNA editing and A→G DNA mutations largely represent promiscuous, erroneous activity that is only slightly deleterious, and therefore in sites where nonfunctional A-to-I RNA editing is tolerated by selection, random A→G mutations are also more likely to be tolerated by selection (Xu and Zhang, 2014); and yet others suggested that selection favors G in the RNA edited sites, and therefore random A→G mutations will be favored by selection there too (Popitsch et al., 2020). Thus, different authors have been led to contradictory conclusions.

In contrast, we have argued that the evolutionary increase through the generations in the rate of RNA editing directly and mechanistically generates an increase in the rate of the corresponding DNA mutation in the corresponding positions (Melamed et al., 2022). Indeed, evidence has been accumulated suggesting a mechanistic connection between RNA editing and DNA mutation either by ADAR acting directly to mutate transcribed DNA (Shiromoto et al., 2021; Jimeno et al., 2021; Zheng et al., 2017; Tsuruoka et al., 2013) or via reverse transcriptase activity of several DNA polymerases (Chandramouly et al., 2021; Su et al., 2019; Franklin et al., 2004) using edited RNA as a template. [1]

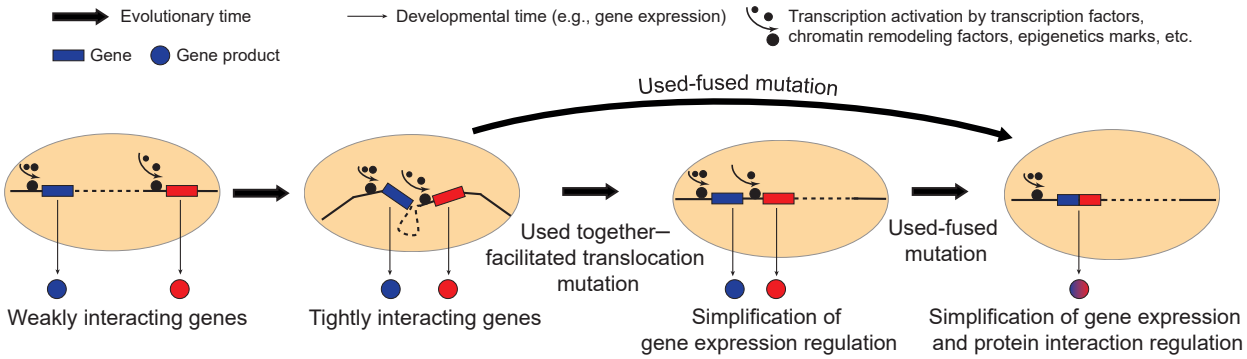
The fact that the correspondence between A-to-I RNA editing on the one hand and A→G DNA substitutions as well as A/G DNA polymorphisms on the other has been observed for both coding and noncoding regions as well as for both synonymous and nonsynonymous A→G changes (Popitsch et al., 2020) is difficult to explain by random mutation and natural selection (rm/ns[2]). In contrast, we argue that the following of A-to-I RNA editing in evolutionary time by A→G mutations in the corresponding DNA positions is not due to rm/ns but is another example of the mutational replacement hypothesis. Following such replacement, regulation at the phenotypic level is no longer needed to obtain the A-to-G outcome, and thus the replacement can be seen as a form of local simplification. Thus, gene fusion, gene duplication, A→G mutations in RNA-edited sites and other mutations mentioned later all fit under a unifying umbrella, that of the replacement hypothesis (Figure 1).

[1] Indications exist that the DNA mutation rate due to RNA editing enzymes, though lower than the RNA editing rate, is still much higher than the human genome-wide average mutation rate (Saraconi et al., 2014; Tasakis et al., 2020).

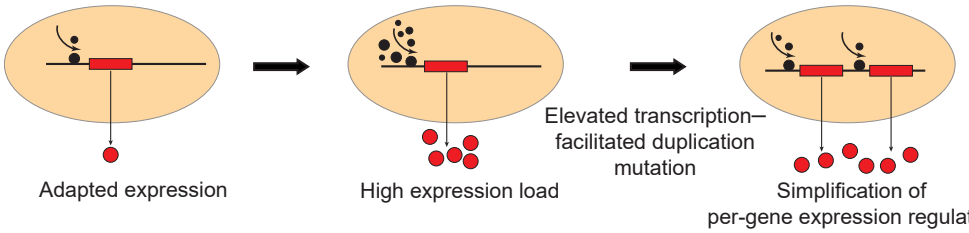
[2] Hereafter, we will denote by rm/ns the view that evolution happens by random mutation and natural selection, including the subsidiary effects of random genetic drift, sexual recombination and migration (Futuyma, 1998).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

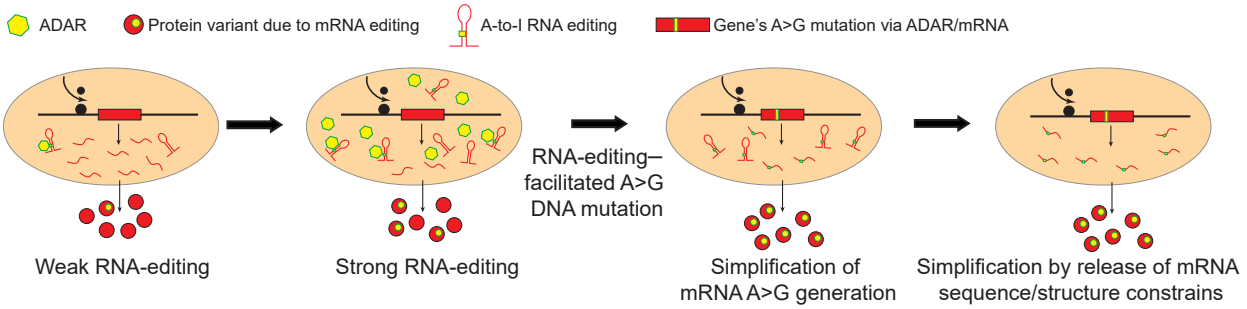
A) Used-fused mechanism of gene fusion



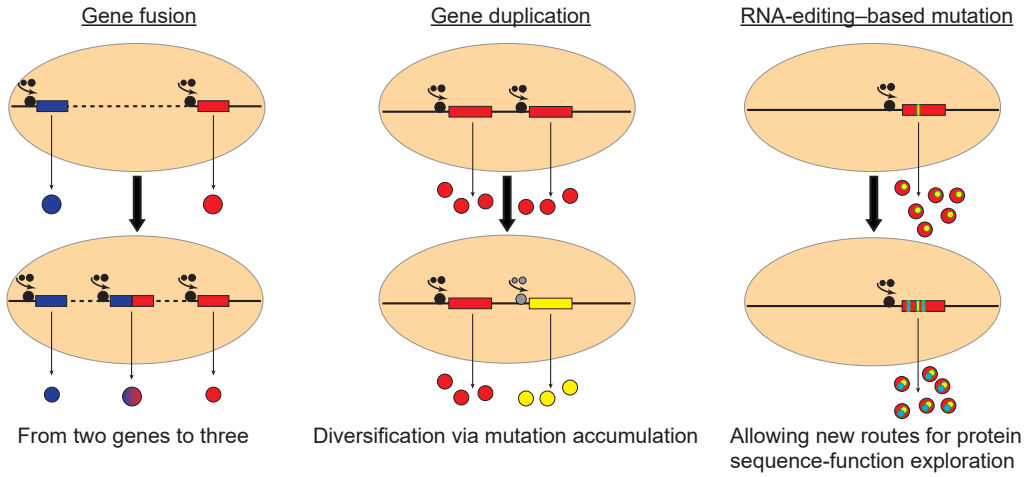
B) Elevated transcription–based gene duplication



C) RNA-editing–based mutation



D) Local simplification leads to a global increase in complexity



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

Figure 1: Gene fusion, gene duplication and RNA-editing–based mutations as examples of the replacement hypothesis. *a*) The gradual evolution of tight interaction between two genes precedes their translocation or fusion. This interaction brings together the two loci with their chromatin open at the same time and place in the nucleus in co-expression domains, topologically associated domains or transcription factories, thus enabling a translocation mutation that makes them neighbors or a fusion mutation that makes them into one gene by various downstream mutational mechanisms. The translocation or fusion mutations obviate preexisting regulatory activity that was needed to bring the two genes together, thus replacing it with a simplified, hardwired, innate state. *b*) Increased expression beyond the norm for a certain gene makes it more likely to undergo a duplication mutation. For CUP1 in yeast, the generation of eccDNA and its reintegration into the genome was implicated. The duplication mutation obviates the need for excessive transcription through regulation, replacing it with a locally simplified, hardwired, innate state. *c*) The evolution of A-to-G RNA editing of a certain site comes together with an increase in the rate of an A→G mutation at the same position in the DNA. This mutation obviates the need for the RNA editing regulatory activity, thus replacing it with a simplified, hardwired, innate state. *d*) According to the theory of Interaction-based Evolution (IBE), the simplification of interactions generates elements that engage in new interactions with other such elements at the system level. Thus, in the context of gene duplication, local simplification leads to a global increase in complexity: the fusion of genes leads from two genes to three, thus increasing the number of genes, the overall extent of interactions between genes, and complexity (left); in the long-term, gene duplication enables the two copies to undergo differential accumulation of mutations, leading to specialization and complexity (center); and the accumulation of point mutations is a part of the evolution of complexity (right).

4 A system-level view of mutation origination

The full meaning of the replacement hypothesis can be understood in the broader context of the theory of Interaction-based Evolution (IBE) (Livnat, 2013, 2017). According to IBE, the heritable changes that drive adaptive evolution under selection are not accidental, because the probability of origination of each heritable change depends on complex information in the genome (i.e., genetic interactions) in a biologically meaningful manner. In addition, they are not Lamarckian, because the complex information that influences them does not come in general directly from the immediate environment but is accumulated in the germline genome through the generations. In turn, this information in the genome that the origination of heritable changes is influenced by has come from previous heritable changes and natural selection (Livnat, 2013, 2017). As a consequence, evolution is driven by the interaction of two sources of information: an external source that is differential survival and reproduction (natural selection), and an internal source that is complex information influencing heritable change in a manner specific to each

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

change, both of which are continually updated: As the organism gradually changes through the generations, so does the selection pressure change, because it depends on the organism and its fit with the environment. As the genome gradually changes, so do the mutation-specific probabilities of mutation origination change, because they respond to the genome in a complex manner. Thus, both sources continually interact as they coevolve, and their interaction drives evolution (Livnat, 2013, 2017).

Demonstrating these principles, according to the replacement hypothesis, the origination of gene fusion, gene duplication and the predicted RNA-editing–based mutations is not due to mere chance unrelated to the biology of the organism, but instead responds in a mechanistic, systematic and mutation-specific manner to phenomena carrying biological information, such as genes being used together, genes being expressed excessively, and certain nucleotides being edited recurrently at the RNA level. In addition, it does not respond in general directly to the environment but to information that has been accumulated in the genome over time—information which is itself not random but reflects the current biological structure and function of the organism, as accumulated under previous nonaccidental heritable changes and selection.

To better understand the meaning of nonaccidental, non-Lamarckian heritable change, one may expand the purview to include the problem of sexual reproduction—"the queen of problems in evolutionary biology" (Bell, 1982). The early 20th century reconciliation of biometry and Mendelism based on the mathematical tools available then (Fisher, 1930) established a theoretical foundation that encouraged conceptualization of genes as separate actors, where selection acts on separate contributions of genes to fitness as opposed to complex wholes (Ewens, 2004; criticized in Wade and Goodnight, 1998; Noble, 2011). This both fit with the notion that genetic change can be treated as a single-locus accidental event—namely the notion of random mutation, which rose to prominence at the same time (Morgan, 1903; Fisher, 1930) [3]—and created the modern problem of the role of sex in evolution (Livnat, 2013): It is often said that sex generates a vast number of different combinations of genes, and because genetic variation is the fuel for natural selection, these combinations facilitate adaptive evolution. However, what is the point of subjecting so many different complex combinations of genes to the test of selec-

[3]The notion of random mutation is not Darwin's own (cf. Darwin, 1859, Ch. 5)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

tion, when just as sex puts them together, it also breaks them down in the next generation? The intuition that the combinations are important has been incomplete.

However, according to IBE, when heritable changes respond to complex information in the genome, even though the combinations themselves disappear, system-level information is transmitted from them to future generations through the heritable changes that are derived from them (Figure 2). This addresses from a unified framework fundamental problems hitherto considered separate—What is the role of sex in evolution? How can selection act effectively on individuals as complex wholes? And what is the fundamental nature of mutation?—by enabling a framework where the basic elements of evolution interact: sexual recombination generates a large number of different combinations of genes; selection acts on these combinations as complex wholes; and heritable changes that are influenced by complex information in the genome transmit information from these combinations to future generations. This information is transmitted precisely in accord with the fitness of the organism as a complex whole, and no Lamarckian transmission of information from soma to germline is involved (Figure 2).

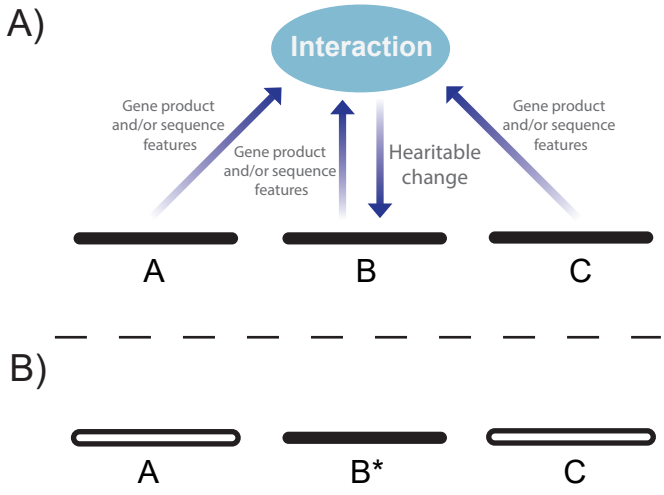


Figure 2: Heritable change puts together information from multiple loci, modified from Livnat (2013). *a*) Without the downward arrow, the figure merely summarizes in schematic form an essential, well-known part of the molecular and cellular biology textbook: genes interact to affect classical traits, like the eye or the ear. However, evidence suggests that there is a heritable-change arrow too: that genes interact in affecting heritable change, both genetic and epigenetic. Skipping the molecular detail, what this means at the high level is that information

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

is transmitted from the multiple interacting alleles at different loci into the locus being changed. This generates from the combination of interacting alleles a new heritable piece of information, such as a new mutation at locus B, denoted B*. This new allele is an elementary unit to the sexual shuffling of the genes—it is not in itself broken down by sexual recombination. Therefore, even if the alleles denoted by black lines at loci A, B and C separate due to meiosis (*b*), the transient, complex whole they constituted does have an effect on future generations through the heritable change that was derived from it. This information is transmitted precisely in accord with the fitness of the organism as a complex whole because transmission depends on survival and reproduction. Only three loci are presented for the sake of simplicity, though in reality many more could affect the interaction.

4.1 A network of information flow

Because each heritable change that is transmitted becomes a part of the complex information in the genome that affects future heritable changes in a manner specific to each change, there is a network of information flow via heritable changes through the genome and the generations: information flows from many genes into any one gene, and from many ancestors into any one descendant (Figure 3).

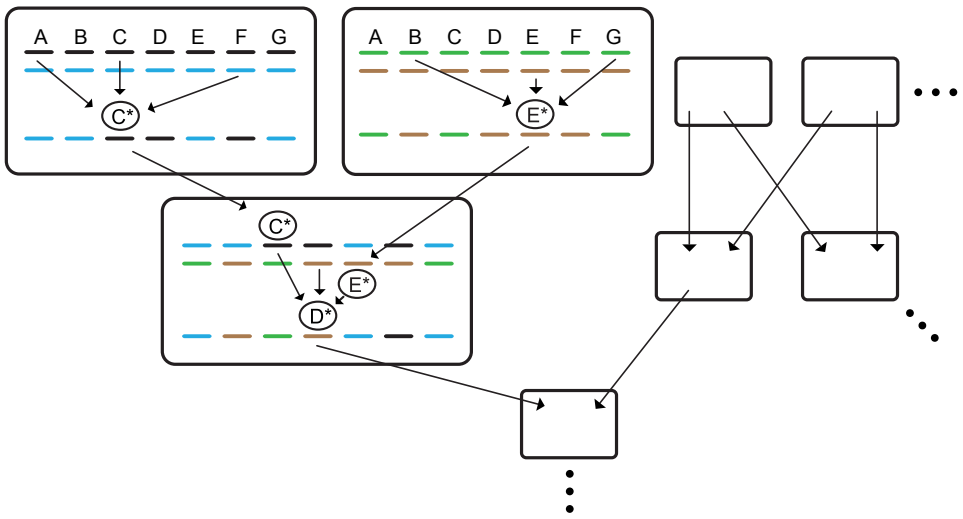


Figure 3: A network of information flow through the genome and generations, from Livnat (2013). Boxes represent individuals. In each box, the top lines represent the genome of the individual and the bottom lines the genome of a gamete. The three large boxes represent two parents and an offspring. Arrows represent information that affects heritable change. When complex information in the genome, genetic and/or epigenetic, affects the probability of heritable change in a change-specific manner, the outcome of an operation generating heritable change in one generation (e.g., C*) affects the outcomes of heritable change operations in future generations (e.g., D*). Therefore, there is a network of information flow via heritable changes through

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

the genome and generations. Even though the genome of each individual is transient, system-level information is transmitted from it to future generations. As a result, sexual recombination, natural selection and heritable changes that respond to complex information in the genome together combine system-level information from many successful individuals over the generations. For simplicity, only one heritable change per individual is shown, though in reality many exist, considering both genetic and epigenetic changes.

Networks of information flow are a deep and fundamental part of nature and key to natural phenomena across the scales of organization, including social insects, the living cell and the brain. As the neuron, by taking a transient combination of input signals from multiple upstream units, operating on them, and transmitting the outcome to downstream units, ties the network together, so heritable changes that are influenced by complex information in the genome, together with sexual recombination, tie heritable changes together into a network. However, while the aforementioned networks operate within the lifetime of an individual, the network of interrelated heritable changes exists on a far larger scale of space and time. According to IBE, a vast array of biological activity affects the probabilities of heritable changes across the genome in each germ cell that passes to the next generation. Thus, the information processing that leads to the evolution of complex adaptations over evolutionary time can involve such cells in each of billions of individuals, each evaluated as a complex whole by natural selection, where the informational changes in each such cell are connected to those of others from past to future. To borrow terms from the mathematical study of networks (Pippenger, 1979), the evolutionary process is wide in space and deep in time, putting together system-level information from many successful individuals over the generations via heritable change operations.

4.2 Principles of information acquisition are apparent in the nature of mutation

This system-level view allows us to examine the mutational mechanisms discussed earlier from a new angle. According to Hebbian learning, when one neuron repeatedly and persistently participates in causing another neuron to fire, the synaptic connection between these two neurons is strengthened (Hebb, 1949). Thus, it has been said that "neurons that fire together wire together" (Löwel and Singer, 1992). Likewise, on the macroscale of brain operation, when actions or pieces of information are repeatedly and consistently used together in a certain context,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

they are fused together or "routinized" into a new action or unit that can be activated or recalled as one—a fundamental principle of cognition and learning called "chunking" (Lindley, 1966; Tulving and Craik, 2005). Given that genes that are used together are mechanistically fused together, a similar principle appears in genetic evolution. According to IBE, the fusion of units that commonly work together is a fundamental principle of learning, and that is why it appears in different realms of information acquisition.

Likewise, evolution as we know it would not have been possible without mutational mechanisms enabling whole-gene duplication. According to the replacement hypothesis, these mechanisms are not activated accidentally with respect to the process of adaptive evolution. Elevated expression pushes genes directly and mechanistically to undergo duplication mutation. Duplication is a critical basic element in other processes of information acquisition (Newell, 1980).

4.3 Evolutionary honing in: a gradual evolutionary process leads to punctuated mutational change

It is empirically clear that heritable changes influence one another over the generations, and that they do so across a spectrum of change types, from commonly occuring, high turnover ones like epigenetic changes, to rare, low turnover ones like large rearrangement mutations (Livnat, 2013, 2017). While gene duplication is mechanistically influenced by the locations of segmental duplications or low copy repeats (SDs/LCRs) (Gu et al., 2008), the generation of SDs/LCRs is influenced by the locations of transposable elements (TEs) (Bailey et al., 2003), the movement of TEs is influenced by sequence characteristics (Graur and Li, 2000), sequence characteristics are influenced by point mutations, and point mutations are influenced by epigenetic changes (Fryxell and Moon, 2005; Qu et al., 2012). While the interaction between genes is affected by chromatin states, epigenetic marks, promoters, enhancers and transcription factors, Bolotin et al. (2021) hypothesized that the interaction between two remote genes makes them more likely to undergo a translocation mutation that brings them to the same neighborhood, and that the interaction between neighboring genes makes them more likely to undergo a fusion mutation. Regarding RNA-editing, heritable changes that affect the RNA folded structure can expose a new effective A-to-I RNA editing site (Gommans et al., 2009), with further heritable

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

changes contributing to RNA stabilization and honing the target site, leading to increased levels of A-to-I editing that can facilitate an A→G mutation according to the replacement hypothesis.

That heritable changes are connected across this spectrum of change types helps to explain the otherwise disruptive, low turnover changes. According to the replacement hypothesis, the evolutionary process based on nonaccidental heritable changes gradually hones in in the long term on genomic regions, specific positions in those regions and specific changes in those positions that are of particular relevance to the currently evolving adaptations. Thus, the accumulation across loci of frequent heritable changes of minor effect under selection leads in the long-term to the increase in the rates of rare, large effect mutations that are more relevant under the current selection pressure and statistically less disruptive than other mutations that could have occurred.

This also clarifies the role of epigenetics in evolution from the perspective of IBE: According to IBE, the general role of epigenetic changes in evolution is not to enable Lamarckian transmission of information from soma to germline (Jablonka and Lamb, 2014; Bonduriansky and Day, 2018) but is in being an important part of the spectrum of heritable change on the high turnover side that meshes with and influences the other, less frequent changes on the intergenerational timescale. Since epigenetic marks often operate en-masse (Duret and Galtier, 2009), and given their nonaccidental, generally non-Lamarckian nature, they enable a connection between the gradual evolutionary process of their accumulation and the more punctuated DNA mutational changes.

4.4 The ultimate source of heritable novelty

IBE attributes heritable novelty in evolution not to random mutation but to emergent interactions between heritable changes (Livnat, 2017). In that, it accords with previous views on the importance of interactions to novelty in evolution (e.g., Müller and Wagner, 1991; Hallgrímsson et al., 2005; Schlosser, 2015; Moczek, 2008). However, it adds that these interactions are simplified by evolution in the long term into new elements through evolutionary honing in and mutational replacement (Livnat, 2017). At the same time, it offers a way of understanding how novel interactions emerge at the system level in the first place: While mutational replacement, which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

operates internally, brings about local simplification, because of natural selection, which operates externally at the same time, the internal organization can only be simplified to the extent that it keeps working (Livnat, 2017). This mutational simplification under the performance pressure of natural selection—related to what is known in other realms of information acquisition as the combination of parsimony and fit—has immense power. According to IBE, it generates from preexisting interactions new elements that not only work better but also have the inherent capacity to come together into novel interactions at the system level (Livnat, 2017): what both is simple and performs, generalizes. Thus, novelty arises from a cycle, where interactions are simplified into elements that engage in novel interactions at the system level—i.e., where local simplification leads to a global increase in complexity (Livnat, 2017). As an example, elevated-transcription–based gene duplication, together with used-fused and other mutational mechanisms, can lead to the increase in the number of genes, the extent of interactions between genes, and biological complexity (Livnat, 2017; Bolotin et al., 2021).

This also offers an explanation for the ubiquitous tendency for cooption in evolution—the phenomenon whereby an element that serves one function gradually comes to serve another by changing its context of interaction with other elements (Gould and Vrba, 1982; Hallgrímsson et al., 2012; Müller and Wagner, 1991; Graur and Li, 2000). According to IBE, the causes of cooptability and of heritable changes are related: mechanisms of heritable change enable simplification, and the combined pressures of simplification and performance underlies the inherent capacity of elements to come together into novel interactions at the system level (Livnat, 2017).

That the genotype influences the origin of mutation supports the importance of the exploration, pursued in evo-devo, of how the phenotype influences the consequences of mutation (Hallgrímsson et al., 2005; Müller, 2007; Wagner, 2014): according to IBE, mechanisms operating in developmental time and in the germ cells partly overlap, affecting and thus tying together both the causes and the consequences of heritable change, as demonstrated in Figure 1 and in the subsection below on the connection between alternative splicing and exon shuffling via mutational replacement.

From rm/ns, cases such as the HbS mutation one fostered a reductionist view, where all that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

was needed was for random mutation to generate the A→T change and then selection has done all the rest. In contrast, we argue that traits arise from interactions of preexisting elements at the system level, and that this principle applies also to the HbS mutation: the HbS mutation did not arise accidentally and did not initiate a process of adaptation but rather arose from preexisting interactions that resulted from a long-term evolutionary process.

5 How to explain the HbS de novo origination patterns

Earlier in this paper we raised the question of how the genome could "know" to increase the rate of the HbS mutation in the gene and in the population where it is of adaptive significance. We then proposed, and demonstrated with examples, that various types of mutations relevant for adaptive evolution under selection could arise mechanistically and directly from previously evolved interactions. Livnat (2013) furthermore hypothesized that the gradual gathering of system-level information from many successful individuals over the generations by the interaction of sexual recombination, nonaccidental mutation and natural selection allows mechanisms of heritable change to converge on the commonality between successful individuals and thus on genetic interactions relevant to the currently evolving adaptations. Thus, from the accumulation of heritable changes of minor interactive effects across loci through the generations, heritable changes of major effect follow mechanistically and directly at relevant base positions and genes (Livnat, 2013). This view connects between the evolution of regulation, considered a rapid and flexible process (Carroll, 2008; Wagner and Lynch, 2010; Jones et al., 2012; Fraser, 2013; Gokhman et al., 2021; Agoglia et al., 2021), and structural mutational changes.

Specifically in the case of malaria, many regulatory and coding regions affect directly and indirectly the within-host environment that the parasite encounters. Thus, we hypothesize in broad outline that small phenotypic variation in the ability to resist malaria that was due to many different genetic causes and had a complex genetic basis was initially present in the population as a result of other evolutionary processes, and that from this variation, a gradual evolutionary process based on the principles of IBE honed in on changes in particularly relevant positions in the hemoglobin and other genes, including the HbS mutation in sub-Saharan Africans.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

While the general process responsible has been outlined above, how it may apply in detail to the HbS mutation remains to be uncovered, and below we provide some observations to help propel the quest for the missing detail. Focusing on hemoglobin genes in saker falcons, Pan et al. (2018) observed a correlation between the expression levels of genes and their mutation rates in blood samples, along with characteristics of transcription-associated mutations (TAM). These single strand mutations become fixed as de novo mutations in the daughter cells after DNA replication during erythropoiesis, giving rise to numerous mutation variants in the soma and to an increased diversity of alternatively spliced mRNA variants due to de novo splice sites (Pan et al., 2018). They argued that this effect is strong in the hemoglobin genes because they are highly expressed and even stronger in a falcon population living at a high altitude, where there is increased oxygen demand. They also found that more than 80% of the hemoglobin mutations were to T, and that the most common mutation type was A→T, on the coding strand (Pan et al., 2018). These observations are of interest given that the HbS mutation is an A→T one and that it was the point mutation of highest de novo rate in the sub-Saharan African hemoglobin subunit beta (*HBB*) gene in the HbS de novo mutation study (Melamed et al., 2022). This raises the possibility that also in humans, a greater amount of relevant RNA diversity may have evolved in African populations that have been subject to intense malaria selection pressures compared to northern European populations, and that the HbS mutation arises as a type of replacement following evolved complex genetic interactions that have not yet been charted.

Importantly, mutation rates in Pan et al's study were more complex than simply being based on transcription. They were lower in gene bodies in methylated regions, highly nonuniform across positions, and appeared particularly high ($\sim 10^{-4}$) at de novo splice sites based on the observation of many spliced variants among small cDNA samples (e.g., Pan et al., 2018, p. 1110). These observations appear to be more consistent with regulatory processes than with the errant, scattered genetic change implied by the random mutation concept. Thus, we hypothesize in broad outline that a complex set of pre-evolved phenomena may be increasing the rates of specific hemoglobin mutations in specific human populations. A DNA hairpin structure due to a local palindrome at the site of the HbS mutation (Alvarez-Dominguez et al., 2013)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

may also be involved in the specificity of the mutations generated in the *HBB* site, although nonlocal information must exist that interacts with these features (Livnat, 2013) in order to explain the observed difference in the HbS mutation rate between the African and northern-European populations (Melamed et al., 2022).

Mutational replacement involving TAM need not involve direct transmission of information from soma to germline, as in the previous cases of replacement discussed. Gradual increases in the rates of specific genetic changes due to complex interactions in the soma, accepted by selection over the generations, may come together with smaller gradual increases of the corresponding mutation-specific rates in the germline due to partly overlapping mutational mechanisms, leading to replacement and simplification. Consistently, Park et al. (2012) noted a significant correspondence between germline TAMs and somatic gene expression. Also as in the previous cases, potential relevance of TAM-involved replacement for facilitating evolution can be seen: transcription beyond the norm for a certain gene may indicate instability following a recent environmental pressure that the organism has not yet evolved to fully counter, and further changes in such a gene may be more likely to be adaptively relevant. Additional factors may focus mutational change on particular mutations and positions within such a gene.

At some point in the evolutionary past, however, there may have been no 20A→T somatic DNA or a 20A-to-U RNA change, even if such a change later appeared and was replaced by the HbS mutation. Therefore, we note that cooption both increases the range of possibilities for replacement and is needed for evolutionary novelty: the ultimate origins of a 6 Glu→Val change could be in cooption of a different but related change. Pre- and post-cooption traits are generally related in their biological meaning (Livnat, 2017), enabling a gradual evolutionary process where interactions that evolve and repeatedly occur over the generations lead directly and mechanistically to heritable changes that may carry some but not necessarily all of the meaning that the previous interactions had and that may take the latter's place.

Thus, while the molecular details of what causes the de novo HbS mutation patterns are purely speculative, of first and foremost importance is the system-level view of mutation origination that is consistent with these patterns. What the details are of the HbS mutation mech-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

anism is not the final question: There is no "homunculus" in the genome that directs different mutations adaptively, and no one mutational mechanism that gives rise to all mutations. According to the system-level view, even just the HbS mutation alone arose after a series of heritable changes, each of which originated due to its own proximal complex causes. There may be an enormous array of interrelated mutational phenomena, themselves continually evolving, of which only a small amount may be known at the present time.

6 Consequences of the system-level view of mutation origination

The explanatory power of the replacement hypothesis affects diverse topics, such as directed mutation, the evolution of genome organization, parallel evolution and the contribution of transposable elements to the evolution of gene regulatory networks.

6.1 Long-term directed mutational responses to specific environmental pressures are possible

Empirical data on the nature of mutation has been collected and interpreted through the lens of the traditional dichotomy between random mutation and Lamarckism. Therefore, the lack of evidence for Lamarckian mutation in Luria and Delbrück's (1943) fluctuation test has been taken as an empirical proof of the random, accidental mutation concept (Futuyma, 1998, p. 282; Lenski and Mittler, 1993). However, according to IBE, the conceptual space of possibilities is greater than that dichotomy.

According to Lamarckism, a direct phenotypic response to an environmental change that occurs within the lifetime of the organism can induce beneficial heritable change in a manner that circumvents natural selection. An example would be a unicellular organism sensing its environment and producing beneficial heritable change in response to that sensing (Shapiro, 2011, pp. 7-12, 137, 143-144). In contrast, IBE holds that it is differential survival and reproduction that provides feedback on the fit between the organism and its environment. However, this differential survival and reproduction is based on heritable changes that respond to complex information in the genome. Because this information is accumulated in the genome over the generations,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

long-term directed mutational responses to specific environmental pressures are possible (Livnat, 2013)[4]. That is, while genetic and epigenetic heritable changes occur at each generation, large-effect, easily observable ones particularly relevant to the new environmental pressure may take multiple generations to arise via evolutionary honing in and mutational replacement.

Therefore, whereas previous studies looked for an immediate, directed mutational response to an environmental challenge (Luria and Delbrück, 1943; Cairns et al., 1988), IBE led us to compare the HbS mutation rate between populations that had been subject to different malarial selection pressures for many generations. Consistently, results showed that the rate was higher both in the gene (comparing 20A→T in *HBB* and *HBD*) and in the population where HbS is of adaptive significance (Melamed et al., 2022). We expect that further studies of mutation rates in nature will be able to generalize our HbS results to other genes and organisms, and that future experimental evolution studies designed according to the principles of IBE (e.g., multigenerational selection pressure) will be able to recapitulate our observations using various model organisms and target genes.

6.2 Mutational replacement and the evolution of genome organization

Gilbert (1978) hypothesized that the intron-exon structure of eukaryotic genes facilitates exon shuffling. From the perspective of rm/ns, this hypothesis could be interpreted to mean that the existence of introns reduces the chance that a random mutation will translocate one exon into another and disrupt the latter, and increases the chance that the randomly translocated exon will fall between other exons and form a new and useful combination of exons there. However, according to the replacement hypothesis, exon shuffling is not driven by accidental mutation: Through mutational replacement, an interaction of exons from afar can lead directly to a translocation which turns those exons into neighbors, and an interaction between neighboring exons can lead directly to a DNA fusion mutation (Livnat, 2017; Bolotin et al., 2021). Furthermore, trans-splicing can lead directly to cis-splicing, and cis-splicing to fusion (Livnat, 2017; Bolotin et al., 2021).

[4]Sometimes, long-term acting mechanisms can also lead to an immediate response, but when such responses are interpreted from a Lamarckian perspective, they lose their generality (Livnat, 2017, p. 178), as in the CUP1 case discussed above.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

Replacement-based exon shuffling provides a simpler explanation than rm/ns to cases where the same exons are trans-spliced in one species or population and cis-spliced in another (Fischer et al., 2008; Labrador and Corces, 2003; Shao et al., 2012; Kong et al., 2015), or where some functions are achieved by multiple single-module proteins in one taxon but by a single multi-module protein in another (Henikoff et al., 1997; Graur and Li, 2000). Likewise, Bolotin et al.'s recent empirical results are consistent with the used-fused mutational replacement hypothesis (Bolotin et al., 2021). Thus, we argue that exon shuffling—a phenomenon in evolutionary time—and alternative splicing—a phenomenon in developmental time—are directly connected via mutational replacement mechanisms, and that the evolution of genome organization can be driven by mutational replacement mechanisms as opposed to random mutation and random genetic drift (cf. Lynch, 2007).

6.3 Mutational replacement and the repeatability of evolution

Parallel adaptive evolution may be not only due to similarities in selection pressures and phenotypic constraints between genetically related species (Blount et al., 2018), but also due to similarities in mutational tendencies: if mutations respond to complex information in the genome, the more genetically related two biological entities are, the more similar their mutational tendencies should be (Livnat, 2013, 2017). This principle of genetic relatedness in mutational tendencies is consistent with the observation that similar malaria-related mutations tend to appear repeatedly on different genetic backgrounds within a human population, while different such within-population-repeating mutations appear in different human populations (Livnat, 2013), and has been supported by the population-level differences in the HbS mutation and other mutation-specific rates demonstrated by the HbS study (Melamed et al., 2022). Further consistent, Bolotin et al. found evidence for extensive parallelism in gene fusion mutations, and that gene fusion mutations occur due to mutational replacement (Bolotin et al., 2021). Thus, under IBE, parallelism may be far more common than previously recognized. The sharing of a mutation by a monophyletic group does not favor the possibility that the common ancestor had the mutation as opposed to having had the genetic background on which that mutation was more likely to arise.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

6.4 Transposable elements, gene regulatory networks and the role of epigenetics in evolution

Despite the fact that transposable elements (TEs) have been perceived as "selfish elements" at the DNA level (Dawkins, 1976; Doolittle and Sapienza, 1980), recent evidence has clarified that their contribution to adaptive evolution is immense (Lynch et al., 2011; Fedoroff, 2012; Chuong et al., 2013; Ellison and Bachtrog, 2013; Notwell et al., 2015; Chuong et al., 2016). This evidence supports the initial proposal that, by inserting multiple copies of itself and its transcription factor binding sites at different loci, one TE can become a master coordinator of multiple genes (Britten and Davidson, 1969; Lynch et al., 2011, 2015). Using the replacement hypothesis framework, we offer an explanation for how and why such insertions occur: A gradual evolutionary process first coopts a set of genes previously active in other organs to interact in a novel network toward a novel, complex function (Lynch et al., 2015). As a result, the chromatin at these cooperating loci is open at the same time—according to the replacement hypothesis, also in the germline (Livnat, 2013). A TE active at the same time in the germline could then insert itself into these open loci. A positive feedback loop could furthermore arise, where this TE's activity becomes more and more focused at the relevant time window and locations in the nucleus as it inserts additional copies of itself into the interacting loci.

The process as a whole is one of mutational replacement and simplification: At the early stage of cooption of genes into a novel network, where a complex interaction emerges at the system-level in a haphazard manner, different genes are activated by different pre-evolved arms of regulation, consistent with the nature of early complex adaptations in general (Livnat, 2017). By contributing the same set of ready-made or cryptic regulatory elements (Emera and Wagner, 2012a,b), the TE then comes to partly or fully replace the previously distinct arms of gene regulation with one control. The result is simplification of network regulation through mutational replacement.

Thus, rather than the contribution of TEs being either to the genesis of novel gene regulatory networks (GRNs) or to the turnover of regulatory elements within an existing GRN (Lynch et al., 2015), these may be not two mutually exclusive possibilities but different aspects of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

same process: the subjecting of a previously disorganized set of coopted cooperating genes, each activated by its own preexisting arm of regulation, to a common control is a part of an evolutionary process of simplification and routinization, exemplifying the IBE principle that a complex adaptation emerges in a fuzzy, disorganized state and is then crystallized as a whole into a clockwork-like state through a gradual evolutionary process of simplification under performance pressure (Livnat, 2017).

TEs also bear on our point that the role of epigenetic changes in evolution in general may be not in enabling Lamarckian transmission but in meshing with and guiding less frequently occurring heritable changes. Linquist and Fullerton (2021) aruged that because of the tight connection between TEs and methylation, as well as between TE activity on the one hand and stress and hypomethylation on the other, phenomena that have been attributed to epigenetics-based Lamarckian inheritance (Jablonka and Lamb, 2014; Bonduriansky and Day, 2018) could also be interpreted as essentially non-Lamarckian germline activity of TEs and associated epigenetic changes in them. Whereas Linquist et al. proposed this while viewing TEs strictly as selfish elements, their argument could be reinterpreted from the IBE perspective, according to which TEs are more than just selfish elements; they are a part of a system-level view of mutation origination—a part that has long-term usefulness for the system as a whole even while having the appearance of selfish elements at smaller spatio-temporal scales (Livnat, 2013).

7 Summary and outlook

Calls have multiplied for a replacement or extension of the modern evolutionary synthesis (Noble, 2008; Pigliucci and Müller, 2010; Noble et al., 2014; Noble, 2015). The theory of Interaction-based Evolution (IBE) provides one such possibility. According to IBE, evolution is driven by the interaction of two forces: an external force of differential survival and reproduction, and an internal force of nonaccidental, non-Lamarckian heritable change. Each heritable change has its own specific origination probability, which depends on complex genetic and epigenetic information accumulated in the genome. This information in turn has come from previous heritable changes and previous selection pressures. Both the external and internal forces

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 June 2022 doi:10.20944/preprints202206.0409.v1

are continually updated through the generations: as the organism changes, so does the selection pressure gradually change; and as the genome changes, so do the rates of heritable changes gradually change across the genome. Thus, these two forces continually interact as they coevolve, and their interaction drives evolution. Furthermore, since heritable changes in one generation affect the origination rates of changes in later generations, a network of information flow exists through the genome and the generations. Thus, natural selection, sexual recombination and heritable changes that respond to complex information in the genome combine system-level information from many individuals that have succeeded in survival and reproduction over the generations.

Topics informed by this view include directed mutation, evolutionary parallelism, genome organization evolution and more: *a*) Because mutation rates respond in a mutation-specific manner to complex information that accumulates in the genome through the generations, mutation-specific responses to specific environmental pressures are in general possible in the long-term. *b*) Genetic relatedness in mutational tendencies exist, with important implications for parallel evolution. *c*) The evolution of genome organization may be largely driven by mutational mechanisms rather than random mutation and random genetic drift. *d*) A special role for epigenetics in evolution may be not in enabling Lamarckian transmission from soma to germline but in being a frequently occurring type of nonaccidental yet non-Lamarckian heritable change, meshing with and affecting the origination of less frequently occurring heritable changes.

A key part of this view—the replacement hypothesis—holds that mutations can follow directly and mechanistically from previously evolved interactions between genes, thus forming a direct mechanistic link between the evolution of regulation and structural mutational changes. Accordingly, we argued that the HbS mutation did not originate at random and did not begin a process of adaptation by natural selection, but rather a long-term evolutionary process preceded it that has led to its increased rate of origination in the gene and population where it is of adaptive significance. According to this view, a vast landscape of mutational phenomena remains to be studied that is essential for our understanding of evolution. We expect that future natural studies will generalize our HbS results to other genes and organisms and that future

experimental studies will be able to recapitulate our observations using various model organisms and target genes by designing platforms for measuring de novo mutation rates at the single mutation resolution under multigenerational artificial selection pressures and the conditions for long-term adaptive evolution specified by IBE.

Competing interests statement

The authors declare no competing interests.

Acknowledgments

We thank Kim Weaver for extensive help. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

Author contributions: AL and DM wrote the paper.

References

Agoglia RM, Sun D, Birey F, Yoon SJ, Miura Y, Sabatini K, Pașca SP, and Fraser HB. 2021. Primate cell fusion disentangles gene regulatory divergence in neurodevelopment. *Nature* **592**: 421–427.

Aguilera A. 2002. The connection between transcription and genomic instability. *The EMBO Journal* **21**: 195–201.

Aguilera A and Gaillard H. 2014. Transcription and recombination: when RNA meets DNA. cold spring harb. perspect. biol. 6: a016543.

Altenberg L, Liberman U, and Feldman MW. 2017. Unified reduction principle for the evolution of mutation, migration, and recombination. *Proceedings of the National Academy of Sciences USA* **114**: E2392–E2400.

 doi:10.20944/preprints202206.0409.v1

Alvarez-Dominguez JR, Amosova O, and Fresco JR. 2013. Self-catalytic DNA depurination underlies human β-globin gene mutations at codon 6 that cause anemias and thalassemias. *Journal of Biological Chemistry* **288**: 11581–11589.

Bailey JA, Liu G, and Eichler EE. 2003. An *Alu* transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* **73**: 823–834.

Bass BL. 2002. RNA editing by adenosine deaminases that act on RNA. *Annual review of biochemistry* **71**: 817–846.

Bell G. 1982. *The Masterpiece of Nature: the Evolution and Genetics of Sexuality*. University of California Press, Berkeley.

Blount ZD, Lenski RE, and Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**: eaam5979.

Bolotin E, Melamed D, and Livnat A. 2021. Genes used together are more likely to be fused together in evolution by mutational mechanisms: A bioinformatic test of the used-fused hypothesis. *bioRxiv 07.31.454590* .

Bonduriansky R and Day T. 2018. *Extended Heredity*. Princeton University Press.

Brennan G, Kozyrev Y, and Hu SL. 2008. TRIMCyp expression in old world primates *Macaca nemestrina* and *Macaca fascicularis*. *P Natl Acad Sci USA* **105**: 3569–3574.

Britten RJ and Davidson EH. 1969. Gene regulation for higher cells: A theory: New facts regarding the organization of the genome provide clues to the nature of gene regulation. *Science* **165**: 349–357.

Cairns J, Overbaugh J, and Miller S. 1988. The origin of mutants. *Nature* **335**: 142–145.

Carlson J, Locke AE, Flickinger M, Zawistowski M, Levy S, Myers RM, Boehnke M, Kang HM, Scott LJ, Li JZ, et al.. 2018. Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *Nature Communications* **9**: 3753.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**: 25–36.

Carvalho CM, Zhang F, and Lupski JR. 2010. Genomic disorders: A window into human gene and genome evolution. *Proceedings of the National Academy of Sciences USA* **107**: 1765–1771.

Chandramouly G, Zhao J, McDevitt S, Rusanov T, Hoang T, Borisonnik N, Treddinick T, Lopezcolorado FW, Kent T, Siddique LA, et al.. 2021. Polθ reverse transcribes RNA and promotes RNA-templated DNA repair. *Science Advances* **7**: eabf1771.

Chen JY, Peng Z, Zhang R, Yang XZ, Tan BCM, Fang H, Liu CJ, Shi M, Ye ZQ, Zhang YE, et al.. 2014. RNA editome in rhesus macaque shaped by purifying selection. *PLoS genetics* **10**: e1004274.

Chicote JU, López-Sánchez M, Marquès-Bonet T, Callizo J, Pérez-Jurado LA, and García-España A. 2020. Circular DNA intermediates in the generation of large human segmental duplications. *BMC Genomics* **21**: 1–11.

Chuong EB, Elde NC, and Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**: 1083–1087.

Chuong EB, Rumi MK, Soares MJ, and Baker JC. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature Genetics* **45**: 325–329.

Danecek P, Nellåker C, McIntyre RE, Buendia-Buendia JE, Bumpstead S, Ponting CP, Flint J, Durbin R, Keane TM, and Adams DJ. 2012. High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biology* **13**: 1–12.

Darwin C. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* John Murray, London, 1st edition.

Dawkins R. 1976. *The Selfish Gene*. Oxford University Press.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380.

Doolittle WF and Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**: 601–603.

Duret L and Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* **10**: 285–311.

Durkin K, Coppieters W, Drögemüller C, Ahariz N, Cambisano N, Druet T, Fasquelle C, Haile A, Horin P, Huang L, et al.. 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* **482**: 81–84.

Edelman LB and Fraser P. 2012. Transcription factories: genetic programming in three dimensions. *Current Opinion in Genetics & Development* **22**: 110–114.

Ellison CE and Bachtrog D. 2013. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**: 846–850.

Emera D and Wagner GP. 2012a. Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. *P Natl Acad Sci USA* **109**: 11246–11251.

Emera D and Wagner GP. 2012b. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Briefings in Functional Genomics* **11**: 267–276.

Ewens WJ. 2004. *Mathematical Population Genetics. I. Theoretical Introduction, 2nd ed.* Springer Science & Business Media.

Fedoroff NV. 2012. Transposable elements, epigenetics, and genome evolution. *Science* **338**: 758–767.

Feldman MW and Liberman U. 1986. An evolutionary reduction principle for genetic modifiers. *Proceedings of the National Academy of Sciences USA* **83**: 4824–4827.

Fischer SE, Butler MD, Pan Q, and Ruvkun G. 2008. Trans-splicing in *C. elegans* generates the negative RNAi regulator ERI-6/7. *Nature* **455**: 491–496.

Fisher RA. 1930. *The Genetical Theory of Natural Selection*. The Clarendon Press, Oxford.

Franklin A, Milburn PJ, Blanden RV, and Steele EJ. 2004. Human DNA polymerase-η, an a-t mutator in somatic hypermutation of rearranged immunoglobulin genes, is a reverse transcriptase. *Immunology and Cell Biology* **82**: 219–225.

Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Research* **23**: 1089–1096.

Freeman S and Herron JC. 1998. *Evolutionary Analysis*. Pearson Benjamin Cummings.

Fryxell KJ and Moon WJ. 2005. CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol* **22**: 650–658.

Futuyma DJ. 1998. *Evolution*. Sinauer Associates, Sunderland, MA, 3rd edition.

Gilbert W. 1978. Why genes in pieces? *Nature* **271**: 501.

Gokhman D, Agoglia RM, Kinnebrew M, Gordon W, Sun D, Bajpai VK, Naqvi S, Chen C, Chan A, Chen C, et al.. 2021. Human–chimpanzee fused cells reveal cis-regulatory divergence underlying skeletal evolution. *Nature Genetics* **53**: 467–476.

Gómez-González B and Aguilera A. 2019. Transcription-mediated replication hindrance: a major driver of genome instability. *Genes & Development* **33**: 1008–1026.

Gommans WM, Mullen SP, and Maas S. 2009. RNA editing: a driving force for aaptive evolution? *Bioessays* **31**: 1137–1145.

Gould SJ and Vrba ES. 1982. Exaptation—a missing term in the science of form. *Paleobiology* **8**: 4–15.

Graur D and Li WH. 2000. *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Grauso M, Reenan R, Culetto E, and Sattelle D. 2002. Novel putative nicotinic acetylcholine receptor subunit genes, d α 5, d α 6 and d α 7, in drosophila melanogaster identify a new and highly conserved target of adenosine deaminase acting on RNA-mediated A-to-I pre-mRNA editing. *Genetics* **160**: 1519–1533.

Gu W, Zhang F, and Lupski JR. 2008. Mechanisms for human genomic rearrangements. *PathoGenetics* **1**: 4.

Haig D. 2007. Weismann rules! Ok? Epigenetics and the Lamarckian temptation. *Biology & Philosophy* **22**: 415–428.

Hallgrímsson B, Brown J, and Hall B. 2005. The study of phenotypic variability: an emerging research agenda for understanding the developmental-genetic architecture underlying phenotypic variation. In *Variation: A Central Concept in Biology* (eds. B Hallgrímsson and B Hall), pp. 525–551. Elsevier, Burlington, MA.

Hallgrímsson B, Jamniczky HA, Young NM, Rolian C, Schmidt-ott U, and Marcucio RS. 2012. The generation of variation and the developmental basis for evolutionary novelty. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **318**: 501–517.

Hamperl S, Bocek MJ, Saldivar JC, Swigut T, and Cimprich KA. 2017. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell* **170**: 774–786.

Harris K. 2015. Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences USA* **112**: 3439–3444.

Harris K and Pritchard JK. 2017. Rapid evolution of the human mutation spectrum. *Elife* **6**.

Hartl DL and Clark AG. 2007. *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts, 4th edition.

Hebb D. 1949. *The Organization of Behavior*. Wiley & Sons, New York.

Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, and Hood L. 1997. Gene families: the taxonomy of protein paralogs and chimeras. *Science* **278**: 609–614.

Hodgkinson A and Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet* **12**: 756–766.

Hull RM, Cruz C, Jack CV, and Houseley J. 2017. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biology* **15**: e2001333.

Jablonka E and Lamb MJ. 2014. *Evolution in Four Dimensions, Revised Edition: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT press.

Jackson DA, Hassan AB, Errington RJ, and Cook PR. 1993. Visualization of focal sites of transcription within human nuclei. *The EMBO Journal* **12**: 1059.

Jimeno S, Prados-Carvajal R, Fernández-Ávila MJ, Silva S, Silvestris DA, Endara-Coll M, Rodríguez-Real G, Domingo-Prim J, Mejías-Navarro F, Romero-Franco A, et al.. 2021. ADAR-mediated RNA editing of DNA: RNA hybrids is required for DNA double strand break repair. *Nature Communications* **12**: 1–15.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al.. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Kaessmann H, Vinckenbosch N, and Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nature Reviews Genetics* **10**: 19–31.

Kleene KC. 2005. Sexual selection, genetic conflict, selfish genes, and the atypical patterns of gene expression in spermatogenic cells. *Developmental biology* **277**: 16–26.

Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* **21**: 12–27.

Kong Y, Zhou H, Yu Y, Chen L, Hao P, and Li X. 2015. The evolutionary landscape of intergenic trans-splicing events in insects. *Nat Commun* **6**: 8734.

 doi:10.20944/preprints202206.0409.v1

Labrador M and Corces VG. 2003. Extensive exon reshuffling over evolutionary time coupled to trans-splicing in drosophila. *Genome research* **13**: 2220–2228.

Leigh Jr EG. 1970. Natural selection and mutability. *Am Nat* **104**: 301–305.

Lenski RE and Mittler JE. 1993. The directed mutation controversy and neo-darwinism. *Science* **259**: 188–194.

Li H, Wang J, Mor G, and Sklar J. 2008. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. *Science* **321**: 1357–1361.

Liao CH, Kuang YQ, Liu HL, Zheng YT, and Su B. 2007. A novel fusion gene, TRIM5-Cyclophilin A in the pig-tailed macaque determines its susceptibility to HIV-1 infection. *AIDS* **21(Suppl 8)**: S19–S26.

Lindley R. 1966. Recording as a function of chunking and meaningfulness. *Psychonomic Science* **6**: 393–394.

Linquist S and Fullerton B. 2021. Transposon dynamics and the epigenetic switch hypothesis. *Theoretical Medicine and Bioethics* **42**: 137–154.

Livnat A. 2013. Interaction-based evolution: how natural selection and nonrandom mutation work together. *Biol Direct* **8**: 24.

Livnat A. 2017. Simplification, innateness, and the absorption of meaning from context: how novelty arises from gradual network evolution. *Evol Biol* **44**: 145–189.

Löwel S and Singer W. 1992. Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science* **255**: 209–212.

Luria SE and Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491.

Lynch M. 2007. *The Origins of Genome Architecture*. Sinauer Associates Sunderland.

Lynch VJ, Leclerc RD, May G, and Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* **43**: 1154–1159.

Lynch VJ, Nnamani MC, Kapusta A, Brayer K, Plaza SL, Mazur EC, Emera D, Sheikh SZ, Grützner F, Bauersachs S, et al.. 2015. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Reports* **10**: 551–561.

Melamed D, Nov Y, Malik A, Yakass MB, Bolotin E, Shemer R, Hiadzi EK, Skorecki KL, and Livnat A. 2022. De novo mutation rates at the single-mutation resolution in a human HBB gene-region associated with adaptation and genetic disease. *Genome Research* **32**: 488–498.

Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, et al.. 2015. The human transcriptome across tissues and individuals. *Science* **348**: 660–665.

Moczek AP. 2008. On the origins of novelty in development and evolution. *BioEssays* **30**: 432–447.

Monroe J, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, Neumann M, Kliebenstein D, et al.. 2022. Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature* pp. 1–5.

Morgan TH. 1903. *Evolution and Adaptation*. The Macmillan Company, New York.

Mouakkad-Montoya L, Murata MM, Sulovari A, Suzuki R, Osia B, Malkova A, Katsumata M, Giuliano AE, Eichler EE, and Tanaka H. 2021. Quantitative assessment reveals the dominance of duplicated sequences in germline-derived extrachromosomal circular DNA. *Proceedings of the National Academy of Sciences* **118**.

Müller GB. 2007. Evo–devo: extending the evolutionary synthesis. *Nature Reviews Genetics* **8**: 943–949.

Müller GB and Wagner GP. 1991. Novelty in evolution: restructuring the concept. *Annual Review of Ecology and Systematics* **22**: 229–256.

Newell A. 1980. Physical symbol systems. *Cognitive Science* **4**: 135–183.

Newman RM, Hall L, Kirmaier A, Pozzi LA, Pery E, Farzan M, O'Neil SP, and Johnson W. 2008. Evolution of a TRIM5-CypA splice isoform in old world monkeys. *PLoS Pathog* **4**: e1000003.

Nisole S, Lynch C, Stoye JP, and Yap MW. 2004. A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. *P Natl Acad Sci USA* **101**: 13324–13328.

Noble D. 2008. *The Music of Life: Biology Beyond Genes*. Oxford University Press.

Noble D. 2011. Neo-darwinism, the modern synthesis and selfish genes: are they of use in physiology? *The Journal of Physiology* **589**: 1007–1015.

Noble D. 2015. Evolution beyond neo-darwinism: a new conceptual framework. *Journal of Experimental Biology* **218**: 7–13.

Noble D, Jablonka E, Joyner MJ, Müller GB, and Omholt SW. 2014. Evolution evolves: physiology returns to centre stage. *The Journal of Physiology* **592**: 2237–2244.

Notwell JH, Chung T, Heavner W, and Bejerano G. 2015. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nature Communications* **6**: 1–7.

Ohlson J, Pedersen JS, Haussler D, and Öhman M. 2007. Editing modifies the GABAA receptor subunit α3. *RNA* **13**: 698–703.

Ohno S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, Heidelberg.

Osborne CS. 2014. Molecular pathways: transcription factories and chromosomal translocations. *Clinical Cancer Research* **20**: 296–300.

Pan S, Bruford MW, Wang Y, Lin Z, Gu Z, Hou X, Deng X, Dixon A, Graves JAM, and Zhan X. 2018. Transcription-associated mutation promotes RNA complexity in highly expressed genes—a major new source of selectable variation. *Molecular Biology and Evolution* **35**: 1104–1119.

Papantonis A and Cook PR. 2013. Transcription factories: genome organization and gene regulation. *Chemical Reviews* **113**: 8683–8705.

Park C, Qian W, and Zhang J. 2012. Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Reports* **13**: 1123–1129.

Pigliucci M and Müller GB. 2010. Elements of an extended evolutionary synthesis. In *Evolution: The Extended Synthesis* (eds. M Pigliucci and GB Müller), pp. 3–17. Massachusetts Institute of Technology Press Cambridge, MA.

Pinto Y, Cohen HY, and Levanon EY. 2014. Mammalian conserved ADAR targets comprise only a small fragment of the human editosome. *Genome Biology* **15**: 1–15.

Pippenger N. 1979. On simultaneous resource bounds. In *20th Annual Symposium on Foundations of Computer Science (SFCS 1979)*, pp. 307–311. IEEE.

Popitsch N, Huber CD, Buchumenski I, Eisenberg E, Jantsch M, Von Haeseler A, and Gallach M. 2020. A-to-I RNA editing uncovers hidden signals of adaptive genome evolution in animals. *Genome Biology and Evolution* **12**: 345–357.

Qu W, Hashimoto S, Shimada A, Nakatani Y, Ichikawa K, Saito TL, Ogoshi K, Matsushima K, Suzuki Y, Sugano S, et al.. 2012. Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns. *Genome Res* **22**: 1419–1425.

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, Dominiczak A, Morris A, Porteous D, Smith B, et al.. 2016. Timing, rates and spectra of human germline mutation. *Nat Genet* **48**: 126.

Saraconi G, Severi F, Sala C, Mattiuz G, and Conticello SG. 2014. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biology* **15**: 1–10.

Sarkar S. 1990. On the possibility of directed mutations in bacteria: Statistical analyses and reductionist strategies. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1990, pp. 111–124. Philosophy of Science Association.

Sayah DM, Sokolskaja E, Berthoux L, and Luban J. 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* **430**: 569–573.

Schlosser G. 2015. Vertebrate cranial placodes as evolutionary innovations – the ancestor's tale. *Current Topics in Developmental Biology* **111**: 235–300.

Shao W, Zhao QY, Wang XY, Xu XY, Tang Q, Li M, Li X, and Xu YZ. 2012. Alternative splicing and trans-splicing events revealed by analysis of the *Bombyx mori* transcriptome. *RNA* **18**: 1395–1407.

Shapiro JA. 2011. *Evolution: A View from the 21st Century*. FT Press.

Shiromoto Y, Sakurai M, Minakuchi M, Ariyoshi K, and Nishikura K. 2021. ADAR1 RNA editing enzyme regulates R-loop formation and genome stability at telomeres in cancer cells. *Nature Communications* **12**: 1–18.

Soler-Oliva ME, Guerrero-Martínez JA, Bachetti V, and Reyes JC. 2017. Analysis of the relationship between coexpression domains and chromatin 3D organization. *PLoS Computational Biology* **13**: e1005708.

Su Y, Ghodke PP, Egli M, Li L, Wang Y, and Guengerich FP. 2019. Human DNA polymerase η has reverse transcriptase activity in cellular environments. *Journal of Biological Chemistry* **294**: 6073–6081.

Takahashi KK and Innan H. 2020. Duplication with structural modification through extrachromosomal circular and lariat DNA in the human genome. *Scientific Reports* **10**: 1–11.

Tasakis RN, Laganà A, Stamkopoulou D, Melnekoff DT, Nedumaran P, Leshchenko V, Pecori R, Parekh S, and Papavasiliou FN. 2020. ADAR1 can drive Multiple Myeloma progression by acting both as an RNA editor of specific transcripts and as a DNA mutator of their cognate genes. *Biorxiv* .

Thomas BJ and Rothstein R. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**: 619–630.

Tian N, Wu X, Zhang Y, and Jin Y. 2008. A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. *RNA* **14**: 211–216.

Tsuruoka N, Arima M, Yoshida N, Okada S, Sakamoto A, Hatano M, Satake H, Arguni E, Wang JY, Yang JH, et al.. 2013. ADAR1 protein induces adenosine-targeted DNA mutations in senescent Bcl6 gene-deficient cells. *Journal of Biological Chemistry* **288**: 826–836.

Tulving E and Craik FI. 2005. *The Oxford Handbook of Memory*. Oxford University Press.

Vinckenbosch N, Dupanloup I, and Kaessmann H. 2006. Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Sciences* **103**: 3220–3225.

Virgen CA, Kratovac Z, Bieniasz PD, and Hatziioannou T. 2008. Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. *P Natl Acad Sci USA* **105**: 3563–3568.

Wade MJ and Goodnight CJ. 1998. Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* **52**: 1537–1553.

Wagner GP. 2014. *Homology, Genes, and Evolutionary Innovation*. Princeton University Press.

Wagner GP and Lynch VJ. 2010. Evolutionary novelties. *Current Biology* **20**: R48–R52.

Walsh B and Lynch M. 2018. *Evolution and Selection of Quantitative Traits*. Oxford University Press, Oxford, UK.

 doi:10.20944/preprints202206.0409.v1

Wilson SJ, Webb BL, Ylinen LM, Verschoor E, Heeney JL, and Towers GJ. 2008. Independent evolution of an antiviral TRIMCyp in rhesus macaques. *P Natl Acad Sci USA* **105**: 3557–3562.

Wilson TE, Arlt MF, Park SH, Rajendran S, Paulsen M, Ljungman M, and Glover TW. 2015. Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome research* **25**: 189–200.

Xia B, Yan Y, Baron M, Wagner F, Barkley D, Chiodin M, Kim SY, Keefe DL, Alukal JP, Boeke JD, et al.. 2020. Widespread transcriptional scanning in the testis modulates gene evolution rates. *Cell* **180**: 248–262.

Xu G and Zhang J. 2014. Human coding RNA editing is generally nonadaptive. *Proceedings of the National Academy of Sciences USA* **111**: 3769–3774.

Zheng Y, Lorenzo C, and Beal PA. 2017. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. *Nucleic Acids Research* **45**: 3369–3377.