Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *Evolutionary Dynamics and Information Hierarchies in Biological Systems*

In Darwinian evolution, feedback from natural selection leads to biased mutations

Lynn Helena Caporale and John Doyle[1]

[1]Control and Dynamical Systems California Institute of Technology, Pasadena, California 91125

Address for correspondence: Lynn Helena Caporale 1 Sherman Square, New York, NY 10023. caporale@usa.net

Natural selection provides feedback through which information about the environment and its recurring challenges is captured, inherited, and accumulated within genomes in the form of variations that contribute to survival. The variation upon which natural selection acts is generally described as "random." Yet evidence has been mounting for decades, from such phenomena as mutation hotspots, horizontal gene transfer, and highly mutable repetitive sequences, that variation is far from the simplifying idealization of random processes as *white* (uniform in space and time and independent of the environment or context). This paper focuses on what is known about the generation and control of mutational variation, emphasizing that it is not uniform across the genome or in time, not unstructured with respect to survival, and is neither memoryless nor independent of the (also far from white) environment. We suggest that, as opposed to frequentist methods, Bayesian analysis could capture the evolution of nonuniform probabilities of distinct classes of mutation, and argue not only that the locations, styles, and timing of real mutations are not correctly modeled as generated by a white noise random process, but that such a process would be inconsistent with evolutionary theory.

Keywords: Darwin; random mutation; natural selection; evolution; feedback

Introduction

While the term *random mutation* is widely used as if it were the original foundation of evolutionary theory, Darwin explicitly stated:

> I have . . . sometimes spoken as if the variations . . . had been due to chance. This, of course, is a wholly incorrect expression, but it serves to acknowledge plainly our ignorance of the cause of each particular variation.[1]

A century before biochemists began to describe the underpinnings of variation among individuals, it was attention to variation that led to the theory of evolution. Charles Darwin and Alfred Russel Wallace were perceptive naturalists and collectors, who noticed variation among individuals of the same species.[a] They proposed that those variants that are most fitted to their environment pass on traits that contributed to fitness in the next generation, leading to descent with modification and adaptation,[2,3] although they did not know how traits were modified or inherited.

With respect to Mendel's observations,[4] Wallace wrote:

> The essential basis of evolution, involving as it does the most minute and all-pervading adaptation to the whole environment, is *extreme and ever-present plasticity, as a condition of survival and adaptation.*[b] But the essence of Mendelian characters is their rigidity. They are transmitted without variation, and therefore, except by the rarest of accidents, can never become adapted to ever varying conditions.[5]

[a]"considering the amount of individual variation that . . . experience as a collector had shown . . . to exist . . . .;"[3] name "this principle, by which each slight variation is selected, Natural Selection."[2]

[b]Emphasis added.

doi: 10.1111/nyas.12235

Mendel's observations were integrated into evolutionary theory[6] through the concept that variation within populations results from different versions (alleles) of genes (Mendel's inherited characters). The statistician Ronald Fisher[c,7] introduced a geometric model with the mathematical assumption that generation of the phenotypic manifestations of variation could be represented as a normally distributed random process. What was implicit in this model is that mutation creates new alleles with white noise increments and then selection alters the prevalence of different alleles. Thus, the concept of *random* was attached to evolutionary theory not by Darwin but over half a century later.

That mutations are random is taught now as an integral part of evolutionary theory. For example, a book targeted to high school teachers, filled with engaging examples of ways to teach evolution, describes "Darwin's law of natural selection" as stating that "descent with modification and adaptation result from the natural selection of heritable random variations."[8] The term *law* is used instead of *theory* due to misunderstanding by those not trained in science about the meaning of *theory*;[d] but, there is at least as much confusion about the meaning of *random*.

There has been a narrowing within physics and science generally of the term *random* to be nearly synonymous with *white noise*. An assumption that variation is a random white noise process would mean that mutations are uniform (or unbiased) with respect to position along the chromosome, stationary (i.e., uniform also with respect to time), memoryless (independent of past changes), and independent of context or environment (i.e., autonomous). These assumptions simplify modeling and analysis because mutations thus restricted would render genome sequences a random walk with white noise increments, and selection simply would favor those fittest to the environment. However, for real biology to preserve this idealization would require selection to have no feedback that could act on the mechanisms that generate variation, as this would likely disrupt its white noise properties.

In the rest of this paper, we will discuss how well-known experiments clearly show that real biology systematically and routinely violates all of the white noise assumptions. We will emphasize that feedback from selection to the biochemical processes that generate variation would make mutation very far from white by increasing the probability that variants will survive (compared to variants generated by a random white noise process). We then briefly end with a discussion of the implications for a more rigorous mathematical formalization of evolutionary theory, which will be pursued in subsequent papers.

Mutation is not uniform with respect to position along the DNA

The term *random mutation* was attached to evolutionary theory before the chemical nature of mutation was understood. Even so, the assumption of uniformity (unbiased) with respect to position along the chromosome was known to be violated[9,10] even before the Avery laboratory demonstrated[11] that the chemical underlying transmission of inherited characters is DNA. Mutation hotspots were observed almost as soon as DNA was understood to be the genetic material; as Seymour Benzer commented in 1961 regarding the genetic map of the phage T4 that "the distribution is nonrandom leaps to the eye."[12] The availability of genome sequences has made the nonuniform distribution of mutations noted by Benzer all the more obvious.[e,13,14]

Even without this evidence, a thoughtful biochemist can predict that it would be highly unlikely for mutations to be uniformly distributed along a DNA sequence. Although routinely treated as text in search algorithms, DNA bases are, of course, not actually letters. DNA double helices are physical chemical entities with distinct properties,[15] such

[c] "The possible positions representing adaptations superior to that represented by A will be enclosed by a sphere passing through A and centred at O. If A is shifted through a fixed distance, r, *in any direction* its translation will improve the adaptation if it is carried to a point within this sphere, but will impair it if the new position is outside." (emphasis added; note that in his discussion shifts can represent changes either in the organism's phenotype or its environment).

[d] As in, "oh, it's just a theory."

[e] "It is clear that the mammalian genome is evolving under the influence of non-uniform local forces."[13]

as the uneven tilt and twist of base pairs (which cause sequence-dependent deviations from the iconic DNA structure),[16] repetitive sequences that tend to misalign and slip, and many noncanonical structures.[17–19] Such sequence context–dependent variations in physical chemical properties result in often dramatic sequence context–dependent effects on the fidelity of the enzymes that repeatedly copy and repair DNA, which in turn affects the evolution of that DNA sequence.

Thus, even if one were to synthesize a computer-generated white noise sequence of nucleotides, when biochemically replicated it would reveal intrinsic sequence context–dependent variations in the probability of distinct classes of mutation. In other words, the probability of distinct classes of mutation would not be expected to be random white noise, and thus not be uniformly distributed along the DNA molecule (or chromosome). This simple description of biochemistry has implications for evolutionary theory, as described in the next section.

Generation of variation can be biased as to whether a mutation will contribute to fitness

Since the early 20th century, it has been argued that most mutations must be deleterious, although these discussions, which began prior to an understanding of the biochemistry of mutation, did not consider local contexts that can affect the probability of distinct classes of mutation; further, as referenced in a discussion by David King, "for mutations arising spontaneously under natural conditions," (i.e., in contrast to mutations created in the laboratory) "the ratio of benefit to harm has never been realistically assessed."[20]

In contrast to our assertion that feedback from natural selection would affect the probability of distinct classes of mutation, it has been argued that natural selection cannot "assist the process of evolutionary change" as "selection lacks foresight and no one has described a plausible way to provide it."[21] Thus, in statements of evolutionary theory, the assertion is made that mutation is random with respect to its probability of being adaptive.[f] However, many classes of environmental challenge recur. Hosts combat pathogens (and pathogens avoid host defenses); predators and prey do battle through biochemical adaptations;[22] bird beaks must pick up and crack[23] available seeds (or insects)—a menu that may change rapidly due, for example, to a drought.

It is important to emphasize that not all random processes are well modeled as white noise. Thus we can ask whether there are biases in random mutation. There are of course rare environmental events, such as meteor strikes, and a wide range of processes that cause mutation with low probability at any site (it is to these events that the historical objections reviewed in Ref. 20 are likely to apply). However, environments change in ways that have structure, which would select for variation that has compensatory structure. For example, if a pathogen's environment contains a host immune system that continually generates new antibodies directed specifically against the pathogen's coat, mechanisms that generate rapid, focused, but still probabilistic variation in coats would be expected to be favored by selection.[24] Of course, the extent to which selection acts at such sites depends on the challenge/opportunity.[25,26]

In fact, since the probability of mutation varies along the DNA, and since the theory of evolution states that selection acts on variation, the theory of evolution actually predicts that variation will become at least somewhat structured, with mutations focused through mechanisms that could not have been taken into account before the biochemistry of mutations was investigated.[27] Selection should act on the biochemistry of genome variation much as it acts on beaks and wings. In other words, due to the repeated action of selection, an assertion that mutation is uniform, memoryless, and stationary, far from being integral to Darwin's insights, would be an implausible idealization that is inconsistent with evolutionary theory.

But in what sense could DNA sequences incorporate the potential to generate variants with an increased likelihood of surviving another round of selection? As more examples are found, our ability to imagine expands. For example selection can deplete mutable sequences,[g] such as repeats that tend to misalign and gain and lose units, from regions where variation is harmful, but such mutable

[f] "Mutation is a random process with respect to the adaptive needs of the species"—T. Dobzhansky.[6]

[g] By selecting against organisms that inherit mutable sequences at loci where they damage essential functions at a high rate.

sequences are enriched[20,28] in proteins involved in interactions with a changing environment. For example, in proteins that affect a pathogen's coat, loss or gain of a unit of a tetrameric repeat (such as CAATCAATCAATCAAT) shifts the reading frame, leading to loss of a coat variant recognized by the immune system (and which thus would have targeted the individual for destruction), at rates that are orders of magnitude above that of the background mutation rate (i.e., genome-wide average nucleotide substitution rate).[29,30]

Gain or loss of even a single unit in repeats such as GGGGGGGGG or GGGGGGGGGGG has a dramatic effect on the strength of binding of transcription factors, as it rotates the position of the –10 and –35 bacterial promoter consensus sequences around the helix relative to each other.[31] In the eukaryote *Saccharomyces cerevisiae*, it has been reported that ~25% of genes have tandem repeats in promoters, affecting expression levels.[32] Thus, there is significant standing variation in populations of individuals descended from a common ancestor bearing such mutable sequences.[20]

Note that a mutation involving loss or gain of units in repetitive sequences is reversible. Because such mutations are reversible, descendants are not trapped on a narrow fitness peak of the moment; rather, any individual has the potential to generate, among its population of descendants, a range of variants, facilitating survival of descendants as they confront the challenges and opportunities of a wider range of environments. Such highly repetitive sequences are found not only in bacteria, but also in eukaryotes,[33,34] including people.[35]

In contrast to recessive alleles, which are explicitly present in the genome, alternative genotypes that arise from slips in repetitive DNA sequences can be viewed as implied by the sequence that encodes them. Nonallelic diversity also is implied by sequences that raise the probability of gene duplication.[36] Thus, the probability of distinct classes of mutations has been biased, making some classes of mutation (such as changes in pathogen coats) more likely than others. Further, for some cases, such as a change of coat, any change may well protect the pathogen. This is in contrast to a change in an antibody-binding site, where, while generation of variation has focused on the variable region, a repertoire must be generated by targeted mutation in order to create the antibody that will bind to the pathogen's new coat. Horizontal gene transfer (HGT) is an obvious and dramatic example of a mechanism that expedites creation of a set of accessible functional genomes that is vastly larger than would be possible with only white noise mutation.[37]

Dobzhansky wrote,[6] "only a vitalist Pangloss can imagine that the genes know how and when it is good for them to mutate." This makes sense when considering whether one isolated individual nucleotide could know whether it might be better to be an A or a G in the next generation. We now understand that each nucleotide is embedded in a context, from genomic to environmental, with correlations emerging via selection over evolutionary time scales, and that this context can affect the fate of that nucleotide.

While locations in the genome may present with different probabilities of beneficial versus deleterious changes relative to genome averages, the direction of an individual mutation nevertheless still can be viewed as essentially randomly generated, but by a process that has been biased by selection. This process generates mutations that are not accurately modeled as random white noise. Note that saying a process is not modeled well by white noise is not equivalent to saying the process is not random (i.e., not probabilistic). Note also the difference between focusing on the probability of any one individual mutation and the probability of a mutation arising in a population. For example, we can say that for one individual bacterium's DNA it is (biased) random whether the change of length of a repetitive sequence takes it in the direction that would contribute to survival; however, given the size of bacterial populations, for loci that mutate at rates as high as 1/1000 and higher,[29] the population would almost certainly contain the variant that would survive the genetically anticipated challenge.

In other words, mechanisms that generate variation can adapt to a recurring nonuniform distribution of challenges, and thus in effect have a type of memory that would generate variation that is still random (i.e., probabilistic) but with an increased probability of generating variants that survive classes of challenges the genome and its descendants are likely to face in the future (i.e., if they are the same classes of challenges that the genome's lineage survived in the past). Thus, the statement that "all mutation is random"—in the sense of unstructured and uniform—is inconsistent not only with a growing body of data, but also with the theory

of evolution, due to the repeated effects of selection on mechanisms that generate mutations.

The mutation rate is neither stationary in time nor independent of the environment

While faithful reproduction of the genome transmits adaptations from generation to generation, those lineages that do not generate any diversity may be vulnerable[6] to, for example, a pathogen, or sudden loss of a food source. Thus, a balance between fidelity and exploration would be expected to evolve. That the rate of mutation does not change over time had been the assumption underlying the use of molecular clocks.[38] But is the probability of each class of mutation really unaffected by environmental or other influences that change over time?

As described in more detail below, the potential adaptive value of variation is constant neither across the genome nor over time. Thus stability (protection of adaptations) and diversity (exploration of new adaptations) can be balanced by an increased probability of variation not only in certain regions of the genome but also at times when the organism finds itself poorly adapted to its environment.

Since generation of variation results from biochemical processes, generation of variation, like biochemical processes, can be regulated. Biochemical mechanisms are available that can focus variation on different regions of the genome by, for example, induction of different sets of enzymes during different times during replication,[14,36] an effect that is increasingly accessible to analysis.[39]

This paper arises from a workshop centered on consideration of information hierarchies in biological systems.[40] Organisms sense, and respond with regulated metabolic changes to, the stress of starvation;[41] similarly, organisms sense and respond to the stress of DNA damage.[42] Thus, as sensing and signaling mechanisms are in place[43] that respond to the type and extent of the stress, biochemical mechanisms that affect genome variation, and thus affect evolution, can evolve connections downstream of signals that sense specific changes in the environment and specific classes of stress. For example, *Escherichia coli* senses and makes genetic changes in response to specific external clues. When it senses it is in the host environment, due to temperature and the presence of specific metabolites, a reversible mutation, inversion,[h] causes phase switching of fimbrae.[45]

The vertebrate immune system also demonstrates that the location and timing of distinct classes of genetic variation can be regulated. Targeted biochemical reactions generate variation in specific cell lineages (e.g., V/D/J rearrangement,[46] hypermutation[47]) and, in response to the environment, induce directed gene rearrangements (e.g., immunoglobulin class switch[48]). There is no reason to assume that such regulated, targeted variation would be unavailable to the germline.

Have connections between genetic variation and stress in fact been observed? *Arabidopsis* stressed by the presence of pathogens were observed to increase somatic recombination.[49] Barbara McClintock observed increased variation in response to the stress of DNA breakage and suggested that a cell is able to sense that it is under stress and that this might set in motion the orderly sequence of events that will mitigate this danger[50] and even trigger genome restructuring.[51] As Richard Jorgensen summarized,[i,52] McClintock proposed "a complex process that integrates information" and that could distinguish among, and mount appropriate distinct responses to, distinct classes of challenges.[53]

In documenting the generation of new regulatory networks and the apparent sudden burst of transposition by *mPing* in rice, under the stress of adapting to a temperate climate, Naito *et al.*[54] suggested that for selfing plants, bursts of transposable elements may generate genetic diversity rapidly, but also suggested that this is not limited to plants, as

[h] A vertebrate example of recurrent mutation involving inversion was revealed in a genomic study of "reuse of globally shared standing genetic variation, including chromosomal inversions, [which] has an important role in repeated evolution of distinct marine and freshwater sticklebacks,"[44] although there is no evidence yet that either addresses the question whether this inversion is induced rather than selected from standing variation or that standing variation is increased at that locus. Dobzhansky found evidence for seasonal variation of the prevalence of chromosome inversion as an adaptive trait in *Drosophila*.[6]

[i] "To paraphrase McClintock (1978), it is time to explore the nature and evolutionary significance of these attentive systems for adaptive genome restructuring in response to stress, "*the consequences of which vary according to the nature of the challenge to be met.*"

evidence for the rapid bursts of miniature inverted-repeat transposable elements is found in "virtually all sequenced eukaryotic genomes."

Increased mutation, observed in bacteria stressed by DNA damage or starvation, depends upon the activation of specific gene products.[55–58] and thus is not simply the result of inability to cope with the damage. Laboratory activation of the SOS DNA-damage and the RpoS-general/starvation stress response was sufficient to trigger a mutagenic mode of DNA break repair and thus increased mutation without an external stress.[58] In other words, the bacterium interprets induction of certain pathways as a biochemical signal that it is stressed. Having sequenced thousands of genomes, and determined expression patterns under varying conditions, it is possible to begin to examine genome wiring to explore whether and how distinct types of stress (and other environmental signals) might connect to pathways that affect distinct classes of variation of genome sequences.[59,60]

While evolution of responses to the environment that occur within the lifetime of an individual are widely acknowledged, there is no reason to limit biochemical responses to the environment to those that affect only a generation, since lineages survive over evolutionary timescales. It is important to note the role of feedback between selection and mechanisms that generate genome variation. Generators of diversity fall under selective pressure owing to the effects on survival of spatial and temporal biases of the classes of mutation that they generate.

Mutation, repair, and recombination depend upon biochemical processes, which can fall under the control of a wide range of regulatory systems. Thus, we cannot assume that mutation is stationary, unaffected by the environment, or constant in time.

What can evolve?

There is much more to understand about evolution than traits observed by naturalists in the field and base-by-base changes in DNA sequences observed in the laboratory. The genome is organized, with hierarchies of recognition and control. An evolutionary perspective is essential to comprehending this organization,[j,61] as is a perspective that includes feedback control and dynamics. In turn, a perspective built upon understanding this organized complexity[62] and its contexts will lead us to a deeper understanding of evolution. As we analyze genomic sequences, attention to structured and nonwhite forms of variation is likely to inform us of challenges that a lineage faces and that have exerted selective pressure during its evolution.

The initial step toward our ability to decipher information carried by DNA was to "crack the genetic code,"[63] but our work is not done. The degeneracy of the table of codons[64] and the existence of extensive[13] nonprotein coding regions leaves room to transmit additional messages underneath and around a protein-coding sequence, including messages that modulate the rate and type of genetic change. For example, the same amino acid sequence can be specified by either a mutable repetitive sequence or a more stable nonrepetitive sequence.[24] Thus, the same sequence can both specify a protein sequence and structure variation by implying mutability (i.e., affect the probability that descendants will be diverse at specific places in the genome). Such intertwined information represents an efficient use of genomic space.

Considering genome organization, self-reference, and behavior, an informative way to describe what can evolve is suggested by applying the term *protocol*[k,65] to genomics, a term used in engineering for sets of rules by which components interact to create new levels of functionality (familiar for enabling transmission of information through the internet). The table of codons is a familiar example of a protocol in biology (shared codons is one of many shared protocols necessary to enable HGT).[66] Protocols provide a useful concept for discussing evolution, including evolution of the genetic code,[67] as well as the labeled fragments with rules for their assembly that structure generation both of somatic diversity in the vertebrate immune system[46] and of the diverse repertoire of trypanosome coat proteins.[68] There are qualitatively distinct forms of information in genomes that may be nonlinear and dependent upon genomic context and relationships among sequences (such as inverted repeats or more complex structures[69]).[14]

[j] "Biological organization will never be understood except as the expression of an underlying evolutionary process."—Woese and Goldenfeld.[61]

[k] From the Greek use of *protocollon*, which referred to a leaf of paper glued to and labeling a manuscript scroll, defining its contents http://www.linfo.org/protocol.html

The profound importance of context, and an astounding sense of the complexity of organization that defines different contexts and behaviors of the genome, jumps to our attention in the formation of the macronucleus in ciliates, with genome-wide and predictable silencing, DNA deletions, inversions, and amplifications, built upon recognition and regulation involving RNA.[70–72]

In evolution, there would be a selective advantage for descendants of an individual that evolved an active framework that focuses exploration, compared to descendants of individuals that have a uniform probability of trying every mutation and every insertion site. This suggests the possibility that the DNA sequence of large gene families may represent a successful evolutionary framework, much as the protein sequence represents a successful functional framework.[73]

The biochemical infrastructure[74] that enables HGT in bacteria enormously increases (over white noise) the probability the bacteria will gain access to life-saving information, compared to if they lacked such infrastructure. Recent sequencing of roughly 100 *E. coli* strains (including subspecies Shigella spp) found that the genes universally shared, the median number per strain, and the total across all strains were approximately 1000, 4000, and 20,000, with the first and last numbers expected to continue to diverge as more strains are sequenced.[75] Further, in the canonical example of acquisition of antibiotic resistance by HGT, bacterial survivors in an antibiotic-rich environment (e.g., a hospital) would be a rich source for sensitive bacteria to tap for horizontal acquisition of genes conferring antibiotic resistance (the environment has structure with respect to the availability of genes accessible through HGT such that, e.g., genes encoding antibiotic resistance are most likely to be accessible from a neighboring bacterium just when and where an antibiotic-sensitive bacterium needs them).

Similarly, for eukaryotic parasites, the ability to vary coats rapidly through site-directed recombination provides a selective advantage compared to a probability of either uniformly distributed base changes or of recombination uniformly distributed in its genome.[68] Another eukaryotic example[20] of an important framework that facilitates generation of nonwhite diversity and exploration of new adaptations, to which significant resources are devoted, is meiosis.

Natural selection has embedded innate knowledge about the world within surviving genomes, embodied in diverse mechanisms. There are deep theories in systems engineering that help explain not just how such mechanisms that incorporate information about the environment work, but why they are necessary for robust performance. A richer evolutionary theory could incorporate analysis of how regularities in the environment's dynamics can become embedded in genomes in the form of dynamics of control circuits, but this has so far been explored in only a few settings.[76,77] We are very familiar with control circuits responsive to regularities of the environment that operate within a generation, such as circadian rhythms.[78] In another familiar example, using innate circuitry devoted to this purpose, *E. coli* swim toward glucose.[79] This circuitry obviously embodies structural models of attractants in the receptor proteins that recognize them, but the dynamics of adaptation circuitry also embodies a model of the structure of the environment (i.e., a model of the direction in which the concentration of the attractant increases). Furthermore, the necessity of these internal models can be made mathematically precise.[77]

The sophistication of internal models appears to increase with greater organism complexity. *Caenorhabditis elegans* inherit both innate chemosensory attractant and avoidance behaviors and mechanisms for individuals to adapt based upon experience.[80] Human sensorimotor control circuits, from basic reflexes to the most sophisticated learned skills, involve mechanisms that vary greatly in speed and flexibility, but all depend on internal models of the dynamics of the body, its extension via tools, and the environment.[81] Thus, internal models of regularities in the dynamics of its environment can be expected to contribute to an organism's fitness. Such models of the environment can also become embedded in the mechanisms that generate variation across generations. For example, innate information that structures the generation of variation underlies genomic mechanisms that facilitate our ability to create an antibody directed against an antigen never previously encountered by an individual or ancestors.[82]

Treating mutations as hypotheses about survival in an environment

Wallace rejected the predictable reassortment of the characters Mendel observed as a mechanism of evolution and instead saw that Mendel's work spoke to the stability of the inheritance of information from generation to generation. In fact, careful in his experimental design, and seeking mathematical laws, Mendel chose the characters he studied in his well-tended pea plants with care, as true breeding with two clearly distinguishable forms.[4] How startled might he have been if, confronted with Barbara McClintock's maize that had been stressed by DNA damage, he had observed the suddenly spotted kernels![53]

Much as the concept of genes as independently assorting fixed units of inheritance was shaped by Mendel's attention to an experimental design focused on stable, easily distinguished characters, evolutionary theory was shaped by statisticians whose work emphasized variance around a mean and random sampling. Sharon Bertsch McGrayne contrasts Fisher's approach to that of Bayesian statistics in her summary of Turing's words[83] regarding Enigma cryptanalysis: "confirming inferences suggested by a hypothesis would make the hypothesis itself more probable."[84] How might a Bayesian perspective be applied to evolutionary theory? Distinct classes of mutation could be modeled differently,[85] but beyond this, suppose the set of mutations (or lack of them) generated in each individual's gametes was treated as a prior model or hypothesis about survival: then survival of descendants is an observation. Assumptions of the model (including the assumptions underlying various understandings of random mutation outlined above) would be tested over many generations, with the model updated based upon observations (i.e., the selection and survival of descendants bearing variation generated along the genome by diverse processes).

Such a Bayesian view would predict that evolution itself would drive genomes away from white noise variations, not merely to nonuniform mutations, but ultimately toward embodied models of environmental regularities. The above catalog of mechanisms more than hints that this is possible. Darwin proposed what now looks like a feedback control engineering theory of evolution, but subsequent formalizations have interpreted it in terms of information theory and statistical physics, with minimal feedback. But variation and selection together represent essential elements in a feedback loop, and variation is not outside that loop, however appealingly simple such an assumption may be. Thus, an essential concept is that feedback from the environment, operating through (selection of) surviving descendants, must inevitably incorporate a worldview into the mechanisms that generate genome variation and genome contexts and DNA sequences that encode them, such that the probability of distinct classes of mutation can become aligned with probable effects on survival.

On reflection, Darwin appears to have been reaching for this concept when he suggested that, "deviations of structure are in some way due to the nature of the conditions of life, to which the parents and their more remote ancestors have been exposed during several generations."[2]

Wallace and Darwin began with attention to variation. Now, we are in a position to focus on the organization and regulation of biochemical mechanisms underlying generation of that variation. In fact, Darwin recognized that "a grand and almost untrodden field of inquiry will be opened, on the causes and laws of variation."[2]

Natural selection has led to the evolution, in genomes, of information that structures exploration and facilitates successful adaptation to likely challenges; thus, the most revealing and intriguing aspect of mutation, the generation of variation, is not that it is random, but rather the ways and extent to which it may become biased by feedback from selection.

Acknowledgments

The authors are deeply indebted to David King for his important and insightful comments on an earlier version of this manuscript, and recommend his recent article, which contributes historical background for this discussion.[20] This work was supported in part by the National Science Foundation under Grant No. PHYS-1066293 and the hospitality of the Aspen Center for Physics. J.D. was in part supported by the NSF, AFOSR, and the Institute for Collaborative Biotechnologies through Grant W911NF-09-0001 from the U.S. Army Research Office. The content does not necessarily reflect the

position or the policy of the Government, and no official endorsement should be inferred.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Darwin, C. 1859. On the origin of the species. http://www.talkorigins.org/faqs/origin.html Chapter 5: Laws of Variation.
2. Darwin, C. 1859. On the origin of the species http://www.talkorigins.org/faqs/origin.html Chapter 3.
3. Wallace, A.R. 1905. *My Life*. New York: Dodd, Mead & Company.
4. Mendel's Paper at MendelWeb. http://www.mendelweb.org/MWpaptoc.html
5. Marchant, J. 1916. *Letters and Reminiscences*, vol **2**. London: Cassell. http://darwin-online.org.uk/converted/published/1916_Marchant_F1592.2.html
6. Dobzhansky, T. 1970. *Genetics of the Evolutionary Process*. New York: Columbia University Press.
7. Fisher, R. 1930. In *The Genetical Theory of Natural Selection*. Oxford at the Clarendon Press. pp 38–39.
8. Pennock, R.T. 2005. "On teaching evolution and the nature of science." In *Evolutionary Science and Society: Educating a New Generation*. J. Cracraft & R. Bybee, Eds.: 7–12. Colorado Springs, CO: BSCS.
9. McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36:** 344–355.
10. Lewis, E.B. 1945. The relation of repeats to position effect in drosophila melanogaster. *Genetics* **30:** 137–166.
11. Avery, O.T., C.M. Macleod & M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **79:** 137–158.
12. Benzer, S. 1961. On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA* **47:** 403–415.
13. Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562.
14. Caporale, L.H. 2012. Overview of the creative genome: effects of genome structure and sequence on the generation of variation and evolution. *Ann. N. Y. Acad. Sci.* **1267:** 1–10.
15. Crothers, D.M. 2006. "Sequence-dependent properties of DNA and their role in function." In *The Implicit Genome*. L. Caporale, Ed.: 23. Oxford: Oxford University Press
16. Dickerson, R.E. 1983. Base sequence and helix structure variation in B and A DNA. *J. Mol. Biol.* **166:** 419–441.
17. Maizels, N. & L.T. Gray. 2013. The G4 Genome. *PLoS Genet.* **9:** e1003468.
18. Srinivasan, A.R., R.R. Sauers, M.O. Fenley, *et al.* 2009. Properties of the nucleic-acid bases in free and Watson-Crick hydrogen-bonded states: computational insights into the sequence-dependent features of double-helical DNA. *Biophys. Rev.* **1:** 13–20.
19. Sinden, R., V.N. Potaman, E.A. Oussatcheva, *et al.* 2002. Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosci.* **27**(Suppl. 1): 53–65.
20. King, D.G. 2012. Indirect selection of implicit mutation protocols. *Ann. N. Y. Acad. Sci.* **1267:** 45–52.
21. Dickinson, W.J. & J. Seger. 1999. Cause and effect in evolution. *Nature* **399:** 30.
22. Olivera, B.M., M. Watkins, P. Bandyopadhyay, *et al.* 2012. Adaptive radiation of venomous marine snail lineages and the accelerated evolution of venom peptide genes. *Ann. N. Y. Acad. Sci.* **1267:** 61–70.
23. Weiner, J. 1994. *The Beak of the Finch : A Story of Evolution in Our Time*. New York: Knopf.
24. Caporale, L.H. 2006. An Overview of the implicit genome in *The Implicit Genome*. L. Caporale, Ed. Oxford: Oxford University Press.
25. Wilkins, A.S. 2005. Book review. Darwin in the genome: molecular strategies in biological evolution. *Bioessays*, **27:** 111–112.
26. Caporale, L. 2003. Darwin in the genome. *Bioessays* **27:** 984.
27. Caporale, L.H. 2002. *Darwin in the Genome*. New York: McGraw Hill.
28. King, D.G., E.N. Trifonov & Y. Kashi. Tuning Knobs in the Genome: Evolution of Simple Sequence Repeats by Indirect Selection in *The Implicit Genome*. L. Caporale, Ed. Oxford: Oxford University Press.
29. Moxon, R., C. Bayliss & D. Hood. 2006. Bacterial contingency loci: the role of simple sequence dna repeats in bacterial adaptation. *Annu. Rev. Genet.* **40:** 307–333.
30. Gemayel, R., M.D. Vinces, M. Legendre & K.J. Verstrepen. 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* **44:** 445–477.
31. Van der Ende, A., C.T. Hopman, S. Zaat, *et al.* 1995. Variable expression of class 1 outer membrane protein in Neisseria meningitides is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J. Bacteriol.* **177:** 2475–2480
32. Vinces, M.D., M. Legendre, M. Caldara, *et al.* 2009. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* **324:** 1213–1216.
33. Fondon, J.W. 3rd & H.R. Garner. 2004. Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. USA* **101:** 18058–18063.
34. Fondon, J.W. 3rd, A. Martin, S. Richards, *et al.* 2012. Analysis of microsatellite variation in *Drosophila melanogaster* with population-scale genome sequencing. *PLoS One* **7:** e33036.
35. McIver, L.J., J.F. McCormick, A. Martin, *et al.* 2013. Population-scale analysis of human microsatellites reveals novel sources of exonic variation. *Gene* **516:** 328–334.
36. Caporale, L.H. 2003. Natural selection and the emergence of a mutation phenotype. *Annu. Rev. Microbiol.* **57:** 467–485.
37. Andam, C.P., G.P. Fournier & J.P. Gogarten. 2011. Multilevel populations and the evolution of antibiotic resistance through horizontal gene transfer. *FEMS Microbiol. Rev.* **35:** 756–767.
38. Rodríguez-Trelles, F., R. Tarrío & F.J. Ayala. 2001. Erratic overdispersion of three molecular clocks: GPDH, SOD, and XDH. *Proc. Natl. Acad. Sci. USA* **98:** 11405–11410.
39. Koren, A., P. Polak, J. Nemesh, *et al.* 2012. Differential relationship of DNA replication timing to different forms of

human mutation and variation. *Am. J. Hum. Genet.* **91:** 1033–1040.
40. Walker, S.I., *et al.* 2013. Evolutionary dynamics and information hierarchies in biological systems. *Ann. N.Y. Acad. Sci.* **XXXX:** XX–XX.
41. Durfee, T., A.-M. Hansen, H. Zhi, *et al.* 2008. Transcription profiling of the stringent response in *Escherichia coli*. *J. Bacteriol.* **190:** 1084–1096.
42. Ciccia, A. & S.J. Elledge. 2010. The DNA damage response: making it safe to play with knives. *Mol. Cell.* **40:** 179–204.
43. Singh, A.H., D.M. Wolf, P. Wang & A.P. Arkin. 2008. Modularity of stress response evolution. *Proc. Natl. Acad. Sci. USA* **105:** 7500–7505.
44. Jones, F.C., M.G. Grabherr, Y.F. Chan, *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484:** 55–61.
45. Gally, D.L., J.A. Bogan, B.I. Eisenstein & I.C. Blomfeld. 1993. Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: effects of temperature and media. *J. Bacteriol.* **175:** 6186–6193.
46. Hsu, E., N. Pulham, L.L. Rumfelt & M.F. Flajnik. 2006. The plasticity of immunoglobulin gene systems in evolution. *Immunol. Rev.* **210:** 8–26.
47. Di Noia, J.M. & M.S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76:** 1–22.
48. Kenter, A.L., S. Feldman, R. Wuerffel, *et al.* 2012. Three-dimensional architecture of the IgH locus facilitates class switch recombination. *Ann. N. Y. Acad. Sci.* **1267:** 86–94.
49. Lucht, J.M., B. Mauch-Mani & H.-Y. Steiner, *et al.* 2002. Pathogen stress increases somatic recombination frequency in *Arabidopsis*. *Nat. Genet.* **30:** 311–314.
50. McClintock, B. 1984. The significance of responses of the genome to challenge. *Science* **226:** 792–801.
51. Shapiro, J.A. 2010. Mobile DNA and evolution in the 21st century. *Mob. DNA* **1:** 4.
52. Jorgensen, R.A. 2004. Restructuring the genome in response to adaptive challenge: McClintock's bold conjecture revisited. *Cold Spring Harb. Symp. Quant. Biol.* **69:** 349–354.
53. McClintock, B. 1978. "Mechanisms that rapidly reorganize the genome." Proceedings Stadler Genetic Symposium 10: 25-47. In *The Dynamic Genome: Barbara McClintock's Ideas in the Century of Genetics*. N. Fedoroff & D. Botstein, Eds. Cold Spring Harbor Laboratory Press, Plainview, NY 1992.
54. Naito, K., F. Zhang, T. Tsukiyama, *et al.* 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461:** 1130–1134.
55. Witkin, E.M. 1967. The radiation sensitivity of *Escherichia coli* B: a hypothesis relating filament formation and prophage induction. *Proc. Natl. Acad. Sci. USA* **57:** 1275–1279.
56. Radman, M. 1975. SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci.* **5A:** 355–367.
57. Foster, P.L. 2005. Stress responses and genetic variation in bacteria. *Mutat. Res.* **569:** 3–11.
58. Ponder, R.G., N.C. Fonville & S.M. Rosenberg. 2005. A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol. Cell.* **16:** 791–804.
59. Shee, C., J.L. Gibson, M.C. Darrow, *et al.* 2011. Impact of a stress-inducible switch to mutagenic repair of DNA breaks on mutation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **108:** 13659–13664.
60. Al Mamun, A.A., M.J. Lombardo, C. Shee, *et al.* 2012. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science* **338:** 1344–1348.
61. Woese, C.R. & N. Goldenfeld. 2006. How the microbial world saved evolution from the scylla of molecular biology and the charybdis of the modern synthesis. *Proc. Natl. Acad. Sci. USA* **103:** 10696–10701.
62. Alderson, D.L. & J.C. Doyle. 2010. Contrasting views of complexity and their implications for network-centric infrastructures. *IEEE Trans Syst. Man Cybernetics-Part A* **40:** 839–852.
63. Brenner, S. Collinearity and the genetic code 1966. *Proc. Roy. Soc. B.* **164:** 170–180.
64. Caporale, L.H. 1984. Is there a higher level genetic code that directs evolution? *Mol. Cell. Biochem.* **64:** 5–13.
65. Doyle, J., M. Csete & L.H. Caporale. 2006. An engineering perspective: the implicit protocols in *The Implicit Genome*. L. Caporale, Ed. Oxford: Oxford University Press.
66. Doyle, J.C. & M.E. Csete. 2011. Architecture, constraints, and behavior. *Proc. Natl. Acad. Sci. USA* **108**(Suppl. 3): 15624–15630.
67. Vetsigian, K., C. Woese & N. Goldenfeld. 2009. Collective evolution and the genetic code. *Microbiol. Mol. Biol. Rev.* **73:** 14–21.
68. Barry, J.D., J.P. Hall & L. Plenderleith. 2012. Genome hyper-evolution and the success of a parasite. *Ann. N. Y. Acad. Sci.* **1267:** 11–17.
69. Breaker, R.R. 2011. Prospects for Riboswitch discovery and analysis. *Mol. Cell* **43:** 867–879.
70. Prescott, D.M. 1999. The evolutionary scrambling and developmental unscrambling of germline genes in hypotrichous ciliates. *Nucleic Acids Res.* **27:** 1243–1250.
71. Swart, E.C., J.R. Bracht, V. Magrini, *et al.* 2013. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol.* **11:** e1001473.
72. Nowacki, M., K. Shetty & L.F. Landweber. 2011. RNA-mediated epigenetic programming of genome rearrangements. *Annu. Rev. Genomics Hum. Genet.* **12:** 367–389.
73. Caporale, L.H. 2000. Mutation is modulated: implications for evolution. *Bioessays* **22:** 388–395.
74. Lambowitz, A.M., N. Craig, R. Gragie & M. Gellert. 2002. *Mobile DNA American Society for Microbiol*; Washington, DC 2nd edition.
75. Lukjancenko, O., T.M. Wassenaar & D.W. Ussery. 2010. Comparison of 61 Sequenced. *Escherichia coli* Genomes. *Microb. Ecol.* **60:** 708–720.
76. Chandra, F., G. Buzi & J.C. Doyle. 2011. Glycolytic oscillations and limits on robust efficiency. *Science* **333:** 187–192.
77. Yi, T.M., Y. Huang, M.I. Simon & J. Doyle. 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. USA* **97:** 4649–4653.

78. Bagheri, N., M.J. Lawson, J. Stelling & F.J. Doyle 3rd. 2008. Modeling the *Drosophila melanogaster* circadian oscillator via phase optimization. *J. Biol. Rhythms* **23:** 525–537.
79. Adler, J., G.L. Hazelbauer & M.M. Dahl. 1973. Chemotaxis toward sugars in *Escherichia coli. J. Bacteriol.* **115:** 824–847.
80. Tsunozaki, M., H.S. Chalasani & C.I. Bargmann. 2008. A behavioral switch: cGMP and PKC signaling in olfactory neurons reverses odor preference in *C. elegans. Neuron* **59:** 959–971.
81. Cisek, P. & J.F. Kalaska. 2010. Neural mechanisms for interacting with a world full of action choices. *Ann. Rev. Neurosci.* **33:** 269–298.
82. Landsteiner, K. 1936. Artificial Conjugated Antigens. Serological Reactions with Simple Chemical Compounds. Chapter V in *The Specificity of Serological Reactions*. Charles C. Thomas: Springfield, Illinois.
83. Turing, A.M. 1942. Report on Cryptographic Machinery Available at Navy Department Washington. http://www.turing.org.uk/sources/washington.html
84. Bertsch McGrayne, S. 2011. *The Theory That Would Not Die How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press. New Haven and London.
85. Gjini, E., D.T. Haydon, J.D. Barry & C.A. Cobbold. 2012. The impact of mutation and gene conversion on the local diversification of antigen genes in African trypanosomes. *Mol. Biol. Evol.* **29:** 3321–331.